

Antwortqualitäten

Zur Vergleichbarkeit von Einstellungsmessungen in
standardisierten Befragungen

Kumulative Arbeit

Inaugural-Dissertation

zur Erlangung der Doktorwürde

der

Philosophischen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität

zu Bonn

vorgelegt von

Alice Barth

aus

Starnberg

Bonn 2018

Gedruckt mit der Genehmigung der Philosophischen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

Zusammensetzung der Prüfungskommission:

Prof. Dr. Clemens Albrecht

(Vorsitzender)

Prof. Dr. Jörg Blasius

(Betreuer und Gutachter)

Prof. Dr. Nina Baur

(Gutachterin)

Prof. Dr. André Beauducel

(weiteres prüfungsberechtigtes Mitglied)

Tag der mündlichen Prüfung: 23.10.2018

Zusammenfassung

Die vorliegende Arbeit befasst sich mit der Frage, inwiefern in standardisierten Befragungen erhobene Einstellungen über Zeit und verschiedene Gruppen hinweg vergleichbar sind. Während ein standardisierter Erhebungsprozess notwendige Bedingung zur Herstellung von Vergleichbarkeit ist, ist die tatsächliche Äquivalenz der Daten damit jedoch keineswegs hinreichend gewährleistet. Unter der Vielzahl erhebungs-, verarbeitungs- und analysebedingter Einschränkungen der Vergleichbarkeit liegt der Fokus hier auf differentiellen Reaktionen der Interviewten auf die Befragung („Antwortqualitäten“), den Möglichkeiten, diese empirisch zu erfassen und den Implikationen für die Umfrageforschung. Vor dem Hintergrund kognitiver Prozesse beim Beantworten von Survey-Fragen sowie den Einflüssen des unmittelbaren und des gesellschaftlichen Kontexts werden in den vier Artikeln, die den Kern der Arbeit darstellen, verschiedene Vergleichbarkeitsprobleme thematisiert. Dazu werden Sekundäranalysen von Daten aus renommierten sozialwissenschaftlichen Erhebungen durchgeführt, wobei mittels struktur-entdeckender multivariater statistischer Verfahren unterschiedliche Antwortqualitäten sichtbar gemacht werden.

Ein Artikel befasst sich mit methodologischen Herausforderungen der Messung von Einstellungswandel anhand des Beispiels von Geschlechterrollen. Die Anwendung der multiplen Korrespondenzanalyse auf Daten aus British Social Attitudes Survey und British Household Panel Study zeigt dabei strukturelle Veränderungen des latenten Konstrukts „Einstellungen zu Geschlechterrollen“ und eine Zunahme methoden-induzierter Variation. Aussagen über Einstellungswandel aufgrund veränderter Mittelwerte werden durch derartige methodische Veränderungen in Frage gestellt. In einem weiteren Artikel geht es um den Zusammenhang zwischen Antwortqualitäten und den ideologischen Dispositionen der Befragten unter Verwendung von Daten aus dem US-amerikanischen General Social Survey. Multiple Indikatoren der Reaktionen auf die Befragung werden mittels latenter Klassenanalyse zu einer Typologie der Antwortqualitäten zusammengefasst. Die Projektion dieser in den mittels multipler Korrespondenzanalyse erstellten Raum ideologisch-politischer Verortung ergibt, dass bestimmte ideologische (Dis)positionen mit spezifischen Reaktionsmustern assoziiert sind, was potentielle Probleme für die Abbildung der „öffentlichen Meinung“ mittels standardisierten Befragungen impliziert. Im nächsten Papier werden im ALLBUS erhobene Einstellungen zu Geschlechterrollen in Ost- und Westdeutschland und über Zeit verglichen. Unter Berufung auf rollentheoretische Ansätze findet dabei die latente Klassenanalyse Verwendung, vor den inhaltlichen Vergleichen erfolgt eine Prüfung auf Messinvarianz. Das letzte Papier schließlich beschäftigt sich mit einem speziellen Problem der Vergleichbarkeit von Einstellungsmessungen über Zeit: von der Befragung

selbst induzierte Veränderungen, bekannt unter dem Begriff *panel conditioning*. Es wird argumentiert, dass die Stärke von Einstellungen durch wiederholte Befragungen zunimmt, was – unter Beachtung individueller Voraussetzungen wie Erfahrung, Interesse und Motivation – sowohl Auswirkungen auf Denken und Handeln als auch auf Antwortqualitäten hat.

Insgesamt spiegeln die Arbeiten fünf Aspekte des Problemkomplexes „Antwortqualitäten“ wider, die teilweise miteinander interagieren: Unterschiede in der Bedeutung und Interpretation von Begriffen, der Verortung auf Skalen, der Struktur von Konstrukten, der Zentralität von Themen und beim Umgang mit der Umfrage an sich stellen Herausforderungen für den Vergleich standardisierter Einstellungsmessungen dar. In diesem Sinne fragt die gemeinsame Diskussion der Artikel nach Strategien des Umgangs mit Vergleichbarkeitsproblemen und plädiert für die Beachtung subjektiver und reaktiver Elemente bei der Analyse standardisierter Einstellungsmessungen.

Inhaltsverzeichnis

Zusammenfassung.....	3
Inhaltsverzeichnis.....	5
1. Einleitung.....	6
2. Die Befragungssituation – theoretische Perspektiven und Forschungsstand	9
2.1. Kognitive Prozesse	11
2.1.1 Verständnis.....	12
2.1.2 Informationsabruf und Urteilsbildung.....	14
2.1.3 Bekanntgabe des Urteils	15
2.2 Unmittelbarer und weiterer Befragungskontext.....	16
2.3 Bedingungen der Produktion von Meinungen	19
2.4 Zwischenfazit: Kontextabhängigkeit des Antwortprozesses.....	20
3. Vergleichbarkeit von Einstellungsmessungen?.....	22
3.1 Methoden.....	23
3.2 Die Bedeutung und Interpretation von Begriffen	24
3.3 Die Verortung auf Skalen.....	27
3.4 Die Struktur des Konstrukts	29
3.5 Die Zentralität von Themen	31
3.6 Der Umgang mit der Umfrage.....	33
4. Diskussion.....	36
5. Literatur	40
6. Artikel	53
6.1 Barth, A. (2016). The Changing Nature of Attitude Constructs - An Application of Multiple Correspondence Analysis on Gender Role Attitudes.....	53
6.2 Barth, A.; Schmitz, A. (2018). Response Quality and Ideological Dispositions. An Integrative Approach using Geometric and Classifying Techniques.....	76
6.3 Barth, A.; Trübner, M. (2018). Structural Stability, quantitative change: A latent class analysis towards gender role attitudes in Germany.....	103
6.4 Bergmann, M.; Barth, A. (2018). What Was I Thinking? A Theoretical Framework for Analysing Panel Conditioning in Attitudes and (Response) Behaviour.....	125

1. Einleitung

Fragen nach Einstellungen und Meinungen haben einen beträchtlichen Anteil an der Datenerhebung mithilfe von standardisierten Befragungen¹. Viele nationale und internationale Studien wie der General Social Survey (GSS), die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS), das International Social Survey Programme (ISSP), der Eurobarometer oder der European Social Survey (ESS) erheben schwerpunktmäßig Einstellungen der Bevölkerung. Inhaltliches Interesse gilt den Fragen nach Einstellungsveränderungen, Unterschieden zwischen verschiedenen Gruppen, Nationen und Kulturen, den Ursachen von Einstellungen und Einstellungswandel, Relationen zwischen verschiedenen Einstellungen sowie dem Zusammenhang zwischen Einstellungen und Handlungen.

Für die Validität der Analysen ist zentral, dass festgestellte Unterschiede und Zusammenhänge auf inhaltlicher, anstatt auf methodischer Variation beruhen: Unterschiedliche Messwerte sollen inhaltliche Unterschiede widerspiegeln, gleiche Messwerte die gleiche Ausprägung der Einstellung anzeigen. In anderen Worten, wenn Messen definiert wird als „die systematische Zuordnung einer Menge von Zahlen oder Symbolen zu den Ausprägungen einer Variablen, mithin auch zu den Objekten“ (Friedrichs 1990[1973]:97), so ist die *Äquivalenz* dieser Zuordnung über Untersuchungsobjekte, Zeit und Raum notwendige Bedingung für die Interpretation von Zusammenhängen und Unterschieden (siehe auch Horn & McArdle 1992; Vandenberg & Lance 2000).

In quantitativ orientierten Studien wird versucht, Äquivalenz durch die *Standardisierung* der Befragung zu erreichen (Baur 2014; Schaeffer 2002). So sollen artifizielle Effekte, die beispielsweise durch Unterschiede in Interviewerverhalten, Frageformulierung und –reihenfolge verursacht werden, vermieden werden. Aus der Forschung zu ebensolchen Artefakten geht jedoch hervor, dass die Befragung kein einfaches Reiz-Reaktions-Experiment, sondern ein komplexer sozialer Prozess ist: Die subjektive Definition der Situation seitens der Beforschten, ihr sozialer und kultureller Hintergrund sowie die differentielle Interaktion mit der Interviewerin oder dem Interviewer bzw. dem Instrument untergraben den Versuch, alle Kontextbedingungen stabil zu

¹ Die Begriffe „Einstellung“ und „Meinung“ werden in dieser Arbeit weitgehend synonym verwendet. Während einige Forscherinnen argumentieren, dass Meinungen eine Oberkategorie darstellen, die Einstellungen und Überzeugungen (*beliefs*) umfasst (Chandler 2008), halten andere Meinungen für die messbaren, manifesten Indikatoren latenter Einstellungen (Mora y Araujo 2011). Ein anderer Forschungszweig betont die evaluative (Braun 2006:73-74) bzw. die affektive Komponente von Einstellungen gegenüber Meinungen (Fishbein & Ajzen 1975:11-13; Oskamp & Schultz 2005:13-14). Bergman (1998) vermutet die Hauptunterscheidung angesichts der Vielzahl variierender Definitionen beider Begriffe in der disziplinären Verwendung: die Sozialpsychologie erforscht Einstellungen, während sich die Politikwissenschaft eher auf Meinungen fokussiert. In einer Arbeit, die sich wesentlich auf Erkenntnisse der interdisziplinär geprägten Survey Methodology stützt, erscheint eine Differenzierung zwischen Einstellung und Meinung daher müßig.

halten (Bachleitner, Aschauer & Weichbold 2010; Braun 2006; Hilgers 1997). Daraus ergibt sich, dass auch standardisierte Befragungen je nach Kontext sehr unterschiedliche Reaktionen bei den Befragten hervorrufen.

Im Bereich der Survey Methodology werden methodische Effekte, die sich aus unterschiedlichen Reaktionen auf einzelne Fragen bzw. die Gesamtbefragung ergeben, häufig als Frage der „Antwortqualität“ beschrieben. Als Qualitätsindikatoren werden dabei unter anderem die Anzahl fehlender Werte (*item nonresponse*) bzw. nicht-substantieller Antworten wie „weiß nicht“, die Variabilität von Antworten in Itembatterien und die Reaktion auf fiktive Fragegegenstände bzw. die (Miss)achtung von Anweisungen verwendet (Berinsky, Margolis & Sances 2014; Galesic & Bosnjak 2009; Häder & Kühne 2010; Heerwegh & Loosveldt 2008). Der Ausdruck Antwortqualität impliziert allerdings, dass es gute und schlechte Qualität von Antworten gibt, wobei die Qualität als umso besser einzustufen ist, je geringer Fehler bzw. Verzerrung sind, d.h. umso näher sie dem „wahren Wert“ im Sinne der Klassischen Testtheorie (vgl. Lord & Novick 2008[1968]) kommt. Gerade in Bezug auf Einstellungen, die nicht direkt beobachtbar sind, sondern ein theoretisches Konstrukt darstellen, ist jedoch fraglich ob ein „wahrer Wert“ existiert, von dem die Abweichungen bestimmt werden können (Converse 2006[1964]; Wilson & Hodges 1992; Zaller & Feldman 1992). Im Sinne der Beschreibung unterschiedlicher Interpretationen und Reaktionen auf die Befragung, die zwar als verschieden, aber nicht per se als falsch und richtig klassifiziert werden können, wird im Folgenden der Begriff der *Antwortqualitäten* verwendet bzw. die Frage der *Vergleichbarkeit* von Einstellungsmessungen akzentuiert.

Im Fokus der vorliegenden Arbeit steht die Frage, wie unterschiedliche Antwortqualitäten in bestehenden Datensätzen erfasst und dargestellt werden können. Weiterhin beschäftigt sie sich mit den Hintergründen ihrer Entstehung und Formen des Umgangs mit Vergleichbarkeitsproblemen, die aus unterschiedlichen Antwortqualitäten resultieren. Dazu werden zunächst kognitive und interaktionale Prozesse bei der Befragung sowie die Einflüsse situativer, sozialer und gesellschaftlicher Rahmenbedingungen erläutert. Dem folgt eine Beschreibung der Komponenten von Antwortqualitäten, die die Vergleichbarkeit von Befragungsdaten gefährden können. Dieser Problembereich umfasst gruppen-, zeit- und kulturspezifische Unterschiede in der Struktur von Einstellungskonstrukten, der Bedeutung von Worten, der Verortung auf Antwortskalen, des thematischen Interesses und, daraus folgend, der Einstellungsstärke sowie der Reaktion auf die Befragungssituation als solche. Die vier Artikel, die Methoden der Erfassung und Interpretation von Antwortqualitäten darstellen, werden in diesem Rahmen verortet.

Barth (2016) befasst sich mit methodologischen Herausforderungen bei der Messung von Veränderungen der Einstellung zu Geschlechterrollen in Großbritannien. Die Visualisierung der

Antwortstrukturen in einer standardisierten Itembatterie mittels multipler Korrespondenzanalyse zeigt, dass die latente Struktur des Konstrukts sich über die Zeit verändert und der Anteil methodischer Variation zunimmt.

Barth und Schmitz (2018) demonstrieren den Zusammenhang der ideologischen Disposition von Befragten mit verschiedenen Antwortqualitäten. Mittels latenter Klassenanalyse wird eine differenzierte Typologie von Antwortqualitäten erstellt und in das korrespondenzanalytisch konstruierte „Feld der ideologischen Konsumtion“ der USA projiziert. Dabei zeigt sich, dass verschiedene Ausprägungen von Reaktionen auf die standardisierte Befragung systematisch im Feld variieren.

Barth und Trübner (2018) schlagen vor, die aus rollentheoretischen Überlegungen abgeleitete Komplexität von Geschlechterrolleneinstellungen empirisch mittels der Analyse latenter Klassen zu prüfen. Dem Vergleich der Rollenerwartungen in Ost- und Westdeutschland sowie zwischen 1991 und 2012 geht dabei eine statistische Prüfung auf Messäquivalenz voraus, um die Effekte verschiedener gesellschaftlicher Rahmenbedingungen auf Frageninterpretation und Antwortverhalten zu prüfen.

Bergmann und Barth (2018) widmen sich einer speziellen Herausforderung zeitvergleichender Einstellungsmessung, dem *panel conditioning*, aus theoretischer Perspektive. Kognitive Prozesse bei der Teilnahme an Panelstudien, insbesondere die Zunahme von Einstellungsstärke und daraus folgende Effekte für die Fragenbeantwortung, werden unter Rückgriff auf ein Modell assoziativer Netzwerke ausgearbeitet. Eine tragende Rolle spielen in diesem Zusammenhang individuelle Voraussetzungen der Befragten wie Erfahrung, Interesse und Motivation.

Die gemeinsame Diskussion der Artikel legt das Augenmerk auf Strategien des Umgangs mit unterschiedlichen Antwortqualitäten und erörtert Möglichkeiten, die Kontextgebundenheit standardisierter Befragungen stärker in den Fokus empirischer Analysen zu nehmen.

2. Die Befragungssituation – theoretische Perspektiven und Forschungsstand

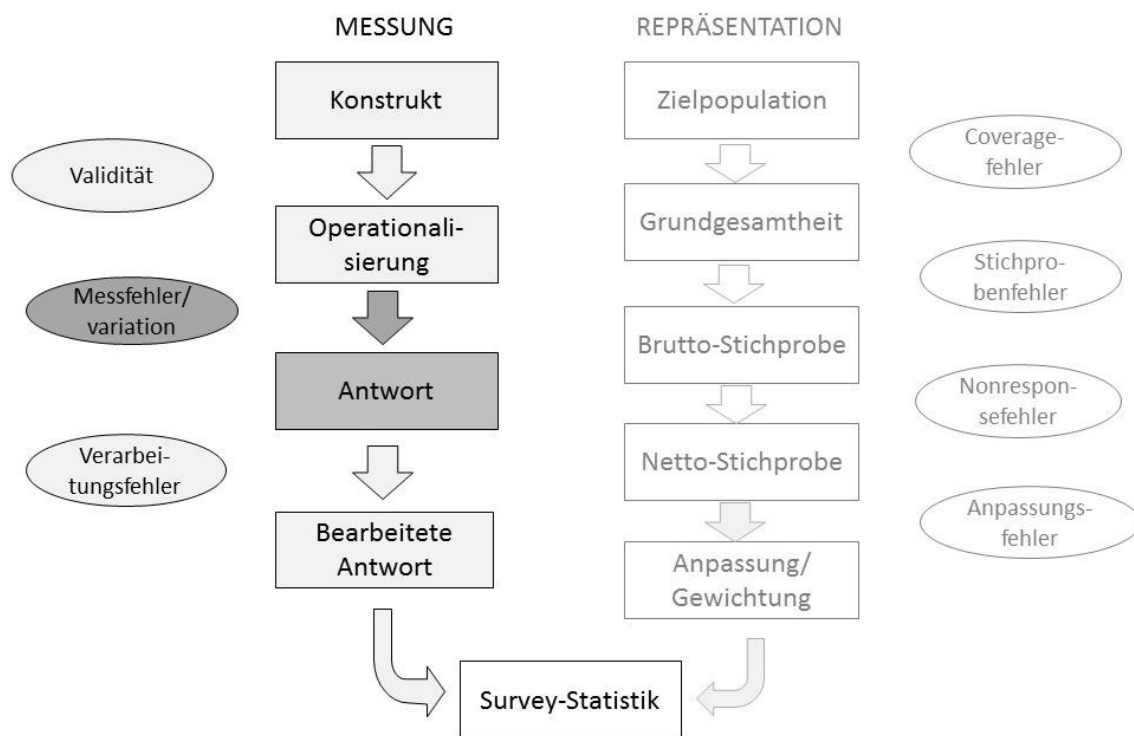
Trotz der ubiquitären Verwendung von Daten aus standardisierten Befragungen in der empirischen Sozialforschung ist immer wieder beklagt worden, dass keine umfassende Theorie der Befragung an sich existiert, sondern meist nur Einzelaspekte betrachtet werden (Bachleitner et al. 2010; Groves 1987). Bachleitner et al. (2010) unterscheiden zwischen den Problembereichen *Teilnahme an Befragungen* und *Antwortverhalten*, die jeweils mit Rückgriff auf unterschiedliche theoretische Ansätze behandelt werden. Im Bereich des ersteren dominieren rationale Handlungstheorien, die die (Nicht-)Teilnahme mit einer auf Kosten-Nutzen-Abwägungen (vgl. Esser 1986; Schnell, Hill & Esser 2005:356-357), sozialen Gegenseitigkeitsvorstellungen (Dillman 2007) oder Entscheidungsheuristiken (Groves, Cialdini & Couper 1992) basierenden Entscheidung seitens der Befragten erklären. Auch die von Groves et al. (Groves, Singer & Corning 2000; Groves et al. 2009) entwickelte Leverage-Saliency-Theorie geht von einer rationalen Entscheidung auf der Basis unterschiedlich gewichteter Aspekte der Befragung (Thema, Incentives, Zeitaufwand usw.) aus. Die Teilnahmebereitschaft hängt somit sowohl von Merkmalen der Befragung, der Befragten und der Interviewenden bzw. des Instruments ab, als auch von allgemeinen kulturellen und gesellschaftlichen Voraussetzungen, wie etwa Einstellungen zu Umfragen und den daraus generierten Ergebnissen (Bachleitner et al. 2010:19).

Das Antwortverhalten hingegen wird vor allem mit Phasenmodellen auf Basis kognitionspsychologischer Erkenntnisse, teilweise mit Bezug auf soziale und Konversationsnormen, erklärt (Sudman, Bradburn & Schwarz 1996; Tourangeau, Rips & Rasinski 2000). Ein weiterer Forschungsbereich sind kommunikative Prozesse im Survey-Interview (Maynard et al. 2002; Suchman & Jordan 1994). Einerseits wird aus ethnomethodologischer Perspektive die Interaktion zwischen Interviewenden und Befragten thematisiert und die Sinnhaftigkeit einer standardisierten Konversation vor dem Hintergrund der „fundamentally social nature of the interview“ (Maynard & Schaeffer 2002:3) hinterfragt. Andererseits geraten in Zeiten sinkender Teilnahmebereitschaft zunehmend kommunikative Strategien bei der Initiierung des Kontakts mit potentiellen Teilnehmerinnen und Teilnehmern in den Fokus surveymethodologischer Forschung (Bradburn 2016). Die Interaktionsperspektive ist somit sowohl im Bereich der Teilnahme an Befragungen als auch beim Antwortverhalten vertreten, jedoch lassen sich die Erkenntnisse nur bedingt auf schriftliche Befragungen ohne direkte Interviewer-Befragten-Interaktion übertragen.

Ein übergreifender Ansatz einer Theorie der Befragung ist das am Ablauf des Befragungsprozesses orientierte Paradigma des *Total Survey Error* (Biemer 2010; Groves et al. 2009). Der Total Survey Error hat zum Ziel, mit der Minimierung verschiedener Fehlerquellen die unter der Bedingung

eingeschränkter finanzieller Ressourcen maximal mögliche Datenqualität zu erreichen (Biemer 2010). Somit werden zwar die verschiedenen Stufen des Forschungsprozesses zusammen betrachtet, der Fokus liegt aber eher auf den Auswirkungen bestimmter Design-Entscheidungen als auf der Offenlegung unterliegender Mechanismen (Bachleitner et al. 2010). Ähnliches gilt für die von Don Dillman entwickelte *Tailored Design Method* (Dillman, Smyth & Christian 2014), die ebenfalls den gesamten Forschungsablauf vor allem im Hinblick auf die praktische Gestaltung des Fragebogens und die Maximierung von Teilnahmebereitschaft und Antwortqualität betrachtet. Hilfreich sind derartige prozessorientierte Modelle, um sich die Herausforderungen, die sich an verschiedenen Stellen des Survey-Prozesses ergeben, zu vergegenwärtigen. Die an Groves et al. (2009) orientierte Abbildung des *Survey Life Cycle* aus einer Qualitätsperspektive (Abb. 1) verdeutlicht die Unterscheidung zwischen Aspekten der *Messung von Konstrukten* und der *Repräsentation von Bevölkerungsmerkmalen*².

Abbildung 1: Modell des Survey Life Cycle aus einer Qualitätsperspektive



Eigene Abbildung, orientiert an Groves et al. (2009), S. 48

Der Fokus der vorliegenden Arbeit liegt auf dem Bereich der Messung, und zwar insbesondere auf dem Übergang zwischen dem operationalisierten Konzept und der Antwort der Befragten – also dem, was in der Ausdrucksweise der Forschung zu Surveyfehlern bzw. Surveyqualität als

² Andere Konzeptualisierungen des Total Survey Error treffen eine Primärunterscheidung zwischen sampling error und nonsampling error (Biemer & Lyberg 2003; Smith 2011).

„Messfehler“ (bzw. „Messvariation“, vgl. Smith 2011) bezeichnet wird. Insgesamt gibt es sechs Komponenten des Messprozesses, die zum gesamten Messfehler einer Befragung beitragen können: Interviewende, Befragte, Modus der Datenerhebung, Fragebogen, Informationssystem (die den Befragten zur Verfügung stehenden Informationen) und schließlich die Umgebung, in der die Befragung stattfindet (Biemer & Lyberg 2003:116-117). Hier erfolgt eine Fokussierung auf die Reaktionen der Befragten, deren Interaktion mit Informationssystem, Interviewenden bzw. Instrument und die Einbettung all dessen in den gesellschaftlichen Kontext. Theoretische Ansätze, die sich auf diese Aspekte beziehen, werden im Folgenden erläutert.

2.1. Kognitive Prozesse

In der kognitivpsychologischen Forschung zum Antwortverhalten in Surveys findet ein Modell, das den Antwortprozess in mehrere, von den Befragten auszuführende Aufgaben bei der Beantwortung von Fragen unterteilt, breite Anwendung (Groves et al. 2009:218-223; Sudman et al. 1996:56-79; Tourangeau et al. 2000:5-16).

Abbildung 2: Modell der kognitiven Prozesse beim Beantworten von Survey-Fragen

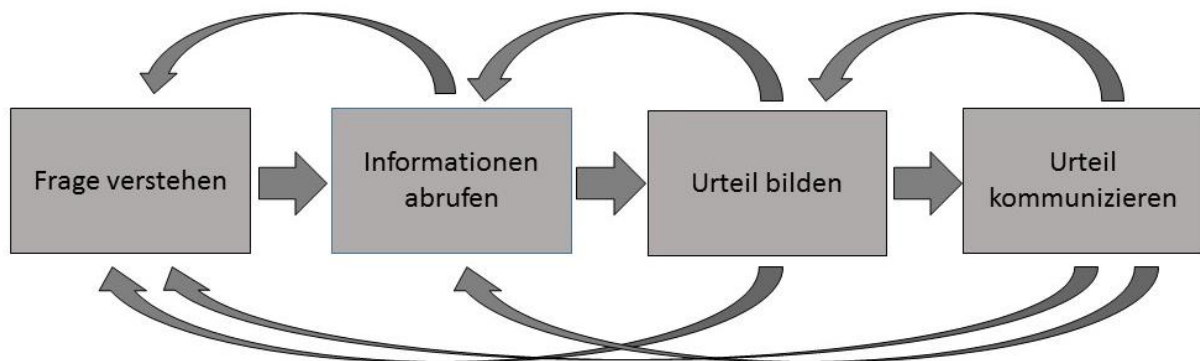


Abbildung nach Groves et al. (2009), S. 218. Siehe auch Sudman et al. (1996); Tourangeau et al. (2000)

Nach dem Verstehen der Frage erfolgen das Abrufen relevanter Informationen, die Bildung eines Urteils und schließlich die Kommunikation dieses Urteils, wobei dieses in die gegebenenfalls zur Verfügung stehenden Antwortoptionen eingepasst wird. Die Unterteilung der beim Beantworten einer Frage ablaufenden kognitiven Prozesse in die genannten Blöcke bietet – auch wenn der Antwortprozess nicht notwendigerweise alle Schritte in dieser Reihenfolge umfasst – die Möglichkeit, Ergebnisse aus Experimenten, die Frageformulierung, Reihenfolge, Antwortoptionen und andere Bedingungen variieren, theoretisch einzuordnen. Die wohl wichtigste Erkenntnis ist hierbei, dass der Antwortprozess nicht nach einem interindividuell gleichen Stimulus-Reaktions-

Modell abläuft, sondern stark kontextabhängig ist (Atteslander 2006:104-121; Schwarz et al. 2009). Sowohl die Eigenschaften des Befragungsinstruments als auch der soziale und kulturelle Hintergrund der Befragten schlagen sich in Antworteffekten nieder, die die Vergleichbarkeit von Einstellungsmessungen verkomplizieren und teils sogar unmöglich machen. Jeder der kognitiven Schritte hält in dieser Hinsicht eigene Herausforderungen bereit:

2.1.1 Verständnis

Ob Befragte den Sinn einer Frage so verstehen, wie es der Intention der Forscherin oder des Forschers entspricht, hängt sowohl vom *semantischen* als auch vom *pragmatischen* Verständnis ab (Porst 2008:18-23; Sudman et al. 1996:59-69). Während es beim semantischen Verständnis darum geht, die wörtliche Bedeutung der Frage bzw. einzelner Begriffe zu erfassen, bezieht sich das pragmatische Verständnis darauf, zu verstehen wie die Frage gemeint ist – was wollen die Fragenden eigentlich wissen? Schwierigkeiten mit der wörtlichen Bedeutung treten auf, wenn Begriffe unbekannt, vage oder mehrdeutig sind und damit Raum für unterschiedliche Interpretationen lassen. So ist beispielsweise empirisch erwiesen, dass ein Großteil der Befragten des EU-weiten Eurobarometers nicht weiß, welche Aufgaben der Europäische Ombudsmann hat. Trotzdem wurde des Öfteren gefragt, wie groß das Vertrauen in diesen ist (Höpner & Jurczyk 2012). Hier ist davon auszugehen, dass andere Einstellungen und Kenntnisse der Befragten, etwa ihre generelle Meinung zu europäischen Institutionen oder ihre Erfahrungen mit Ombudspersonen in anderen Kontexten, die Beantwortung der Frage beeinflussen. In ähnlicher Weise zeigen Befragungsexperimente immer wieder, dass ein beträchtlicher Teil der Befragten die Kompetenzen erfundener Politiker bewertet (Reuband 2000) oder eine dezidierte Meinung zu nicht existenten Gesetzen äußert (Schuman & Presser 1981:147-160; Sturgis & Smith 2010). Im Gegensatz dazu dürfte beispielsweise der Begriff „Familie“ allen Befragten bekannt sein, es hängt aber stark vom kulturellen Hintergrund ab, wie viele Personen welchen Verwandtschaftsgrades in die Definition mit einfließen (Baur 2014). Das semantische Verständnis umfasst auch die Syntax der Frage – so können komplexe oder mehrdeutige Satzstrukturen sowie doppelte Verneinungen zu Missverständnissen führen (Tourangeau et al. 2000:38-50).

Die Forschung zum pragmatischen Frageverständnis geht davon aus, dass das Survey-Interview eine Spezialform einer Konversation darstellt (Suchman & Jordan 1990). Konversationen im Alltag sind in der Regel dadurch gekennzeichnet, dass alle Beteiligten an ihrem Gelingen im Sinne eines gerichteten Austauschs (im Gegensatz zu einer Folge unzusammenhängender Äußerungen) interessiert sind und dementsprechend implizite Regeln beachten (Grice 1975; Clark & Schober 1992). Der Sprachphilosoph Paul Grice identifizierte – in Anlehnung an Kant – vier *Maximen der kooperativen Kommunikation*, die er unter den Begriffen *Quantity*, *Quality*, *Relation* und *Manner*

subsumiert. Im Einzelnen besagen diese Maximen, (1) dass so viele Informationen wie nötig, aber nicht mehr als erforderlich bereitgestellt werden, (2) dass Gesprächspartner sich im Allgemeinen an der Wahrheit (bzw. dem, was sie dafür halten) orientieren, (3) dass die jeweiligen Beiträge zum Thema der Konversation passen, und schließlich, (4) dass man sich möglichst klar und verständlich ausdrückt (Grice 1975). Diese Maximen kommen entsprechend auch in Survey-Interviews zur Anwendung (Clark & Schober 1992; Sudman et al. 1996).

Die Befragten gehen also davon aus, dass die gestellten Fragen einen Sinn haben und prinzipiell beantwortbar sind. Falls die Intention nicht direkt aus dem Wortlaut der Frage zu erkennen ist, werden sie versuchen, aus dem Kontext – Antwortoptionen, vorangegangene Fragen und ihre eigenen Antworten – zu erschließen, welche Informationen von Interesse sind, bzw. was die „richtige“ Antwort sein könnte (Sudman et al. 1996). So kann beispielsweise die Antwortskala einen „Anker“ darstellen, aus dem Befragte schließen, was die intendierte Bedeutung vager Termini wie „Erfolg im Leben“ (Schwarz et al. 1991) oder „Wut“ (Schwarz et al. 1988; Winkielmann, Knäuper & Schwarz 1998) ist. Verstärkt wird die Tendenz zum Rückgriff auf den Fragenkontext dadurch, dass die Möglichkeit, Nachfragen zu stellen in selbst-administrierten Interviews (online, paper-and-pencil) in aller Regel nicht vorhanden ist. In face-to-face oder telefonischen Befragungen sind Interviewende häufig angewiesen, nicht inhaltlich auf Nachfragen einzugehen. Stattdessen wird die standardisierte Frage wiederholt oder die Interpretation dem/der Befragten selbst überlassen: „Whatever it means to you“ (Schober & Conrad 2002:71; vgl. auch Maynard & Schaeffer 2002; Schwarz et al. 2009).

Umgekehrt kann jedoch auch zu viel Information bereitgestellt werden, etwa wenn lange und komplexe Definitionen die kognitiven Kapazitäten der Befragten überschreiten oder Präsuppositionen, also in der Fragestellung enthaltene Vorannahmen, nicht auf alle Befragten zutreffen. Als Beispiel für letzteres nennen Tourangeau et al. (2000) die Aussage „Das Familienleben leidet oft, weil Männer sich zu stark auf ihre Arbeit konzentrieren“, die lange Zeit Bestandteil der Geschlechterrollen-Batterie des amerikanischen General Social Survey (GSS) wie auch des International Social Survey Programme (ISSP) war. Befragte, die der Vorannahme, dass Männer sich zu stark auf ihre Arbeit konzentrieren, nicht zustimmen, haben Schwierigkeiten, ihre Einstellung zu Geschlechterrollen in diesem Rahmen zu kommunizieren.

Auch aus der Perspektive ethnomethodologischer Forschung zum Frageverständnis wird betont, dass Sprache per se mehrdeutig ist (Houtkoop-Steenstra 2000; Suchman & Jordan 1994). Fragen und Antworten in Surveys erlangen Sinn nur vor dem Hintergrund der alltagspraktischen Erfahrungen und soziokulturellen Bedeutungszusammenhänge der Befragten (Baur 2009; Cicourel 1974; Maynard & Schaeffer 2002). Bei den Forscherinnen und Forschern wirken beim Erstellen

des Fragebogens und der späteren Kodierung und Analyse der Antworten ebenfalls implizite Annahmen über geteilte Wissensbestände (Mishler 1986) und Vorstellungen von „Common-Sense-Handlungsabläufe[n]“ (Cicourel 1974:315). Schwierigkeiten ergeben sich, wenn diese stillschweigenden Annahmen nicht zutreffen, Befragte und Interviewende oder Forschende also mit unterschiedlichen Bedeutungshorizonten operieren. Während die Kontextgebundenheit von Fragen in Querschnittsanalysen der eigenen Herkunftsgesellschaft nicht auffallen mag, besteht spätestens in kultur- und zeitvergleichenden Studien die Gefahr, falsche Schlüsse zu ziehen, weil Forscherinnen und Forscher nicht mit den spezifischen Implikationen von Konzepten vertraut sind oder den Antworten unterschiedliche Interpretationen seitens der Befragten zugrunde liegen (Baur 2004; 2009; Braun 2003).

In Anlehnung an Tourangeau et al. (2000:60) lassen sich die Überlegungen zum Fragenverständnis in zwei Problemkomplexe unterteilen: Zum einen kann es aufgrund semantischer oder pragmatischer Unklarheiten dazu kommen, dass Befragte grundlegende Schwierigkeiten haben, den Sinn der Frage zu erfassen und infolgedessen nicht in der Lage sind, weitere zur Beantwortung nötige kognitive Schritte zu unternehmen. Die Frage wird dann entweder gar nicht, mit einer Ausweichkategorie („weiß nicht“, „keine Angabe“ etc.) oder mit einer beliebigen Kategorie aus den zur Auswahl stehenden beantwortet. Zum anderen kann die – Kommunikationssituationen natürlicherweise innewohnende – Neigung der Befragten, den Fragenkontext mit zur Interpretation heranzuziehen, und die Unterschiedlichkeit kultur- und gruppenspezifischer Deutungsmuster zu Verständnisunterschieden führen. Da in diesen Fällen zumeist eine substantielle Antwortkategorie gewählt wird, sind derartige Verzerrungen ohne zusätzliche Informationen schwer nachzuweisen (Baur 2009; Braun 2003).

2.1.2 Informationsabruf und Urteilsbildung

Sind sich Befragte darüber klargeworden, worauf sich eine Frage bezieht, wird als nächster Schritt der Abruf von relevanten Informationen aus dem Gedächtnis angenommen, auf deren Basis ein Urteil gebildet wird. Welche Informationen dabei relevant sind, hängt vor allem von deren *momentaner Verfügbarkeit* ab – aus zeit- und ressourcenökonomischen Gründen werden Befragte in der Regel nicht alle potentiell zugehörigen Aspekte (u.a. allgemeine Eindrücke, generelle Werte und Normen, spezifische Überzeugungen sowie frühere Urteile, siehe Tourangeau et al. 2000:172-196) aus dem Gedächtnis abrufen und sorgsam abwägen, sondern den Suchprozess stoppen, sobald ein akzeptabel scheinendes Ergebnis erreicht ist (Sudman et al. 1996:70-71).

Die Bewertung der Ergebnisse dieses Prozesses in Bezug auf Einstellungsfragen ist stark davon abhängig, wie Einstellungen als solche definiert werden (Schwarz 2007; Tourangeau et al. 2000:165-196). Werden Einstellungen als relativ dauerhaftes, positiv oder negativ gerichtetes Gefühl bezogen

auf eine bestimmte Entität verstanden (Petty & Cacioppo 1996[1981]:7; Eagly & Chaiken 1993:1), so sind „flüchtige“, momentane Einflüsse eine Fehlerquelle, die das Messen der „wahren“ Einstellung erschwert. In der kognitivpsychologischen Forschung im Rahmen der Survey Methodology hat sich allerdings eher die Ansicht durchgesetzt, dass Einstellungen „temporäre Konstruktionen“ (Tourangeau et al. 2000:197) sind, die zwar auf dauerhaften Überzeugungen und Werten beruhen können, aber meist spontan in Antwort auf eine konkrete Frage aus gerade verfügbaren Assoziationen erstellt werden (Schwarz 2007; Wilson & Hodges 1992). Diese Erkenntnis speist sich unter anderem aus den Ergebnissen mannigfaltiger Experimente, aus denen hervorgeht, dass berichtete Einstellungen erheblichen zeitlichen Schwankungen unterworfen sind (Converse 2006; Zaller & Feldman 1992) und stark von Frageformulierung und –platzierung im Fragebogen sowie den verfügbaren Antwortkategorien beeinflusst werden (Schuman & Presser 1981; Schwarz & Bless 1992). Aus dieser Perspektive sind Einflüsse des Kontexts keine Störelemente, sondern immanenter Teil des Prozesses, in dem aus den verfügbaren Informationen ein zusammenfassendes Urteil gebildet wird. Dabei gilt, dass je häufiger Befragte mit dem betreffenden Einstellungsobjekt in Berührung gekommen sind (durch persönliche Erfahrung, aber auch durch wiederholtes Nachdenken darüber im Kontext vorheriger Befragungen), je konsistenter die Informationen sind und je zentraler die Einstellung für die Befragten ist („Einstellungsstärke“, siehe Howe & Krosnick 2017; Krosnick & Petty 1995), umso schneller läuft der Prozess des Informationsabrufs und der Urteilsbildung ab und umso stabiler, d.h. kontextunabhängiger, erscheint die geäußerte Einstellung. Die Implikationen dieser Überlegung für die Einstellungsmessung in Panelstudien werden ausführlich in Bergmann & Barth 2018 (siehe auch Kapitel 3.5) dargestellt.

2.1.3 Bekanntgabe des Urteils

Sind Befragte aufgrund des internen Abruf- und Auswahlprozesses zu einem Urteil gelangt, wird dieses im nächsten Schritt in einer der Befragung angemessenen Art und Weise kommuniziert. Bei geschlossenen Fragen impliziert dies die Einpassung in das vorgegebene Antwortformat. Idealtypisch wird die Antwort dabei in eine Zahl oder Position auf einer Ratingskala „übersetzt“. Aus der Forschung zu Fragebogeneffekten geht hervor, dass Befragte sich bei diesem Schritt häufig an der Reichweite und der Richtung der Skala orientieren (Schuman & Presser 1981; Schwarz & Bless 1992; Schwarz et al. 1988). Bei Einstellungsfragen hängt die Verortung zusätzlich davon ab, mit wem Befragte sich vergleichen, um ihre eigene Position zu bestimmen (Couper, Conrad & Tourangeau, 2007; Mussweiler 2003), und kann sich durch subjektive Erfahrungen verändern. So wurden beispielsweise erstaunlich geringe Unterschiede in der Selbstbewertung von Lebensqualität und Gesundheitszustand zwischen Patientinnen und Patienten mit schweren chronischen Krankheiten und Gesunden auf standardisierten Ratingskalen festgestellt (Breetvelt & van Dam

1991; Rapkin & Schwartz 2004). Zudem scheint der kulturelle Hintergrund den Umgang mit Skalen – insbesondere die Tendenz zu extremen Antworten und Akquieszenz – zu beeinflussen (Marín, Gamba & Marín 1992; Smith 2004).

Bei der Bekanntgabe des Urteils spielen wiederum Konversationsnormen eine Rolle: in der Wahl ihrer Antworten werden Befragte berücksichtigen, welche Informationen für die Forscherin oder den Forscher relevant sein könnten, also nichts berichten was nicht Teil der vorgegebenen Antwortoptionen ist, ihrer Meinung nach „nichts zur Sache tut“ oder bereits abgefragt wurde (Porst 2008:23-27). Ebenso kann es sein, dass Befragte ihr Urteil „bearbeiten“, bevor sie es kommunizieren, weil sie um ihre Selbstpräsentation besorgt sind. Dieser Effekt sozialer Erwünschtheit ist eine Folge der sozialen Normen, die Befragte subjektiv wahrnehmen, und ist dementsprechend situations- und kulturabhängig (Johnson & Van de Vijver 2003).

Insgesamt stellt das bewusste Durchlaufen aller Schritte einen idealisierten Verlauf des Antwortprozesses dar – tatsächlich ist davon auszugehen, dass häufig einige oder alle der angenommenen Schritte schlampig ausgeführt oder ganz übersprungen werden. Jon Krosnick hat für „suboptimales“ Antwortverhalten, das auf Mangel an Motivation bzw. Fähigkeiten der Befragten und die Schwierigkeit der Fragestellung zurückzuführen ist, den Begriff *satisficing* geprägt (Krosnick 1991). Er versteht darunter die Tendenz von Befragten, die durch die Befragung erzeugte kognitive Belastung durch verschiedene Strategien zu reduzieren, z.B. indem die erstbeste plausibel erscheinende Antwort gewählt wird, beim Bewerten mehrerer Objekte auf Ratingskalen nicht differenziert wird oder indem die Wahl einer Antwort zufallsgeleitet erfolgt.

Auch die Unterscheidung zwischen *satisficing* und *optimizing* ist jedoch eine idealtypische, wie Tourangeau et al. (2000:17) sowie Groves et al. (2009:224) argumentieren: Realitätsgerechter sei die Vorstellung eines Kontinuums zwischen Sorgfalt und Nachlässigkeit, auf dem sich die Befragten bewegen. Wichtiger als die Frage nach dem genauen Ablauf der kognitiven Prozesse ist für die vergleichende Forschung die Erkenntnis, dass die inneren Vorgänge während des Antwortprozesses entscheidend durch Kontextwissen und soziale Normen der Befragten geprägt werden. Damit spielen Aspekte, die in standardisierten Befragungen meist nicht direkt adressiert werden (können), eine große Rolle für das Antwortverhalten.

2.2 Unmittelbarer und weiterer Befragungskontext

Über die kognitiven Prozesse hinaus muss das Interview als soziale Situation verstanden werden. Maynard und Schaeffer (2002:35) betonen, dass die Praxis der Befragung auf interaktionalen „taken-for-granted skills“ beruht, die, in ähnlicher Weise wie die von Grice identifizierten

Konversationsnormen, eine zielgerichtete Kommunikation erst ermöglichen. Dies gilt vor allem für persönliche und telefonische Befragungen, bei denen eine direkte Kommunikation mit den Interviewenden erfolgt, aber auch für schriftliche Befragungen, bei denen die Befragten auf vorformulierte Fragen und Aussagen reagieren.

Suchman und Jordan (1994) argumentieren, dass es sich bei standardisierten Befragungen um Konversationen handelt, die sich in vielfacher Hinsicht von alltäglichen Kommunikationssituationen unterscheiden: Während in „normalen“ Konversationen eine beidseitige, „lokale“ Kontrolle über Themen und Gesprächsverlauf besteht, werden diese in der Befragung im Vorhinein extern vorgegeben und einseitig von den Interviewenden bzw. dem Befragungsinstrument implementiert. Zusätzlich wird das in Gesprächen übliche Interaktionsprinzip dadurch verletzt, dass die Interviewenden möglichst nicht von der vorgegebenen Frageformulierung abweichen sollen und Nachfragen seitens der Befragten, wenn überhaupt, nur in begrenztem Maße möglich sind. Während dies einerseits der Vermeidung von Interviewereffekten dient, wird andererseits die Verständigung auf einen gemeinsamen Deutungshorizont erschwert.

Für die Generierung valider Daten ist es notwendig, dass Befragte eine Frage so verstehen, wie es von der Forscherin oder dem Forscher als dritter, nicht anwesender Partei intendiert war (Suchman & Jordan 1990). Aus etlichen Studien des Interaktionsverhaltens von Interviewenden und Befragten geht jedoch hervor, dass dies beileibe nicht immer der Fall ist. Die Fallstricke der Standardisierung reichen von Missverständnissen bis hin zu komplett sinnentleerten Kommunikationsakten – die Konversationsanalysen fördern jedoch auch häufig Abweichungen von der vorgeschriebenen Prozedur zu Tage, die die Beteiligten vornehmen um den intendierten Sinn der Konversation zu wahren (Houtkoop-Steenstra 2000; Suchman & Jordan 1994). Als Konsequenz wird von diesem Forschungszweig vorgeschlagen, den Interviewenden mehr Freiheit in der Gesprächsführung zu gewähren und so ihre Rolle als „the principal mediator between the intended meaning of the question and the interpretations of the respondent“ (Suchman & Jordan 1990:240; vgl. auch Gobo & Mauceri 2014) zu stärken. Die Frage, wie bei diesem Vorgehen verstärkte Interviewereffekte vermieden werden können, bleibt allerdings offen.

Von der erfolgreichen Initiierung der Interaktion hängt nicht nur der Verlauf des Interviews ab, sondern auch die Frage ob sich jemand überhaupt zur Teilnahme bereit erklärt. Letzteres beinhaltet vor allem die schnelle Etablierung einer gemeinsamen Basis („common ground“, Bradburn 2016).

Für eine möglichst umfassende Untersuchung der Einflussfaktoren in einer Befragungssituation müssen zusätzlich zu kognitiven und interaktionalen Prozessen noch der lokale und der gesellschaftliche Kontext, in den die Befragung eingebettet ist, betrachtet werden. Bachleitner et al.

(2010) weisen darauf hin, dass nicht nur Instrument und Interviewende, sondern auch der unmittelbare räumliche und zeitliche Kontext, in dem das Interview stattfindet, die Befragungsergebnisse beeinflussen. So kann die Atmosphäre des Befragungsortes inhaltlich relevant sein, wie an einer Studie zum Sicherheitsempfinden, die an verschiedenen innerstädtischen Orten durchgeführt wurde, belegt wird (Bachleitner et al. 2010:97-100). Auch die emotionale Befindlichkeit der Befragten, ein Faktor der nur selten als Kontextvariable in Betracht gezogen wird, kann einen starken Einfluss auf das Antwortverhalten haben, etwa bei der Bewertung verschiedener Aspekte eines Museums (Bachleitner et al. 2010:103-105) oder der Einschätzung der eigenen Lebenszufriedenheit (Schwarz & Clore 1983).

Schließlich gibt es bestimmte gesamtgesellschaftliche Voraussetzungen, die das Durchführen standardisierter Befragungen erleichtern oder überhaupt erst ermöglichen. Die Vertrautheit mit den formalen Rollenanforderungen des/der „Interviewten“ bzw. „Interviewenden“ ist eine relativ neue Errungenschaft industrialisierter Gesellschaften (Cisneros-Puebla, Faux & Mey 2004) – ein Resultat der Verbreitung des „modern temper“ (Riesmann & Benney 1956a:7). Im Einzelnen umfasst dies einen relativ hohen Grad an individueller Autonomie, sozialer Differenzierung und Rollenheterogenität (Esser 1975a; Riesmann & Benney 1956b) sowie formalisierte Kommunikationsmedien, die „zur situationsneutralen Übertragung kognitiver Inhalte“ (Esser 1975a: 318) geeignet sind. Spezifischer benennt Bradburn (2016) das Vertrauen in den Schutz der Privatsphäre der Befragten sowie die Bereitschaft, wahre Angaben zu machen als interindividuelle Voraussetzungen für die Durchführung von Befragungen. Daraus folgen besondere Schwierigkeiten bei der kultur- und gruppenvergleichenden Forschung, wenn diese Bedingungen in unterschiedlich hohem Maße vorliegen.

Aus mehreren Studien (Curtin, Presser & Singer 2005; de Leeuw & Heer 2002; Kim et al. 2011) geht hervor, dass die Bereitschaft zur Teilnahme an Befragungen in westlichen Industrienationen in den letzten Jahrzehnten abgenommen hat. Dies wird teilweise auf Änderungen des *survey-taking climate* – definiert als Bevölkerungseinstellungen gegenüber Meinungsumfragen (Loosveldt & Storms 2008; Yan & Datta 2015) – zurückgeführt. So zeigen Kim et al. (2011) für die Vereinigten Staaten, dass das Vertrauen in die Meinungsforschung von 1998 bis 2006 deutlich abgenommen hat. Die Aussage, „Answering questions in polls or research surveys is an interesting experience“ erfährt in den 2000er Jahren deutlich weniger Zustimmung als in den 1980er Jahren, wobei die Mehrheit dennoch weiterhin von der Sinnhaftigkeit von Umfragen überzeugt ist. Zu einem Versuch, das *survey-taking climate* mittels einer Medienkampagne zu beeinflussen, berichten Yan und Datta (2015), dass gesteigertes Wissen über Ziele und Zwecke des US-amerikanischen Census zu positiveren Einstellungen gegenüber diesem, und damit auch zu einer höheren

Teilnahmewahrscheinlichkeit, führten. Die erkenntnistheoretische Einschränkung derartiger „Surveys on Surveys“ ist allerdings offensichtlich (Goyder 1986).

2.3 Bedingungen der Produktion von Meinungen

Weitergehend wird die Frage impliziter Voraussetzungen von Meinungsumfragen von Pierre Bourdieu (2013a[1973]; 2013b[1977]) diskutiert. Vor allem in Bezug auf politische Aussagen konstatiert er eine Diskrepanz zwischen der Annahme, „dass die Produktion einer Meinung in jedermanns Reichweite liegt“ (Bourdieu 2013a:243), welche der Anwendung standardisierter Befragungsinstrumente zugrunde liegt, und realen sozialen Gegebenheiten. Er weist diesbezüglich nach, dass der Anteil an „Enthaltungen“ systematisch je nach der differentiellen Nähe der Befragten zu der angesprochenen Problematik variiert – bei politischen Fragen differenzieren vor allem Geschlecht, Bildungsniveau und Klassenzugehörigkeit (Bourdieu 2013a, 2013b; siehe auch Bergström 2012). In Übereinstimmung mit Bourdieus Thesen zeigt Laurison (2015), dass Amerikanerinnen und Amerikaner in unteren Einkommensschichten auch unter Konstanthaltung von technischer Kompetenz und Einstellung zur Befragung eine höhere Wahrscheinlichkeit haben, auf politische Fragen mit „weiß nicht“ zu antworten. Die Wahrscheinlichkeit, eine Meinung zu einem bestimmten Problem zu haben, ist dementsprechend eine Funktion des Interesses an dem spezifischen Problemkomplex (womit auch das Verständnis der Frage einhergehen kann) sowie des Gefühls der Legitimität und Kompetenz, seine Meinung zu äußern.

Die Rate der ausgedrückten Meinungslosigkeit in Umfragen variiert nicht nur zwischen verschiedenen Gruppen, sondern auch zwischen Ländern (Reuband 1990; Sicinski 1970), was ebenso wie Variationen in Teilnahme- bzw. Rücklaufzeiten (Couper & de Leeuw 2003; Johnson et al. 2002) auf unterschiedliche Strategien und Gewohnheiten im Umgang mit Befragungen hindeutet. Bourdieu weist in diesem Zusammenhang darauf hin, dass nicht nur die manifesten Antwortstrategien, sondern auch die unterliegenden Prinzipien der Meinungsproduktion abhängig vom sozialen Hintergrund der Befragten sind. So gibt es je nach sozialer Position große Unterschiede in der Feinheit der Wahrnehmung politischer Unterschiede, wobei untere Klassen dazu tendieren, politische Probleme als ethische Fragen wahrzunehmen und entsprechend zu beantworten (Bourdieu 2013a:247-254). Die unterschiedlichen Problemkontexte, auf die sich die Befragten in ihren Antworten beziehen, werden jedoch in der Auswertung nicht beachtet. In ähnlicher Weise stellte Philip Converse bereits 1964 fest, dass nur auf der Ebene der politischen Elite zusammenhängende, abstrakte ideologische Prinzipien vorherrschen, während die Behandlung politischer Probleme am anderen Ende der „Informationsdimension“ auf konkreten,

alltagsbezogenen Überlegungen fußt, die in keinem übergreifenden Denksystem organisiert sind. Daraus ergibt sich auch, dass die zeitliche Stabilität geäußerter Meinungen der „Massenöffentlichkeit“ umso geringer ist, je weniger das jeweilige Thema bzw. Einstellungsobjekt einen direkten Bezug zur Alltagswelt der Befragten hat.

Problematisch bei der Erhebung der „öffentlichen Meinung“ mittels standardisierter Befragungen ist somit auch, dass die *Auswahl* der Fragen von den Forscherinnen und Forschern bzw. den Auftraggebenden getroffen wird. Daher ist es nicht unüblich, dass eine Meinungsumfrage „die Menschen unter Zugzwang setzt, auf eine Frage zu antworten, die sie sich nicht gestellt haben“ (Bourdieu 2013a:246). Die Erhebung und Veröffentlichung von Meinungsumfragen suggeriert also, es bestehe ein Konsens über die Probleme, über das, was „fragwürdig“ ist. Damit werden die Machtverhältnisse, die den Diskurs prägen, verschleiert: „in realen Situationen sind Meinungen Mächte und Meinungsverhältnisse Machtkonflikte zwischen sozialen Gruppen“ (Bourdieu 2013a:252). Die Kumulation der erfassten Meinungen unter der Annahme ihrer Gleichwertigkeit repräsentiere somit nicht die „öffentliche Meinung“, sondern stelle ein sinnfreies Artefakt dar³ (siehe auch Blumer 1948). Die Fähigkeit politischer Umfragen, Wahlergebnisse vorauszusagen, ist dieser Auffassung entsprechend kein Validitätskriterium, sondern basiert auf der „shared artificiality“ der Verfahren (Perrin & McFarland 2011:89).

2.4 Zwischenfazit: Kontextabhängigkeit des Antwortprozesses

Zusammenfassend lässt sich konstatieren, dass der Antwortprozess in standardisierten Befragungen auf mehreren Ebenen kontextabhängig ist. Zunächst liefert der unmittelbare Fragekontext (vorhergehende Fragen, Einleitung der Frage, Antwortoptionen etc.) den Befragten Hinweise zur Interpretation und Beantwortung. Die Befragungssituation (Raum, Zeit, emotionale Verfasstheit, ggf. Charakteristika und Verhalten der Interviewenden) hält weitere Kontextinformationen bereit, die bei der Erschließung der Fragenbedeutung und der Verfügbarkeit relevanter Assoziationen eine Rolle spielen können. Weiterhin führen persönliche Erfahrungen, Alltagswissen und Einstellungen der Befragten nicht nur zu den – erwünschten – inhaltlich unterschiedlichen Antworten, sondern bedingen auch die Reaktion auf die Befragungssituation an sich. Der soziale Hintergrund der Befragten hat Auswirkungen darauf, welche Prinzipien die Meinungsbildung bestimmen und ob überhaupt eine – den Vorgaben der

³ Die Diskussion, was „öffentliche Meinung“ ist und in welcher Weise, wenn überhaupt, diese zu erfassen wäre kann im Rahmen dieser Arbeit nicht umfassend diskutiert werden. Siehe zu diesem Thema u.a. Binkley (1928), Blondiaux (1998), Blumer (1948), Clark (1933), Converse (1987), Herbst (1998) Lazarsfeld (1957), Wilson (2013[1962]).

Befragung gemäß kommunizierbare – Meinung vorliegt. Die Signifikanz kultureller bzw. gesellschaftlicher Kontexte zeigt sich einerseits in differentiellen Narrativen und Bedeutungshorizonten, die beeinflussen wie Fragen verstanden werden und inwiefern Einstellungen zu bestimmte Konzepten zugänglich und aktivierbar sind (Braun 2006:55; Verba 1969:69). Andererseits wirken sie sich in Form sozialer Normen auf die „Surveytauglichkeit“, Kooperationsbereitschaft und Kommunikationsverhalten der Befragten aus. Es kann auch zu Interaktionseffekten zwischen verschiedenen Kontextebenen kommen, etwa wenn sich die Richtung von Kontexteffekten durch das Instrument je nach dem kulturellen Hintergrund der Befragten unterscheidet (Braun 2003; Schwarz 2003).

Eine große Herausforderung bei der Messung von Einstellungen und Meinungen besteht also darin, vor dem Hintergrund unterschiedlicher Kontexte Vergleichbarkeit bzw. Äquivalenz zwischen den Antworten unterschiedlicher Individuen, Gruppen oder Nationen sicherzustellen. Während auf der einen Seite die Standardisierung der Instrumente und Abläufe dazu dienen soll, Verzerrungen durch externe Einflüsse wie beispielsweise Unterschiede in der Itemformulierung, Interviewerverhalten oder institutionelle Praktiken weitgehend auszuschließen⁴, kann auf der anderen Seite gerade ein hoher Grad an Standardisierung dazu führen, dass Unterschiede im Fragenverständnis und dem Umgang mit der Befragung als solcher in den erhobenen Daten nicht auf den ersten Blick erkennbar sind. Wenn die gemessenen Einstellungen aber nicht nur aufgrund inhaltlicher Unterschiede variieren, sondern auch durch unterschiedliche „Bedeutungshintergründe“ und Antwortqualitäten, kann die Nichtbeachtung letzterer in empirischen Analysen zu Fehlinterpretationen führen. Wie genau äußern sich derartige Effekte bei der Messung von Einstellungen? Welche Möglichkeiten der Entdeckung und Analyse gibt es? Diese Fragen werden in den Papieren, die den Kern der vorliegenden kumulativen Dissertation bilden, anhand von Beispielen verhandelt. Das folgende Kapitel bietet einen Überblick über mögliche Folgen unterschiedlicher Kontexte bei der Messung von Einstellungen. In den so entwickelten Rahmen werden die Forschungsergebnisse der Papiere eingeordnet und erläutert.

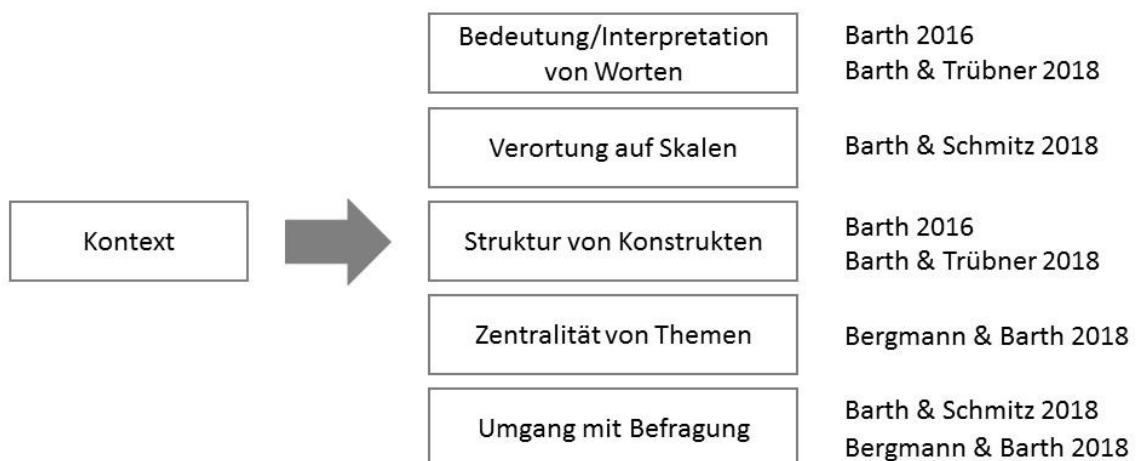
⁴ Ob diese Standards in der Praxis eingehalten werden, ist eine andere Frage. Zur Prüfung auf Interviewfälschungen und institutionelles Fehlverhalten mit ähnlichen Methoden wie den hier vorgeschlagenen siehe z.B. Blasius & Thiessen 2012; 2015; Bredl, Winker & Kötschau 2012; Winker, Menold & Porst 2013.

3. Vergleichbarkeit von Einstellungsmessungen?

Die Kontextabhängigkeit des Antwortprozesses in standardisierten Befragungen kann sich bei der Einstellungsmessung in unterschiedlichen Effekten niederschlagen. Unter Rückgriff auf empirische Arbeiten zur Vergleichbarkeit von Einstellungsmessungen werden im Folgenden mögliche Folgen unterschiedlicher Kontexteinflüsse erläutert (siehe Abbildung 3):

Zum ersten können Worte bzw. Sätze oder Satzteile je nach Kontext der Befragung bzw. der Befragten unterschiedlich interpretiert werden. Zudem verorten sich Befragte nicht nur gemäß ihrer substantiellen Einstellung auf Skalen, sondern können unterschiedliche Vorstellungen von der Ausdehnung bzw. den Endpunkten einer Skala haben, nur einen begrenzten Bereich der Skala nutzen oder unterschiedliche Vergleichsstandards anlegen. Darüber hinaus kann sich die Struktur von latenten, mithilfe mehrerer manifester Items gemessenen Einstellungen über Gruppen oder im Zeitvergleich unterscheiden, was in der Konsequenz bedeuten kann, dass – im übertragenen Sinne – Äpfel mit Birnen verglichen werden. Schließlich sind je nach subjektiver Vertrautheit und Zentralität bestimmter Themen die Einstellungen der Befragten stärker oder schwächer, was differentielle Antwortqualitäten zur Folge haben kann. Zuletzt können auch generalisierte Einstellungen gegenüber Befragungen zu verschiedenen Reaktionen, und damit Qualitätsunterschieden in den Antworten, führen. Diese Aspekte werden jeweils in einem oder mehreren der dieser Dissertation zugrundeliegenden Artikel aufgegriffen, wie die Übersicht in Abbildung 3 verdeutlicht. Es kann auch zu einer Interaktion der Effekte kommen, etwa wenn unterschiedliche Wortbedeutungen die Struktur der zugehörigen Konstrukte beeinflussen, oder wenn die Einstellungstärke der Befragten gegenüber bestimmten Themen in Relation zu ihrem Umgang mit der Befragung im Allgemeinen steht.

Abbildung 3: Verortung der Papiere im Forschungszusammenhang



Die Herausforderungen stellen sich in ähnlicher Weise in Studien, die über Zeit (Längsschnitt- oder wiederholte Querschnittsbefragungen) und über Gruppen (soziale, kulturelle, nationale etc.) vergleichen, weshalb jedes der folgenden Unterkapitel beide Vergleichsmöglichkeiten mit einem oder mehreren Beispielen abdeckt.

3.1 Methoden

Die in den vier Artikeln verwendeten Daten stammen aus großangelegten und wissenschaftlich streng überprüften sozialwissenschaftlichen Erhebungen, in denen prinzipiell von einer sehr guten Datenqualität auszugehen ist. Allerdings nimmt die Frage nach unterschiedlichen Antwortqualitäten bislang selten eine prominente Stellung in großen quantitativen Erhebungen ein. Die Verwendung von Sekundärdaten impliziert somit, dass in der Regel keine oder wenig Variablen vorhanden sind, die unterschiedliche Bedeutungshintergründe oder Reaktionen auf die Befragung als solche direkt abbilden. Daher wird in der vorliegenden Arbeit hauptsächlich auf explorative multivariate Verfahren zurückgegriffen, die mittels der Darstellung von Zusammenhängen zwischen verschiedenen Einstellungsvariablen Rückschlüsse auf das Verhältnis zwischen inhaltlicher und methodischer Variation erlauben. Insbesondere sind dies multiple Korrespondenzanalyse (Multiple Correspondence Analysis, im weiteren MCA) und Analyse latenter Klassen (Latent Class Analysis, im weiteren LCA).

Die MCA ist ein exploratives multivariates Skalierungsverfahren, das vorwiegend zur Entdeckung latenter Strukturen in umfangreichen Kontingenztabelle verwendet wird. Aus einer Vielzahl manifester kategorialer Variablen werden die wichtigsten Dimensionen extrahiert. Das Verfahren wird daher häufig als Hauptkomponentenanalyse mit kategorialen Daten beschrieben (Benzécri et col. 1973; Blasius 2001:6). Wie die Hauptkomponentenanalyse bringt die MCA eine numerische Lösung mit Faktorladungen und Eigenwerten hervor. Das Hauptaugenmerk der MCA liegt aber meist auf der Visualisierung der Dimensionen, die als Achsen einen niedrigdimensionalen euklidischen Raum aufspannen. In diesem Koordinatensystem werden Variablenausprägungen und Individuen bzw. Objekte verortet, wobei die relativen Distanzen dabei als Maßstab für die Ähnlichkeit bzw. Unähnlichkeit der Antwortmuster fungieren: Kategorien, die häufig gemeinsam auftreten, sind in dem von der MCA aufgespannten Raum nahe beieinander, und umgekehrt. Das Verfahren ist damit zentraler Teil des Methodenarsenals der geometrischen Datenanalyse (Greenacre & Blasius 2006; Le Roux & Rouanet 2004).

Die LCA ist ein multivariates Klassifikationsverfahren, bei dem die Ausprägungen mehrerer manifester Variablen einer oder mehreren latenten Variablen zugeordnet werden. Im Fall der LCA

haben die latenten Variablen kategoriales Skalenniveau, daher wird von latenten Klassen gesprochen. Die latenten Klassen fassen Individuen mit ähnlichen Antwortmustern zusammen, wobei die Zuordnung von Individuen zu Klassen auf Wahrscheinlichkeiten beruht. Die Passung des Modells zu den Daten kann mithilfe statistischer Kennwerte, allen voran Chi-Quadrat-Tests, Likelihood-Ratio-Tests und informationstheoretische Maße (z.B. AIC, BIC, CAIC), getestet werden. Als modellbasiertes Klassifikationsverfahren lässt sich die LCA der Familie finiter Mischverteilungsmodelle zuordnen (Vermunt & Magidson 2004).

Mit einem Paneldatenmodell mit fixen Effekten (FE-Regression) findet bei Bergmann und Barth (2018) auch ein hypothesenprüfendes Verfahren Verwendung. Die Besonderheit der FE-Regression ist, dass personenspezifische Heterogenität eliminiert wird, indem bei der Modellschätzung der individuenspezifische Mittelwert jeder Variable von allen Ausprägungen derselben subtrahiert wird. Damit sind nur noch intra-individuelle Veränderungen zu sehen (Giesselmann & Windzio 2012). Im vorliegenden Fall ermöglicht die Modellierung, den Effekt vorangegangener Befragungen auf Einstellungsstärke und Antwortqualität zu isolieren.

3.2 Die Bedeutung und Interpretation von Begriffen

Survey-Items sollten interindividuell gleich verstanden werden, um die Antworten adäquat vergleichen zu können. Die Forschung zeigt jedoch, dass ein und derselbe Begriff bzw. Satz je nach dem Hintergrund der Befragten unterschiedlich aufgefasst werden kann (Warner & Hoffmeyer-Zlotnik 2009). Während in multilingualen Kontexten seit langem ein Problembewusstsein für mögliche Bedeutungsunterschiede besteht (Brislin 1986; Harkness 2003; Harkness, Villar & Edwards 2010), wird dies bei Befragungen im gleichen Sprachraum weit seltener thematisiert. Auch dort können jedoch unterschiedliche soziale und institutionelle Kontexte dazu führen, dass Fragen differenziell verstanden werden. Ein Beispiel, das die Probleme unterschiedlicher Bedeutungskontexte sowohl im Zeit- als auch im Gruppenvergleich illustriert, ist die Messung von Einstellungen gegenüber Geschlechterrollen. Geschlechterrollen sind sowohl starkem sozialen Wandel unterworfen als auch regional durch unterschiedliche institutionelle, soziale und normative Rahmenbedingungen geprägt. Eine der populärsten Itembatterien zur Messung von Einstellungen gegenüber Geschlechterrollen fand und findet in teils leicht abgewandelter Form unter anderem in GSS, ISSP, BHPS und dessen Nachfolger Understanding Society, BSA und ALLBUS Verwendung (Braun 1998; Braun 2006; Walter 2018). Das Instrument enthält unter anderem das Item „Alles in allem: Das Familienleben leidet darunter, wenn die Frau voll berufstätig ist“ (*Familie leidet*), mit dem die Wahrnehmung negativer Folgen der weiblichen Berufstätigkeit für die Familie erfasst werden

soll. Bei der Formulierung bleibt offen, inwieweit der Vater berufstätig ist und in welchem Alter die betroffenen Kinder sind. Diese Informationen werden von den Befragten jedoch häufig implizit ergänzt, wobei eine Orientierung an gesellschaftlichen Rahmenbedingungen erfolgt (Braun 2003; Braun & Harkness 2005).

Braun (2003:62-63) berichtet die Ergebnisse eines Experiments aus dem Jahr 1998, welches das unterschiedliche Funktionieren des Items in West- und Ostdeutschland zeigt. Um die impliziten Bedeutungshintergründe offenzulegen wurden die Befragten zunächst gebeten, einzuschätzen inwieweit ein dreijähriges Kind leidet, wenn beide Eltern Vollzeit berufstätig sind, die Mutter Vollzeit und der Vater Teilzeit, oder die Mutter Vollzeit und der Vater nicht arbeitet. Die Antworten wurden in Abhängigkeit von der Antwort auf das ISSP-Item *Familie leidet* ausgewertet. In der Bewertung der spezifischen Situation des dreijährigen Kindes zeigten sich große Unterschiede zwischen Befragten in Ost- und Westdeutschland, die bei *Familie leidet* ähnliche Kategorien gewählt hatten. Befragte in Westdeutschland, die die Aussage zum Leiden der Familie betont ablehnten, gaben mehrheitlich an, dass ein dreijähriges Kind leidet, wenn beide Eltern Vollzeit arbeiten. Sie nahmen bei der Beantwortung des Items *Familie leidet* also offenbar an, dass der Mann in diesem Fall nicht oder nur in Teilzeit erwerbstätig ist oder dass die Kinder bereits älter sind. Zusätzlich vermutet Braun, dass die Spezifikation, dass die Mutter Vollzeit arbeitet, von einigen Befragten ignoriert wurde, weil sie zwar ihre nicht-traditionale Einstellung betonen wollten (und das Item daher ablehnten), sich aber aufgrund der gesellschaftlichen Rahmenbedingungen nicht vorstellen konnten, dass die Vollzeit-Berufstätigkeit beider Eltern und Kinderbetreuung kompatibel seien. Für ostdeutsche Befragte schien hingegen die Berufstätigkeit des Vaters nicht von Belang zu sein, insgesamt wurde die Situation des dreijährigen Kindes deutlich weniger problematisch bewertet. Die unterschiedlichen sozialen und institutionellen Kontextbedingungen, die in der eher von einer traditionellen Geschlechterrollenaufteilung geprägten westdeutschen Gesellschaft im Gegensatz zum ehemals sozialistischen Ostdeutschland herrschen (Pfau-Effinger & Smidt 2011) führen somit dazu, dass eine gleich gestellte Frage unterschiedlich verstanden wird.

Während die Bedeutungsunterschiede in diesem Fall durch den Vergleich mit zusätzlich erhobenen präzisierten Einstellungsmessungen ermittelt wurden, liegen solche weiterführenden Informationen in der Regel nicht vor. Jedoch lassen auch multivariate Analyseverfahren, die Einstellungsstrukturen in verschiedenen Gruppen bzw. zu verschiedenen Zeitpunkten darstellen, erste Rückschlüsse auf unterschiedliche Interpretationen von Items zu. So stellen Barth & Trübner (2018) bei einer Analyse der Geschlechterrollen-Erwartungen in Deutschland mittels LCA fest, dass das Item „Für eine Frau ist es wichtiger, ihrem Mann bei seiner Karriere zu helfen, als selbst Karriere zu machen“, das eine traditionell geprägte Rollenerwartung symbolisiert, in bestimmten

Gruppen wenig Differenzierungskraft aufweist. So liegt die Wahrscheinlichkeit, diese Aussage abzulehnen, in der traditionellsten Subgruppe bei fast 40%. Da diese Gruppe auch das niedrigste Einkommens- und Bildungsniveau aufweist, vermuten die Autorinnen, dass für einige Befragte der Begriff der Karriere wenig alltägliche Relevanz hat, woraus sich inkonsistente Antwortmuster ergeben.

Barth (2016) beobachtet bei der Analyse der Geschlechterrollen-Itematterie in Großbritannien zwischen Anfang der 1990er und Mitte der 2000er Jahre, dass der Varianzanteil von Methodenartefakten steigt: in der Visualisierung der Zusammenhänge zwischen den Items mittels MCA gewinnt die Dimension, die zwischen extremen und moderaten Antwortkategorien differenziert, an Erklärungskraft. Dieser Effekt zeigt sich sowohl in einer Panelstudie (BHPS) als auch in einer wiederholten Querschnittsbefragung (BSA). Daher wird angenommen, dass diese Veränderung darauf hinweist, dass die Befragten zunehmend Schwierigkeiten haben, sich inhaltlich sinnvoll zu verorten. Vor dem Hintergrund der gesellschaftlichen Veränderungen im Untersuchungszeitraum, etwa der zunehmenden Arbeitsmarktpartizipation von Frauen, scheint eine abnehmende Passung zwischen der Alltagswelt der Befragten und den größtenteils in den 1960er Jahren entstandenen Items (Braun 1998; 2006; Walter 2018) wahrscheinlich. In diesem Fall stellt dann weniger die semantische als die pragmatische Bedeutung das Problem dar – die zur Interpretation der Items herangezogenen kontextuellen Bedingungen unterscheiden sich und erschweren damit unmittelbare Einstellungsvergleiche.

Während in der kulturvergleichenden Forschung die Erkenntnis, dass die Wahrung der pragmatischen Bedeutung wichtiger ist als eine möglichst genaue Replikation bzw. Übersetzung der semantischen Struktur, seit längerem etabliert ist (Brislin 1986; Braun & Harkness 2005), wird in Analysen zum Einstellungswandel, die sich auf wortgetreu replizierte Items stützen, selten überprüft ob die Bedeutung der verwendeten Formulierungen gleich bleibt (Baur 2004; Riordan et al. 2001). Beim Zeitvergleich besteht die Schwierigkeit, dass Methoden zur direkten Überprüfung der Interpretation wie *probing* (Behr et al. 2012; Oksenberg, Cannell & Kalton 1991) oder kognitive Interviews (Beatty & Willis 2007; Willis 2004) an früheren Messzeitpunkten nicht nachträglich implementiert werden können und dass (zusätzliche) Referenzitems oder externe Validierungskriterien (Braun 2006) häufig nicht verfügbar sind. In diesen Fällen sind daher strukturen-entdeckende bzw. -vergleichende Verfahren wie Faktoranalyse, MCA oder LCA ein sinnvolles Hilfsmittel, um Unstimmigkeiten zu ermitteln.

3.3 Die Verortung auf Skalen

Ein weiteres Vergleichbarkeitsproblem entsteht, wenn Befragte Antwortoptionen auf unterschiedliche Weise wahrnehmen und nutzen. Insbesondere bei der weit verbreiteten Messung von Einstellungen mithilfe mehrerer Likert-skaliertes Items kann es zu einer Reihe von methodeninduzierten Effekten kommen. Seit langem bekannt sind in dieser Hinsicht *response styles*: Antworttendenzen, die unabhängig vom Inhalt der Frage bestehen (Cronbach 1946; Paulhus 1991). Obwohl in der Literatur keine Einigkeit über die Anzahl möglicher Antworttendenzen besteht, lässt sich feststellen, dass die meist zitierten Akquieszenz (bzw. Disakquieszenz), die Wahl von Extremkategorien oder der Mittelkategorie und inkonsistentes bzw. zufälliges Antworten sind (Baumgartner & Steenkamp 2001; Blasius & Thiessen 2012; Van Vaerenbergh & Thomas 2013). Zu den populärsten Erklärungsansätzen für das Auftreten von *response styles* gehören neben situationalen Faktoren wie der Beschaffenheit des Instruments (Kieruj & Moors 2010; 2013; Weijters, Cabooter & Schillewaert 2010) oder der Erhebungsmethode (Weijters, Schillewaert & Geuens 2008) Charakteristika der Befragten, allen voran Persönlichkeitsmerkmale (He & Van de Vijver 2013; Kieruj & Moors 2013; Knowles & Nathan 1997) und kognitive Fähigkeiten (Lechner & Rammstedt 2015; Light, Zax & Gardiner 1965; Weijters, Geuens & Schillewaert 2010), aber auch die soziale Position (Meisenberg & Williams 2008; Ross & Mirowsky 1984) und der ethnische bzw. kulturelle Hintergrund der Befragten (Baron-Epel et al. 2010; Johnson et al. 2005; Smith 2004; Weech-Maldonado et al. 2008).

Auf diesen Ergebnissen aufbauend stellen Barth und Schmitz (2018) fest, dass bestimmte Antworttendenzen bei der Beantwortung (politischer) Einstellungsfragen auch mit der ideologischen Disposition der Befragten zusammenhängen. Eine auf dem US-amerikanischen GSS (2010-2014) basierende LCA verschiedener *response-style*-Indikatoren ergibt 10 Klassen, die unterschiedliche Ausprägungen und Kombinationen von Antworttendenzen darstellen: Während die Mehrheit der Befragten keine oder nur leichte Antwortverzerrungen aufweist, zeigen sich mehrere kleinere Klassen, die durch deutliche Antworttendenzen einzeln oder in Kombination auffallen. So gibt es jeweils eine Klasse, die sich durch Tendenz zur Mitte, zu Extremen und zu Akquieszenz auszeichnet, weiterhin treten die Kombinationen *midpoint-responding* und Disakquieszenz sowie *extreme responding* und Akquieszenz auf. Die Ergebnisse weisen damit auf die bisher wenig beachtete Möglichkeit unterschiedlicher Antworttendenz-Kombinationen in spezifischen Bevölkerungsgruppen hin, was für die Verwendung multipler Indikatoren sowie eine relationale Modellierung spricht.

Eine Verortung der Antworttendenz-Klassen im mittels MCA konstruierten US-amerikanischen „Feld der ideologischen Konsumption“⁵ zeigt sodann, dass bestimmte Arten der Verortung auf Skalen mit der ideologischen Disposition der Befragten zusammenhängen. So sind etwa Akquieszenz und die Tendenz zu Extremen mit traditionellen Wertvorstellungen sowie der Zustimmung zu wohlfahrtsstaatlichen Maßnahmen verbunden. Auf der anderen Seite ist die Tendenz zur Disakquieszenz mit liberalen bzw. libertären Werthaltungen verknüpft. Aus der Analyse geht demnach hervor, dass substantielle Einstellungsunterschiede mit einem unterschiedlichen Umgang mit Antwortskalen einhergehen, was zu einer verzerrten Repräsentation bestimmter Bereiche des ideologischen Feldes führen kann.

Für die Messung von Antworttendenzen gibt es eine Reihe empirischer Ansätze, die vom Zählen verwendeter Antwortkategorien in einer Reihe von Items heterogenen Inhalts und einer darauf beruhenden Indikatorenbildung (Baumgartner & Steenkamp 2001; Reynolds & Smith 2010) über das Hinzufügen spezieller „Prüf-Items“ zum Fragebogen (Couch & Keniston 1960; Weijters et al. 2008) bis hin zu komplexen Modellierungen, etwa durch einen oder mehrere Methoden-Faktoren in konfirmatorischen Faktoranalysen (Billiet & McClendon 2000) bzw. Latent-Class Faktoranalysen (Kieruj & Moors 2013; Moors 2012) oder speziellen Modellen wie *latent-class bilinear multinomial logit model* (Van Rosmalen, Van Herk & Groenen 2010), *multidimensional ordinal item response theory model* (De Jong & Steenkamp 2010), *constrained dual scaling* (Schoonees, van de Velden & Groenen 2015) und *calibrated sigma method* (Weijters, Baumgartner & Geuens 2016) reichen. Insbesondere letztere zeichnen sich dadurch aus, dass mit der Modellierung auch direkt die Korrektur von möglichen Verzerrungen durch Antworttendenzen erfolgt.

Auf andere Art stellt sich die Verortung auf Skalen als problematisch dar, wenn sich die Vergleichsstandards der Befragten unterscheiden bzw. ändern. Bei der Erfassung von Einstellungswandel im Zeitverlauf wird dieses Phänomen unter anderem als *response shift* (Rapkin & Schwartz 2004; Schwartz et al. 2006) oder als *beta change* (Golembiewski, Yeager & Billingsley 1976; Riordan et al. 2001) bezeichnet. Die Bezeichnung *response shift* findet aktuell vor allem im Bereich der gesundheitsbezogenen Lebensqualität Verwendung (Schwartz et al. 2006; Sprangers & Schwartz 1999). Dort wurde wiederholt festgestellt, dass sich die mit standardisierten Befragungsinstrumenten gemessene Lebensqualität von Patienten mit schweren chronischen

⁵ Die Bezeichnung „Feld der ideologischen Konsumption“ ist aus der Bourdieu’schen Terminologie abgeleitet. Bourdieu spricht in Bezug auf Umfrageinstitute, politische Akteurinnen und Akteure sowie Medien vom *Feld der Ideologieproduktion*, in dem „das verfügbare begriffliche Instrumentarium zur Erkenntnis der sozialen Welt“ erarbeitet wird (Bourdieu 1987:623). Die in der Meinungsforschung Befragten werden mit den aus diesem Feld hervorgegangenen Diskursen konfrontiert, während sie selbst die Begrifflichkeiten und Denkschemata nicht aktiv mitgestalten, sondern lediglich im Rahmen ihrer Möglichkeiten bewerten. In diesem Sinne können sie als Konsumentinnen und Konsumenten der ideologischen Dispositionen verstanden werden.

Krankheiten kaum von der weniger schwer erkrankter bzw. gesunder Probandinnen und Probanden unterscheidet (Breetvelt & Van Dam 1991; Stensman 1985). Dies wird auf eine Veränderung der subjektiven Bewertungsmaßstäbe und Prioritäten zurückgeführt, die im Krankheitsverlauf oder durch bestimmte Interventionen bzw. Lebensereignisse erfolgt (Rapkin & Schwartz 2004; Schwartz et al. 2006).

Aus der Organisationsforschung stammt die Bezeichnung einer subjektiven Re-Kalibrierung von Skalen nach einer Intervention als *beta change* – der Begriff wird dort in Abgrenzung von *alpha change* (der absoluten quantitativen Veränderung einer Variable auf einer konstanten Skala) verwendet (Golembiewski et al. 1976; Riordan et al. 2001). Empirisch findet in der *response shift*-Forschung vor allem der *then-test*, also der Vergleich einer retrospektiven Einschätzung der Lebensqualität mit dem tatsächlichen Messwert zum damaligen Zeitpunkt, Verwendung (Schwartz et al. 2006). Dieses Verfahren wurde auch in Bezug auf *beta change* vorgeschlagen (Terborg, Howard & Maxwell 1980), verbreiteter sind in diesem Bereich jedoch konfirmatorische Faktorenanalyse und latente Wachstumskurvenmodelle (Riordan et al. 2001). Signifikante Veränderungen der Faktorvarianzen oder Faktorladungen weisen auf *beta change* hin (Schmitt 1982; Vandenberg & Self 1993). Somit kann die Veränderung der Skalenverortung auch in das Konzept der Messäquivalenz (Horn & McArdle 1992; Vandenberg & Lance 2000) eingeordnet werden: die als *beta change* klassifizierte Veränderung von Faktorvarianzen bzw. -ladungen lässt sich mit dem Test auf *metric invariance* vergleichen (Davidov et al. 2014; Steenkamp & Baumgartner 1998).

3.4 Die Struktur des Konstrukts

Eine noch schwerwiegendere Einschränkung des Einstellungsvergleichs als differentielle Skalennutzung stellen strukturelle Unterschiede des gemessenen Konstrukts über Zeit bzw. Gruppen dar. Bei der Prüfung auf Messäquivalenz steht die Frage, ob die Struktur eines Konstrukts sich im Gruppenvergleich als äquivalent erweist, ob also *configural equivalence* (Steenkamp & Baumgartner 1998; Vandenberg & Lance 2000) vorliegt, am Anfang jeder Testreihe. Im Allgemeinen gehen Verfahren zur Prüfung von Messäquivalenz, sei es im Rahmen konfirmatorischer Faktorenanalyse multipler Gruppen (MGCFA), *item response theory* (IRT) oder LCA, von der Existenz latenter Variablen aus, die mit mehreren manifesten Indikatoren gemessen werden (Davidov et al. 2014; Kankaraš, Vermunt & Moors 2011). Wird ein metrisches Skalenniveau der latenten Variable angenommen, bedeutet *configural equivalence*, dass die Faktorstruktur gleich ist; bei kategorialem Skalenniveau ist die gleiche Anzahl an Kategorien der latenten Variable entscheidend (Kankaraš, Moors & Vermunt 2011). Trifft diese Voraussetzung

nicht zu, wird nicht das gleiche latente Konstrukt gemessen und weitere Äquivalenzttests, geschweige denn der Vergleich von Zusammenhängen oder Mittelwerten, sind nicht sinnvoll. In der oben vorgestellten Klassifikation des Einstellungswandels wird die Veränderung von Konstrukten als *gamma change* bezeichnet. Golembiewski et al. (1976:138) verstehen darunter „a quantum shift in ways of conceptualizing salient dimensions of reality“.

Barth & Trübner (2018) argumentieren, bezogen auf Geschlechterrollen-Einstellungen in Deutschland, dass eine unterschiedliche Kombination und Struktur von Rollenerwartungen in verschiedenen Bevölkerungsgruppen zu erwarten ist, die sich nicht auf einem eindimensionalen Kontinuum von „traditional“ nach „egalitär“ verorten lässt. Aufgrund der unterschiedlichen historischen und institutionellen Rahmenbedingungen in Ost- und Westdeutschland und der gesellschaftlichen Veränderungen und politischen Maßnahmen im Untersuchungszeitraum (1990-2012) ist zudem fraglich, ob die Struktur des Konstrukts (in diesem Fall die Klassenprofile) in Ost und West sowie über Zeit äquivalent ist. Die Autorinnen modellieren das Konstrukt daher mit LCA und testen auf Messäquivalenz zwischen Ost- und Westdeutschland sowie über Zeit. Die Analyse ergibt fünf Klassen, die unterschiedliche Mischverhältnisse von Rollenerwartungen gegenüber Frauen in ihren Rollen als (berufstätige) Mutter und Ehefrau in variierender Intensität abbilden. Die Klassen erweisen sich als strukturell stabil. Da somit die Voraussetzung der regionalen und zeitlichen Messäquivalenz erfüllt ist, kann die Größe der Klassen sinnvoll verglichen werden.

Barth (2016) zeigt mittels MCA, dass sich die Struktur des in standardisierten Befragungen gemessenen Konstrukts „Geschlechterrolleneinstellung“ von Anfang 1990 bis Mitte 2000 in Großbritannien gewandelt hat. Während das Konstrukt am Anfang des Messzeitraums eindeutig eindimensional ist, wird die Struktur im Laufe der Zeit komplexer und methoden-induzierte Varianz gewinnt an Bedeutung. Die Tatsache, dass die beschriebenen Effekte sowohl in einer Längsschnitt- als auch in einer wiederholten Querschnittstudie auftreten (BHIPS und BSA) deutet auf einen Periodeneffekt als Ursache hin. Das heißt, Veränderungen des sozialen, kulturellen und gesellschaftlichen Kontexts beeinflussen, wie die Items verstanden und beantwortet werden und wie einzelne Aspekte geschlechtsspezifischer Rollen mit dem von der Itematterie abgebildeten Gesamtkonzept in Zusammenhang stehen. Es wird vermutet, dass die kognitive Repräsentation des Konzepts komplexer wird, beispielsweise indem unterschiedliche Rollenerwartungen für Ehefrauen und Mütter gelten (vgl. dazu Barth & Trübner 2018). Gleichzeitig sind stärkere Tendenzen zum *straightlining* und zur Wahl von Mittelkategorien zu beobachten, was darauf hindeutet, dass es für etliche Befragte zunehmend schwieriger wird, sich bei den Items inhaltlich sinnvoll zu verorten. Die Analyse in Barth (2016) verdeutlicht somit auch, dass die in der

vorliegenden Arbeit vorgenommene Differenzierung zwischen Bedeutungsunterschieden, Skalennutzung und Strukturänderung des latenten Konstrukts auf empirischer Ebene schwierig sein kann, da eine Interaktion der Aspekte wahrscheinlich ist.

Ähnlich verhält es sich in einer Studie von Blasius und Thiessen (2006), die zum Testen der Vergleichbarkeit und Datenqualität der Geschlechterrollen-Itematterie im ISSP 1994 in 24 Ländern ebenfalls die MCA anwenden. Sie identifizieren fünf Länder-Cluster, in denen das latente Konstrukt jeweils eine ähnliche Struktur aufweist, die sich jedoch von den anderen Clustern deutlich unterscheidet. Während es einerseits Länder mit relativ guter Datenqualität gibt, in denen die Struktur entweder ein- oder zweidimensional ist, wird andererseits auch ein Cluster beschrieben, in dem der größte Varianzanteil durch die Unterscheidung zwischen extremen und moderaten Antwortkategorien erklärt wird und erst die zweite Dimension inhaltlich interpretierbar ist, sowie Fälle in denen die ersten beiden Dimensionen ausschließlich durch methodische Artefakte bestimmt sind. Zusätzlich lässt die Anordnung der Antwortkategorien im zweidimensionalen Raum Rückschlüsse auf mögliche Übersetzungs- bzw. Datenerhebungsprobleme zu, wenn die zu erwartende ordinale Reihenfolge von „stimme voll zu“ bis „stimme gar nicht zu“ nicht zutrifft.

Die MCA erweist sich in diesem Sinne als gut geeignet für das Screening kategorialer Daten auf vergleichbare Konstruktstruktur und Datenqualität. Wie bei allen statistischen Verfahren zur Prüfung auf Messäquivalenz gilt jedoch auch hier, dass zwar die Äquivalenz bzw. deren Fehlen festgestellt werden kann, zur Erklärung der Ursachen inäquivalenter Messungen aber weitergehende Verfahren bzw. zusätzliche Daten benötigt werden. Wird vermutet, dass das unterschiedliche Funktionieren von Items mit gesellschaftlichen Kontextvariablen zusammenhängt, können diese beispielsweise in einem Multilevel-Modell als Prädiktoren integriert werden (Davidov et al. 2012; 2018). Ein anderer Ansatz ist die Nutzung von Mixed Methods: so verwenden Lugtig, Boeije und Lensvelt-Mulders (2012) Informationen aus qualitativen Interviews mit Studierenden, um die in einer quantitativen Studie festgestellte Inäquivalenz des Konzepts „Studienmotivation“ zu Studienbeginn und sechs Monate später zu erklären.

3.5 Die Zentralität von Themen

Wie weiter oben ausgeführt, ist die Fähigkeit, Meinungen zu bestimmten Themen zu haben bzw. zu produzieren von der Nähe der Befragten zu der jeweiligen Problematik abhängig. Eine Reihe empirischer Studien deutet in diesem Zusammenhang darauf hin, dass die unterschiedliche subjektive Salienz abgefragter Themen die Messung und den Vergleich von Einstellungen in inhaltlicher und methodischer Hinsicht beschränken kann. Groves, Presser und Dipko (2004)

weisen nach, dass bereits die Entscheidung, an einer Umfrage teilzunehmen, deutlich durch das thematische Interesse beeinflusst wird. Zu diesem Ergebnis kommen auch Zillmann et al. (2014), die anhand einer in eine Dating-Website eingebundenen Online-Umfrage feststellen, dass eine Befragung zum Thema „Online-Dating“ insbesondere für diejenigen mit niedrigen Erfolgchancen bei der digitalen Partnersuche hohe Relevanz hat. Nicht nur die Teilnahmewahrscheinlichkeit, sondern auch die Bereitschaft, offene Fragen zu beantworten sowie die Qualität der dort gegebenen Antworten sind interessensabhängig (Holland & Christian 2009).

Blasius und Thiessen (2001; 2012) vergleichen die Antwortmuster von Befragten mit hohem und niedrigem politischen Interesse in einer Itematterie, die das politische Vertrauen und Selbstwirksamkeit (*political efficacy*) messen soll. Sie stellen fest, dass die weniger Interessierten nicht nur häufiger die Mittelkategorie und „weiß nicht“ wählten, sondern – wie die Anwendung von *subset* MCA zeigt – auch weitaus größere Probleme mit der Beantwortung von Fragen mit doppelter Negation und generell komplizierter Formulierung hatten als die politisch interessierte Gruppe. Die häufig festgestellte positive Korrelation zwischen politischem Interesse, politischem Vertrauen und Selbstwirksamkeit könnte in diesem Sinne auch ein Methodenartefakt sein, das dadurch entsteht, dass politisch uninteressierte Befragte den Statements pauschal zustimmen.

Ganz allgemein stellen die Überrepräsentation der Antworten thematisch Interessierter, bzw. systematische Ausfälle und Qualitätsprobleme bei weniger Involvierten, ein Problem in der Auswertung von Surveydaten dar. Spezifisch für Paneldaten ist hingegen die Problematik, dass die wiederholte Befragung selbst die Meinungen, bzw. das Interesse und die Einstellungsstärke der Befragten beeinflussen kann (Lazarsfeld 1940; Warren & Halpern-Manners 2012). Einer theoretischen Konzeptualisierung dieses als *panel conditioning* bekannten Phänomens widmen sich Bergmann und Barth (2018). Vor dem Hintergrund widersprüchlicher Forschungsergebnisse – etliche Studien präsentieren empirische Belege für Effekte wiederholter Befragungen (Bergmann 2015; Halpern-Manners, Warren & Torche 2017; Sturgis, Allum & Brunton-Smith 2009), während andere Forscher nur marginale oder keine Effekte feststellen (Axinn, Jennings & Couper 2015, Barber et al. 2016; Mann 2005) – argumentieren die Autoren, dass zur Einordnung der komplizierten Datenlage nicht nur angemessene Analysemethoden, sondern auch eine genauere Kenntnis der Wirkmechanismen von *panel conditioning* vonnöten sind. Sie gehen davon aus, dass eine wiederholte Umfrageteilnahme die kognitive Informationsverarbeitung von Befragten beeinflusst: Durch die mehrmalige Konfrontation mit den gleichen Themen erhöhen sich Zugänglichkeit, interne Konsistenz und Extremität von Einstellungen. Die Erhöhung der Einstellungsstärke wiederum verstärkt die Stabilität der betreffenden Einstellungen und steigert ihren Einfluss auf Denken und Handeln (Krosnick & Petty 1995). Die stärksten Effekte sind dabei

gerade bei denjenigen zu erwarten, die zu Beginn der Studie relativ schwach ausgeprägte Einstellungen aufweisen. Bergmann und Barth testen den angenommenen Mechanismus mit einer FE-Regression am Beispiel der Unentschlossenheit in Bezug auf die Wahlentscheidung für eine bestimmte Partei bei der nächsten Wahl. Als Datengrundlage werden sechs Panelwellen der German Longitudinal Election Study 2009 verwendet. Es zeigt sich, dass die politische Unentschlossenheit bei den Panelbefragten wesentlich stärker abnimmt als in der allgemeinen Bevölkerung (Kontrolle durch Querschnittsbefragungen). Der Effekt wird dabei von der initialen Zugänglichkeit und Extremität parteibezogener Einstellungen moderiert: Befragte mit anfänglich schwachen Einstellungen zeigen größere Veränderungen. Daraus ergibt sich, dass Panelstudien tendenziell die Stärke, Stabilität und Handlungsrelevanz wiederholt abgefragter Einstellungen überschätzen, da diese Eigenschaften teils durch die Studienteilnahme selbst geprägt werden und damit nicht repräsentativ für die Gesamtpopulation sind.

Empirisch kann die Einstellungsstärke entweder direkt durch die Abfrage der Wichtigkeit der Einstellung bzw. des entsprechenden Objekts (Gopinath & Nyer 2009; Howe & Krosnick 2017) oder indirekt mithilfe operativer Indikatoren (Bassili 1996) gemessen werden. Bei letzterer Methode wird die Zugänglichkeit von Einstellungen häufig durch Antwortzeiten (*response latency*) operationalisiert (Fazio 1990; Powell & Fazio 1984), während etwa die interne Konsistenz durch die Aufrechnung positiver und negativer Äußerungen bezüglich des Einstellungsobjekts (Thompson, Zanna & Griffin 1995) festgestellt werden kann.

3.6 Der Umgang mit der Umfrage

Über die Reaktion auf einzelne Worte, Fragen und Konstrukte hinaus kann die generalisierte Reaktion der Befragten auf die Befragungssituation an sich im Sinne verschiedener Antwortqualitäten betrachtet werden. Barth und Schmitz (2018) stellen diesbezüglich fest, dass nicht nur Antworttendenzen je nach ideologischer Disposition unterschiedlich verteilt sind, sondern dass sich auch die Konformität mit den impliziten Normen, die in einer Befragung wirksam werden, unterscheidet. Dazu werden auf der Datenbasis des GSS (2010-2014) die Relationen zwischen verschiedenen manifesten Indikatoren, etwa die Anzahl „nicht-substantieller“ Antworten wie „weiß nicht“/„keine Angabe“ oder die Bewertung der Kooperation der Befragten durch die Interviewerin oder den Interviewer, mittels LCA analysiert. Diese Typologie der *survey compliance* ergibt mehrere Klassen von Befragten, die sich jeweils in ihrem Umgang mit der Befragung unterscheiden. Der größte Teil der Befragten (etwa 80%) kann als „funktional“ im Sinne einer hohen Kooperationsbereitschaft und wenigen fehlenden Werten kategorisiert werden. Daneben gibt es jedoch auch eine Klasse, in der gehäuft Verständnisprobleme, gepaart mit einer

überdurchschnittlich langen Interviewdauer und häufigem Gebrauch der „weiß nicht“-Option auftreten (9.5%). Hier ist von einer grundlegenden Kooperationsbereitschaft, aber auftretenden Schwierigkeiten mit den konkreten Anforderungen auszugehen. Eine weitere Klasse ist durch sehr kurze Interviews, die vergleichsweise höchste Rate an „weiß nicht“-Antworten, viele fehlende Werte sowie eine hohe Wahrscheinlichkeit, für weitere Studien nicht erreichbar zu sein, gekennzeichnet (9.4%). Dieses Muster kann so interpretiert werden, dass die Befragten versuchen, sich der Befragung – bewusst oder unbewusst – zu entziehen. Schließlich fällt eine sehr kleine Klasse von Befragten (1.4%) auf, in der sowohl etliche Angaben als auch die Teilnahme an weiteren Studien mit hoher Wahrscheinlichkeit offen verweigert werden, also eine deutliche Opposition zu den impliziten Befragungsnormen erkennbar wird.

Diese Typen von Antwortqualitäten hängen wiederum mit der Position im US-amerikanischen „Feld der ideologischen Konsumtion“ zusammen. Alle Arten „dysfunktionaler“ Reaktionen weisen eine Nähe zum Pol traditioneller Werthaltungen auf, wobei jedoch eine Differenzierung je nach Einstellung zu den Aufgaben des Staates deutlich wird: Diejenigen Befragten, die staatlichen Eingriffen grundsätzlich eher kritisch gegenüberstehen, gehören eher zu den „offenen Verweigerern“, während im Bereich hoher Zustimmungswerte zu wohlfahrtsstaatlichen Maßnahmen Verständnisprobleme (und, wie oben erläutert, Akquieszenz und Tendenz zu Extremwerten) vorherrschen. Insgesamt betrachtet fällt die Abwesenheit jeglicher Qualitätsprobleme in dem Bereich auf, wo liberale Werte und eine pro-wohlfahrtsstaatliche Haltung gemeinsam auftreten. Die Ergebnisse konnten in ähnlicher Weise für den deutschen Raum (auf Basis des World Values Survey 2013 in Deutschland) repliziert werden (Schmitz & Barth 2018). Daraus lässt sich schließen, dass auch scheinbar ideologisch neutrale Erhebungsinstrumente unterschiedliche Reaktionen bei Befragten hervorrufen, die systematisch mit deren ideologischer Positionierung zusammenhängen. In anderen Worten liegt eine bisher wenig beachtete Form der Verzerrung durch die differentielle „kulturelle Distanz“ zwischen Instrument und Befragten vor (vgl. Van de Vijver & Poortinga 1997). Für die Repräsentation der „öffentlichen Meinung“ durch Surveys bedeutet dies – in Einklang mit der Bourdieuschen Argumentation – dass bestimmte Meinungsspektren verzerrter abgebildet werden als andere.

Eine ganz andere Problematik ergibt sich, wenn der Umgang mit der Befragung durch die Befragung selbst – das heißt, durch vorhergehende Fragen bzw. Panel-Wellen – beeinflusst wird. Neben den Effekten auf die Einstellungsstärke kann die frühere Umfrageteilnahme auch dazu führen, dass sich Befragte an die Struktur des Fragebogens erinnern und mit diesem Wissen Antworten, die zu Anschlussfragen führen, vermeiden (Nancarrow & Cartwright 2007; Toepoel, Das & van Soest 2008; Warren & Halpern-Manners 2012). Andererseits kann die Erfahrung aber

auch hilfreich sein, um komplizierte Fragen exakter beantworten zu können (Fisher 2016; Waterton & Lievesley 1989) und das Vertrauen in die durchführende Institution bzw. die Interviewerin oder den Interviewer zu stärken, was zu einer größeren Bereitschaft zur Preisgabe sensibler Informationen bzw. einem Rückgang des Effekts sozialer Erwünschtheit führt (Bailar 1975; Struminskaya 2016). Bergmann und Barth (2018) nehmen mit Rückgriff auf Jon Krosnicks Theorie des *survey satisficing* (Krosnick 1991) an, dass die Motivation der Befragten eine Schlüsselrolle bei der Vorhersage von Effekten der wiederholten Befragung auf das Antwortverhalten spielt. *Ceteris paribus* sollten größere Vertrautheit mit der Befragungsprozedur und gestiegene Einstellungsstärke dazu führen, dass die Beantwortung bekannter Fragen als einfacher empfunden wird und die Befragten sich zunehmend als kompetente Informantinnen und Informanten erleben, was laut Krosnicks Theorie gute Bedingungen für *optimizing* (sorgfältiges und wohlüberlegtes Beantworten der Fragen) bietet. Sinkt jedoch die Motivation der Befragten infolge von Länge und/oder Eintönigkeit der Befragung, so wird die Anwendung von Strategien zur Reduktion der kognitiven Belastung (*satisficing*) wahrscheinlicher.

Unterschiede im Umgang mit der Befragung können also zum einen durch individuelle Kontextvariablen wie die generalisierte Einstellung zu Befragungen oder die ideologische Disposition, zum anderen durch den Prozess der Datenerhebung selbst bedingt sein, wobei eine Interaktion dieser Aspekte insbesondere in Panelbefragungen eher die Regel als die Ausnahme sein dürfte (Lipps 2014). Empirisch kann der Umgang mit der Befragung direkt durch die Art der Fragenbeantwortung oder auf der Meta-Ebene durch Selbsteinschätzung der Sorgfalt und Motivation in Bezug auf die Befragung bzw. deren Einschätzung durch Interviewende erfasst werden. Bei der erstgenannten Methode ist, analog zur Messung von Antworttendenzen, zwischen der Nutzung bestehender, eigentlich substantieller Fragen und dem Hinzufügen spezieller Fragen zum Fragebogen (z.B. *instructional manipulation checks*, mit denen getestet wird, ob Befragte Fragentext bzw. Anweisungen aufmerksam lesen, siehe Berinsky et al. 2014; Oppenheimer, Meyvis & Davidenko 2009) zu unterscheiden. Aus der Analyse von Barth & Schmitz (2018) sowie anderen Publikationen (Hess & Stathopoulos 2013; Kaminska, McCutcheon & Billiet 2010) geht in dieser Hinsicht hervor, dass aufgrund der facettenreichen Natur von Antwortqualitäten eine relationale Kombination mehrerer Indikatoren sinnvoll ist.

4. Diskussion

In Befragungen gemessene Einstellungen können nur dann valide verglichen werden, wenn alle Befragten die gleichen Fragen beantworten, dabei keinen unterschiedlichen Einflüssen ausgesetzt sind und ihre Position einer der verfügbaren Antwortoptionen zuordnen können. Die Standardisierung von Instrument und Erhebungsprozess dient dazu, diese Bedingungen der Vergleichbarkeit zu gewährleisten. Eine Betrachtung der kognitiven und interaktionalen Prozesse während der Befragung, sowie der Einbettung des Phänomens „Survey“ in einen größeren gesellschaftlichen Zusammenhang zeigt jedoch, dass vollständige Vergleichbarkeit kaum zu erreichen ist. Auch – und gerade – in standardisierten Befragungen haben Kontextbedingungen der Befragung wie der Befragten vielgestaltige Auswirkungen darauf, wie Einstellungsfragen beantwortet werden.

Aufbauend auf dieser Erkenntnis wurden in dieser Arbeit Aspekte unterschiedlicher Antwortqualitäten vorgestellt und an empirischen Beispielen erläutert. Das Hauptaugenmerk lag dabei auf der Identifikation von Problemen der Vergleichbarkeit, die erst durch die Prüfung multivariater Zusammenhänge deutlich werden. Analytisch wurden fünf Kategorien unterschieden: Die Bedeutung und Interpretation von Begriffen, die Verortung auf Skalen, die Struktur des unterliegenden Konstrukts, die Zentralität von Themen sowie der Umgang mit der Umfrage an sich. Tatsächlich lässt sich jedoch häufig eine Interaktion mehrerer Aspekte beobachten, wie auch aus der Einordnung der Forschungsartikel in jeweils zwei Hauptkategorien deutlich wird. So können sich etwa Bedeutungsunterschiede in einzelnen Items auf die Struktur des Konstrukts auswirken oder die subjektive Zentralität eines Themas auch den Umgang mit der gesamten Umfrage beeinflussen. Weiterhin ist festzustellen, dass methodische und inhaltliche Effekte meist nur schwer voneinander zu trennen sind, beispielsweise wenn systematische Unterschiede in der subjektiven Zentralität der Umfragethemen zwischen Befragten bestehen, die zu unterschiedlichen Teilnahmewahrscheinlichkeiten und Antwortqualitäten führen. In diesem Fall hängen zentrale, inhaltliche Variablen (z.B. politisches Interesse, ideologische Position) untrennbar mit verschiedenen, meist als Methodenartefakte klassifizierten Antworttendenzen zusammen. Umgekehrt kann die Befragung selbst substantielle Änderungen in Einstellungen und Verhalten hervorrufen. Vor diesem Hintergrund stellt sich die Frage, wie in der Forschungspraxis mit der geschilderten Problematik umzugehen ist. Grob gesagt lassen sich dabei drei Strategien identifizieren:

1. Ignorieren: Obwohl die Diskussion der hier geschilderten Probleme weit über ein Nischenpublikum hinausgeht und der größte Teil möglicher Kontexteinflüsse schon seit langem bekannt ist (Esser 1975b; Lazarsfeld 1935; Scheuch 1965; Scheuch 1993[1989]), werden in

empirischen Analysen häufig keine Konsequenzen aus diesem Wissen gezogen. So wird bei Analysen von Befragungsdaten aus ein und demselben Sprachraum selten die Möglichkeit unterschiedlichen Fragenverständnisses als Ursache festgestellter Variation erforscht, und selbst in multinationalen Kontexten ist die systematische Prüfung der funktionalen Äquivalenz bezüglich semantischer und pragmatischer Bedeutung keine Selbstverständlichkeit (Mohler & Johnson 2010). Auch Skalenverortung und unterschiedliche Konstruktstrukturen werden in den meisten vergleichenden Untersuchungen nicht thematisiert (Riordan et al. 2001; Vandenberg & Lance 2000). „Nicht-substantielle“ Antworten wie „weiß nicht“ oder „keine Angabe“ werden aus inhaltlichen Analysen häufig einfach ausgeschlossen, ohne mögliche Zusammenhänge zwischen Antwortqualitäten und der untersuchten Fragestellung zu betrachten. Unter Umständen kann diese Praxis zu veritablen Fehleinschätzungen führen, etwa wenn die Antworten politisch uninteressierter bzw. befragungsskeptischer Bürgerinnen und Bürger in politischen Meinungsumfragen unterrepräsentiert sind (Barth & Schmitz 2018), oder kulturelle Unterschiede in der Bereitschaft, Unwissen bzw. Meinungslosigkeit (Sicinski 1970) oder Vorurteile (Weins 2009) zuzugeben, unentdeckt bleiben.

2. Kontrollieren. Angesichts der Risiken, die mit dem Ignorieren von Vergleichbarkeitsproblemen verbunden sind, gewinnen verschiedene Kontrollstrategien zunehmend an Popularität. Auf der Ebene des Erhebungsprozesses wird durch die Etablierung einheitlicher Qualitätsstandards versucht, Vergleichbarkeit zwischen verschiedenen Erhebungskontexten herzustellen (Baur 2014; Lyberg & Stukel 2010; Weichbold 2009). Dabei bewegt man sich auf einem schmalen Grat zwischen Flexibilität, um möglichst viele Erhebungskontexte und Fälle einbeziehen zu können, und der Einhaltung fester Standards – eine zu starke Tendenz in jede der beiden Richtungen kann in Bezug auf die Vergleichbarkeit der resultierenden Daten mehr schaden als nützen. Auf der Ebene der Datenauswertung kann mit verschiedenen Methoden die Vergleichbarkeit bzw. Messäquivalenz der Daten geprüft werden (Barth & Trübner 2018; Braun & Johnson 2010; Kankaraš et al. 2011; Van de Vijver & Leung 1997). Es werden verschiedene Äquivalenzniveaus unterschieden – je stärker die Äquivalenz, umso mehr Kennwerte können verglichen werden, wobei komplette Äquivalenz eher ein hypothetisches Ideal als ein praktisch zu erreichendes Ziel darstellt (Mohler & Johnson 2010; Verba, Nie & Kim 1978). Es gilt auch hier: Je unterschiedlicher die Kontexte, desto relevanter wird die Abwägung zwischen der Strenge der Äquivalenzforderungen und der inhaltlichen Validität der Instrumente (Rippl & Seipel 2008). Werden Äquivalenztests angewendet, was immer noch eher die Ausnahme als die Regel ist, so stellt sich meist heraus, dass die getesteten Konzepte nicht völlig messäquivalent über alle Gruppen hinweg sind (Davidov et al. 2014). Aber auch wenn nur partielle Messinvarianz vorhanden ist, können Vergleiche in eingeschränkter Form erfolgen (Byrne, Shavelson & Muthén 1989;

Steenkamp & Baumgartner 1998). Weiterhin gibt es einige multivariate Methoden zur Messung und gleichzeitigen Kontrolle von Antworttendenzen (z.B. Van Rosmalen et al. 2010; Welkenhuysen-Gybels, Billiet & Cambré 2003), die jedoch ebenfalls eher selten verwendet werden.

3. Informativ nutzen. Die dritte mögliche Strategie ist es, die Nicht-Vergleichbarkeit selbst zum primären Forschungsproblem zu erheben und nach den Ursachen der unterschiedlichen Herangehensweisen und Antwortqualitäten zu suchen. Zum einen kann dies durch eine genauere Prüfung der strukturellen Unterschiede und Antwortqualitäten mittels explorativer multivariater Methoden erfolgen (Barth 2016; Barth & Schmitz 2018). Zum anderen können zusätzliche Variablen, z.B. Sprache, ethnische Gruppe, Alter oder ideologische Disposition zur Erklärung unterschiedlicher Antwortqualitäten herangezogen werden (Davidov et al. 2014; Barth & Schmitz 2018). Werden die Gründe für das unterschiedliche Funktionieren von Items oder verschiedene Antwortqualitäten auf gesellschaftlicher (i.d.R. nationaler) Ebene vermutet, bietet sich die Verwendung von Makro-Prädiktoren in einem Mehrebenenmodell an (Davidov et al. 2012; 2018; He et al. 2014). Ist die Befragung selbst ein möglicher Verzerrungsfaktor, sollten Paneleffekte beachtet werden (Bergmann & Barth 2018). In der vorliegenden Arbeit wurde zudem für ein relationales Verständnis von Antwortqualitäten plädiert, was mit der gemeinsamen Verortung von Reaktionen und Charakteristika der Befragten in einem mehrdimensionalen Raum operationalisiert werden kann (Barth & Schmitz 2018). Während hier also vor allem Methoden der informativen Nutzung von Vergleichbarkeitsproblemen diskutiert wurden, die mit (Sekundär-)Daten aus standardisierten Befragungen operieren, besteht eine weitere vielversprechende Strategie in qualitativ orientierten Verfahren zur Aufdeckung von kontextbedingten Unterschieden. Dies kann durch ergänzende offene Interviews (Knappertsbusch 2017; Lugtig et al. 2012), kognitive Interviews zur Überprüfung des Gesamtfragebogens (Beatty & Willis 2007; Willis 2004) oder Nachhaken (*probing*) bei einzelnen Items (Behr et al. 2012; Braun, Behr & Díez Medrano 2018) verwirklicht werden.

Die komplementäre Ergänzung standardisierter Einstellungsmessungen durch qualitative Methoden hat aber nicht nur auf methodischer, sondern auch auf konzeptioneller Ebene Potential. So wird die Kontextgebundenheit der erhobenen Daten, für deren Beachtung in standardisierten Befragungen die vorliegende Arbeit argumentiert, in der qualitativen Forschung von Anfang an systematisch mit einbezogen. Die Reflexion der inhärenten Subjektivität sozialer Daten stellt dabei ein zentrales Gütekriterium dar (Helfferrich 2011). Dies bedeutet im Umkehrschluss nicht, wie von einigen Kritikern quantitativer Methoden geäußert, dass die standardisierte Befragung zwar möglicherweise reliable, aber niemals (ökologisch) valide Daten generieren kann (Cicourel 1974; Mishler 1986). Im Gegenteil lässt sich argumentieren, dass gerade die Praxis der Standardisierung

die Auswirkungen von Kontexteffekten auf die Validität herausstellt, indem zentrale Aspekte (Stimuli, Antwortoptionen, Anweisungen, Fragenreihenfolge etc.) konstant gehalten werden (Schaeffer 2002:96). Dennoch – oder vielleicht gerade deswegen – fehlt bei der Konzeption, Durchführung und Auswertung von standardisierten Befragungen oft das Bewusstsein für die Subjektivität und Reaktivität, die bei der Forschung mit Menschen unweigerlich eine Rolle spielen. Diese Aspekte theoretisch stärker zu reflektieren und praktisch – durch die Forschung zu kontextuellen Einflüssen und Mixed-Methods-Designs – umzusetzen, stellt eine seit langem bekannte, aber immer noch sehr aktuelle Herausforderung der Survey Methodology dar.

5. Literatur

- Atteslander, Peter (2006): *Methoden der empirischen Sozialforschung*. Berlin [u.a.]: Schmidt.
- Axinn, William G.; Jennings, Elyse A.; Couper, Mick P. (2015): Response of sensitive behaviors to frequent measurement. In: *Social Science Research* 49, S. 1-15. DOI: 10.1016/j.ssresearch.2014.07.002.
- Bachleitner, Reinhard; Aschauer, Wolfgang; Weichbold, Martin (2010): *Die Befragung im Kontext von Raum, Zeit und Befindlichkeit. Beiträge zu einer prozessorientierten Theorie der Umfrageforschung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bailar, Barbara A. (1975): The Effects of Rotation Group Bias on Estimates from Panel Surveys. In: *Journal of the American Statistical Association* 70 (349), S. 23-30. DOI: 10.2307/2285370.
- Barber, Jennifer S.; Gatny, Heather H.; Kusunoki, Yasamin; Schulz, Paul (2016): Effects of intensive longitudinal data collection on pregnancy and contraceptive use. In: *International Journal of Social Research Methodology* 19 (2), S. 205-222. DOI: 10.1080/13645579.2014.979717.
- Baron-Epel, Orna; Kaplan, Giora; Weinstein, Ruth; Green, Manfred S. (2010): Extreme and acquiescence bias in a bi-ethnic population. In: *European Journal of Public Health* 20 (5), S. 543-548. DOI: 10.1093/eurpub/ckq052.
- Barth, Alice (2016): The Changing Nature of Attitude Constructs - An Application of Multiple Correspondence Analysis on Gender Role Attitudes. In: *Quality & Quantity*, 50 (4), S. 1507-1523. DOI:10.1007/s11135-015-0218-9
- Barth, Alice; Schmitz, Andreas (2018): Response quality and ideological dispositions: an integrative approach using geometric and classifying techniques. In: *Quality & Quantity* 52 (1), S. 175-194. DOI: 10.1007/s11135-016-0458-3
- Barth, Alice; Trübner, Miriam (2018): Structural Stability, Quantitative Change: A Latent Class Analysis Approach towards Gender Role Attitudes in Germany. In: *Social Science Research* 72, S. 183-193. DOI: 10.1016/j.ssresearch.2018.02.008
- Bassili, John N. (1996): Meta-judgmental versus operative indexes of psychological attributes. The case of measures of attitude strength. In: *Journal of Personality and Social Psychology* 71 (4), S. 637-653. DOI: 10.1037/0022-3514.71.4.637.
- Baumgartner, Hans; Steenkamp, Jan-Benedict E.M. (2001): Response Styles in Marketing Research - A Cross-National Investigation. In: *Journal of Marketing Research* 38, S. 143-156. DOI: 10.1509/jmkr.38.2.143.18840.
- Baur, Nina (2004): Wo liegen die Grenzen quantitativer Längsschnittanalysen? In: Gerhard Schulze und Nina Baur (Hg.): *Bamberger Beiträge zur empirischen Sozialforschung* 23/2004. Bamberg.
- Baur, Nina (2009): Memory and Data. Methodological Implications of Ignoring Narratives in Survey Research. In: Noel Packard (Hg.): *Sociology of Memory. Papers from the Spectrum*. Newcastle: Cambridge Scholars Publishing, S. 289-312.
- Baur, Nina (2014): Comparing Societies and Cultures. In: *Historical Social Research* 39 (2), S. 257-291. DOI: 10.12759/hsr.39.2014.2.257-291.
- Beatty, Paul C.; Willis, Gordon B. (2007): Research Synthesis. The Practice of Cognitive Interviewing. In: *Public Opinion Quarterly* 71 (2), S. 287-311. DOI: 10.1093/poq/nfm006.
- Behr, Dorothée; Braun, Michael; Kaczmirek, Lars; Bandilla, Wolfgang (2012): Testing the Validity of Gender Ideology Items by Implementing Probing Questions in Web Surveys. In: *Field Methods* 25 (2), S. 124-141. DOI: 10.1177/1525822X12462525.

- Bergmann, Michael; Barth, Alice (2018). What was I thinking? A Theoretical Framework for Analysing Panel Conditioning in Attitudes and (Response) Behaviour. In: *International Journal of Social Research Methodology* 21(3), S. 333-345. DOI: 10.1080/13645579.2017.1399622
- Benzécri, Jean-Paul et col. (1973): *L'analyse des données*. Paris: Dunod.
- Bergman, Manfred Max (1998): A Theoretical Note on the Differences Between Attitudes, Opinions, and Values. In: *Swiss Political Science Review* 4 (2), S. 81-93. DOI: 10.1002/j.1662-6370.1998.tb00239.x.
- Bergmann, Michael (2015): *Panel Conditioning. Wirkungsmechanismen und Konsequenzen wiederholter Befragungen*. Baden-Baden: Nomos.
- Bergström, Ylva (2012): Educational and Social Dimensions of Political Participation: Producing a Political Opinion. In: *Praktische Grunde* 4, S. 23-34.
- Berinsky, Adam J.; Margolis, Michele F.; Sances, Michael W. (2014): Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys. In: *American Journal of Political Science* 58 (3), S. 739-753. DOI: 10.1111/ajps.12081.
- Biemer, Paul P. (2010): Total Survey Error. Design, Implementation, and Evaluation. In: *Public Opinion Quarterly* 74 (5), S. 817-848. DOI: 10.1093/poq/nfq058.
- Biemer, Paul P.; Lyberg, Lars E. (2003): *Introduction to survey quality*. Hoboken, NJ [u.a.]: Wiley.
- Billiet, Jaak B.; McClendon, McKee J. (2000): Modeling Acquiescence in Measurement Models for Two Balanced Sets of Items. In: *Structural Equation Modeling: A Multidisciplinary Journal* 7 (4), S. 608-628. DOI: 10.1207/S15328007SEM0704_5.
- Binkley, Robert C. (1928): The Concept of Public Opinion in the Social Sciences. In: *Social Forces* 6 (3), S. 389-396. DOI: 10.2307/3004859.
- Blasius, Jörg (2001): *Korrespondenzanalyse*. München: Oldenbourg.
- Blasius, Jörg; Thiessen, Victor (2001): Methodological Artifacts in Measures of Political Efficacy and Trust. A Multiple Correspondence Analysis. In: *Political Analysis* 9 (1), S. 1-20. DOI: 10.1093/oxfordjournals.pan.a004862.
- Blasius, Jörg; Thiessen, Victor (2006): Assessing Data Quality and Construct Comparability in Cross-National Surveys. In: *European Sociological Review* 22 (3), S. 229-242. DOI: 10.1093/esr/jci054.
- Blasius, Jörg; Thiessen, Victor (2012): *Assessing the Quality of Survey Data*. Los Angeles: Sage.
- Blasius, Jörg; Thiessen, Victor (2015): Should we trust survey data? Assessing response simplification and data fabrication. In: *Social Science Research* 52, S. 479-493. DOI: 10.1016/j.ssresearch.2015.03.006.
- Blondiaux, Loïc (1998): *La fabrique de l'opinion. Une histoire sociale des sondages*. Paris: Seuil.
- Blumer, Herbert (1948): Public Opinion and Public Opinion Polling. In: *American Sociological Review* 13 (5), S. 542-549. DOI: 10.2307/2087146.
- Bourdieu, Pierre (1987): *Die feinen Unterschiede. Kritik der gesellschaftlichen Urteilskraft*. Frankfurt am Main: Suhrkamp.
- Bourdieu, Pierre (2013a) [frz. Erstveröffentlichung 1973]: Die öffentliche Meinung gibt es nicht. In: Pierre Bourdieu: *Schriften zur Politischen Ökonomie*. Hg. v. Franz Schultheis und Stephan Egger. Berlin: Suhrkamp, S. 243-254.
- Bourdieu, Pierre (2013b) [frz. Erstveröffentlichung 1977]: Politische Fragen. In: Pierre Bourdieu: *Schriften zur Politischen Ökonomie*. Hg. v. Franz Schultheis und Stephan Egger. Berlin: Suhrkamp, S. 125-208.

- Bradburn, Norman M. (2016): Surveys as Social Interactions. In: *Journal of Survey Statistics and Methodology* 4 (1), S. 94-109. DOI: 10.1093/jssam/smv037.
- Braun, Michael (1998): Gender Roles. In: Jan W. van Deth (Hg.): *Comparative politics. The problem of equivalence*. Colchester, UK: ECPR Press, S. 111-134.
- Braun, Michael (2003): Communication and Social Cognition. In: Janet A. Harkness, Fons J.R. Van de Vijver und Peter Ph. Mohler (Hg.): *Cross-cultural survey methods*. Hoboken, NJ: Wiley, S. 57-68.
- Braun, Michael (2006): *Funktionale Äquivalenz in interkulturell vergleichenden Umfragen. Mythos und Realität*. Mannheim: ZUMA.
- Braun, Michael; Behr, Dorothée; Díez Medrano, Juan (2018): What do respondents mean when they report to be “citizens of the world”? Using probing questions to elucidate international differences in cosmopolitanism. In: *Quality & Quantity* 52 (3), S. 1121-1135. DOI: 10.1007/s11135-017-0507-6.
- Braun, Michael; Harkness, Janet A. (2005): Text and context: challenges to comparability in survey questions. In: Jürgen H. P. Hoffmeyer-Zlotnik und Janet A. Harkness (Hg.): *Methodological Aspects in Cross-National Research*. Mannheim: ZUMA, S. 95-107.
- Braun, Michael; Johnson, Timothy P. (2010): An Illustrative Review for Detecting Inequivalences. In: Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars E. Lyberg und Peter Ph. Mohler (Hg.): *Survey methods in multinational, multiregional, and multicultural contexts*. Hoboken, NJ: Wiley, S. 375-394.
- Bredl, Sebastian; Winker, Peter; Kötschau, Kerstin (2012): A statistical approach to detect interviewer falsification of survey data. In: *Survey Methodology* 38 (1), S. 1-10.
- Breetvelt, Iris S.; van Dam, Fritz S.A.M. (1991): Underreporting by cancer patients. The case of response-shift. In: *Social Science & Medicine* 32 (9), S. 981-987. DOI: 10.1016/0277-9536(91)90156-7.
- Brislin, Richard W. (1986): The Wording and Translation of Research Instruments. In: Walter J. Lonner und John W. Berry (Hg.): *Field methods in cross-cultural research*. Beverly Hills, CA [u.a.]: Sage, S. 137-164.
- Byrne, Barbara M.; Shavelson, Richard J.; Muthén, Bengt (1989): Testing for the equivalence of factor covariance and mean structures. The issue of partial measurement invariance. In: *Psychological Bulletin* 105 (3), S. 456-466. DOI: 10.1037/0033-2909.105.3.456.
- Chandler, Jamie Patrick (2008): Opinions. In: Paul J. Lavrakas (Hg.): *Encyclopedia of Survey Research Methods*. Thousand Oaks, CA: Sage, S. 552-553.
- Cicourel, Adam V. (1974): *Methode und Messung in der Soziologie*. Berlin: Suhrkamp.
- Cisneros-Puebla, Cesar; Faux, Robert; Mey, Günter (2004): Qualitative Researchers - Stories Told, Stories Shared: The Storied Nature of Qualitative Research. An Introduction to the Special Issue: FQS Interviews I. In: *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 5 (3). DOI: 10.17169/fqs-5.3.547.
- Clark, Carroll D. (1933): The Concept of the Public. In: *The Southwestern Social Science Quarterly* 13 (4), S. 311-320.
- Clark, Herbert H.; Schober, Michael F. (1992): Asking Questions and Influencing Answers. In: Judith M. Tanur (Hg.): *Questions about questions. Inquiries into the cognitive bases of surveys*. New York: Russell Sage Foundation, S. 15-48.
- Converse, Philip E. (1987): Changing Conceptions of Public Opinion in the Political Process. In: *Public Opinion Quarterly* 51 (2), S. S12-S24.

- Converse, Philip E. (2006) [Erstveröffentlichung 1964]: The nature of belief systems in mass publics. In: *Critical Review* 18 (1-3), S. 1-74. DOI: 10.1080/08913810608443650.
- Couch, Arthur; Keniston, Kenneth (1960): Yeasayers and naysayers. Agreeing response set as a personality variable. In: *The Journal of Abnormal and Social Psychology* 60 (2), S. 151-174. DOI: 10.1037/h0040372.
- Couper, Mick P.; Conrad, Frederick G.; Tourangeau, Roger (2007): Visual Context Effects in Web Surveys. In: *Public Opinion Quarterly* 71 (4), S. 623-634. DOI: 10.1093/poq/nfm044.
- Couper, Mick P.; Leeuw, Edith de (2003): Nonresponse in Cross-Cultural and Cross-National Surveys. In: Janet A. Harkness, Fons J.R. Van de Vijver und Peter Ph. Mohler (Hg.): *Cross-cultural survey methods*. Hoboken, NJ: Wiley, S. 157-178.
- Cronbach, Lee J. (1946): Response Sets and Test Validity. *Educational and Psychological Measurement* 6(4), S. 475-494.
- Curtin, Richard; Presser, Stanley; Singer, Eleanor (2005): Changes in Telephone Survey Nonresponse over the Past Quarter Century. In: *Public Opinion Quarterly* 69 (1), S. 87-98. DOI: 10.1093/poq/nfi002.
- Davidov, Eldad; Dülmer, Hermann; Ciecuch, Jan; Kuntz, Anabel; Seddig, Daniel; Schmidt, Peter (2018): Explaining Measurement Nonequivalence Using Multilevel Structural Equation Modeling. In: *Sociological Methods & Research* 47 (4), S. 729-760. DOI: 10.1177/0049124116672678.
- Davidov, Eldad; Dülmer, Hermann; Schlüter, Elmar; Schmidt, Peter; Meuleman, Bart (2012): Using a Multilevel Structural Equation Modeling Approach to Explain Cross-Cultural Measurement Noninvariance. In: *Journal of Cross-Cultural Psychology* 43 (4), S. 558-575. DOI: 10.1177/0022022112438397.
- Davidov, Eldad; Meuleman, Bart; Ciecuch, Jan; Schmidt, Peter; Billiet, Jaak (2014): Measurement Equivalence in Cross-National Research. In: *Annual Review of Sociology* 40 (1), S. 55-75. DOI: 10.1146/annurev-soc-071913-043137.
- Dillman, Don A. (2007): *Mail and internet surveys. The tailored design method, with new internet, visual and mixed-mode guide*. Hoboken, NJ: Wiley.
- Dillman, Don A.; Smyth, Jolene D.; Christian, Leah Melani (2014): *Internet, phone, mail, and mixed-mode surveys. The tailored design method*. Hoboken, NJ: Wiley.
- Eagly, Alice H.; Chaiken, Shelly (1993): *The Psychology of Attitudes*. Orlando, FL: Harcourt Brace Jovanovich College Publishers.
- Esser, Hartmut (1975a): Differenzierung und Integration sozialer Systeme als Voraussetzungen der Umfrageforschung. In: *Zeitschrift für Soziologie* 4 (4), S. 316-334.
- Esser, Hartmut (1975b): *Soziale Regelmäßigkeiten des Befragtenverhaltens*. Meisenheim am Glan: Hain.
- Esser, Hartmut (1986): Können Befragte lügen? Zum Konzept des „wahren Wertes“ im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung, *ZUMA-Arbeitsbericht, 1986/02*, Mannheim: ZUMA.
- Fazio, Russell H. (1990): A Practical Guide to the Use of Response Latency in Social Psychological Research. In: Clyde Hendrick und Margaret S. Clark (Hg.): *Review of Personality and Social Psychology* 11. *Research Methods in Personality and Social Psychology*. Newbury Park, CL: Sage, S. 74-97.
- Fishbein, Martin; Ajzen, Icek (1975): *Belief, attitude, intention, and behavior. An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Fisher, Paul (2016): Does repeated measurement improve income data quality? *ISER Working Paper Series* 11. Essex: Institute for Social and Economic Research (ISER).

- Friedrichs, Jürgen (1990) [Erstauflage 1973]: *Methoden empirischer Sozialforschung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Galesic, Mirta; Bosnjak, Michael (2009): Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. In: *Public Opinion Quarterly* 73 (2), S. 349-360. DOI: 10.1093/poq/nfp031.
- Giesselmann, Marco; Windzio, Michael (2012): *Regressionsmodelle zur Analyse von Paneldaten*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gobo, Giampietro; Mauceri, Sergio (2014): *Constructing Survey Data: An Interactional Approach*. London [u.a.]: Sage.
- Golembiewski, Robert T.; Billingsley, Keith; Yeager, Samuel (1976): Measuring Change and Persistence in Human Affairs. Types of Change Generated by OD Designs. In: *The Journal of Applied Behavioral Science* 12 (2), S. 133-157. DOI: 10.1177/002188637601200201.
- Gopinath, Mahesh; Nyer, Prashanth U. (2009): The effect of public commitment on resistance to persuasion. The influence of attitude certainty, issue importance, susceptibility to normative influence, preference for consistency and source proximity. In: *International Journal of Research in Marketing* 26 (1), S. 60–68. DOI: 10.1016/j.ijresmar.2008.08.003.
- Goyder, John (1986): Surveys on Surveys: Limitations and Potentialities. In: *Public Opinion Quarterly* 50 (1), S. 27-41.
- Greenacre, Michael; Blasius, Jörg (Hg.) (2006): *Multiple correspondence analysis and related methods*. Boca Raton, FL: Chapman & Hall/CRC.
- Grice, H. Paul (1975): The logic of conversation. In: Peter Cole, Jerry Morgan und John P. Kimball (Hg.): *Syntax and semantics*, Bd. 3. New York: Academic Press, S. 305-315.
- Groves, Robert M. (1987): Research on Survey Data Quality. In: *Public Opinion Quarterly* 51 (2), S. S156-S172.
- Groves, Robert M.; Cialdini, Robert B.; Couper, Mick P. (1992): Understanding the Decision to Participate in a Survey. In: *Public Opinion Quarterly* 56 (4), S. 475-495. DOI: 10.1086/269338.
- Groves, Robert M.; Fowler, Floyd J.; Couper, Mick P.; Lepkowski, James M.; Singer, Eleanor; Tourangeau, Roger (2009): *Survey methodology*. Hoboken, NJ: Wiley.
- Groves, Robert M.; Presser, Stanley; Dipko, Sarah (2004): The Role of Topic Interest in Survey Participation Decisions. In: *Public Opinion Quarterly* 68 (1), S. 2-31. DOI: 10.1093/poq/nfh002.
- Groves, Robert M.; Singer, Eleanor; Corning, Amy (2000): Leverage-Saliency Theory of Survey Participation: Description and an Illustration. In: *Public Opinion Quarterly* 64 (3), S. 299-308.
- Häder, Michael; Kühne, Mike (2010): Mobiltelefonerfahrung und Antwortqualität bei Umfragen. In: *Methoden, Daten, Analysen (mda)* 4 (2), S. 105-125.
- Halpern-Manners, Andrew; Warren, John Robert; Torche, Florencia (2017): Panel Conditioning in the General Social Survey. In: *Sociological Methods & Research* 46 (1), S. 103-124. DOI: 10.1177/0049124114532445.
- Harkness, Janet A. (2003): Questionnaire translation. In: Janet A. Harkness, Fons J.R. Van de Vijver und Peter Ph. Mohler (Hg.): *Cross-cultural survey methods*. Hoboken, NJ: Wiley, S. 35-56.
- Harkness, Janet A.; Villar, Ana; Edwards, Brad (2010): Translation, adaptation, and design. In: Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars E. Lyberg, Peter Ph. Mohler et al. (Hg.): *Survey methods in multinational, multiregional, and multicultural contexts*. Hoboken, NJ: Wiley, S. 115-140.

- He, Jia; Van de Vijver, Fons J.R. (2013): A general response style factor. Evidence from a multi-ethnic study in the Netherlands. In: *Personality and Individual Differences* 55 (7), S. 794-800. DOI: 10.1016/j.paid.2013.06.017.
- He, Jia; Van de Vijver, Fons J.R.; Espinosa, Alejandra Dominguez; Mui, Phoebe H. C. (2014): Toward a unification of acquiescent, extreme, and midpoint response styles. In: *International Journal of Cross Cultural Management* 14 (3), S. 306-322. DOI: 10.1177/1470595814541424.
- Heerwegh, Dirk; Loosveldt, Geert (2008): Face-to-Face versus Web Surveying in a High-Internet-Coverage Population. Differences in Response Quality. In: *Public Opinion Quarterly* 72 (5), S. 836-846. DOI: 10.1093/poq/nfn045.
- Helfferich, Cornelia (2014): Leitfaden- und Experteninterviews. In: Nina Baur und Jörg Blasius (Hg.): *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Springer, S. 559-574.
- Herbst, Susan (1998): *Reading public opinion: how political actors view the democratic process*. Chicago: University of Chicago Press.
- Hess, Stephane; Stathopoulos, Amanda (2013): Linking response quality to survey engagement. A combined random scale and latent variable approach. In: *Journal of Choice Modelling* 7, S. 1-12. DOI: 10.1016/j.jocm.2013.03.005.
- Hilgers, Andrea (1997): *Artefakt und empirische Sozialforschung. Genese und Analyse der Kritik*. Berlin: Duncker und Humblot.
- Holland, Jennifer L.; Christian, Leah Melani (2009): The Influence of Topic Interest and Interactive Probing on Responses to Open-Ended Questions in Web Surveys. In: *Social Science Computer Review* 27 (2), S. 196-212. DOI: 10.1177/0894439308327481.
- Höpner, Martin; Jurczyk, Bojan (2012): Kritik des Eurobarometers. Über die Verwischung der Grenze zwischen seriöser Demoskopie und interessengeleiteter Propaganda. In: *Leviathan* 40 (3), S. 326-349.
- Horn, John L.; McArdle, J. Jack (1992): A practical and theoretical guide to measurement invariance in aging research. In: *Experimental aging research* 18 (3-4), S. 117-144. DOI: 10.1080/03610739208253916.
- Houtkoop-Steenstra, Hanneke (2009): *Interaction and the standardized survey interview. The living questionnaire*. Cambridge: Cambridge University Press.
- Howe, Lauren C.; Krosnick, Jon A. (2017): Attitude Strength. In: *Annual Review of Psychology* 68, S. 327-351. DOI: 10.1146/annurev-psych-122414-033600.
- Johnson, Timothy; Kulesa, Patrick; Cho, Young Ik; Shavitt, Sharon (2005): The Relation Between Culture and Response Styles. In: *Journal of Cross-Cultural Psychology* 36 (2), S. 264-277. DOI: 10.1177/0022022104272905.
- Johnson, Timothy P.; O'Rourke, Diane; Burris, Jane; Owens, Linda (2002): Culture and Survey Nonresponse. In: Robert M. Groves, Don A. Dillman, John L. Eltinge und Roderick J.A. Little (Hg.): *Survey nonresponse*. New York, Chichester: Wiley, S. 55-69.
- Johnson, Timothy P.; Van de Vijver, Fons J.R. (2003): Social Desirability in Cross-Cultural Research. In: Janet A. Harkness, Fons J.R. Van de Vijver und Peter Ph. Mohler (Hg.): *Cross-cultural survey methods*. Hoboken, NJ: Wiley, S. 195-206.
- Jong, Martijn G. de; Steenkamp, Jan-Benedict E. M. (2010): Finite Mixture Multilevel Multidimensional Ordinal IRT Models for Large Scale Cross-Cultural Research. In: *Psychometrika* 75 (1), S. 3-32. DOI: 10.1007/s11336-009-9134-z.

- Kaminska, Olena; McCutcheon, Allan L.; Billiet, Jaak (2010): Satisficing Among Reluctant Respondents in a Cross-National Context. In: *Public Opinion Quarterly* 74 (5), S. 956-984. DOI: 10.1093/poq/nfq062.
- Kankaraš, Miloš; Moors, Guy; Vermunt, Jeroen K. (2011): Testing for Measurement Invariance with Latent Class Analysis. In: Eldad Davidov, Peter Schmidt und Jaak Billiet (Hg.): *Cross-cultural analysis. Methods and applications*. New York: Routledge, S. 359-384.
- Kankaraš, Miloš; Vermunt, Jeroen K.; Moors, Guy (2011): Measurement Equivalence of Ordinal Items. A Comparison of Factor Analytic, Item Response Theory, and Latent Class Approaches. In: *Sociological Methods & Research* 40 (2), S. 279-310. DOI: 10.1177/0049124111405301.
- Kieruj, Natalia D.; Moors, Guy (2010): Variations in Response Style Behavior by Response Scale Format in Attitude Research. In: *International Journal of Public Opinion Research* 22 (3), S. 320-342. DOI: 10.1093/ijpor/edq001.
- Kieruj, Natalia D.; Moors, Guy (2013): Response style behavior. Question format dependent or personal style? In: *Quality & Quantity* 47 (1), S. 193-211. DOI: 10.1007/s11135-011-9511-4.
- Kim, Jibum; Gershenson, Carl; Glaser, Patrick; Smith, Tom W. (2011): The Polls-Trends. Trends In Surveys On Surveys. In: *Public Opinion Quarterly* 75 (1), S. 165-191. DOI: 10.1093/poq/nfq047.
- Knappertsbusch, Felix (2017): Ökologische Validität durch Mixed-Methods-Designs. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 69 (S2), S. 337-360. DOI: 10.1007/s11577-017-0467-9.
- Knowles, Eric S.; Nathan, Kobi T. (1997): Acquiescent Responding in Self-Reports: Cognitive Style or Social Concern? In: *Journal of Research in Personality* 31, S. 293-301. DOI: 10.1006/jrpe.1997.2180.
- Krosnick, Jon A. (1991): Response strategies for coping with the cognitive demands of attitude measures in surveys. In: *Applied Cognitive Psychology* 5 (3), S. 213-236. DOI: 10.1002/acp.2350050305.
- Krosnick, Jon A.; Petty, Richard E. (1995): Attitude Strength: An Overview. In: Richard E. Petty und Jon A. Krosnick (Hg.): *Attitude strength. Antecedents and consequences*. Mahwah, NJ: Lawrence Erlbaum Associates, S. 1-24.
- Laurison, Daniel (2015): The Willingness to State an Opinion. Inequality, Don't Know Responses, and Political Participation. In: *Sociological Forum* 30 (4), S. 925-948. DOI: 10.1111/socf.12202.
- Lazarsfeld, Paul F. (1935): The Art of Asking WHY in Marketing Research: Three Principles Underlying the Formulation of Questionnaires. In: *National Marketing Review* 1 (1), S. 26-38.
- Lazarsfeld, Paul F. (1940): "Panel" Studies. In: *Public Opinion Quarterly* 4 (1), S. 122-128.
- Lazarsfeld, Paul F. (1957): Public Opinion and the Classical Tradition. In: *Public Opinion Quarterly* 21 (1), S. 39-53. DOI: 10.1086/266685.
- Lechner, Clemens M.; Rammstedt, Beatrice (2015): Cognitive ability, acquiescence, and the structure of personality in a sample of older adults. In: *Psychological assessment* 27 (4), S. 1301-1311. DOI: 10.1037/pas0000151.
- Leeuw, Edith de; Heer, Wim de (2002): Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. In: Robert M. Groves, Don A. Dillman, John L. Eltinge und Roderick J.A. Little (Hg.): *Survey nonresponse*. New York, Chichester: Wiley, S. 41-54.
- Le Roux, Brigitte; Rouanet, Henry (2004): *Geometric Data Analysis. From Correspondence Analysis to Structured Data Analysis*. Dordrecht: Springer Science + Business Media, Inc.

- Light, Carole S.; Zax, Melvin; Gardiner, Dwight H. (1965): Relationship of age, sex and intelligence level to extreme response style. In: *Journal of Personality and Social Psychology* 2 (6), S. 907-909.
- Lipps, Oliver (2014): Learning, understanding, and motivation effects on “don’t know” in panel surveys. FORS Working Paper Series 2014-1, Lausanne: FORS.
- Loosveldt, Geert; Storms, Vicky (2008): Measuring Public Opinions About Surveys. In: *International Journal of Public Opinion Research* 20 (1), S. 74-89. DOI: 10.1093/ijpor/edn006.
- Lord, Frederic M.; Novick, Melvin R. (2008) [Erstveröffentlichung 1968]: Statistical theories of mental test scores. Charlotte, NC: Information Age Publishing.
- Lugtig, Peter; Boeije, Hennie R.; Lensvelt-Mulders, Gerty J. L. M. (2012): Change? What Change? In: *Methodology* 8 (3), S. 115-123. DOI: 10.1027/1614-2241/a000043.
- Lyberg, Lars E.; Stukel, Diana M. (2010): Quality Assurance and Quality Control in Cross-National Comparative Studies. In: Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars E. Lyberg und Peter Ph. Mohler (Hg.): *Survey methods in multinational, multiregional, and multicultural contexts*. Hoboken, NJ: Wiley, S. 227-249.
- Mann, Christopher B. (2005): Unintentional Voter Mobilization. Does Participation in Preelection Surveys Increase Voter Turnout? In: *The ANNALS of the American Academy of Political and Social Science* 601 (1), S. 155-168. DOI: 10.1177/0002716205278151.
- Marín, Gerardo; Gamba, Raymond J.; Marín, Barbara V. (1992): Extreme Response Style and Acquiescence among Hispanics. The Role of Acculturation and Education. In: *Journal of Cross-Cultural Psychology* 23 (4), S. 498-509. DOI: 10.1177/0022022192234006.
- Maynard, Douglas W.; Schaeffer, Nora Cate (2002): Standardization and its Discontents. In: Douglas W. Maynard, Hanneke Houtkoop-Steenstra, Nora Cate Schaeffer und Johannes van der Zouwen (Hg.): *Standardization and tacit knowledge. Interaction and practice in the survey interview*. New York: Wiley, S. 3-46.
- Maynard, Douglas W.; Houtkoop-Steenstra, Hanneke; Schaeffer, Nora Cate; van der Zouwen, Johannes (Hg.) (2002): *Standardization and tacit knowledge. Interaction and practice in the survey interview*. New York: Wiley.
- Meisenberg, Gerhard; Williams, Amandy (2008): Are acquiescent and extreme response styles related to low intelligence and education? In: *Personality and Individual Differences* 44 (7), S. 1539-1550. DOI: 10.1016/j.paid.2008.01.010.
- Mishler, Elliot G. (1986): *Research Interviewing. Context and Narrative*. Cambridge: Harvard University Press.
- Mohler, Peter Ph.; Johnson, Timothy P. (2010): Equivalence, Comparability, and Methodological Progress. In: Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars E. Lyberg und Peter Ph. Mohler (Hg.): *Survey methods in multinational, multiregional, and multicultural contexts*. Hoboken, NJ: Wiley, S. 17-29.
- Moors, Guy (2012): The effect of response style bias on the measurement of transformational, transactional, and laissez-faire leadership. In: *European Journal of Work and Organizational Psychology* 21 (2), S. 271-298. DOI: 10.1080/1359432X.2010.550680.
- Mora y Araujo, Manuel (2011): Political Attitudes. In: Bertrand Badie, Dirk Berg-Schlosser und Leonardo Morlino (Hg.): *International Encyclopedia of Political Science*. Los Angeles: Sage, S. 95-99.
- Mussweiler, Thomas (2003): Comparison processes in social judgment. Mechanisms and consequences. In: *Psychological Review* 110 (3), S. 472-489. DOI: 10.1037/0033-295X.110.3.472.

- Nancarrow, Clive; Cartwright, Trixie (2007): Online access panels and tracking research. The conditioning issue. In: *International Journal of Market Research* 49 (5), S. 573-594. DOI: 10.1177/147078530704900505.
- Oksenberg, Lois; Cannell, Charles; Kalton, Graham (1991): New Strategies for Pretesting Survey Questions. In: *Journal of Official Statistics* 7 (3), S. 349-365.
- Oppenheimer, Daniel M.; Meyvis, Tom; Davidenko, Nicolas (2009): Instructional manipulation checks. Detecting satisficing to increase statistical power. In: *Journal of Experimental Social Psychology* 45 (4), S. 867-872. DOI: 10.1016/j.jesp.2009.03.009.
- Oskamp, Stuart; Schultz, P. Wesley (2005): *Attitudes and Opinions*. New York: Psychology Press.
- Paulhus, Delroy L. (1991): Measurement and Control of Response Bias. In: John P. Robinson, Phillip R. Shaver und Lawrence S. Wrightsman (Hg.): *Measures of personality and social psychological attitudes*. San Diego, CA: Academic Press, S. 17-59.
- Perrin, Andrew J.; McFarland, Katherine (2011): Social Theory and Public Opinion. In: *Annual Review of Sociology* 37 (1), S. 87-107. DOI: 10.1146/annurev.soc.012809.102659.
- Petty, Richard E.; Cacioppo, John T. (1996) [Erstveröffentlichung 1981]: *Attitudes and persuasion. Classic and contemporary approaches*. Boulder, CO: Westview Press.
- Pfau-Effinger, Birgit; Smidt, Maike (2011): Differences in women's employment patterns and family policies. Eastern and western Germany. In: *Community, Work & Family* 14 (2), S. 217-232. DOI: 10.1080/13668803.2011.571401.
- Porst, Rolf (2011): *Fragebogen. Ein Arbeitsbuch*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Powell, Martha C.; Fazio, Russell H. (1984): Attitude Accessibility as a Function of Repeated Attitudinal Expression. In: *Personality and Social Psychology Bulletin* 10 (1), S. 139-148. DOI: 10.1177/0146167284101016.
- Rapkin, Bruce D.; Schwartz, Carolyn E. (2004): Toward a theoretical model of quality-of-life appraisal. Implications of findings from studies of response shift. In: *Health and Quality of Life Outcomes* 2 (1), S. 1-12. DOI: 10.1186/1477-7525-2-14.
- Reuband, Karl-Heinz (1990): Meinungslosigkeit im Interview. In: *Zeitschrift für Soziologie* 19 (6), S. 428-443. DOI: 10.1515/zfsoz-1990-0603.
- Reuband, Karl-Heinz (2000): „Pseudo-Opinions“ in Bevölkerungsumfragen: wie die Bürger fiktive Politiker beurteilen. In: *ZA-Information/Zentralarchiv für empirische Sozialforschung* 46, S. 26-38.
- Reynolds, Nina; Smith, Anne (2010): Assessing the Impact of Response Styles on Cross-Cultural Service Quality Evaluation. A Simplified Approach to Eliminating the Problem. In: *Journal of Service Research* 13 (2), S. 230-243. DOI: 10.1177/1094670509360408.
- Riesman, David; Benney, Mark (1956a): The Sociology of the Interview. In: *The Midwest Sociologist* 18 (1), S. 3-15.
- Riesman, David; Benney, Mark (1956b): Asking and Answering. In: *The Journal of Business* 29 (4), S. 225-236.
- Riordan, Christine M.; Richardson, Hettie A.; Schaffer, Bryan S.; Vandenberg, Robert J. (2001): Alpha, Beta, and Gamma Change. A Review of Past Research with Recommendations for New Directions. In: Chester Schriesheim und Linda L. Neider (Hg.): *Equivalence in measurement*. Greenwich, CT: Information Age Publishing, S. 51-97.
- Rippl, Susanne; Seipel, Christian (2008): *Methoden kulturvergleichender Sozialforschung. Eine Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.

- Ross, Catherine E.; Mirowsky, John (1984): Socially-Desirable Response and Acquiescence in a Cross-Cultural Survey of Mental Health. In: *Journal of Health and Social Behavior* 25, S. 189-197. DOI: 10.2307/2136668.
- Schaeffer, Nora Cate (2002): Conversation with a Purpose – or Conversation? Interaction in the Standardized Interview. In: Douglas W. Maynard, Hanneke Houtkoop-Steenstra, Nora Cate Schaeffer und Johannes van der Zouwen (Hg.): *Standardization and tacit knowledge. Interaction and practice in the survey interview*. New York: Wiley, S. 95-124.
- Scheuch, Erwin K. (1965): Sozialer Wandel und Sozialforschung. Über die Beziehungen zwischen Gesellschaft und empirischer Sozialforschung. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 17 (1), S. 1-48.
- Scheuch, Erwin K. (1993) [Erstveröffentlichung 1989]: Theoretical implications of comparative survey research. Why the wheel of cross-cultural methodology keeps on being reinvented. In: *Historical Social Research* 18 (2), S. 172-195. DOI: 10.12759/hsr.18.1993.2.172-195.
- Schmitt, Neal (1982): The Use Of Analysis Of Covariance Structures To Assess Beta And Gamma Change. In: *Multivariate Behavioral Research* 17 (3), S. 343-358. DOI: 10.1207/s15327906mbr1703_3.
- Schmitz, Andreas; Barth, Alice (2018): Objektivierung der Kritik, Kritik der Objektivierung. *Manuskript zur Begutachtung eingereicht*.
- Schnell, Rainer; Hill, Paul Bernhard; Esser, Elke (2005): *Methoden der empirischen Sozialforschung*. München: Oldenbourg.
- Schober, Michael F.; Conrad, Frederick G. (2002): A Collaborative View of Standardized Survey Interviews. In: Douglas W. Maynard, Hanneke Houtkoop-Steenstra, Nora Cate Schaeffer und Johannes van der Zouwen (Hg.): *Standardization and tacit knowledge. Interaction and practice in the survey interview*. New York: Wiley, S. 67-94.
- Schoonees, Pieter C.; van de Velden, Michel; Groenen, Patrick J. F. (2015): Constrained Dual Scaling for Detecting Response Styles in Categorical Data. In: *Psychometrika* 80 (4), S. 968-994. DOI: 10.1007/s11336-015-9458-9.
- Schuman, Howard; Presser, Stanley (1981): *Questions and answers in attitude surveys. Experiments on question form, wording, and context*. San Diego [u.a.]: Academic Press.
- Schwartz, Carolyn E.; Bode, Rita; Repucci, Nicholas; Becker, Janine; Sprangers, Mirjam A. G.; Fayers, Peter M. (2006): The clinical significance of adaptation to changing health. A meta-analysis of response shift. In: *Quality of life research* 15 (9), S. 1533-1550. DOI: 10.1007/s11136-006-0025-9.
- Schwarz, Norbert (2003): Culture-sensitive context effects. A challenge for cross-cultural surveys. In: Janet A. Harkness, Fons J.R. Van de Vijver und Peter Ph. Mohler (Hg.): *Cross-cultural survey methods*. Hoboken, NJ: Wiley, S. 93-100.
- Schwarz, Norbert (2007): Attitude Construction. Evaluation in Context. In: *Social Cognition* 25 (5), S. 638-656. DOI: 10.1521/soco.2007.25.5.638.
- Schwarz, Norbert; Bless, Herbert (1992): Assimilation and Contrast Effects in Attitude Measurement: An Inclusion/Exclusion Model. In: *Advances in Consumer Research* 19 (1), S. 72-77.
- Schwarz, Norbert; Clore, Gerald L. (1983): Mood, misattribution, and judgments of well-being. Informative and directive functions of affective states. In: *Journal of Personality and Social Psychology* 45 (3), S. 513-523. DOI: 10.1037/0022-3514.45.3.513.
- Schwarz, Norbert; Knäuper, Barbel; Hippler, Hans-J.; Noelle-Neumann, Elisabeth; Clark, Leslie (1991): Rating Scales. Numeric Values May Change the Meaning of Scale Labels. In: *Public Opinion Quarterly* 55 (4), S. 570-582. DOI: 10.1086/269282.

- Schwarz, Norbert; Knäuper, Bärbel; Oyserman, Daphna; Stich, Christine (2009): The psychology of asking questions. In: Edith D. de Leeuw, Don A. Dillman und Joop J. Hox (Hg.): *International Handbook of Survey Methodology*. London: CRC Press, S. 18-34.
- Schwarz, Norbert; Strack, Fritz; Müller, Gesine; Chassein, Brigitte (1988): The Range of Response Alternatives May Determine the Meaning of the Question. Further Evidence on Informative Functions of Response Alternatives. In: *Social Cognition* 6 (2), S. 107-117. DOI: 10.1521/soco.1988.6.2.107.
- Sicinski, Adam (1970): "Don't Know" Answers in Cross-National Surveys. In: *Public Opinion Quarterly* 34 (1), S. 126-129.
- Smith, Peter B. (2004): Acquiescent Response Bias as an Aspect of Cultural Communication Style. In: *Journal of Cross-Cultural Psychology* 35 (1), S. 50-61. DOI: 10.1177/0022022103260380.
- Smith, Tom W. (2011): Refining the Total Survey Error Perspective. In: *International Journal of Public Opinion Research* 23 (4), S. 464-484. DOI: 10.1093/ijpor/edq052.
- Sprangers, Mirjam A. G.; Schwartz, Carolyn E. (1999): Integrating response shift into health-related quality of life research. A theoretical model. In: *Social Science & Medicine* 48 (11), S. 1507-1515.
- Steenkamp, Jan-Benedict E. M.; Baumgartner, Hans (1998): Assessing Measurement Invariance in Cross-National Consumer Research. In: *Journal of Consumer Research* 25 (1), S. 78-107. DOI: 10.1086/209528.
- Stensman, Richard (1985): Severely mobility-disabled people assess the quality of their lives. In: *Scandinavian Journal of Rehabilitation Medicine* 17 (2), S. 87-99.
- Struminskaya, Bella (2016): Respondent Conditioning in Online Panel Surveys. In: *Social Science Computer Review* 34 (1), S. 95-115. DOI: 10.1177/0894439315574022.
- Sturgis, Patrick; Allum, Nick; Brunton-Smith, Ian (2009): Attitudes Over Time: The Psychology of Panel Conditioning. In: Peter Lynn (Hg.): *Methodology of Longitudinal Surveys*. New York, Chichester: Wiley, S. 113-126.
- Sturgis, Patrick; Smith, Patten (2010): Fictitious Issues Revisited. Political Interest, Knowledge and the Generation of Nonattitudes. In: *Political Studies* 58 (1), S. 66-84. DOI: 10.1111/j.1467-9248.2008.00773.x.
- Suchman, Lucy; Jordan, Brigitte (1990): Interactional Troubles in Face-to-Face Survey Interviews. In: *Journal of the American Statistical Association* 85 (409), S. 232-241.
- Suchman, Lucy; Jordan, Brigitte (1994): Validity and the Collaborative Construction of Meaning in Face-to-Face Surveys. In: Judith M. Tanur (Hg.): *Questions about questions. Inquiries into the cognitive bases of surveys*. New York: Russell Sage Foundation, S. 241-267.
- Sudman, Seymour; Bradburn, Norman M.; Schwarz, Norbert (1996): *Thinking about answers. The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.
- Terborg, James R.; Howard, George S.; Maxwell, Scott E. (1980): Evaluating Planned Organizational Change. A Method for Assessing Alpha, Beta, and Gamma Change. In: *Academy of Management Review* 5 (1), S. 109-121. DOI: 10.5465/AMR.1980.4288933.
- Thompson, Megan M.; Zanna, Mark P.; Griffin, Dale W. (1995): Let's Not Be Indifferent About (Attitudinal) Ambivalence. In: Richard E. Petty und Jon A. Krosnick (Hg.): *Attitude strength. Antecedents and consequences*. Mahwah, NJ: Lawrence Erlbaum Associates, S. 361-386.
- Toepoel, Vera; Das, Marcel; van Soest, Arthur (2008): Effects of Design in Web Surveys. In: *Public Opinion Quarterly* 72 (5), S. 985-1007. DOI: 10.1093/poq/nfn060.

- Tourangeau, Roger; Rips, Lance J.; Rasinski, Kenneth (2000): *The psychology of survey response*. Cambridge: Cambridge University Press.
- Van de Vijver, Fons J. R.; Leung, Kwok (1997): *Methods and data analysis for cross-cultural research*. Thousand Oaks [u.a.]: Sage.
- Van de Vijver, Fons J. R.; Poortinga, Ype H. (1997): Towards an Integrated Analysis of Bias in Cross-Cultural Assessment. In: *European Journal of Psychological Assessment* 13 (1), S. 29-37. DOI: 10.1027/1015-5759.13.1.29.
- Van Rosmalen, Joost; van Herk, Hester; Groenen, Patrick J.F (2010): Identifying Response Styles. A Latent-Class Bilinear Multinomial Logit Model. In: *Journal of Marketing Research* 47 (1), S. 157-172. DOI: 10.1509/jmkr.47.1.157.
- Van Vaerenbergh, Y.; Thomas, T. D. (2013): Response Styles in Survey Research. A Literature Review of Antecedents, Consequences, and Remedies. In: *International Journal of Public Opinion Research* 25 (2), S. 195-217. DOI: 10.1093/ijpor/eds021.
- Vandenberg, Robert J.; Lance, Charles E. (2000): A Review and Synthesis of the Measurement Invariance Literature. Suggestions, Practices, and Recommendations for Organizational Research. In: *Organizational Research Methods* 3 (1), S. 4-70. DOI: 10.1177/109442810031002.
- Vandenberg, Robert J.; Self, Robin M. (1993): Assessing newcomers' changing commitments to the organization during the first 6 months of work. In: *Journal of Applied Psychology* 78 (4), S. 557-568. DOI: 10.1037/0021-9010.78.4.557.
- Verba, Sidney (1969): The uses of survey research in the study of comparative politics: Issues and Strategies. In: Stein Rokkan, Sidney Verba, Jean Viet und Elina Almasy (Hg.): *Comparative Survey Analysis*. Den Haag: Mouton, S. 56-106.
- Verba, Sidney; Nie, Norman H.; Kim, Jae-On (1978): *Participation and Political Equality: A Seven-Nation Comparison*. Cambridge: Cambridge University Press.
- Vermunt, Jeroen K.; Magidson, Jay (2004): Latent class models. In: David Kaplan (Hg.): *The Sage Handbook of Quantitative Methodology for the Social Sciences*. Thousand Oaks, London, New Delhi: Sage, S. 175-198.
- Walter, Jessica G. (2018): The adequacy of measures of gender roles attitudes. A review of current measures in omnibus surveys. In: *Quality & Quantity* 52 (2), S. 829-848. DOI: 10.1007/s11135-017-0491-x.
- Warner, Uwe; Hoffmeyer-Zlotnik, Jürgen H. P. (2009): Das Konzept des „privaten Haushalts“ in international vergleichenden sozialwissenschaftlichen Umfragen und dessen Operationalisierung. In: Birgit Pfau-Effinger, Sladana Sakac Magdalenic und Christof Wolf (Hg.): *Ansätze und Messkonzepte der international vergleichenden Sozialforschung unter den Bedingungen der Globalisierung*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 207-228.
- Warren, John Robert; Halpern-Manners, Andrew (2012): Panel Conditioning in Longitudinal Social Science Surveys. In: *Sociological Methods & Research* 41 (4), S. 491-534. DOI: 10.1177/0049124112460374.
- Waterton, Jennifer; Lievesley, Denise (1989): Evidence of Conditioning Effects in the British Social Attitudes Panel. In: Daniel Kasprzyk, Greg Duncan, Graham Kalton und M. P. Singh (Hg.): *Panel surveys*. New York: Wiley, S. 319-339.
- Weech-Maldonado, Robert; Elliott, Marc N.; Oluwole, Adetokunbo; Schiller, K. Cameron; Hays, Ron D. (2008): Survey response style and differential use of CAHPS rating scales by Hispanics. In: *Medical care* 46 (9), S. 963-968. DOI: 10.1097/MLR.0b013e3181791924.

- Weichbold, Martin (2009): Zur Bestimmung und Sicherung der „Qualität“ von Umfragen. In: Martin Weichbold, Johann Bacher und Christof Wolf (Hg.): *Umfrageforschung. Herausforderungen und Grenzen*. Wiesbaden: VS Verlag für Sozialwissenschaften / GWV Fachverlage GmbH, S. 553-570.
- Weijters, Bert; Baumgartner, Hans; Geuens, Maggie (2016): The calibrated sigma method. An efficient remedy for between-group differences in response category use on Likert scales. In: *International Journal of Research in Marketing* 33 (4), S. 944-960. DOI: 10.1016/j.ijresmar.2016.05.003.
- Weijters, Bert; Cabooter, Elke; Schillewaert, Niels (2010): The effect of rating scale format on response styles. The number of response categories and response category labels. In: *International Journal of Research in Marketing* 27 (3), S. 236-247. DOI: 10.1016/j.ijresmar.2010.02.004.
- Weijters, Bert; Geuens, Maggie; Schillewaert, Niels (2010): The stability of individual response styles. In: *Psychological Methods* 15 (1), S. 96-110. DOI: 10.1037/a0018721.
- Weijters, Bert; Schillewaert, Niels; Geuens, Maggie (2008): Assessing response styles across modes of data collection. In: *Journal of the Academy of Marketing Science* 36 (3), S. 409-422. DOI: 10.1007/s11747-007-0077-6.
- Weins, Cornelia (2009): Messinstrumente im internationalen Vergleich. Möglichkeiten und Grenzen der Analyse von Vorurteilen. In: Birgit Pfau-Effinger, Sladana Sakac Magdalenic und Christof Wolf (Hg.): *Ansätze und Messkonzepte der international vergleichenden Sozialforschung unter den Bedingungen der Globalisierung*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 129-152.
- Welkenhuysen-Gybels, Jerry; Billiet, Jaak; Cambré, Bart (2003): Adjustment for Acquiescence in the Assessment of the Construct Equivalence of Likert-Type Score Items. In: *Journal of Cross-Cultural Psychology* 34 (6), S. 702-722. DOI: 10.1177/0022022103257070.
- Willis, Gordon B. (2004): *Cognitive Interviewing: A tool for improving questionnaire design*. Thousand Oaks: Sage.
- Wilson, Francis Graham (2013) [Erstveröffentlichung 1962]: *A theory of public opinion*. New Brunswick, N.J., London: Transaction Publishers.
- Wilson, Timothy D.; Hodges, Sarah M. (1992): Attitudes as Temporary Constructions. In: Leonard L. Martin und Abraham Tesser (Hg.): *The construction of social judgments*. Hillsdale, NJ [u.a.]: Erlbaum, S. 37-65.
- Winker, Peter; Menold, Natalja; Porst, Rolf (2013): *Interviewers' deviations in surveys. Impact, reasons, detection and prevention*. Frankfurt am Main: Peter Lang.
- Winkielman, Piotr; Knäuper, Bärbel; Schwarz, Norbert (1998): Looking back at anger. Reference periods change the interpretation of emotion frequency questions. In: *Journal of Personality and Social Psychology* 75 (3), S. 719-728. DOI: 10.1037/0022-3514.75.3.719.
- Yan, Ting; Datta, A. Rupa (2015): Altering the Survey-taking Climate: The Case of the 2010 U.S. Census. In: *Survey Methods: Insights from the Field*. DOI: 10.13094/SMIF-2015-00014.
- Zaller, John; Feldman, Stanley (1992): A Simple Theory of the Survey Response: Answering Questions versus Revealing Preferences. In: *American Journal of Political Science* 36 (3), S. 579-616.
- Zillmann, Doreen; Schmitz, Andreas; Skopek, Jan; Blossfeld, Hans-Peter (2014): Survey topic and unit nonresponse. In: *Quality & Quantity* 48 (4), S. 2069-2088. DOI: 10.1007/s11135-013-9880-y.

6. Artikel

The Changing Nature of Attitude Constructs

An Application of Multiple Correspondence Analysis on Gender Role Attitudes

Alice Barth, University of Bonn

© 2016. Post-print manuscript version made available under the CC-BY-NC-ND 4.0 license
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

The original research article is published as:

Barth, Alice (2016). The Changing Nature of Attitude Constructs - An Application of Multiple Correspondence Analysis on Gender Role Attitudes. *Quality & Quantity*, 50(4), 1507-1523.
doi:10.1007/s11135-015-0218-9

Abstract:

Attitude change is frequently measured by comparing respondents' scores on the same instrument at different points in time. However, there is a variety of methodological challenges in measuring attitude change: when respondents' handling of the questionnaire or their understanding of the items change, the comparability of constructs is threatened. This paper proposes the investigation of systematic methodological variation over time by multiple correspondence analysis. Visualizing respondents' 'cognitive maps' facilitates the exploration of both changes in the underlying structure of attitude constructs – that is, changes in meaning – as well as data quality. The approach is illustrated with the analysis of two item batteries on gender role attitude from the BHPS and the BSA from the beginning of the 1990s to the mid-2000s. Both data sets exhibit similar structural changes - more methodological variation and increasing complexity of the attitude construct. The comparison of latent structures over time provides useful information about the nature of change in social constructs.

1. Introduction

The temporal dynamics of values and attitudes is a topic with increasing relevance in modern social sciences utilizing survey techniques (e.g. Alwin & Scott 1996; Bardi & Goodwin 2011; Meuleman, Davidov & Billiet 2007). Longitudinal analyses rely on a number of implicit assumptions about the nature of change, the very mechanism they often want to grasp. In surveys, attitude change is commonly measured by repeatedly administering the same questionnaire to respondents – when numerical values change significantly over time, it is inferred that attitude change has occurred. In doing so, it is assumed that while respondents' attitudes change, the underlying concept stays the same and can therefore adequately be measured and compared at different times using the same instrument. However, this perception is challenged by the notion that change may take different forms: in addition to changes in level, shifts in the meaning of response categories as well as a reconceptualization of the construct of interest can occur, which compromise the comparability of measurement (Golembiewski, Billingsley & Yeager 1976; Riordan, Richardson, Schaffer & Vandenberg 2001). Furthermore, systematic methodological variation over time may be generated by mere changes in response behaviour, e.g. due to panel conditioning effects (Cantor 2008; Warren & Halpern-Manners 2012) or general public trends, such as a decreasing willingness to cooperate in surveys (Loosveldt & Storms 2008).

Investigating the complexities of attitude change is important for two reasons: Firstly, neglecting methodological variation over repeated measurement occasions can lead to erroneous conclusions about the direction and magnitude of attitude change. In the worst case, the computation of intra-individual trajectories or the comparison of mean levels is inherently meaningless when response behaviour or the meaning of indicators have been subject to changes over time (e.g. Chan 1998; Davidov 2011). Accordingly, the first part of the paper discusses causes and consequences of several methodological challenges in measuring attitude change.

Secondly, it is argued that measurement variance over time is not just a statistical nuisance, but focusing on changes in response patterns and structure of attitude scales can provide useful information both about time-varying patterns in respondents' handling of questionnaires as well as potential changes in the meaning of established indicators and scales. These points are illustrated in the second part of the paper by the application of multiple correspondence analysis (MCA) to a set of gender role attitude items from two renowned British surveys, the British Household Panel Survey (BHPS) and the British Social Attitudes Survey (BSA). Both have repeatedly administered a set of items concerning gender role attitude from the early 1990s to the mid-2000s. By comparing results from a panel to a repeated cross-section survey, we can differentiate between aging, panel experience, cohort succession and period effects as potential causes for change in latent structures.

The use of MCA is proposed as this technique visualizes the most important patterns in the data, thereby facilitating a direct comparison of the underlying structures in a set of items over time. MCA operates without making distributional assumptions, thus it also displays non-linear relationships between items which helps differentiating between substantive and methodological variation (e.g. response styles, misunderstandings of questions etc.). This approach is in line with Golembiewski and colleagues who noted in their threefold typology of change that “scaling techniques less sensitive to metric-level assumptions also could profitably be used to seek underlying structures in this type of analysis” (Golembiewski et al. 1976: 154). It will be demonstrated how the information from MCA maps can be used to draw conclusions about the comparability and validity of constructs over time.

The paper will show that in the example of gender role attitudes, the construct’s structure becomes more complex and the amount of methodological variation considerably increases towards the end of the investigation period. As similar structural changes emerge both in the panel and in the cross-section data set, it is argued that the effect is most likely related to period. From the locations of single response categories in the latent space, it can be inferred that respondents’ understanding of the attitude construct has changed and that they experienced increasing problems in adequately handling the item battery.

2. Methodological challenges of measuring attitude change

In most research designs using survey data, individual as well as aggregate-level attitude change is assessed via a comparison of self-report evaluative items measured at one occasion with the same items captured at some later point(s) in time. As attitudes are mostly operationalized as latent constructs, the use of multiple indicators is generally deemed the most reliable form of measurement (Alwin & Scott 1996; Saris & Gallhofer 2007). Though seldom stated explicitly, change is usually conceptualized as an absolute quantitative variation in a given variable from one measurement occasion to the next. Golembiewski et al. (1976) term this kind of variation ‘alpha change’ under the conditions that the instrument is constantly calibrated and related to a stable construct. If alpha change applies, absolute differences in responses can be meaningfully interpreted as true change (Millsap & Hartog 1988; Riordan et al. 2001). However, there are several challenges to this ideal methodological condition of measuring change.

Firstly, the assessment of ‘true’ change can be impeded due to methodologically induced changes in respondents’ handling of the instrument (Frick, Goebel, Schechtman, Wagner & Yitzhaki 2006; Cantor 2008). In the case of panel surveys, it has been argued that repeated participation can

stimulate respondents' motivation and foster their understanding of questions, evoking more precise or honest answers and less item non-response (Sturgis, Allum & Brunton-Smith 2009; Waterton & Lievesley 1989; Yan & Eckman 2012). On the other hand, panel respondents may become bored of repeatedly answering the same questions, this loss of interest resulting in worse response quality such as avoidance of follow-up questions or non-differentiation (Meurs, van Wissen & Visser 1989; Toepoel, Das & van Soest 2009). Irrespective of whether the quality of responses turns for the better or for the worse, the assessment of true change is impeded by any variation that reflects alterations in response behaviour instead of real changes in respondents' attitudes.

Secondly, attitude change may assume shapes that are more fundamental than a variation in level. Golembiewski et al. (1976) draw attention to the different conceptual facets of change by introducing 'beta' and 'gamma' change as counterparts to the traditional notion of absolute quantitative variation in level or 'alpha change'. Beta change involves a subjective recalibration of some intervals of the measurement instrument by respondents. Thus, the anchoring of an underlying scale does not remain constant over time, which may bias substantive analyses of attitude change (Chan 1998; Riordan et al. 2001). The term gamma change refers to a major conceptual change in the domain of interest, a redefinition of the whole construct under study. While beta change entails stretching or shrinking of scale intervals, under the condition of gamma change a subsequent measure is 'off the scale' because the frame of reference is substantively different (Golembiewski et al. 1976). For example, Lugtig, Boeije & Lensvelt-Mulders (2012) compared study motivation of freshman students in their first week at university and in a follow-up at the end of their first year. It turned out that the hypothesized two-factor structure (internal and external motivation) could be confirmed only in the second wave; on the first measurement occasion, the concept of study motivation was very diffuse. Furthermore, the relative importance of some indicators changed – after some experience with academic life, intrinsic motives of study motivation became more relevant, whereas feelings of introjection were less associated with motivational structures. It is evident that in the presence of gamma change, comparisons of mean values are not meaningful, or, as Vandenberg and Lance (2000:9) put it, they may be “tantamount to comparing apples and spark plugs”. Hence, both beta and gamma change threaten the validity of analyses conducted under the assumption of alpha change (Golembiewski et al. 1976; Riordan et al. 2001). Alterations in respondents' subjective metric or a redefinition of the construct under study mean that neither mean-level differences nor correlations with external variables can be unambiguously attributed to true changes in the population of interest (Horn & McArdle 1992; Vandenberg & Lance 2000). Despite these potentially harmful consequences, few researchers are concerned with the possibility of different types of change. Riordan et al. (2001) reviewed 266

longitudinal studies in organizational research; most of these focused solely on alpha change, whereas 6% examined planned beta/gamma change (e.g. evaluations of the effectiveness of raters' trainings) and another 6% checked for unintentional beta/gamma change.

Thirdly, social transformations can affect the properties of measurement instruments. Whereas the methodological problems hitherto discussed are mainly connected to influences by the research design (repeated interviewing, experimental interventions) or properties of the sample (maturation, aging), similar challenges exist in the assessment of attitude change in the general population by repeated cross-section designs. From a sociological perspective, the formation and change of attitudes as evaluative judgements is related to the social context (Voas 2014). Thus, developments in social, economic or institutional frameworks can effectuate substantive attitude change, but they may also complicate research into the temporal dynamics of attitudes. It is obvious that questions sometimes become outdated due to institutional changes – after the demise of the Soviet Union, surveying attitudes towards communist countries became quite obsolete. In addition, the propensity of measurement instruments to exhibit ceiling or floor effects may increase due to social changes, which inhibits the representation of further change at the upper or lower end of the scale. Societal transformations or formative events may also change the frame of reference in which questions are understood, leading to the phenomenon of gamma change discussed above. For example, the notion of security may change drastically after a terrorist attack.

Fourthly, an issue that relates to methodological problems is the general deterioration of the 'survey climate' in most Western societies. The willingness to cooperate in survey interviews is generally decreasing (Curtin, Presser & Singer 2005; Loosveldt & Storms 2008) while there is a growing distrust of pollsters (Kim, Gershenson, Glaser & Smith 2011). Though there is no consistent relationship between response rates and nonresponse bias (Groves 2006), there are concerns that reluctant respondents provide data of worse quality (Kaminska, Cutcheon & Billiet 2010) as they are less motivated and thus more inclined to 'satisfice' (Krosnick 1991; Krosnick, Narayan & Smith 1996). These considerations demonstrate that methodological challenges in the conceptualization of change do not only occur in experimental studies or panel designs, but relate to the study of attitude change in a larger context.

Finally, substantive changes and the incidence of methodological problems can be interdependent. For example, if a certain item seems increasingly abstract, ambiguous or complicated because its connection to respondents' real life experiences has diminished over the course of time, the cognitive demand to provide a substantive answer becomes higher. Consequently, the propensity to take 'mental shortcuts' such as acquiescence, non-differentiation, saying "don't know" or randomly choosing an answer is likely to increase (Krosnick 1991). Thereby the amount of non-

substantive variation varies over time, which means a change in error structures that can affect substantive conclusions (Smith 2011).

In survey research, the most popular solution for tackling the various methodological problems associated with measuring attitude change is testing for measurement invariance/equivalence (Billiet 2003; Davidov 2008; Millsap 2011). Typically, variants of confirmatory factor analysis (CFA) are used to evaluate whether the same instrument produces equivalent measures across groups or over time (Poznyak, Meuleman, Abts & Bishop 2014; Vandenberg 2002); some researchers also propose item response theory (IRT) (Maurer, Raju & Collins 1998; Meade, Lautenschlager & Hecht 2005). In a review of 75 papers on the assessment of measurement invariance using CFA, Schmitt & Kuljanin (2008) concluded that full measurement invariance seldom holds. Modelling partial invariance is sometimes suggested as a solution (Byrne, Shavelson & Muthén 1989; Steenkamp & Baumgartner 1998); however, this approach may prove cumbersome as well as theoretically elusive when parameter modifications are mainly motivated by statistical instead of theoretical reasons (Vandenberg & Lance 2000).

As the methods used in the context of measurement equivalence primarily seek to establish the condition of sufficiently invariant measures for the assessment of 'true', i.e. alpha change, signs of differential response behaviour or structural changes are basically considered nuisances that hinder comparison over time (Horn & McArdle 1992). However, it can be argued that the very deviations from measurement equivalence provide valuable insights about the changing nature of social constructs. The occurrence of systematic methodological variation may indicate individual or social changes which are worth exploring for a better understanding of what social science instruments really measure and how respondents' perceptions of certain constructs develop. Therefore, we propose the application of multiple correspondence analysis (MCA) for the study of change in attitude constructs, as a methodology which does not aim at assessing the stability of a certain statistical model over time, but is sensitive to methodological changes.

3. Method

MCA is a multivariate scaling technique that visualizes underlying structures in data tables. It transforms a set of categorical observed variables to a smaller set of latent variables and is therefore often referred to as principal component analysis (PCA) with categorical data (Blasius & Thiessen 2012:46; Le Roux & Rouanet 2004:180). Analogous to PCA, MCA provides eigenvalues and factor loadings, but its emphasis is on the geometric representation of data structures. Objects (respondents) and variables (variable categories) are represented as points in a low-dimensional

weighted Euclidean space, so that spatial proximity of categories or individuals indicates similarity and vice versa.

In geometric data analysis, “the model should follow the data, not the inverse” (Jean-Paul Benzécri, cited after Blasius & Greenacre 2006:6). Instead of testing whether the data fit a pre-specified statistical model, MCA visualizes the most important patterns in the data. As it does not restrict the data to be linear, it is especially suitable to detect non-linear relationships that are often associated with methodological effects such as response styles (Blasius & Thiessen 2012; Schoonees, van de Velden & Groenen, in press). MCA is thus more sensitive to structural differences and method-induced variance than the classical methods for the assessment of measurement equivalence, most notably CFA and IRT, which makes it an ideal method for our purpose of investigating the nature of change in attitudes over time.

MCA has been successfully applied to examine the comparability of constructs, e.g. by Blasius and Thiessen (2006) who test for cross-cultural construct equivalence of a set of ISSP items or Fernández, de Rada Igúzquiza, Lautre & Calvo (2012) who compare whether a face-to-face and a telephone survey represent the same universe. König (2010) demonstrates how changing social categories in a changing society can be assessed with multiple or joint correspondence analysis. Apart from investigating construct comparability and change, MCA can be used to assess the quality of data. The geometrical representation of every single response category enables the researcher to detect irregularities which may point to problems such as misunderstanding of items, response styles or processing errors (Blasius & Thiessen 2012). When the input data are ordinal, such as Likert-scaled items, it is an indicator of good response quality when the response categories retain their successive order in the latent space. Further, the maximum of variance should be explained by the substantive concept(s), in contrast to methodological artefacts. Gender role attitude, the example used here, is often conceptualized as a continuum from a traditional role model of the male breadwinner and female housewife to an egalitarian point of view that supports equality in all domains (Mickelson, Claffey & Williams 2006; Read & Grundy 2011; Sweeting, Bhaskar, Benzeval, Popham & Hunt 2014). Consequently, the first latent dimension in MCA is expected to reflect this polarity if the construct is indeed one-dimensional. A different structure, by contrast, would point towards a number of substantive latent dimensions greater than one or the prevalence of methodological artefacts.

The analysis is carried out in two parts: Firstly, changes in the latent structure of gender role attitude items in the British Household Panel Study (BHPS) are analysed. With this data set it is assessed how response structures develop within a stable group of respondents over time, not aiming at representativeness. The MCA solutions of three panel waves (first, ninth and seventeenth) are

discussed. In the second part, the same analysis is carried out for four waves of the British Social Attitudes Survey (BSA), thereby enabling a comparison of structural changes in a sample of long-term panel respondents to repeated cross-section surveys in the same period in Great Britain. Whereas cohort succession and attrition are controlled for in the panel sample, the cross-section sample rules out effects of ageing and panel experience. This allows for a differentiated consideration of potential causes for methodological changes.

4. Data

The first analysis is based on a set of six gender role attitude items from the BHPS. The respective items were asked biennially from 1991 to 2007 in a self-completion questionnaire that was administered after the main face-to-face interview, along with a variety of other attitudinal and health-related questions. Thus, there are nine waves in which the wording and context of the items remained the same. Similar to many other social science surveys, the assessment of gender role attitudes in the BHPS centres on questions of female labour force participation and its consequences for the family. The item battery is balanced with three items that embrace a traditional role assignment and three egalitarian statements (see table 1). In order to rule out effects of non-random attrition, the sample used here is restricted to respondents who took part in at least eight of the nine panel waves. Additionally, proxy respondents were removed from the data set, which leaves us with about 4,700 to 5,000 cases in each wave.

Table 1 Wording of items used in the analysis

Do you personally agree or disagree...	BHPS	BSA
A: A pre-school child is likely to suffer if his or her mother works.	yes	yes
B: All in all, family life suffers when the woman has a full-time job.	yes	yes
C: A woman and her family would all be happier if she goes out to work.	yes	no
D: Both the husband and wife should contribute to the household income.	yes	yes
E: Having a fulltime job is the best way for a woman to be an independent person.	yes	no
F: A husband's job is to earn money; a wife's job is to look after the home and family.	yes	yes
G: A working mother can establish just as warm and secure a relationship with her children as a mother who does not work	no	yes
H: Being a housewife is just as fulfilling as working for pay	no	yes

In the second part, the BHPS solutions are compared to the gender role attitude structure in the BSA, a yearly, nationally representative survey of about 3,000 individuals in Great Britain. Gender role attitudes were part of the BSA in 1989, 1994, 2002 and 2006. They were asked in a self-completion questionnaire which was given to respondents after a face-to-face interview – the same survey technique as in the BHPS. The size of the sample varies between approximately 1,000 and 2,000 respondents as the BSA usually employs three different versions of their questionnaire, so the gender role module was asked either of a random third or two-thirds of the full sample. To ensure comparability, the analysis is also based on a six-item scale that was constantly used on all four time points. While four of the BSA items are exactly the same as in the BHPS (items a, b, d, f), two are slightly different (in the following referred to as g and h, see table 1) but related to female employment and its consequences for the family as well.

The gender role items from both surveys have often been used in the form of a scale ranging from a conservative, male breadwinner model on one end to the support of complete gender equality on the other end (Berrington et al. 2008; Kan 2007; Read & Grundy 2011). The use of a one-dimensional scale has usually been justified by theoretical considerations as well as reasonably high values of Cronbach's alpha. However, it was criticized that in individualized societies, the rejection of a traditional male-breadwinner model is not necessarily synonymous with embracing gender equality in all domains, i.e. that the construct of gender egalitarianism is more complex than just the reverse of traditionalism (Behr, Braun, Kaczmirek & Bandilla 2012; Braun 2008). Comparing the ISSP's item battery on family and gender roles in various countries, Blasius & Thiessen (2006) found that support for single-earner couples does not necessarily imply opposition towards dual-earner households in a substantial number of countries, including Britain. Accordingly, one aim of the present analysis is to test the factorial structure of the gender role items in the BHPS and BSA and to assess whether the dimensionality of gender role attitude remains the same throughout the investigation period.

Though the item battery or parts of it have often served the purpose of comparing gender role attitude over time (e.g. Crompton, Brockman & Lyonette 2005; Kan 2007; Sweeting et al. 2014), construct comparability and response quality over time have not been assessed so far using the data from Great Britain. A recurring finding is that the variables' power of differentiation seems to have slightly diminished over time: Kan (2007) reports that the discrepancy in gender-role attitudes among women with different status in the labour force narrows between 1991 and 1999, Berrington et al. (2008) indicate that the effect of gender role attitude on manifest variables such as female labour force participation is decreasing in later waves and Crompton et al. (2005) state that the

significant association between gender role attitude and division of domestic labour is present in 1994, but disappears in 2002. Whereas the authors mainly diagnose a diminishing link between attitudes and behaviour, methodological changes in conceptualization or response behaviour concerning the gender role item battery could be an alternative explanation for these findings.

5. Results

As a first approximation towards the temporal dynamics of the gender role attitude, table 2 reports some properties of the items in the analysis. A sum score of the respective six items in the BHPS and BSA was constructed, using the mean of the scores (egalitarian items were reverse coded). The possible range is 1 to 5; a higher score indicates a more egalitarian attitude. The sample mean in the BHPS is stable over the period of 17 years, whereas in the BSA egalitarian positions increase from 1989 to 1994 and remain static afterwards. In both surveys, the sum score's standard deviation is declining. This trend suggests that extreme positions decrease as time progresses, which could be a first hint towards a change in meaning of the items or differences in response behaviour. The factor analytical reliability coefficient, Cronbach's alpha, is between .76 and .69, which is similar to other studies assessing this kind of gender role scale (e.g. Mickelson et al. 2006; Van de Vijver 2007; Sweeting et al. 2014). The scale is thus deemed to exhibit acceptable internal consistency.

Table 2 Properties of the six-item gender role scale over time

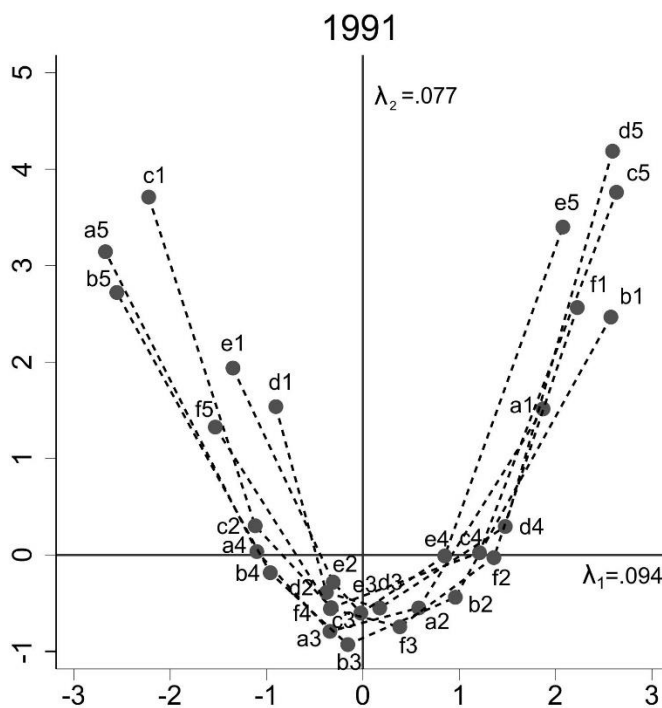
Survey	BHPS			BSA			
	1991	1999	2007	1989	1994	2002	2006
N	4,728	5,062	4,702	1,187	877	1,767	1,621
Mean (sum score)	3.14	3.13	3.11	3.16	3.3	3.3	3.31
SD (sum score)	.65	.61	.57	.76	.73	.70	.62
Cronbach's alpha	.71	.71	.69	.76	.75	.71	.70

Note that these measures are only valid under the assumptions of metric level of measurement, alpha change and one-dimensionality of the scale. The following analyses of the six items using MCA demonstrates, inter alia, that there are serious doubts as to whether these assumptions really hold.

5.1. Effects in a sample of long-term panel respondents (BHPS)

Fig. 1 shows the first and second dimension of the analysis of 4,728 long-term respondents in the BHPS' first wave in 1991. Each response category is represented by a dot (e.g. the dot labelled 'b1' represents strong agreement with statement b: "All in all, family life suffers when the woman has a full-time job"). To simplify interpretation, the successive categories of each item have been connected by lines. In the case at hand, all items retain the ordinal order of their categories when projected onto the first (horizontal) axis, thus showing consistent response patterns. While the input data was not restricted with regard to measurement level, the outcome proves that the internal structure of the six items is indeed ordinal (though not metric – note the differences in distance on dimension 1, for example the distance between a4-a5 is twice as large as a3-a4).

Fig. 1 MCA map of the first two dimensions of gender role attitude items in 1991 (1st wave)



The main interpretation of the MCA map¹ is carried out by evaluating the contributions of the category points to each axis. The first dimension (horizontal axis) mirrors the contrast between strong support of female occupation on the negative part and a conservative role model on the positive part. For example, c1 (strong agreement towards "a woman and her family would all be happier if she goes out to work") is located close to a5 and b5 (strong disagreement towards "a pre-school child/family life suffers if the mother works") on the left. On the far right, the opposite

¹ MCA also provides a numerical solution, which is not shown here as the essential information is conveyed by the respective figures.

categories can be found: strong agreement towards the items promoting a ‘male breadwinner’-model (a1, f1, b1) and strong disagreement towards the egalitarian statements (c5, d5, e5). The moderate categories are located near the origin in the correct order from left to right. Thus, the item set on gender role attitude in the first panel wave proves to be a one-dimensional construct.

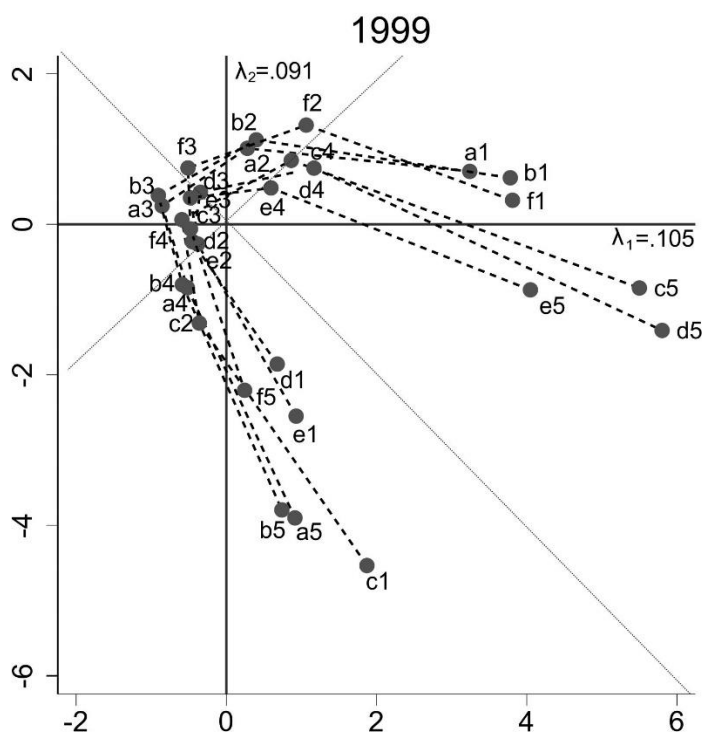
The second dimension, when projected on the vertical axis, separates extreme values on top from moderate responses below the cross of the axes. The resulting parabolic shape or ‘horse shoe’ is a typical structure in MCA when analyzing ordinal data that has a one-dimensional structure. It is a method-induced effect that reflects the associations between the single categories of the data, in our case confirming once more that the items are articulated around a hierarchical scale (for more details on the horseshoe effect, see Greenacre 1984; Le Roux & Rouanet 2004).

Apart from the examination of the dimensions, distances are a central aspect of interpretation. The more distant a category point is located from the centroid, the less it represents the average profile and the more it differentiates between respondents. Accordingly, it can be concluded that strong disagreement with the statement that a pre-school child is likely to suffer when his or her mother works (a5) is considered more extreme than, for example, strong agreement that both partners should contribute to the household income (d1), although they represent the same attitudinal direction. Finally, the explanatory power of the dimensions is regarded, which is conceptually similar to eigenvalues in factor analysis. Applying the variance adjustment suggested by Greenacre (2006), the inertia explained by the first dimension is $\lambda_1 = 0.094$, which means that it accounts for 39% of the total variance ($\lambda_T = 0.241$). The second dimension captures another 32%, while the explanatory power of the third dimension is considerably lower (7.5%). Therefore, only the first two dimensions are shown here. All in all, the geometric space of gender role attitude in the first panel wave indicates high response quality and a one-dimensional construct: the maximum of variance is explained by a single substantive factor, there are no striking anomalies in response patterns and the order of the Likert categories is perfectly retained.

In Fig. 2, results from the 9th wave (N=5,062) are shown. Here, the picture is quite a different one; it is apparent that the distinctive parabolic shape has changed its direction. Accordingly, the interpretation of the axes is different. Whereas the positive part of the first dimension is again mainly determined by a strongly conservative view on gender roles (a1, b1, f1, e5, c5, d5), the negative part is dominated by moderate answers, especially the category 3 (“neither-nor”). Projecting the answer categories which are expected to form the egalitarian end of the scale (a5, b5, c1, d1, e1, f5) onto the x-axis shows that they are located on the positive part as well, next to their immediate counterparts. Thus, the order of the categories is no longer maintained in the first dimension. The second dimension mirrors the contrast between strongly egalitarian views at the

bottom and moderately conservative attitudes (a2, b2, c4, d4, e4, f2) as well as middle categories in the positive part. In contrast to the one-dimensional result in the first wave, a two-dimensional space is now needed to adequately reproduce respondents' gender role attitude. Accordingly, the explained variance of the first dimension has decreased to 34%, the second dimension accounts for 29%. If we consider both dimensions simultaneously and superimpose a diagonal axis through the cross of the axes (see the dotted line from bottom left to top right), we can roughly identify the previous order of the categories by projecting them onto that line². The contrast between moderate and extreme responses can be captured by another line orthogonal to the first one.

Fig. 2 MCA map of first two dimensions of gender role attitude items in 1999 (9th wave)

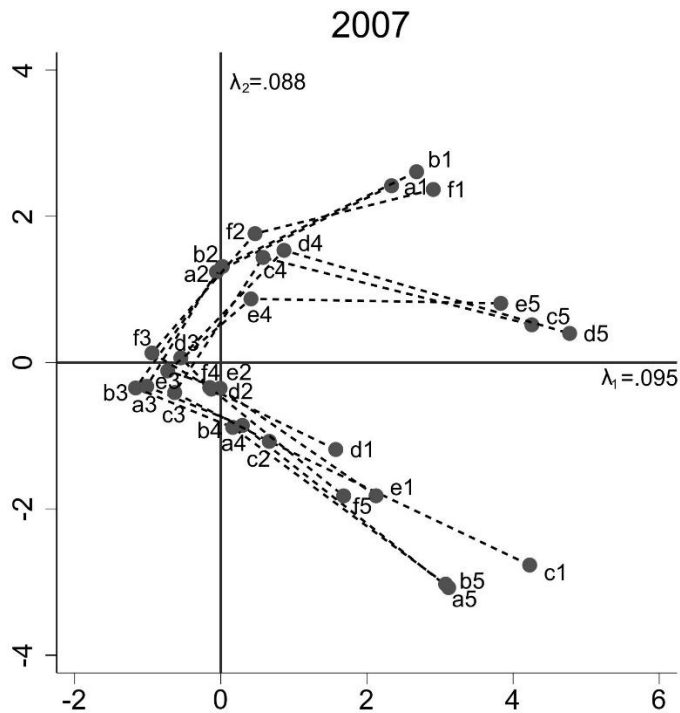


A closer look at the second dimension indicates attitudinal diversification: the cluster of strong agreement with the items evoking a conservative division of labour (a1, b1, f1) is separated from strong disagreement with the egalitarian items (e5, c5, d5): respondents who are very convinced that it is a woman's duty to stay at home do not necessarily disagree with the notion that a job can have positive consequences for a woman and her family. Following the line of thought of Blasius and Thiessen (2006) and Braun (2008), the observed change in response structure can be interpreted as a substantive effect: support for single-earner structures has become more independent from the rejection of dual-earner couples, i.e. the concept of gender roles has become more complex compared to 1991. However, there may also be a methodological effect involved:

² This additional axis is not obtained through statistical rotation; it is just a visual aid to interpretation.

the fact that extreme responses of both egalitarian and traditional role concept are located opposite moderate categories in the first dimension suggests that response styles such as acquiescence or mid-point responding are more prevalent than in 1991. Indeed the frequency of ‘straight-lining’ has increased: whereas only 1.8% of respondents chose the same answer category in all six statements in the first wave (remember that the scale is balanced, so this is very probably a non-substantive response pattern), in wave 9 there are 4.9% who did so, most of which used only the middle category.

Fig. 3 MCA map of first two dimensions of gender role attitude items in 2007 (17th wave)



In the third example, the MCA map of the 17th wave (Fig. 3), the response structure had changed again and even more pronounced than in 1999. In 2007 (N=4,702), the first dimension is entirely defined by the contrast between moderate categories on the left and strong agreement as well as strong disagreement towards all items on the right. Thus, the maximum of variance is no longer explained by substantive content, but rather by the manner of responding: the biggest difference between respondents is whether they tend to choose extremes in contrast to moderate values. The appearance of a ‘method factor’ on the first dimension raises serious doubts about the quality of the data, suggesting that a substantial number of respondents had problems in responding to the items in a substantively meaningful manner. The second dimension approximately mirrors the consecutive order of the categories, showing that at least the second-most part of the variation is still captured by the intended scale, i.e. respondents’ position towards female employment and the division of labour in the family. Yet, the cluster of categories which represent strong disagreement

towards positive effects of female labour participation (c5, d5, e5) does not perfectly fit into the ordinal structure of the y-axis and moved even further away from the rest of the ‘traditional’ end of the scale (a1, b1, f1). Thus, the tendency towards a two-dimensional gender role concept that was already visible in 1999 has intensified.

Considering the distribution of principal inertia, it becomes apparent that higher dimensions have become more important. The first two dimensions taken together merely cover 58.5% of the total variance (first: 30.5%, second: 28%), whereas the third dimension accounts for 15.5%, a doubling compared to the first wave of data collection. The third dimension mainly contrasts categories 1, 3 and 5 with 2 and 4, thereby representing another non-substantive response pattern (not shown). The number of straight-liners has increased again, amounting to 7.5% in 2007.

Comparing the results of these three waves, it is obvious that response structures are not equivalent over time³. Though no considerable changes are visible in the mean values of the scale and the alpha coefficients, MCA has revealed structural transformations in the data. Whereas the hypothesized gender-role scale was perfectly reproduced in a one-dimensional solution in the early 1990s, later waves show increasing explanatory power of the differentiation between moderate and extreme categories and a decomposition of the one-dimensional structure. As the visualization of response categories with MCA can be described as a representation of respondents’ ‘cognitive map’ (Blasius & Thiessen 2012: 11), we conclude that respondents’ perception of gender role attitude, or at least their handling of the scale, has undergone serious changes over the period of investigation, resulting in a considerable loss in data quality.

In the light of our theory, there are two possible explanations for this development. On the one hand, it might essentially represent a deterioration in response quality due to characteristics of the sample of long-term respondents. Respondents’ motivation might decrease with increasing panel experience, resulting in a less diligent selection of response categories, i.e. an increase in satisficing behaviour (Krosnick 1991). In addition, the sample is aging, which could also contribute to the increase in method-induced variation. Older people often show a higher tendency towards response styles than younger respondents (Andrews & Herzog 1986; Weijters, Geuens & Schillewaert 2010). In this case, the meaning of the items has not changed essentially, but they are treated differently over time by the sample of panel respondents.

³ The MCA maps of the waves in between are not shown here, but yielded similar results that confirm the hypothesis of a development towards the predominance of method-induced variance. It has to be kept in mind that we are looking at the same respondents at the three time points shown here, so that the change cannot be attributed to sampling variance.

On the other hand, the structural change of the item battery in this example could be caused by a substantive conceptual change of gender role attitude within the population (or specific strata), reflecting a change in social reality. In the period of investigation, women's participation in the labour market has increased and family structures have undergone changes. These developments were accompanied by legislative reforms and changes in welfare policies, all of which may directly or indirectly affect people's attitudes towards gender roles. Accordingly, the structural changes in the data could be interpreted as a reflection of social transformations which have led to a reconceptualization of gender role attitudes into a more complex, multidimensional construct. As a further consequence of an increasing mismatch between the intended properties of the measurement instrument and social reality, respondents' handling of the item battery might be affected, leading to the observed increase in method-induced variation. Following this reasoning, the decrease in data quality over time points to a change in the properties of the measurement instrument rather than a sample-induced effect.

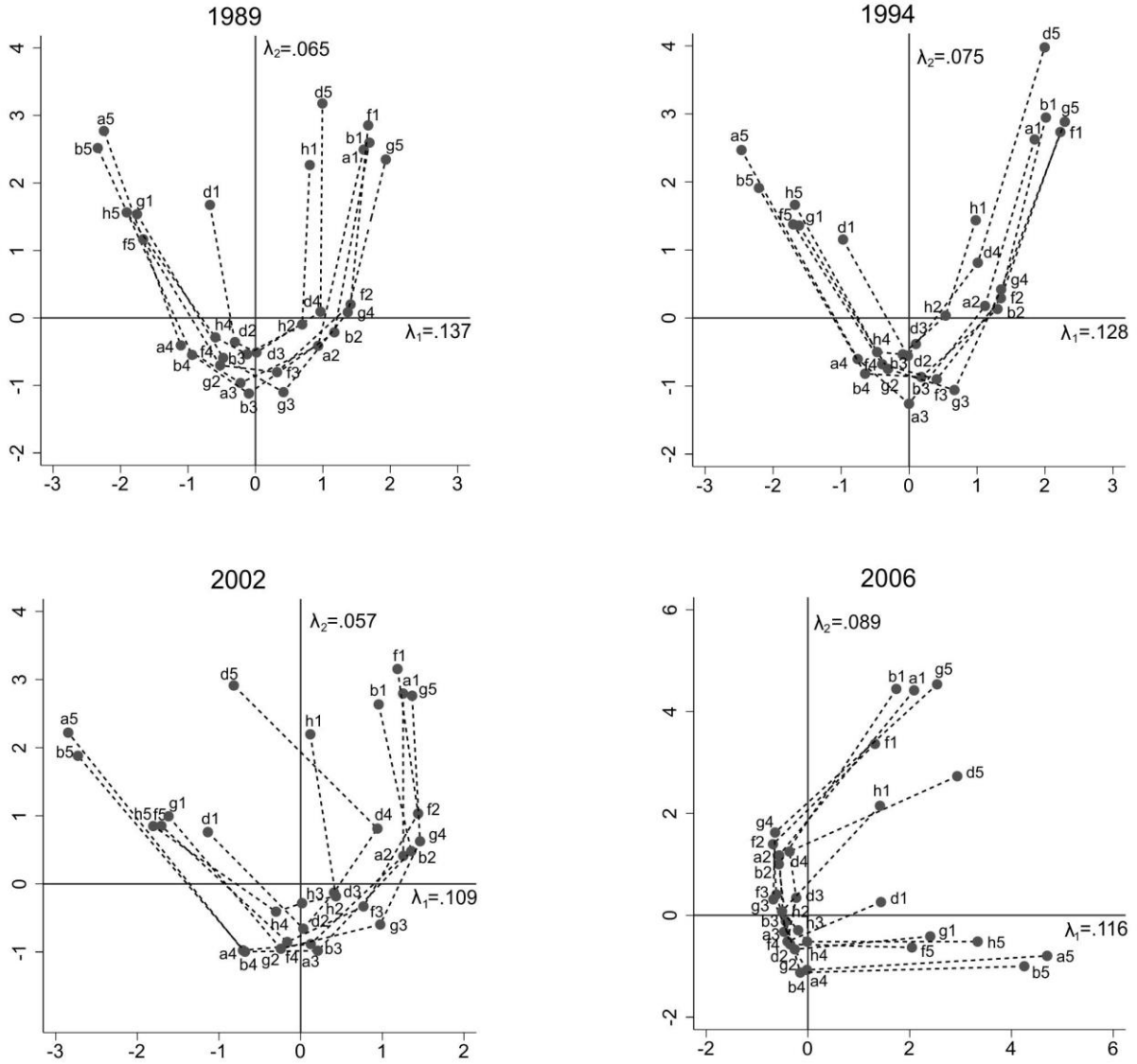
In order to find out which of the two explanations holds, the next step consists of a comparison of the development of gender role attitude structure within the BHPS to the BSA, a repeated cross-sectional survey in the UK. If there are no structural changes visible in the cross-section data over time, we can conclude that the effects observed in the BHPS are due to characteristics of the sample. However, if similar effects emerge in both samples, panel experience and aging can be rejected as reasons for structural changes and the existence of a period effect has to be considered as the most probable explanation for the observed decrease in data quality.

5.2. Effects in a cross-section survey (BSA)

The MCA maps of the four measurement points in the BSA are shown in Fig. 4. From the graphical solution, one can tell immediately that there is a change in structure similar to the one in the panel sample. The interpretation of the axes is carried out equivalently to the previous example. Beginning with a largely one-dimensional latent structure in 1989, where the first principal axis represents the latent gender role scale, the picture becomes increasingly contorted. In 2006, the most part of the existing variation is again explained by the difference between moderate and extreme responses, while the substantive scale is mainly reflected in the second dimension.

The increase in methodological variance is also reflected in the share of total inertia explained by the respective dimensions: In 1989, the inertia of the first dimension is $\lambda_1 = 0.137$, which means that it accounts for 52% of the total variance ($\lambda_1 = 0.266$); the second dimension, which mainly expresses methodological variance, explains a further 25% ($\lambda_2 = 0.065$).

Fig. 4 MCA maps of first two dimensions of gender role attitude items in the BSA in 1989, 1994, 2002 and 2006



In 2006, the explanatory power of the difference between extreme and moderate responses, now reflected on the first dimension, has increased to 39%, while the substantive scale on the second dimension only accounts for 30% of total inertia ($\lambda_T = 0.299$). Moreover, the differentiation of gender role attitude that was diagnosed in the BHPS is visible in the BSA as well. Whereas the items a, b, f and g cluster together, the items d and h increasingly deviate from the rest at the strongly traditional end of the scale, suggesting different response patterns that do not fit in with a one-dimensional scale. The beginning of this process can be observed in 2002 already, corroborating the hypothesis that the structural change in gender-roles is not a contingent phenomenon, but a directional trend.

The fact that differentiation of attitude and decrease in data quality occur in both panel and cross-section study means that our first explanation has to be discarded in favour of the second one: the structural change is not primarily due to characteristics of the sample but is most likely a period effect instead. Irrespective of panel experience, respondents' handling of the gender role attitude scale changes systematically over time. There are indications of changes in respondents' 'cognitive maps' as well as an increasing prevalence of method artefacts in both data sets.

6. Discussion

This study focused on investigating the nature of change in attitude constructs over time. Arguing that in addition to change in level, attitude measurement may be subject to changes both in construct and response behaviour over time, MCA was proposed as a method to investigate methodological changes in the latent structure of attitudes measured by a set of Likert items. As an example, the changing structure of gender role attitude in Great Britain was assessed using data from both a panel survey (BHPS) and a repeated cross-section study (BSA). The analysis demonstrated that a considerable structural change, accompanied by an increase in methodological variation, has occurred in the attitude construct over the course of 17 years. Statistical analyses that compare gender role attitude levels over time should be exercised with a lot of caution as substantive conclusions may be harmed by the diversification of the attitude structure and the growing importance of non-linear relationships in the data. However, focusing on the interpretation and comparison of latent structures with MCA demonstrated that measurement variance is not simply a statistical nuisance, but an important source of information about the nature of change in social constructs.

Whereas gender role attitude was one-dimensional at the beginning of the 1990s in our example, in the mid-2000s the primary difference between respondents was no longer their stance on gender roles, but whether they were more inclined to pick extreme or moderate values, i.e. a proliferation of non-substantive response patterns. Further, a differentiation of gender role attitude was diagnosed, suggesting that respondents' 'cognitive maps' have become more complex. The fact that similar effects have occurred in a sample of long-term panel respondents as well as in a cross-sectional survey leads to the conclusion that neither cohort succession nor attrition, aging or panel conditioning are to be considered primary causes. By the use of data from internationally renowned surveys, one can also be fairly confident that neither survey architecture nor institutional practices varied over the years in a way that could account for the observed effects. Therefore, a period effect seems to be the most parsimonious explanation in this case. Taking into account the societal changes in the realm of gender roles, such as an increase in women's labour market participation,

socio-political actions to improve the compatibility of job and family and a wide public debate on the gendered division of labour, it is plausible that attitudinal structures have become more complex. The increase in methodological variation, which suggests that respondents had more and more problems in handling the item battery, is probably connected to the rising complexity of the concept as respondents did not find their attitudes adequately represented anymore by items that were originally designed to measure a one-dimensional construct.

The methodology presented here can be used to study changes in any item battery of ordered categorical data. Considering that gender relations is but one example of domains where social, institutional and economic frameworks are changing at a rapid pace, affecting the structure of people's values and attitudes, and that the willingness to cooperate in surveys is generally declining, research on methodological changes in attitude constructs over time is warranted in order to prevent nonsensical conclusions and to enhance scientific knowledge on possible changes in the properties of established instruments. While we analysed respondents' perceptions of attitude constructs at the macrolevel, we also agree with Lugtig et al. (2012) who propose that longitudinal studies would profit from a mixed-methods approach for studying the reasons for measurement variance over time at the microlevel.

7. References

- Alwin, Duane F. & Jacqueline Scott (1996). Attitude change: its measurement and interpretation using longitudinal surveys. In: Taylor, Bridget & Katarina Thomson (eds.). *Understanding Change in Social Attitudes*. Aldershot: Dartmouth. 75-106.
- Andrews, Frank M. & Regula A. Herzog (1986). The Quality of Survey Data as Related to Age of Respondent. *Journal of the American Statistical Association*, 81(394), 403-410.
- Bardi, Anat & Robin Goodwin (2011). The dual route to value change: Individual processes and cultural moderators. *Journal of cross-cultural psychology*, 42(2), 271-287.
- Behr, Dorothée, Michael Braun, Lars Kaczmarek & Wolfgang Bandilla (2012). Testing the Validity of Gender Ideology Items by Implementing Probing Questions in Web Surveys. *Field Methods*, 25(2), 124-141.
- Berrington, Ann, Yongjian Hu, Peter W. F. Smith & Patrick Sturgis (2008). A graphical chain model for reciprocal relationships between women's gender role attitudes and labour force participation. *Journal of the Royal Statistical Society*, 171, 89-108.
- Billiet, Jaak (2003). Cross-cultural equivalence with structural equation modeling. In: Harkness Janet, Fons Van de Vijver & Peter Mohler (Eds.). *Cross-cultural survey methods*. New York: Wiley. 247-265.
- Blasius, Jörg & Michael Greenacre (eds.) (2006). *Multiple Correspondence Analysis and Related Methods*. London: Chapman & Hall.
- Blasius, Jörg & Victor Thiessen (2006). Assessing Data Quality and Construct Comparability in Cross-National Surveys. *European Sociological Review*, 22(3), 229-242.
- Blasius, Jörg & Victor Thiessen (2012). *Assessing the Quality of Survey Data*. London: Sage.
- Braun, Michael (2008). Using Egalitarian Items to Measure Men's and Women's Family Roles. *Sex Roles*, 59, 644-656.
- Byrne, Barbara M., Richard J. Shavelson & Bengt Muthén (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105, 456-466.
- Cantor, David (2008). A review and summary of studies on panel conditioning. In: Menard, Scott W. (ed.). *Handbook of longitudinal research*. Burlington, MA: Elsevier/Academic Press. 123-138.
- Chan, David (1998). The Conceptualization and Analysis of Change over Time: An Integrative Approach Incorporating Longitudinal Mean and Covariance Structures Analysis (LMACS) and Multiple Indicator Latent Growth Modeling (MLGM). *Organizational Research Methods*, 1(4), 412-483.
- Crompton, Rosemary, Michaela Brockmann & Clare Lyonette (2005). Attitudes, women's employment and the domestic division of labour: a cross-national analysis in two waves. *Work, Employment and Society*, 19(2), 213-233.
- Curtin, Richard, Stanley Presser & Eleanor Singer (2005). Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly*, 69(1), 87-98.
- Davidov, Eldad (2008). A Cross-Country and Cross-Time Comparison of the Human Values Measurements with the Second Round of the European Social Survey. *Survey Research Methods*, 2(1), 33-46.
- Davidov, Eldad (2011). Nationalism and constructive patriotism: A longitudinal test of comparability in 22 countries with the ISSP. *International Journal of Public Opinion Research*, 23(1), 88-103.

- Fernández, Elena Abascal, Vidal Díaz de Rada Igúzquiza, Ignacio Garcia Lautre & M. Isabel Landaluce Calvo (2012). Face to face and telephone surveys in terms of sampling representativeness: a multidimensional analysis. *Quality & Quantity*, 46, 303-313.
- Frick, Joachim R., Jan Goebel, Edna Schechtman, Gert Wagner & Shlomo Yitzhaki (2006). Using Analysis of Gini (ANoGi) for Detecting Whether Two Sub-Samples Represent the Same Universe: The SOEP Experience. *Sociological Methods and Research*, 34(4), 427-468.
- Golembiewski, Robert T., Keith Billingsley & Samuel Yeager (1976). Measuring Change and Persistence in Human Affairs: Types of Change Generated by OD Designs. *Journal of Applied Behavioral Science*, 12, 133-157.
- Greenacre, Michael J. (1984). *Theory and Applications of Correspondence Analysis*. London: Academic Press.
- Greenacre, Michael J. (2006). From simple to multiple correspondence analysis. In: Blasius, Jörg & Michael Greenacre (eds.). *Multiple Correspondence Analysis and Related Methods*. London: Chapman & Hall. 41-76.
- Groves, Robert M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, 70(5), 646-675.
- Horn, John L. & John J. McArdle (1992). A Practical and Theoretical Guide to Measurement Invariance in Aging Research. *Experimental Aging Research*, 18(3), 117-144.
- Kaminska, Olena, Allan L. McCutcheon & Jaak Billiet (2010). Satisficing among reluctant respondents in a cross-national context. *Public Opinion Quarterly*, 74(5), 956-984.
- Kan, Man Yee (2007). Work orientation and wives' employment careers. *Work and Occupations*, 34(4), 430-462.
- Kim, Jiburn, Carl Gershenson, Patrick Glaser & Tom W. Smith (2011). The Polls – Trends in Surveys On Surveys. *Public Opinion Quarterly*, 75(1), 165-191.
- Konig, Ruben P. (2010). Changing social categories in a changing society: studying trends with correspondence analysis. *Quality & Quantity*, 44, 409-425.
- Krosnick, Jon A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3), 213-236.
- Krosnick, Jon A., Sowmya Narayan & Wendy R. Smith (1996). Satisficing in surveys: Initial evidence. *New directions for evaluation*, 1996(70), 29-44.
- Le Roux, Brigitte & Henry Rouanet (2004). *Geometric Data Analysis – From Correspondence Analysis to Structured Data Analysis*. Dordrecht, Boston, London: Kluwer Academic Publishers.
- Loosveldt, Geert, & Vicky Storms (2008). Measuring public opinions about surveys. *International Journal of Public Opinion Research*, 20(1), 74-89.
- Lugtig, Peter, Hennie R. Boeije & Gerty J. L. M. Lensvelt-Mulders (2012). Change? What Change? An Exploration of the Use of Mixed-Methods Research to Understand Longitudinal Measurement Variance. *Methodology*, 8(3), 115-123.
- Maurer, Todd J., Naumbury S. Raju & William C. Collins (1998). Peer and subordinate performance appraisal measurement equivalence. *Journal of Applied Psychology*, 83(5), 693-702.
- Meade, Adam W., Gary J. Lautenschlager & Janet E. Hecht (2005). Establishing Measurement Equivalence and Invariance in Longitudinal Data with Item Response Theory. *International Journal of Testing*, 5(3), 279-300.

- Meuleman, Bart, Eldad Davidov & Jaak Billiet (2009). Changing attitudes toward immigration in Europe, 2002–2007: A dynamic group conflict theory approach. *Social Science Research*, 38(2), 352–365.
- Meurs, Henk, Leo van Wissen & Jacqueline Visser (1989). Measurement biases in panel data. *Transportation*, 16, 175–194.
- Mickelson, Kristin D., Sharon T. Claffey & Stacey L. Williams (2006). The Moderating Role of Gender and Gender Role Attitudes on the Link between Spousal Support and Marital Quality. *Sex Roles*, 55, 73–82.
- Millsap, Roger E. & Sandra B. Hartog (1988). Alpha, beta, and gamma change in evaluation research: A structural equation approach. *Journal of Applied Psychology*, 73(3), 574–584.
- Millsap, Roger E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.
- Poznyak, Dmitriy, Bart Meuleman, Koen Abts & George F. Bishop (2014). Trust in American government: longitudinal measurement equivalence in the ANES, 1964–2008. *Social Indicators Research*, 118(2), 741–758.
- Read, Sanna & Emily Grundy (2011). Mental health among older married couples: the role of gender and family life. *Journal of Social Psychiatry and Psychiatric Epidemiology*, 46(4), 331–341.
- Riordan, Christine M., Hettie A. Richardson, Bryan S. Schaffer & Robert J. Vandenberg (2001). Alpha, Beta and Gamma Change – A Review of Past Research with Recommendations for New Directions. In: Schriesheim, Chester & Linda L. Neider (Eds.). *Equivalence in Measurement*. IAP. 51–97.
- Saris, Willem E., & Irmtraud N. Gallhofer (2007). *Design, evaluation, and analysis of questionnaires for survey research*. John Wiley & Sons.
- Schmitt, Neal & Goran Kuljanin (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, 18, 210–222.
- Schoonees, Pieter C., Michel van de Velden & Patrick J. F. Groenen (in press). Constrained Dual Scaling for Detecting Response Styles in Categorical Data. *Psychometrika*.
- Smith, Tom W. (2011). Refining the total survey error perspective. *International Journal of Public Opinion Research*, 23(4), 464–484.
- Steenkamp, Jan-Benedict E., & Baumgartner, Hans (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25(1), 78–107.
- Sturgis, Patrick, Nick Allum & Ian Brunton-Smith (2009). Attitudes Over Time - The Psychology of Panel Conditioning. In: Peter Lynn (ed.). *Methodology of longitudinal surveys*. Chichester, UK: John Wiley & Sons. 113–126.
- Sweeting, Helen, Abhita Bhaskar, Michaela Benzeval, Frank Popham, & Kate Hunt (2014). Changing gender roles and their implications for well-being around the new millennium. *Social Psychiatry and Psychiatric Epidemiology*, 49, 791–809.
- Toepoel, Vera, Marcel Das & Arthur van Soest (2009). Effects of Design in Web Surveys: Comparing Trained and Fresh Respondents. *Public Opinion Quarterly* 72(5), 985–1007.
- Vandenberg, Robert J. (2002). Toward a further understanding of and improvement in measurement invariance methods and procedures. *Organizational Research Methods*, 5, 139–158.
- Vandenberg, Robert J. & Charles E. Lance (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–70.

Van De Vijver, Fons J. R. (2007). Cultural and gender differences in gender-role beliefs, sharing household task and child-care responsibilities, and well-being among immigrants and majority members in the Netherlands. *Sex Roles*, 57 (11-12), 813-824.

Voas, David (2014). Towards a Sociology of Attitudes. *Sociological Research Online*, 19(1), doi: 10.5153/sro.3289.

Warren, John Robert, & Andrew Halpern-Manners (2012). Panel conditioning in longitudinal social science surveys. *Sociological Methods & Research*, 41(4), 491-534.

Waterton, Jennifer & Denise Lievesley (1989). Evidence of Conditioning Effects in the British Social Attitudes Panel. In: Greg Duncan, Graham Kalton, Daniel Kasprzyk & M. P. Singh (eds.): *Panel Surveys*. New York: Wiley. 318–339.

Weijters, Bert, Maggie Geuens & Niels Schillewaert (2010). The stability of individual response styles. *Psychological methods*, 15(1), 96-110.

Yan, Ting & Stephanie Eckman (2012). Panel Conditioning: Change in True Value versus Change in Self-Report. In: *Proceedings of the Survey Methods Research Section*. Alexandria, VA: American Statistical Association.

Response Quality and Ideological Dispositions

An Integrative Approach using Geometric and Classifying Techniques

Alice Barth, University of Bonn

Andreas Schmitz, University of Bonn

© 2016. Post-print manuscript version made available under the CC-BY-NC-ND 4.0 license
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

The original research article is published as:

Barth, Alice; Schmitz, Andreas (2018). Response Quality and Ideological Dispositions. An Integrative Approach using Geometric and Classifying Techniques. *Quality & Quantity*, 52(1), 175-194. doi: 10.1007/s11135-016-0458-3

Abstract:

When analyzing survey data, response quality has consequential implications for substantial conclusions. Differences in response quality are usually explained by personality, or socio-demographic or cognitive characteristics. Little, however, is known about how respondents' political attitudes, values, and opinions impact on quality aspects. This is a striking analytical omission, as potential associations between political values and various forms of response biases and artefacts call into question surveys' ability to represent 'public opinion'. In this contribution, response quality is traced back to respondents' political and ideological dispositions. For this purpose, a *relational* understanding of response quality is applied that takes into account different aspects of response behaviors, as well as the interrelations between these indicators. Using data from the US General Social Survey (2010-2014), an empirical typology of response quality is created via finite mixture analysis. The resulting classes are then related to positions in the US field of ideological dispositions, constructed via multiple correspondence analysis. The analyses reveal that there are (1) different combinations of response patterns and thus different empirical response types, and (2) that these types of response quality systematically vary with regard to the respondents' political and ideological (dis)positions. Implications of the findings for public opinion surveys are discussed.

1. Introduction

Why do polling institutes often fail to predict the actual winner in elections? Popular explanations for recent polling mistakes such as the 2015 General Election in the UK (Sturgis et al. 2016), the number of seats for the Likud Party in Israel 2015 (Rahat, Hazan & Bloom 2016), Great Britain's vote to leave the European Union ('Brexit') or Donald Trump's victory in the 2016 US presidential elections (Katz 2016) refer to biased samples (Sturgis et al. 2016) and variations in weighting procedures (Traugott et al. 2009), in the context of social strata. In this contribution, we argue that a respondent's ideological background systematically manifests itself in the form of differential reactions towards public opinion surveys. In particular, we investigate the association between respondents' political and ideological dispositions and the quality of response in the technical sense.

In general terms, 'response quality' refers to the ways in which respondents interact with the questionnaire. High quality is associated with answers reflecting the respondents' actual substantive positions with little to no distortion. Response tendencies such as acquiescence, extreme responding, non-differentiation, saying 'no opinion', 'don't know' or refusing to answer altogether, as well as 'speeding' through the questionnaire and ignoring written instructions, in contrast, are interpreted as symptoms of low response quality (Grauenhorst, Blohm & Koch 2016; Kieruj & Moors 2013; Revilla 2016).

Two perspectives in analyzing response quality can be distinguished: On the one hand, non-substantial response tendencies are treated as measurement errors that need to be avoided or corrected as best as possible (Van Vaerenbergh & Thomas 2013). Explanations for low response quality are thus sought in respondents' deficits in terms of cognitive sophistication or motivation (Lechner & Rammstedt 2015; Revilla 2016), in the survey mode or questionnaire design (Heerwegh & Loosveldt 2002; Schuman & Presser 1981;), or in the interaction between the respondents' and the survey's characteristics (Narayan & Krosnick 1996).

On the other hand, differences in response behavior are seen as temporally stable ways of communication, grounded in personality traits (Kieruj & Moors 2013; Knowles & Nathan 1997), culture (Smith 2004), or socio-economic status (Ross & Mirowsky 1984). From this perspective, response quality constitutes a source of information in itself. It allows researchers to draw conclusions about the respondents' mode of and willingness to express their opinions (He & Van de Vijver 2013; Laurison 2015), and therefore about the adequacy of the questionnaire instrument in different social strata (Bourdieu 1979). Accordingly, if the substantive character of 'non-substantive responses' is to be assessed, the relation of response quality (or rather qualities) to meaningful information must be investigated. While researchers from different disciplines have

been investigating the relation between response quality aspects and non-artefacts, the impact of respondents' ideological dispositions on response qualities has, until now, not gained much attention. This seems rather surprising, as political opinions represent one of the main subjects of interest in social sciences in general and in public opinion surveys in particular. Furthermore, the possibility of systematic relations between the quality of response and respondents' political and ideological positions poses a serious threat to the extrapolation of a representative picture of the 'public opinion' from surveys.

In this paper, we address this deficit by proposing a relational understanding of response qualities. Response biases are usually analyzed in isolation (Cole, McCormick & Gonyea 2012; Greenleaf 1992b) or operationalized one-dimensionally (He, Van de Vijver, Espinosa & Mui 2014; Kaminska, McCutcheon & Billiet 2010). In applying a relational view, response biases are considered as interrelated manifestations of response practices. In this way, different respondents can be characterized by specific forms and levels of biased response. On this basis, we can study how different reactions to the questionnaire survey correspond to ideological positions within US society.

To this end, we identify different types of response behavior by applying finite mixture modeling to various quality indicators derived from the US General Social Survey (GSS; 2010-2014). For the operationalization of ideological dispositions as a relational phenomenon, we then construct a field of ideological positions (Bourdieu 1993; Schmitz, Witte & Gengnagel 2016). There is a close link between field theory and the technique of multiple correspondence analysis (Blasius & Schmitz 2014), a dimension reduction method for categorical data, which is consequently applied for the construction of the field of ideological dispositions in the US. Finally, the response quality classes as derived from finite mixture modeling are then projected into the ideological field. This enables a joint representation of respondents' positions in the field and specific forms of response behavior, thus revealing that the technical quality of responses corresponds to the actor's ideological and political backgrounds. We conclude with an outlook on practical implications for public opinion surveys and future research.

2. State of research

2.1 The concept of response quality

The quality of response has been a common concern since the earliest days of surveys in scientific research (Cronbach 1946; Rorer 1965). In recent times, it has met with growing interest as an important aspect of data quality in general (Blasius & Thiessen 2012). In the current literature, two main conceptualizations of response quality can be identified: The first one addresses response sets – that is, “systematic ways of answering which are not directly related to the question content, but which represent typical behavioural characteristics of the respondents” (Oskamp & Schultz 2005:58). Typically, these ‘styles’ include tendencies such as limited response differentiation, acquiescence, and midpoint or extreme responding (Baumgartner & Steenkamp 2001). They are commonly operationalized by assessing the respondents’ choice of distinct categories in a variety of attitude items of heterogeneous content (He et al. 2014; Meisenberg & Williams 2008; Thomas, Abts & Vander Weyden 2014). Secondly, response quality is operationalized using indicators of general compliance towards the survey. In this sense, the number of ‘non-substantive’ answers (‘don’t know’, ‘no opinion’, ‘refuse’) (Colsher & Wallace 1989; Deutskens, de Ruyter, Wetzels & Oosterveld 2004; Kleiner, Lipps & Ferrez 2015), response times (Galesic & Bosnjak 2009; Zhang & Conrad 2013), number and length of answers to open-ended questions (Medway & Tourangeau 2015; Revilla 2016), and the amount of missing data (Deutskens et al. 2004; Herweegh & Loosveldt 2002) have been used for the assessment of response quality. These aspects are often complemented with respondents’ ratings of the survey experience itself (Meade & Craig 2012), assessments of respondents’ cooperation and comprehension by the interviewer (Medway & Tourangeau, 2015), and inconsistency and instructional manipulation checks¹ (Colsher & Wallace 1989; Revilla 2016). Despite recent approaches towards a unification of several response styles (cp. He et al. 2014; Van Rosmalen, Van Herk & Groenen 2010) and the consideration of differential quality aspects in a framework of ‘survey satisficing’ (Krosnick 1991; Revilla 2016), little is known about interrelations of response quality aspects.

2.2. Traditional explanations of differences in response quality

Explanations of differences in response quality have been addressed in several disciplines. A core discipline in this regard is psychological research, where the finding that response tendencies are relatively stable both across time (Bachmann & O’Malley 1984; Billiet & Davidov 2008) and across

¹ E.g, the request to select a certain answer category as proof that instructions were read (Revilla 2016; Meade & Craig 2012).

different domains and scale formats (He & Van de Vijver 2016; Kieruj & Moors 2013) has led to the assumption that certain response behaviors can be interpreted as a manifestation of underlying personality traits. For example, several studies find that extraversion is positively related to extreme responding and acquiescence (Harzing 2006; Kieruj & Moors 2013). He and Van de Vijver (2013) construct a general response style factor, with extremity and socially desirable responding in the positive part, and negative loadings of acquiescence and midpoint responding. This factor is positively associated with extraversion, conscientiousness, agreeableness, and openness, but negatively related to neuroticism. Knowles & Nathan (1997, p. 299) find that “acquiescent responders are cognitively simple, rigid in their mental organization, and intolerant of alternatives, whereas oppositional responders are more cognitively complex, disordered, forgetful, and welcoming of dissent.” Naemi, Beal, and Payne (2009) also claim that intolerance of ambiguity and ‘simplistic thinking’ are positively related to the probability of extreme responding.

In a different vein, researchers argue that one of the main causes of differences in response quality is respondents’ cognitive ability. It has been demonstrated that lower cognitive ability is associated with heightened acquiescence (Zhou & McClendon 1999), straight-lining (Cole et al. 2012), extreme responding (Light, Zax & Gardiner 1965), and more item non-response, as well as task simplification strategies such as random responses and non-differentiation (Thiessen & Blasius 2008), and a higher number of ‘don’t know’ answers (Colsher & Wallace 1989; Knäuper, Belli, Hill & Herzog 1997). In the same line of argumentation, educational attainment has been suggested as a proxy for both cognitive skills and familiarity with standardized tests and similar exercises (e.g. Krosnick 1992). Accordingly, there is considerable evidence in the literature that educational attainment is negatively related to indicators of low response quality such as acquiescence (Meisenberg & Williams 2008; Weijters, Geuens, & Schillewaert 2010), extreme response style (Greenleaf 1992b; Meisenberg & Williams 2008) midpoint response style (Weijters et al. 2010), effects of response and question order (Narayan & Krosnick 1996), and the heightened use of ‘no opinion’ or ‘don’t know’ categories (Narayan & Krosnick 1996; Krosnick et al. 2002).

It has also been suggested that response quality is lower for elderly people, as both ageing and cohort effects may be related to cognitive skills and expertise in dealing with standardized tests. For example, Holbrook, Cho, and Johnson (2006) demonstrate that older respondents experience more difficulties in mapping response options, and a considerable number of studies show that older respondents are more prone to both item nonresponse and acquiescence (de Leeuw, Hox & Huisman 2003; Meisenberg & Williams 2008). In sum, there are strong indications that response behavior is indeed partially rooted in the respondents’ cognitive schemes and personality.

From a sociological perspective, it is argued that respondents' position in society impacts on response behavior, especially with regard to class, race, and gender. For example, rejecting a direct effect of cognitive skills, Meisenberg & Williams (2008) attribute the result that poorer and more uneducated respondents tend to agree more with assertions to a possible lack of self-confidence in underprivileged population strata. Similarly, Ross & Mirowsky (1984) assume that their findings concerning higher levels of socially desirable responding and acquiescence in groups with lower socio-economic status can be explained by heightened impression management in powerless groups. Greenleaf also demonstrates a negative relationship between household income and both acquiescence (1992a) and extreme responding (1992b). In a recent study, Laurison (2015) addresses the issue of political mechanisms in response behavior and shows that, net of cognitive competencies, low income is associated with higher rates of 'don't know' answers to political survey questions.

Regarding race, numerous studies have found differences in response quality between ethnic groups. In the US, it has long been recognized that African Americans and Hispanics exhibit more acquiescence (Ross & Mirowsky 1984; Winkler, Kanouse & Ware 1982) and extreme responding (Bachmann & O'Malley 1984; Weech-Maldonado, Elliott, Oluwole, Schiller, & Hays 2008). Similar results were obtained for the Arab minority population in Israel when compared to the Jewish majority (Baron-Epel et al. 2010). Holbrook et al. (2006), who demonstrate that comprehension difficulties were higher in all racial minority groups than in non-Hispanic white Americans in their study, stress the importance of culture in survey response, suggesting that "questions that are written from the perspective of the dominant cultural group seem to be difficult for members of minority cultural groups" (p. 587). Similar conclusions are reached in studies that compare response quality across different countries, where it has been shown that response biases accumulate the more culturally and linguistically distant the survey population is from the society or group where the instrument was originally developed (Blasius & Thiessen 2012; Van de Vijver & Poortinga 1997). In a similar vein, Kleiner et al. (2015, p. 20) remark that in the same country, "those [foreign] groups that are more distant culturally and linguistically will be more likely to provide responses of poorer quality".

Gender has also been examined as a determinant of response quality, albeit with inconclusive results. While in the studies of Harzing (2006) and Meisenberg and Williams (2008), male respondents were more prone to extreme responding, De Jong, Steenkamp, Fox, and Baumgartner (2005), and Weijters et al. (2010) report contrary findings, and both Greenleaf (1992b) and Moors (2008) find no consistent effects of gender. Similar conflicting results have been reported with regard to acquiescence (Thomas et al. 2014). With regard to expressions of political attitude,

Atkeson and Rapoport (2003) state that women are more likely to say ‘don’t know’, and provide fewer open-ended comments.

2.3. Ideological Dimension of Response Quality

Overall, there is considerable evidence that response quality is affected both by individual characteristics and the actor’s position in society. However, the fact has largely been neglected that there is systematic variation in respondents’ ability and willingness to produce as well as express opinions on certain topics (Bergström 2012; Laurison 2015; Bourdieu 1979). In more foundational terms, research has also overlooked the fact that actors differ in their attitude towards the institution of surveys in general, which affects both the likelihood of participating in a study at all (Bates, Dahlhamer & Singer 2008; Stähli & Joye 2013; Zillmann et al. 2014) and the quality of responses given (Rogelberg et al. 2001; Stocké 2006). This raises the question as to whether the quality of response is affected by people’s political and social attitudes, values, and opinions.

Several findings suggest that the perception of and reaction to surveys is indeed not independent from respondents’ political dispositions. Milbrath (1962) found that Republicans are more prone to disagree with assertions than Democrats. Assessing differences in survey mode, Esaiasson & Granberg (1993) showed that the evaluation of political parties and party leaders was affected. Regarding the level of the survey institute and its staff, Desrosières and Thévenot (2002) demonstrated – based on the example of the development of occupational categories – that classification processes in survey research are anything but independent from political positions and ideological dispositions.

Moreover, virtually all of the mechanisms of response quality mentioned here have been utilized as predictors in research on ideological dispositions, too. A robust relationship between personality traits and political ideology has been established (Feldman & Johnston 2014; Gerber, Huber, Doherty, Dowling & Ha 2010). There is also a reliable negative relationship between cognitive ability and conservative social and political attitudes (Van Hiel, Onraet & De Pauw 2010); also, socio-demographic variables such as class position, educational attainment, age, and gender are common predictors of ideological differences (Evans 2000; Inglehart & Norris 2000; Schoon, Cheng, Gale, Batty & Deary 2010; Stubager 2010). Ultimately, different cultures (within and between countries) do show a political dimension (Inglehart 1997; Pfau-Effinger 2005).

In sum, systematic reasons indicate that positions in the political spectrum systematically correspond to differently biased practices of response. More specifically, we expect respondents at the extremes of the political spectrum to be more affected by response biases, as their world view

is less compatible with the categories provided in standardized interviews. At the same time – and perhaps more significantly – it can be assumed that the greater the ideological distance between respondents and the survey instrument, the more reluctant the former are to participate, resulting in lower response quality.

In the following analysis, ideological and political (dis)positions will be conceptualized via field theory approach (Bourdieu 1993; Schmitz et al. 2016). This paradigm bears a close resemblance to the quantitative construction of societal spheres and the localization of actors. Therefore, it has also received increasing attention in the analysis of political and ideological contexts (Fourcade, Lande & Schofer 2016). Constructing a field of ideological dispositions thus enables a relational analysis of the correspondence between positions in the field and different manifestations of response bias.

As a consequence of the emphasis on the relations of (dis)positions in an ideological field, research cannot be content with isolated indicators, but requires a broader understanding of response quality. Such an approach must take into account a wide range of biases, as respondents may react to the same instrument in different ways. In fact, previous efforts to unify several response styles have shown that these biases are not independent from each other, but can be assumed to appear as systematic patterns of reactions towards the questionnaire (e.g. He et al. 2014; Van Rosmalen et al. 2010). Nevertheless, existing studies mainly assess one or several singular indicators of response quality (e.g. Deutskens et al. 2004; Grauenhorst et al. 2016), sometimes combined to create an index (e.g. Revilla 2016), thus assuming the existence of a singular underlying dimension of response quality. We, in contrast, assess different types of reactions towards a questionnaire by studying a variety of central indicators, as well as the overarching patterns of response quality they constitute. This strategy facilitates the differentiation of various forms of response bias, and their systematic relationships with respondents' ideological backgrounds.

3. Data and methods

We will use the General Social Survey (GSS), a face-to-face-interview survey that has been conducted in the US since 1972 (Smith et al., 2014). Since 2006, the GSS has implemented a combined repeating cross-section and panel-component design. In this paper, we analyze data that was collected in 2010, 2012, and 2014 in a three-wave panel. The sample size of the panel component was 2,044 in 2010, dropping to 1,304 in 2014 due to panel attrition. The GSS provides

a promising data structure for the questions raised here, as it gathers data on a wide variety of attitudes, behaviors, and socio-demographic factors.

3.1. Modeling response quality

For constructing response style indicators, we use 360 attitudinal items measured on 4-point, 5-point and 7-point Likert scales. In order to avoid spurious correlations between the indicators, non-overlapping proportions of the data set – each containing items from different scales, topical areas, and survey years – are utilized for the construction of the different response style indicators. The indicators are operationalized as the number of respective answer categories (e.g. the middle category in odd scales for midpoint-responding) divided by the respondent's total number of valid answers in the respective subset of items. Indicators for each response style thus represent the respondents' likelihood of choosing a distinct value or range of values on a scale in a large subset of items (the number of items per indicator ranges from 28 to 97). Acquiescence and disacquiescence are operationalized using items on agree-disagree scales, whereas for midpoint and extreme responding, other scale values (such as very true – not at all true; completely wrong – not wrong at all etc.) are used. Average pairwise correlations in the subsets of items used for the construction of the single indicators lie in the range of .08 to .10, which can be judged as sufficient content heterogeneity (Baumgartner & Steenkamp 2001; He et al. 2014).

Respondents' compliance with the survey's demands – that is, their overall reaction towards the questionnaire – is measured by several indicators as well. For this purpose, the proportion of 'don't know' and 'no answer' categories in relation to a respondent's total number of answered questions is calculated. Further indicators are interview length and the respondent's comprehension of questions and cooperation during the survey interview, as judged by the interviewer. Attrition is considered by taking panel status into account, which is categorized as 'reinterviewed', 'eligible, but not reinterviewed' or 'ineligible' in the second and third panel wave. Additionally, we control for the number of system missing (not available) values in first-wave interviews.

In order to simultaneously assess these different indicators and their interrelations, they are analyzed using finite mixture modeling. Finite mixtures of distributions provide a probability-based approach to identifying unobserved heterogeneity within a population (McLachlan & Peel 2004). In the case at hand, it is assumed that the actors' response qualities exhibit a latent categorical structure which can be assessed using the indicators described above. The identification of subgroups according to respondent's values on the response quality indicators, where similar

combinations are likely to be grouped together in a class, yields an empirical typology of response qualities.

In order to ensure adequate model fit and a differentiated analysis of latent classes, response styles and survey compliance are modeled separately, resulting in two empirical typologies. The optimal solution in terms of the best-fitting number of latent groups is determined by statistical information criteria, namely the BIC (Nylund, Asparouhov & Muthén 2007).

3.2 Constructing the ideological field

In order to construct a field of ideological and political attitudes, which will subsequently be used as a reference for locating the empirical typologies of response styles and survey compliance, we apply multiple correspondence analysis (MCA). MCA belongs to the family of techniques used in geometric data analysis (Le Roux & Rouanet 2004). It allows for the extraction of the most important dimensions in a set of categorical variables, and the graphical representation of variable categories and individuals relative to each other in a coordinate system. Distances between categories as well as individuals can be interpreted as a measure of (dis)similarity: If categories often co-appear in individual's responses, they are located close together in the space produced by MCA. Rare co-appearances, accordingly, result in a larger distance between the respective categories. Furthermore, illustrative variables can passively be projected into the field, a technique that has been termed 'visual regression analysis' (Lebart, Morineau & Warwick 1984). Whereas the space is determined by the distances between the categories of active variables, passive variables do not alter the structure of the constructed field, but appear in their average and hence most likely position.

We explore the dimensional structure of ideological dispositions in US society (as represented by the GSS data in 2010) by assessing the interrelations between a comprehensive set of categorical items. The items cover diverse domestic policy issues such as abortion, suicide, pre-marital sex, homosexual marriage, racial discrimination, civil liberties, police violence, capital punishment, secularism, the role of government, and the perceived importance of certain issues for America today. The construction of latent ideological parameters from a large set of categorical items on different scale levels (dichotomous, four-point, and five-point Likert scales) deals with the problem that attitudinal measures themselves may be affected by response bias. Further, the non-linear approach of MCA allows us to detect response patterns (Blasius & Thiessen 2012; Barth 2016), which means that non-substantive variation is disclosed and can be adequately addressed. As we are interested in the distribution of both opinions and non-opinions in the ideological field, we use

‘non-substantive’ response categories – notably ‘don’t know’, ‘no answer’ and ‘refusal’ – as active categories in the construction of the field². The GSS data contain a distinctive source of artificial variation in inter-item analyses, as respondents were randomly assigned to three ballots within the survey. Most of the items in topical modules appeared in two of the three ballots, that is, in two thirds of the whole sample. While increasing the number of variables, this internal rotation structure complicates the assessment of relations between variables that are part of different ballots, because respondents’ non-missing responses on one item are inextricably linked with system missing values in other items. This problem can be addressed, however, because the three-partite structure means that the variation due to the split-ballot design is completely representable in two dimensions. These constitute the first two axes of the MCA solution. Therefore, design variation will be partialized out in the following analysis by restricting substantive analyses to higher dimensions, which are unaffected by the split-ballot effect³.

3.3 Response quality in the ideological field

In the final step, the relationship between ideological disposition and response quality is assessed by projecting the empirical classes of response styles and survey compliance into the field as passive variables. If response qualities were not related to ideological dispositions, we would expect to find the classes of the two empirical typologies at the origin of ordinates in the ideological field. If, however, respondents’ ideological positioning does impact on response behavior, we would expect the classes to systematically vary over the field. Taking up results from previous research, socio-demographic characteristics will be used as illustrative variables, too.

4. Findings

4.1. Class structure of response quality

The typology of response styles is assessed by subjecting the indicators for acquiescence, disacquiescence, extreme and midpoint-responding to finite mixture modeling. In this model, the BIC is minimized in the 10-class solution, which is selected accordingly. The first and largest class, comprising almost one third of respondents (see Table 1), exhibits an average profile with no

² Due to coding inconsistencies and the small number of cases with missing values in some variables, ‘don’t know’ and ‘no answer’ are treated as a single category in the analysis.

³ Applying MCA to data from one ballot only yields a similar spatial structure to the one reported in our results. Regarding future analyses, the recent development of Specific MCA constitutes a promising and adequate alternative to the procedure chosen here.

distinctive features – only the propensity for choosing the middle category in 7-point scales is slightly below average. The second class is characterized by a disproportionately high propensity to disagree with assertions, while respondents are also comparatively strongly inclined to choose middle categories and to avoid extremes. The third class, in contrast, demonstrates a slight aversion to middle categories in favor of extreme answers, while the other parameters exhibit no remarkable tendencies. Class four has the highest relative propensity for expressing disagreement, while at the same time favoring response categories at the extreme ends of scales. This tendency is especially pronounced in 4-point scales; however, the use of middle and extreme categories in odd-numbered scales is unexceptional, making this class profile an example for differential reactions resulting from questionnaire characteristics such as the format of scales. Respondents in the fifth class strongly tend to midpoint-responding and avoid extremes. Additionally, their response behavior entails below-average propensities for both acquiescence and disacquiescence, suggesting a general dislike for making strongly opinionated statements. The sixth class is characterized by a strong revealed preference for extreme values, combined with a disproportionately low number of midpoint responses and a slight tendency to disagree with assertions. Class seven is characterized by respondents' pronounced acquiescence tendencies, underlined by a very low propensity to disagree. The eighth class is distinguished by a very high propensity for choosing the middle category in 7-point Likert scales, whereas the number of middle categories in 5-point scales is closer to the average profile. This class might be considered a methodological artefact, but could also express respondents' limited ability to cope with a high number of response categories. Class nine combines very pronounced tendencies towards both acquiescence and extreme values, while disagreement and middle categories are avoided. Finally, respondents in class ten exhibit the highest relative propensity for choosing the middle category, and avoid extremes in all uneven scale formats. Moreover, their tendency to agree with assertions is well below average, whereas disacquiescence is quite pronounced. Table 1 gives an overview on the size and labelling of the classes.

Indicators for reaction towards the instrument were also analyzed with finite mixture modeling. In this case, the 6-class solution of survey compliance indicators has the lowest BIC value. The biggest class, which accounts for almost half of the sample (see table 2), can be termed 'highly functional', as they exhibit no anomalies: Their propensity to use the 'no answer' or 'don't know' categories is below average, both cooperation and comprehension are almost always rated as good by the interviewers, and panel attrition is disproportionately low. The profile of the second class is quite similar, but both the attrition rate and the average number of missing values are somewhat higher than in the previous class. The third class is mainly distinguished by a long interview duration, a low number of missing values, and a high propensity to take part in at least a second panel wave,

thus also representing a class of comparatively high response quality. Class four is also characterized by a long interview, but shows a higher number of missing values and high usage of the ‘don’t know’ response option. The probability that respondents’ comprehension of questions is only rated ‘fair’ or even ‘poor’ by the interviewer is 0.6, the highest value in all classes. Cooperation is also rated worse than in the previous classes, and more than half of the respondents in this class drop out or become ineligible before wave three. The profile of class five exhibits the highest number of ‘don’t know’ responses of all classes and a comparatively high probability of dropping out due to ineligibility before wave three. The average interview length in this class is very short, with a disproportionately high number of system missing values. Comprehension and cooperation are below average. Finally, the last and smallest class exhibits a very high number of ‘no answer’ categories and a high propensity to leave the study due to refusal. As the comprehension of these respondents is rated as fairly good, but cooperation is below average, we assume that the ‘no answers’ do not point to problems in understanding, but are a sign of unwillingness to provide information.

Table 1: Response style classes (RS)

<i>No.</i>	<i>Class size in percent</i>	<i>Label</i>
1	30.3	average
2	23.2	disagree/middle
3	14.8	slightly extreme
4	12.5	disagree/extreme
5	5.3	middle
6	4.9	extreme
7	3.0	agree
8	2.9	middle (7pt)
9	2.0	agree/extreme
10	1.3	middle/disagree

Table 2: Survey compliance classes (SC)

<i>No.</i>	<i>Class size in percent</i>	<i>Label</i>
1	44.0	highly functional
2	26.2	functional
3	9.5	long/functional
4	9.5	Long/low comprehension
5	9.4	Short/evasive
6	1.4	disinclined

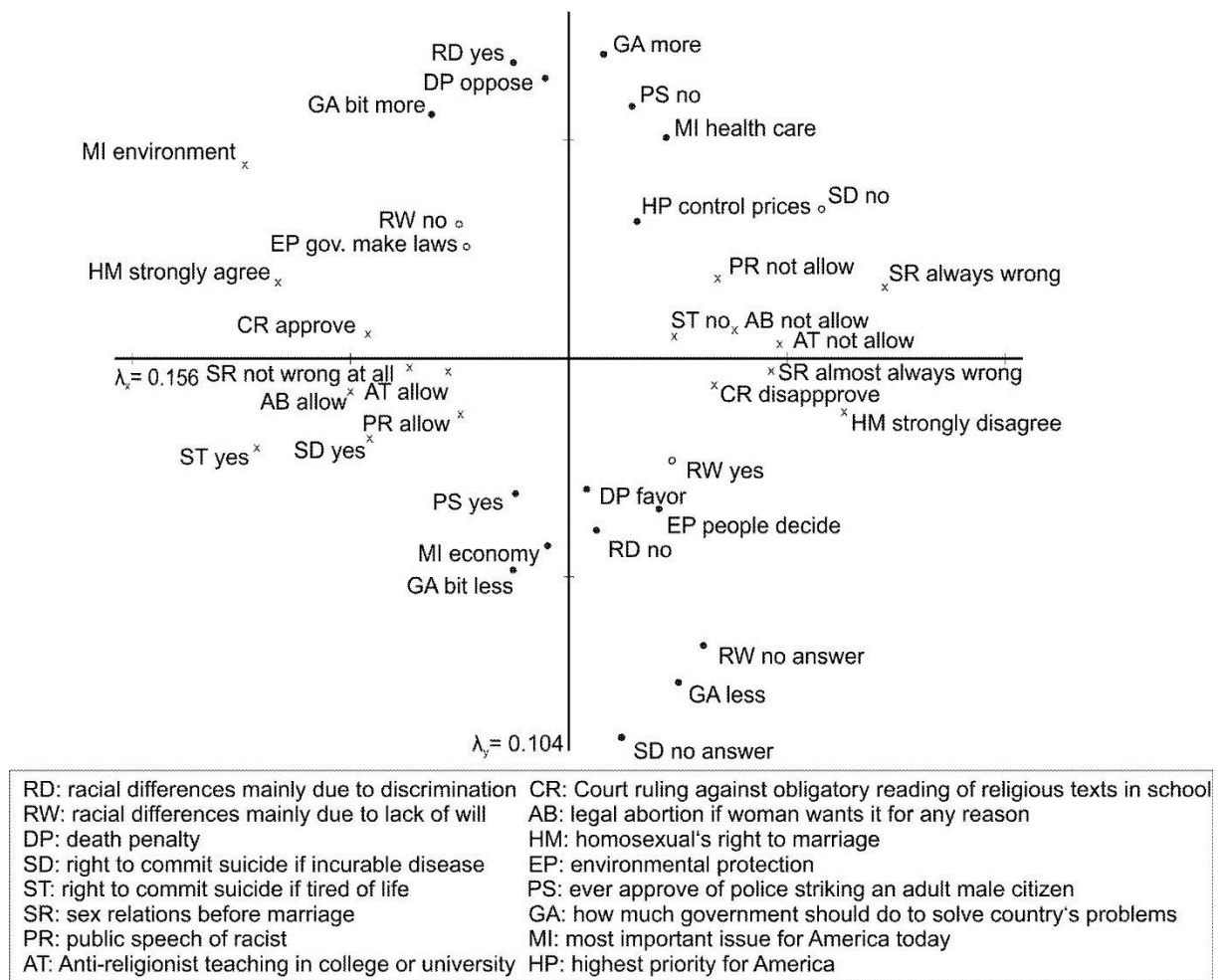
According to the two empirical typologies, it turns out that there are substantial differences in the ways respondents react to a standardized questionnaire. Regarding response styles, the biggest classes are characterized by no or only slight response tendencies, while there are several smaller classes which systematically exhibit a single distinctive style, or a combination such as acquiescence and extreme responding. In terms of survey compliance, a majority of functional respondents can be differentiated in several classes that show distinct patterns of quality problems. These results

confirm that any one-dimensional conceptualization of response quality will overlook important aspects.

4.2. Response quality within the US Field of Ideological Dispositions

The US ideological field is constructed by subjecting a wide variety of political, economic, and social attitude items to MCA. In this interpretation, we consider only dimensions from the third one upwards, as the first two axes represent artificial variation introduced by the GSS's three-ballot design. The fourth dimension mainly contrasts non-substantive categories to all others; therefore, the substantive dimensions three and five graphically represent the ideological field. In Figure 1, the active categories whose contribution to the respective axes is above average are depicted⁴.

Fig. 1 The US field of ideological dispositions (MCA dimensions 3 and 5)



⁴ Cross signifies above average contribution to horizontal axis; filled dot signifies above average contribution to vertical axis; empty dot signifies above average contribution to both axes.

On the horizontal axis, we find traditional social values on the right and liberal, progressive positions on the left. For example, respondents on the right state that sex before marriage is wrong, strongly disagree with homosexuals' right to marriage, and are against abortion and the right to commit suicide. Located on the left are the contrary positions, where individuals' rights to abortion, suicide, sex before marriage and marriage between homosexuals are supported. Furthermore, this dimension is associated with the interrelation between state and religion: On the left, secular positions are held, whereas on the right, respondents think that atheists should not be allowed to teach in colleges or universities, and disapprove of the Supreme Court's ruling that the reading of religious texts may not be required in public schools. The importance of religious elements in daily life is hence associated with a traditional system of values. On the whole, this axis can be understood as a dimension of liberalism versus conservatism in terms of social and cultural issues.

The vertical axis separates supporters of the welfare state, on top, from advocates of deregulation, who mainly stress the state's monopoly on the use of force, on the bottom. In particular, in the upper part, the governments' responsibility to solve the country's problems is stressed, while the death penalty and police violence are opposed, health care is deemed a very salient issue, and the lower socio-economic status of minorities is ascribed to racial discrimination. In the lower part, in contrast, both the death penalty and the use of force by the police are supported. The state is mainly seen as a provider of basic security functions, associated with the opinion that the government should leave more decisions to individuals and private businesses. The US economy is a major concern. Hence, this axis can be understood as contrasting two forms of perceived state responsibility.

This interpretation can be validated by current findings from political science, which display a two-dimensional structure in US mass ideologies (Carmines, Ensley & Wagner 2012a; Shafer & Claggett 1995). Separated into liberal or conservative positions on social and cultural issues on one dimension, and economic and social-welfare issues on the other, the population can empirically be classified into four distinct ideological groups (Claggett, Engle & Shafer 2014). The division between the core positions of the two main parties runs along the diagonal, with the representation of Republican positions in the lower right, and Democrats in the upper left. This interpretation is further corroborated by the passive projection of party identification onto the dimensions (see Figure 2). The lower left quadrant can be described as libertarian, whereas the upper right section represents socially conservative, pro-welfare state positions that have been referred to as 'communitarian' (Carmines, Ensley & Wagner 2012b) or 'populist' (Claggett, Engle & Shafer 2014).

In the final step, the response quality classes discussed above, as well as respondents' ethnicity, educational level, gender, age group, and party identification are passively projected into the ideological field ⁵(see Figure 2). In this way, our main research question – the correspondence of ideological positions to modes of responding to a public opinion survey – can be investigated.

Fig. 2 Response quality classes and selected socio-demographics in the field of ideological dispositions

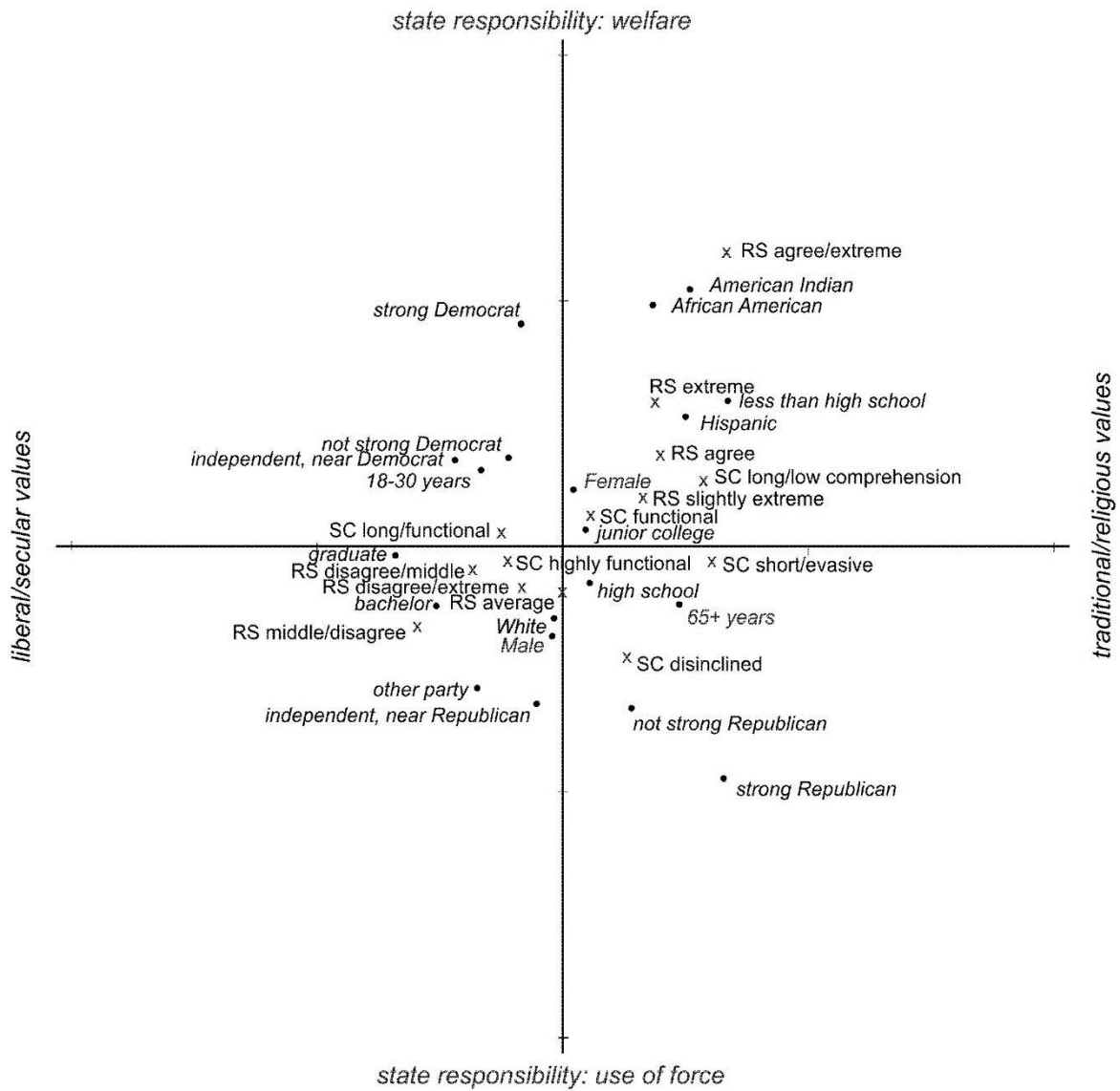


Figure 2 shows the locations of the classes of survey compliance (SC) and those of response style (RS) in the field of ideological dispositions. In terms of response styles, remarkably, acquiescence and extreme responding are located in the upper right part. Thus, these tendencies are strongly

⁵ Categories comprising less than one percent of the sample are not displayed. The response style classes 'middle' and 'middle 7pt,' as well as middle age groups, are located very close to the centroid and thus not displayed for reasons of clarity.

associated with support of the welfare state and traditional social values. Moreover, the observation that acquiescence and a tendency towards extreme values constitute a common combination of response bias, which was also suggested by the existence of an 'agree/extreme' class, is confirmed by the finding that the respective response style classes are located in relative proximity to one another within the field.

The overall tendency to disagree with assertions is located in the opposite quadrant, in the lower left of the ideological field, where a strong endorsement of individual liberties is expressed. On the whole, extreme responding is more common in the upper section, suggesting that supporting the welfare state is associated with favoring the extreme ends of rating scales, whereas midpoint-responding is mostly located in the center of the coordinate system, and thus not associated with a specific dispositional direction. The combination of choosing middle categories and disacquiescence, though, is prevalent in the libertarian part of the field. The opposition between a socially conservative, pro-welfare state and a socially liberal, anti-interventionist stance is thus reproduced in the different ways of responding, with acquiescence and extreme values on one side versus disacquiescence and midpoint-responding on the other.

The passive projection of socio-demographic variables corroborates previous research, as it is evident that the 'communitarian' ideology, which is associated with acquiescence and extreme responding, is disproportionately held by ethnic minorities, namely African Americans, American Indians, and Hispanics, and corresponds to low educational attainment. Conversely, respondents in the libertarian part of the field, who disproportionately disagree with survey statements, display above-average educational levels.

The horizontal axis, showing liberal to traditional social values, is the main differentiating factor among survey compliance classes: Liberal attitudes are mainly associated with functional ways of responding, whereas respondents in the traditional spectrum are more prone to encounter various problems. Consistent with previous research, educational level corresponds to the position on this dimension, insofar as liberal and secular social attitudes are more pronounced among those with greater educational attainment (Schoon et al. 2010). In general, the disproportionate representation of 'dysfunctional' classes in the spectrum of conservative attitudes is suggestive of a mismatch between respondents' mindsets and the categories provided by the questionnaire.

The different manifestations of 'dysfunctional' responding, however, vary considerably along the dimension of attitudes towards the responsibility of the state. In particular, 'long/low comprehension' is located in the upper, pro-welfare part of the axis, whereas 'disinclined'

respondents can be found in the lower quadrant. This finding can be interpreted as a manifestation of different behavioral strategies associated with ideological position: In the 'long/low comprehension' class, the long average duration of survey interviews points to respondents who make an effort despite comprehension problems, or are at least unwilling to refuse answers or break off. On the other hand, the 'disinclined' class comprises those whose problems mainly manifest in both item non-response and attrition due to refusal. We assume that the ideological disposition of strong Republicans, particularly the rejection of governmental intervention, affects how public opinion surveys are perceived by this group of respondents, thus leading to a higher likelihood of disinclined response behavior. In contrast, the position of Democrat partisans in the upper left quadrant is characterized by the absence of any problems in terms of response quality. This may well indicate that the 'fit' between the instrument and respondents' cognitive representation of issues is particularly high in this area.

Additionally, these opposing political positions are associated with age: The passive projection of age shows a linear course from the upper left to the lower right, with the younger more prone to Democrat positions, whereas the Republican ideology – and the likelihood of lower response quality – is more represented in elderly people. Respondents' gender is associated only with the dimension of state responsibility: Whereas women are more supportive of the welfare state, men particularly value the state's monopoly on force. In terms of response quality, this finding suggests that general conclusions about whether men or women are 'better' respondents are unwarranted, as the average number of problematic tendencies is very similar. However, there are some qualitative differences: Women are more prone to acquiescence, but less likely to refuse answers than men.

The concentration of the average of 'functional' response classes close to the centroid is not surprising considering their large size. It could also, however, be interpreted as a sign that respondents holding 'mainstream' opinions experience less problems in handling survey questions, as the correspondence of their 'mental map' to the provided answer categories is higher than in actors in more extreme regions of the field of political dispositions. Overall, the analysis confirms the hypothesis that response quality systematically varies according to positions in the spectrum of political and social opinions.

5. Discussion

Scientists from different fields have long been aware that the validity and reliability of survey data is strongly affected by systematic biases in response behavior (Baumgartner & Steenkamp, 2001; Blasius & Thiessen 2012; Van Vaerenbergh & Thomas, 2013). In this contribution, we argued that current research on response quality is deficient in two respects: Firstly, research on the determinants of response quality focuses on either the psychological or socio-demographic conditions of response biases, but tends to neglect the role of political stances. If, however, different attitudes towards political, ideological, and social issues are connected to particular ways of expressing or holding back one's opinion (Bourdieu 1979; Laurison 2015), the potentially different ways of 'non-substantive' answering, as well as their interrelations, must be investigated. Secondly, little is known about the interplay of different aspects of response quality. Thus, we proposed a relational perspective on response quality, in order to show how different forms of response behaviors relate to each other, and, ultimately, to ideological positions, using the example of US society.

In the first step of our analysis, we constructed a typology of response quality by assessing various response style and survey compliance indicators – and their interactions – via finite mixture modeling, using survey data from the GSS (2010-2014). The classification revealed that, while a majority of respondents is not – or is only slightly – affected by biases, there are several subgroups that exhibit distinct non-substantial response behaviors. The tendencies to agree or disagree, on the one hand, and the disproportionate choice of middle or extreme categories, on the other, appeared in various combinations. Moreover, quality issues in terms of low cooperation and understanding, a high probability of attrition, and a large number of non-substantive answers were jointly represented in several classes, albeit in different ratios that suggest different ways of dealing with the demands of a survey interview. Our approach thus corroborates the assertion that response quality is a multi-faceted construct, entailing more than a simple dichotomy between 'good' and 'bad' quality.

In the second part of our analysis, we demonstrated that these response quality classes systematically correspond to positions within a field of political ideological dispositions, which was constructed using geometric data analysis. The two-dimensional structure of the field was in line with current conceptualizations of the ideological landscape in the US (Carmines, Ensley & Wagner 2012a; Shafer & Claggett 1995). The non-parametric approach of MCA enabled us to partial out design-induced variance and extract latent ideological dimensions from a wide variety of categorical items on political and social attitudes and values, thus capitalizing on the GSS's potential for the

representation of opinions. In doing so, the analysis demonstrated that these and other response patterns are systematically associated with ideological positions. In this sense, respondents holding conservative views, in particular, tended to have problems responding to surveys, whereas liberal individuals were less affected. Moreover, manifestations of low response quality differed according to respondents' opinions towards the state's responsibility. Whereas supporters of the welfare state tended to exhibit certain response styles, especially acquiescence and extreme responding, opponents of governmental interventions were characterized by disinclination towards the survey in general. The relative absence of response quality problems among liberal, pro-welfare state opinions suggested that this field position is mostly composed of respondents who were both able and willing to respond in the way originally intended by the researchers.

There are several implications that can be derived from these findings. In theoretical terms, future research must try to disentangle the causal impact of ideological background from the impact of personality traits and class position (cp. Laurison 2015). The strategy developed throughout this work represents a promising starting point for such studies. For example, within regression models, the constructed typologies can be used as dependent variables, and ideological dimensions as explanatory variables. Apart from such questions of the single variables' particular causal impacts on response behavior, the relational approach proposed here also brings a pragmatic issue to light. The findings bear important implications for survey research, as the quality of response affects several steps of the 'chain of statistical treatment' (Desrosières & Thévenot 2002: 35): Low response quality can be used as a criterion for the removal of whole cases from the data set (Tabachnick & Fidell 2007), missing or non-substantive values are often excluded from substantive analyses, and the presence of response bias affects both univariate and multivariate distributions (Baumgartner & Steenkamp 2001; Van Vaerenbergh & Thomas 2013). As our analyses show that respondents in certain ideological positions exhibit higher average response quality levels than others, the consequence is that the assumption of unbiased representation of 'public opinion' by means of surveys is challenged. In the analysis at hand, the absence of any problems in the realm of Democrat positions suggests that these ideological dispositions will be portrayed without distortion, or even be over-represented in the picture of 'public opinion' drawn by the survey, whereas the conservative pole was characterized by lower quality that may ultimately lead to a misrepresentation of these positions in analyses and published results. We consider the finding of differential response behavior according to ideological position to be generalizable; however, the actual manifestations of the phenomenon – that is, which dispositions and response qualities appear together – are expected to vary across survey organizations, nation states, and over time.

Our results thereby illustrate the meaning of Bourdieu's statement that "public opinion does not exist" (Bourdieu 1979). Surveys cannot be understood solely as neutral technical instruments for the depiction of opinions, but they are – sometimes to a considerable amount – constitutively involved in the very creation of opinions. Bourdieu conceived survey institutes and the media as a 'field of ideological production' (Bourdieu 1984, p. 399) that produces and transfers schemes of thinking about political issues to political 'laymen'. Public opinion researchers approaching this field do not (just) neutrally measure the objective state of opinions. Inevitably, they also measure the effects generated by political discourses on the voters (Champagne 1991). In this respect, the 'field of ideological dispositions' may well be understood as a 'field of ideological consumption'.

In sum, the findings imply that the problem of differential response quality in survey research cannot be ignored. While it is widely known that differences in response quality are not randomly distributed, but depend on socio-demographic characteristics, personality attributes, and cultural background, their relation to ideological dispositions presents a specific challenge to questionnaire-based research on political opinions and attitudes. As we have been using data from the GSS, an established and internationally renowned social science survey that is subject to ongoing methodological assessments (Smith et al. 2014), the results of this analysis are likely to represent the lower boundary of possible interrelations between ideological position and response quality. In the light of our findings, one must state that surveys are anything but "designed so that each citizen has an equal chance to participate and an equal voice when participating" (Verba 1996: 3). Therefore, we strongly recommend the development of methodological approaches in future research that are sensitive to the inherent limitations of survey research in representing 'public opinion'. While our analysis presented the use of geometric and classifying techniques in a processed data set, it would be advisable to assess the relationship between response quality and ideology during the field phase in order to allow for a possible re-sampling of specific political strata. In a similar vein, systematic differences in the reaction to invitations to participate in a survey should be investigated in order to prevent differential nonresponse emanating from ideology.

6. References

- Atkeson, L.R., Rapoport, R.B.: The more things change the more they stay the same: Examining gender differences in political attitude expression, 1952–2000. *Public Opin. Quart.*, 67(4), 495-521 (2003)
- Bachman, J.G., O'Malley, P.M.: Yea-Saying, Nay-Saying, and Going to Extremes: Black-White Differences in Response Styles. *Public Opin. Quart.* 48(2), 491-509 (1984)
- Baron-Epel, O., Kaplan, G., Weinstein, R., Green, M.S.: Extreme and acquiescence bias in a bi-ethnic population. *Eur. J. Public Health* 20(5), 543–548 (2010)
- Barth, A.: The Changing Nature of Attitude Constructs - An Application of Multiple Correspondence Analysis on Gender Role Attitudes. *Qual. Quant.*, 50(4), 1507-1523 (2016)
- Bates, N., Dahlhamer, J., Singer, E.: Privacy concerns, too busy, or just not interested: Using doorstep concerns to predict survey nonresponse. *J. Off. Stat.*, 24(4), 591-612 (2008)
- Baumgartner, H., Steenkamp, J.-B.E.: Response Styles in Marketing Research: A Cross-National Investigation. *J. Marketing Res.* 38(2), 143–156 (2001)
- Bergström, Y.: Educational and Social Dimensions of Political Participation: Producing a Political Opinion. *Praktiske Grunde. Nordisk tidsskrift for kultur-og samfundsvidevidenskab*, 4, 23-34 (2012)
- Billiet, J.B., Davidov, E.: Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociol. Method. Res.* 36(4), 542-562 (2008)
- Blasius, J., Thiessen, V.: *Assessing the quality of survey data*. Sage, London (2012)
- Blasius, J., Schmitz, A.: The Empirical Construction of Bourdieu's Social Space. In: Greenacre, M., Blasius, J. (eds.): *The Visualization and Verbalization of Data*, pp. 205-222. Chapman & Hall, London (2014)
- Bourdieu, P.: Public opinion does not exist. In: Siegelau, S., Mattelart, A. (eds): *Communication and class struggle* 1, pp. 124-130. IG/IMMRC, New York/Bagnolet (1979)
- Bourdieu, P.: *Distinction: A social critique of the judgement of taste*. Harvard University Press, Cambridge (1984)
- Bourdieu, P.: Some Properties of Fields. In: Bourdieu, P., *Sociology in question*, pp. 72-77. Sage, London (1993)
- Carmines, E.G., Ensley, M.J., Wagner, M.W.: Political Ideology in American Politics: One, Two, or None? *The Forum* (10)3 (2012a)
- Carmines, E.G., Ensley, M.J., Wagner, M.W.: Who fits the left-right divide? Partisan polarization in the American electorate. *Am. Behav. Sci.* 56(12), 1631-1653 (2012b)
- Claggett, W.J., Engle, P.J., Shafer, B.E.: The Evolution of Mass Ideologies in Modern American Politics. *The Forum*, 12(2), 223-256 (2014)

- Cole, J.S., McCormick, A.C., Gonyea, R.M.: Respondent use of straight-lining as a response strategy in education survey research: Prevalence and implications. In: Annual Meeting of the American Educational Research Association (AERA) (2012)
- Colsher, P.L., Wallace, R.B.: Data Quality and Age: Health and Psychobehavioral Correlates of Item Nonresponse and Inconsistent Responses. *Journal of Gerontology*, 44(2), 45-52 (1989)
- Cronbach, L.J.: Response sets and test validity. *Educ. Psychol. Meas.* 6(4), 475-494 (1946)
- De Jong, M.G., Steenkamp, J.B.E., Fox, J.P., Baumgartner, H.: Using item response theory to measure extreme response style in marketing research: A global investigation. *J. Marketing Res.* 45(1), 104-115 (2008)
- de Leeuw, D., Hox, J., Huisman, M.: Prevention and treatment of item nonresponse. *Journal of Official Statistics*, 19(2), 153-176 (2003)
- Desrosières, A., Thévenot, L.: *Les catégories socio-professionnelles*. Editions La Découverte, Paris (2002)
- Deutskens, E., de Ruyter, K., Wetzels, M., Oosterveld, P.: Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study. *Marketing Letters*, 15(1), 21–36 (2004)
- Esaiasson, P., Granberg, D.: Hidden negativism: Evaluation of Swedish parties and their leaders under different survey methods. *Int. J. Public Opin. Res.* 5(3), 265-277 (1993)
- Evans, G.: The Continued Significance of Class Voting. *Annu. Rev. Polit. Sci.* 3, 401-417 (2000)
- Feldman, S., Johnston, C.: Understanding the determinants of political ideology: Implications of structural complexity. *Polit. Psychol.* 35(3), 337-358 (2014)
- Fourcade, M., Lande, B., Schofer, E.: Political space and the space of polities: doing politics across nations. *Poetics*, 55, 1-18 (2016)
- Galesic, M., Bosnjak, M.: Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opin. Quart.* 73(2), 349–360 (2009)
- Gerber, A.S., Huber, G.A., Doherty, D., Dowling, C.M., Ha, S.E.: Personality and political attitudes: Relationships across issue domains and political contexts. *Am. Polit. Sci. Rev.* 104(1), 111-133 (2010)
- Grauenhorst, T., Blohm, M., & Koch, A.: Respondent Incentives in a National Face-to-face Survey - Do They Affect Response Quality? *Field Methods*, 28(3), 266-283 (2016)
- Greenleaf, E.A.: Improving Rating Scale Measures by Detecting and Correcting Bias Components in Some Response Styles. *J. Market. Res.* 29(2), 176-188 (1992a)
- Greenleaf, E.A.: Measuring Extreme Response Style. *Public Opin. Quart.* 56(3), 328-351 (1992b)
- Harzing, A.W.: Response styles in cross-national survey research - A 26-country study. *Int. J. Cross Cult. Manag.* 6(2), 243-266 (2006)

- He, J., Van de Vijver, F.J.: A general response style factor: Evidence from a multi-ethnic study in the Netherlands. *Pers. Individ. Differ* 55(7), 794-800 (2013)
- He, J., Van de Vijver, F.J., Espinosa, A.D., Mui, P.H.: Toward a unification of acquiescent, extreme, and midpoint response styles - A multilevel study. *Int. J. Cross Cult. Manag.* 14(3), 306-322 (2014)
- He, J., Van de Vijver, F.J.: Response styles in factual items: Personal, contextual and cultural correlates. *Int. J. Psychol.* 1-8 DOI: 10.1002/ijop.12263 (2016)
- Heerwegh, D., Loosveldt, G.: An Evaluation of the Effect of Response Formats on Data Quality in Web Surveys. *Soc. Sci. Comp. Rev.* 20(4), 471-484 (2002)
- Holbrook, A., Cho, Y.I., Johnson, T.: The Impact of Question and Respondent Characteristics on Comprehension and Mapping Difficulties. *Public Opin. Quart.* 70(4), 565-595 (2006)
- Inglehart, R.: *Modernization and postmodernization: Cultural, economic, and political change in 43 societies* (Vol. 19). Princeton University Press, Princeton NJ (1997)
- Inglehart, R., Norris, P.: The Developmental Theory of the Gender Gap: Women's and Men's Voting Behavior in Global Perspective. *Int. Pol. Sci. Rev.* 21(4), 441-463 (2000)
- Kaminska, O., McCutcheon, A.L., Billiet, J.: Satisficing among reluctant respondents in a cross-national context. *Public Opin. Quart.* 74(5), 956-984 (2010)
- Katz, J.: Who will be president? The New York Times, retrieved from <http://www.nytimes.com/interactive/2016/upshot/presidential-polls-forecast.html> (2016, November 8)
- Kieruj, N.D., Moors, G.: Response style behavior: Question format dependent or personal style? *Qual. Quant.* 47(1), 193-211 (2013)
- Kleiner, B., Lipps, O., Ferrez, E.: Language Ability and Motivation Among Foreigners in Survey Responding. *J. Surv. Stat. Methodol.* 3(3), 339-360 (2015)
- Knäuper, B., Belli, R.F., Hill, D.H., Herzog, A.R.: Question difficulty and respondents' cognitive ability: The effect on data quality. *J. Off. Stat.* 13(2), 181-199 (1997)
- Knowles, E.S., Nathan, K.T.: Acquiescent Responding in Self-Reports: Cognitive Style or Social Concern? *J. Res. Pers.* 31(2), 293-301 (1997)
- Krosnick, J.A.: Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology*, 5(3), 213-236 (1991)
- Krosnick, J.A.: The impact of cognitive sophistication and attitude importance on response-order and question-order effects. In: Schwarz, N., Sudman, S. (eds) *Context effects in social and psychological research*, pp. 203-218. Springer, New York (1992)
- Krosnick, J.A., Holbrook, A.L., Berent, M.K., Carson, R.T., Hanemann, W.M., ... Moody, W.R.: The impact of "no opinion" response options on data quality: Non-attitude reduction or an invitation to satisfice? *Public Opin. Quart.* 66(3), 371-403 (2002)

- Laurison, D.: The Willingness to State an Opinion: Inequality, Don't Know Responses, and Political Participation. *Sociol. Forum* 30(4) 925-948 (2015)
- Lebart, L., Morineau, A., Warwick, K.M.: *Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices*. Wiley, New York (1984)
- Lechner, C.M., Rammstedt, B.: Cognitive ability, acquiescence, and the structure of personality in a sample of older adults. *Psychol. Assessment*, 27(4), 1301–1311 (2015)
- Le Roux, B., Rouanet, H.: *Geometric data analysis: from correspondence analysis to structured data analysis*. Kluwer, Dordrecht (2004)
- Light, C.S., Zax, M., Gardiner, D.H.: Relationship of age, sex, and intelligence level to extreme response style. *J. Pers. Soc. Psychol.* 2(6), 907-909 (1965)
- McLachlan, G., Peel, D.: *Finite mixture models*. John Wiley & Sons, New York (2004)
- Meade, A.W., Craig, S.B.: Identifying Careless Responses in Survey Data. *Psychol. Meth.* 17(3), 437-455 (2012)
- Medway, R.L., Tourangeau, R.: Response Quality in Telephone Surveys: Do Prepaid Cash Incentives Make a Difference? *Public Opin. Quart.* 79(2), 524–543 (2015)
- Meisenberg, G., Williams, A.: Are acquiescent and extreme response styles related to low intelligence and education? *Pers. Individ. Differ.* 44(7), 1539–1550 (2008)
- Milbrath, L.W.: Latent origins of liberalism-conservatism and party identification: A research note. *J. Polit.* 24(4), 679-688 (1962)
- Moors, G.: Exploring the effect of a middle response category on response style in attitude measurement. *Qual. Quant.* 42(6), 779–794 (2008)
- Naemi, B.D., Beal, D.J., Payne, S.C.: Personality predictors of extreme response style. *J. Pers.* 77(1), 261-286 (2009)
- Narayan, S., Krosnick, J.A.: Education Moderates Some Response Effects in Attitude Measurement. *Public Opin. Quart.* 60(1), 58-88 (1996)
- Nylund, K.L., Asparouhov, T., Muthén, B.O.: Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Struct. Equ. Modeling*, 14(4), 535-569 (2007)
- Oskamp, S., Schultz, P.W.: *Attitudes and opinions*. Psychology Press, New York, London (2005)
- Pfau-Effinger, B.: Culture and welfare state policies: reflections on a complex interrelation. *J. Soc. Policy*, 34(1), 3-20 (2005)
- Rahat, G., Hazan, R. Y., Bloom, P.B.N.: Stable Blocs and Multiple Identities: The 2015 Elections in Israel. Manuscript submitted for publication (2016)
- Revilla, M.: Impact of raising awareness of respondents on the measurement quality in a web survey. *Qual. Quant.* 50(4), 1469-1486 (2016)

- Rogelberg, S.G., Fisher, G.G., Maynard, D.C., Hakel, M.D., Horvath, M.: Attitudes toward surveys: Development of a measure and its relationship to respondent behavior. *Organ. Res. Meth.* 4(1), 3-25 (2001)
- Rorer, L.G.: The great response-style myth. *Psychol. Bull.* 63(3), 129-156 (1965)
- Ross, C.E., Mirowsky, J.: Socially-desirable response and acquiescence in a cross-cultural survey of mental health. *J. Health Soc. Behav.* 25(2), 189-197 (1984)
- Schmitz, A., Witte, D., Gengnagel, V.: Pluralizing field analysis: Toward a relational understanding of the field of power. *Social Science Information/Information sur les sciences sociales*, 56(1), 1-25 (2016)
- Schoon, I., Cheng, H., Gale, C.R., Batty, G.D., Deary, I.J.: Social status, cognitive ability, and educational attainment as predictors of liberal social attitudes and political trust. *Intelligence*, 38(1), 144-150 (2010)
- Schuman, H., Presser, S.: *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context.* Academic Press, New York (1981)
- Shafer, B.E., Claggett, W.J.: *The Two Majorities: The Issue Context of Modern American Politics.* Johns Hopkins University Press, Baltimore (1995)
- Smith, P.B.: Acquiescent response bias as an aspect of cultural communication style. *J. Cross Cult. Psychol.* 35(1), 50-61 (2004)
- Smith, T., Marsden, P.V., Hout, M., Son, J., Schapiro, B., Mai, N., Milliff, A.: *General social surveys, 1972–2014 Cumulative Codebook.* National Opinion Research Centre, Chicago (2014)
- Stähli, M.E., Joye, D.: Non-respondent surveys: pertinence and feasibility. *The Survey Statistician* 68, 16-22 (2013)
- Stocké, V.: Attitudes toward surveys, attitude accessibility and the effect on respondents' susceptibility to nonresponse. *Qual. Quant.* 40(2), 259-288 (2006)
- Stubager, R.: The development of the education cleavage: Denmark as a critical case. *West European Politics*, 33(3), 505-533 (2010)
- Sturgis, P., Baker, N., Callegaro, M., Fisher, S., Green, J., ..., Smith, P.: *Report of the Inquiry into the 2015 British general election opinion polls.* Market Research Society and British Polling Council, London (2016)
- Tabachnick, B.G., & Fidell, L.S.: *Using multivariate statistics* (5th ed). Pearson/Allyn & Bacon, Boston MA (2007)
- Thomas, T.D., Abts, K., & Vander Weyden, P.: Response Styles and the Rural–Urban Divide. *Educ. Psychol. Meas.* 74(1), 97-115 (2014)
- Thiessen, V., Blasius, J.: Mathematics achievement and mathematics learning strategies: Cognitive competencies and construct differentiation. *Int. J. Educ. Res.*, 47(6), 362-371 (2008)

- Traugott, M., Bolger, G., Davis, D.W., Franklin, C., Groves, R.M., ..., Wlezien, C.: An Evaluation of the Methodology of the 2008 Pre-Election Primary Polls. AAPOR Ad Hoc Committee on the 2008 Presidential Primary Polling. AAPOR (2009)
- Van de Vijver, F.J., Poortinga, Y.H.: Towards an integrated analysis of bias in cross-cultural assessment. *Eur. J. Psychol. Ass.* 13(1), 29-37 (1997)
- Van Hiel, A., Onraet, E., De Pauw, S.: The Relationship between Social-Cultural Attitudes and Behavioural Measures of Cognitive Style: A Meta-Analytic Integration of Studies. *J. Pers.* 78(6), 1765-1800 (2010)
- Van Rosmalen, J., Van Herk, H., Groenen, P.: Identifying Response Styles: A Latent-Class Bilinear Multinomial Logit Model. *J. Market. Res.* 47(1), 157-172 (2010)
- Van Vaerenbergh, Y., Thomas, T.D.: Response styles in survey research: A literature review of antecedents, consequences, and remedies. *Int. J. Public Opin. Res.* 25(2), 195-217 (2013)
- Verba, S.: The citizen as respondent: sample surveys and American democracy presidential address. *Am. Polit. Sci. Rev.*, 90(1), 1-7 (1996)
- Weech-Maldonado, R., Elliott, M.N., Oluwole, A., Schiller, K.C., Hays, R.D.: Survey Response Style and Differential Use of CAHPS Rating Scales by Hispanics. *Medical Care* 46(9), 963-968 (2008)
- Weijters, B., Geuens, M., Schillewaert, N.: The stability of individual response styles. *Psychol. Meth.* 15(1), 96-110 (2010)
- Winkler, J.D., Kanouse, D.E., Ware, J.E.: Controlling for Acquiescence Response Set in scale development. *J. Appl. Psychol.* 67(5), 555-561 (1982)
- Zhang, C., Conrad, F.: Speeding in web surveys: The tendency to answer very fast and its association with straightlining. *Surv. Res. Meth.* 8(2), 127-135 (2014)
- Zhou, B., McClendon, M.J.: Cognitive Ability and Acquiescence. In: *Proceedings of the Survey Research Methods Section*, 1003-1012 (1999)
- Zillmann, D., Schmitz, A., Skopek, J., Blossfeld, H.P.: Survey Topic and Unit Nonresponse. Evidence from an Online Survey on Mating. *Qual. Quant.* 48(4), 2069-2088 (2014)

Structural Stability, Quantitative Change

A Latent Class Analysis Approach towards Gender Role Attitudes in Germany

Alice Barth, University of Bonn

Miriam Trübner, University of Bonn

© 2018. Post-print manuscript version made available under the CC-BY-NC-ND 4.0 license
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

The original research article is published as:

Barth, Alice; Trübner, Miriam (2018). Structural Stability, Quantitative Change: A Latent Class Analysis Approach towards Gender Role Attitudes in Germany. *Social Science Research*, 72, 183-193. doi: 10.1016/j.ssresearch.2018.02.008

Abstract

Since the 1960s, social science surveys have aimed to assess respondents' attitudes towards gender roles. In this paper, a model-based clustering approach towards gender role attitudes is proposed in contrast to commonly used dimensional methods. Working from a role theory perspective, we expect different profiles in the population when it comes to role expectations. Using data from the German General Social Survey in 1991 and 2012, we identify attitude patterns via multi-group latent class analysis, taking differences between Eastern and Western Germany into account. Five latent classes representing different combinations of role expectations towards couples and parents and varying levels of intensity are distinguished. Attitude change is assessed by comparing latent class prevalence over time, while statistically testing for measurement equivalence. The analysis reveals a regionally variable decrease in traditional role models: Eastern Germany exhibits a faster pace of 'de-traditionalization' and less potential for role conflicts regarding working mothers.

1. Introduction

Change and persistence in gender roles – beliefs, norms, and expectations of ‘appropriate’ behavior based on sex – have been widely discussed topics in the social sciences over the last decades (e.g., Cotter, Hermsen & Vannemann 2011; Lindsey 2015; Mason, Czajka & Arber 1976). In many social science surveys, e.g., the U.S. General Social Survey, the British Understanding Society Panel, the International Social Survey Programme or the German General Social Survey, respondents’ attitudes towards gender roles are assessed via several items that mainly involve attitudes towards the division of labor in the family and the consequences of women working. Analyses of gender role attitudes usually assume a dimensional structure, whereas clustering approaches are relatively rare. In particular, the construct has often been operationalized as one-dimensional, ranging from a traditional, male breadwinner model to an egalitarian point of view (e.g., Cotter et al. 2011; Knudsen & Wærness 2001). Other researchers assume two dimensions, arguing that support for single-earner households does not necessarily imply opposition towards dual-earner arrangements (Blasius & Thiessen 2006). In doing so they distinguish between items which concern women’s nurturant and instrumental roles (Scott 2008) or differentiate between the perceived consequences of women participating in paid labor for families, and general gender norms (Blohm & Walter 2016; Sjöberg 2004). Irrespective of the number and designation of dimensions, the approaches share the general assumption of a dimensional structure of gender role attitudes that holds true for the whole population.

In the dimensional conceptualization, attitude change is assessed as a shift on a latent continuum. Though seldom stated explicitly, researchers generally suppose that the scale of this continuum is time-invariant. It is (implicitly) assumed that changes in response distributions can unanimously be attributed to attitude changes, i.e. that the results are not compromised by time-related non-substantive variation. Operating under these conditions, a large body of research affirms that, over the last decades of the 20th century, there is – albeit at different paces and without a strictly linear course – a global trend towards egalitarian attitudes (Braun & Scott 2009; Dorius & Alwin 2012; Scott 2008). However, the question as to whether the assumption of the construct’s invariance holds true – that is, whether measurement equivalence is given – is rarely assessed.

In this paper we propose a model-based clustering approach towards gender roles as an alternative to this widespread dimensional analysis strategy. Instead of extracting dimensions from a set of items, we apply latent class analysis (LCA), identifying groups (latent classes) with different ‘attitude profiles’. Theoretically, the clustering approach draws on classic sociological conceptualizations of the term ‘role’ (Dahrendorf 1964; Linton 1936; Merton 1957; Parsons 1951). In this perspective, gender role attitudes involve role expectations pertaining to distinct statuses,

such as mother, employee, or wife. This complex of – potentially conflicting – expectations regarding the behavior of men and women in diverse familial and organizational roles suggests that there are different latent classes that comprise distinct combinations of gender role expectations (Grunow, Begall & Buchler 2018; Knight & Brinton 2017).

In a latent class framework, attitude change is assessed by comparing the number, composition, and size of classes at different time points. In this context, we explicitly take the possibility of time-related structural changes and non-substantive variation into account. Therefore, it is considered a necessary precondition to test for measurement equivalence before drawing any substantive conclusions about attitude change, i.e., an equal number of classes and equivalence of class-specific response probabilities in the case of LCA. While dimensional methods also allow researchers to test for measurement equivalence, this has hardly been realized in the context of gender role attitudes (Blasius & Thiessen 2006; Constantin & Voicu 2015). We exemplify our approach by using data on gender role attitudes from the German General Social Survey (GGSS) of 1991 and 2012.

The article is structured as follows: We start with a discussion of the theoretical and methodological conditions of the clustering approach, followed by some information on gender relations in Germany. Then, we apply LCA to six categorical gender role items from the GGSS. We interpret the resulting classes and conduct tests for measurement equivalence. Subsequently, attitude change is assessed by comparing class prevalence over time. We conclude with a discussion of our findings and the methodological implications of our approach.

2. Background

2.1 Gender roles as social roles

The methodological decision in favor of a clustering approach can be illustrated by situating gender roles within the sociological paradigm of social roles (Lindsey 2015). Linton (1936) describes social roles as the behavior of occupants of a certain social status that is oriented towards the patterned expectations of others. In modern societies, individuals occupy multiple statuses at the same time, such as mother, employee, and wife. In turn, each status is characterized by an array of roles related to the social circle of the status occupant. Such a ‘role set’ (Merton 1957) may, for example, incorporate a mother’s interaction with her children, but also with the children’s teachers or their father. Inter-role conflicts occur when a person’s status-set includes roles which are mutually incompatible (Dahrendorf 1964). For example, a father is expected to spend time with his children, an activity that may collide with his perceived duty to work in order to support the family.

These considerations imply that the *structure* and *combination* of role expectations can be expected to differ between individuals. For example, some people's expectations of a mother's behavior towards her children may have implications for their perception of all women in paid labor, while for others these subjects are entirely unconnected. In this perspective, a clustering model investigating segments of the population with different combinations of role expectations is more in line with role-theoretical implications than the postulation of a dimensional structure that is expected to hold for all individuals. A clustering approach that is directly based on single items facilitates a differentiated analysis of heterogeneity in answer patterns. At the same time, it preserves as much information as possible on role expectations towards different statuses and reflects the interplay of various role expectations in distinct groups.

Role expectations are learned and internalized via socialization processes in social institutions such as the family and the education system (Dahrendorf 1964; Parsons 1951) and are characterized by values and norms which evolved in a specific social context (Parsons 1951). Social roles can thus be understood as a linkage between the individual, its social circle, and society at large. In the context of gender, these roles are currently undergoing major changes in contemporary societies (McHugh & Frieze 1997). Lopata (2006) argues that social differentiation in advanced industrial societies leads to more complex role sets, as well as to an increased likelihood of role conflicts. Role expectations are not static, but can be expected to differentiate across time, for example when the 'role-set' (Merton 1957) of a wife is gradually perceived as more independent from the one of a mother. These changes cannot be conceptualized bipolarly on a continuum from traditionality to modernity (cp. Gusfield 1967), or, as in the context of gender roles, egalitarianism (Knight & Brinton 2017). In addition, social roles change in different ways: for instance, the role definitions of wives and mothers have been undergoing more substantial changes in the last decades than those of husbands and fathers. Lopata (2006) also points out that family roles change to varying degrees depending on aspects such as ethnicity, religion, and social class and their interweaving. Accordingly, we assume that gender role expectations will not only have different structures in different segments of the population, these structures may also be subject to differential changes over time. These considerations imply that testing for measurement invariance is necessary.

2.2 Measuring gender role attitudes

The measurement of gender role attitudes by means of item batteries has been an integral part of major social science surveys for many years (Beere 1990; Braun 1998). The survey items which are in the focus of the present article were mainly introduced in American studies in the 1960s (Braun 2008; Mason et al. 1976), whereas the complete item battery has been part of the US General Social

Survey (GSS) since the 1970s (Smith et al. 2013). This set of questions centers on questions of female labor force participation and its consequences for the family. The item battery has been incorporated (partly in an adapted or augmented form) into many well-known social science surveys, among them the International Social Survey Programme, the British Household Panel Study and its successor Understanding Society, and the GGSS (Braun 1998; Lee, Alwin & Tufiş 2007).

The items represent various statuses of women – mother, employee, housewife, or wife – with corresponding role sets, therefore reflecting some of the complexity of expectations towards gender roles. Studies applying a dimensional representation of the gender role items mainly postulate a two-dimensional structure, differentiating between consequences of women working and gender ideology (Lee et al. 2007) or gendered division of labor (Blohm & Walter 2016). Using only a subset of three or four traditionally-slanted items, Mays (2012) and Dirksmeier (2015) employ a one-dimensional scale. Grunow and colleagues (2018) and Knight and Brinton (2017) apply LCA to slightly different gender role item sets used in the European Values Study (EVS) and the World Values Survey (WVS), respectively. They both find a multidimensional structure with several classes that reflect different facets of gender egalitarianism, e.g., the promotion of role symmetry versus a preference for individual choice.

The embeddedness of roles in the social context implies that the methodological problem of temporal validity is particularly relevant in the domain of measuring change in gender role attitudes in both cross-section and panel studies. On the one hand, valid analyses of social change necessitate the use of an unchanging instrument to ensure that substantial changes are not confused with methodological effects, for example, due to differences in item wording (Smith 2005). On the other hand, the use of the same instrument may result in floor or ceiling effects (e.g., when nearly everyone is convinced that a working mother can establish a good relationship with her children, irrespective of other attitudes), or respondents' understanding of terms or concepts may change in the face of an altered social context (Lugtig, Boeijs & Lensvelt-Mulders 2011; McHugh & Frieze 1997), thus compromising the validity of temporal comparisons. For example, the meaning attached to the term 'working mother' depends on occupation and circumstances – respondents may visualize a secretary, a farmer, or a journalist, working full-time or part-time, which can lead to different interpretations of questions. Finally, several studies have criticized the fact that most items currently used in social science surveys represent a traditional view (Braun 2008; Walter 2017) and focus exclusively on women's roles (Braun 1998; Mason & Lu 1988), as many items have remained unchanged since the 1960s. Therefore, emerging differentiations in the realm of non-traditional role expectations are hard to identify, and respondents whose opinions do not fit into offered categories have been shown to exhibit seemingly erroneous response patterns. For

example, Behr et al. (2012) note an emerging preference for individual solutions, which becomes evident in the rejection of both traditional and (specific) egalitarian role models. In addition, when survey questions are perceived as outdated, obscure, or ill-defined by respondents – for example because they are unsure whether a ‘working mother’ refers to a full-time or part-time job – a higher amount of measurement error due to non-substantive response behaviors (e.g. indiscriminately choosing middle categories) is likely to occur. In line with this concern, Barth (2016) demonstrated that the share of method-induced variation – as opposed to substantial one - had been increasing in the gender role instrument in Great Britain from the 1990s to the mid-2000s.

In sum, these observations indicate that in the assessment of changing gender role attitudes, particular attention should be paid to the question of whether the instrument actually measures the same construct in the same way at different time points – that is, whether measurement equivalence is guaranteed (Eid, Langeheine & Diener 2003; Vandenberg & Lance 2000). Although measurement equivalence is a necessary condition for a valid assessment of differences (e.g., Van de Vijver & Leung 1997), and there are a number of indications implying that the meaning of some items is not invariant in time and space, the equivalence of gender role attitudes has only been assessed occasionally. Using multidimensional scaling, Braun (1998) demonstrates systematic intercultural differences in the interpretation of items used in the International Social Survey Programme (ISSP). Constantin and Voicu (2015) claim that gender role scales in the ISSP and the World Value Survey exhibit configural and metric – but not scalar – invariance. This result of confirmatory factor analysis means that the item batteries essentially measure the same concept, but do not permit the comparison of levels in gender egalitarianism across countries. Knight and Brinton (2017) come to a similar conclusion by applying LCA to a pooled data set of several waves from the EVS and WVS: they find that a partially homogeneous model with intercepts varying by country fits the data best, which is conceptually similar to metric invariance (Kankaraš, Moors & Vermunt 2010). While they also test whether their estimates of change over time are dependent on model choice, the question of the class profiles’ temporal invariance is not discussed. Assessing measurement invariance by region (East-West), gender, and time (1982-2004, only Western sample) in Germany, Lee et al. (2007) conclude – despite significant differences between Eastern and Western Germany – that a two-factor model with coefficients constrained to be equal exhibits acceptable fit, thus assuming that measurement invariance holds across all (sub)samples. To our knowledge, the latter is the only formal test of this gender role item battery’s equivalence over time in the literature, although the rootedness of the concept in sociohistorical context and the methodological challenges of measuring social change give cause for concern that gender role measures might not be fully equivalent over time. Due to these considerations, a statistical test of whether the meaning of latent classes remains the same over time will precede our interpretation

of class profiles and quantitative changes. As Germany is characterized by different 'gender arrangements' in the East and West, which might effectuate qualitative differences in attitude patterns, measurement equivalence in terms of region will also be assessed.

2.3 Gender role attitudes in Germany

Germany is an interesting case study for two reasons: First, both social norms and state policies regarding women's labor market participation were fundamentally different between the former Federal Republic of Germany (FRG) in the West, and the German Democratic Republic (GDR) in the East. In the FRG, a traditional family model was prevalent up to the 1970s, with women's duties mainly in the private sphere of the household, particularly childcare, and men as the sole, full-time employed breadwinners (Pfau-Effinger & Smidt 2011). The gendered division of labor was also strongly supported by welfare-state arrangements (Rosenfeld, Trappe & Gornick 2004). This dominant family model started to change in the 1970s, slowly giving way to a male breadwinner/female part-time care-giver model. Nevertheless, a strong emphasis was placed on mothers caring (full-time) for their children in the family during the first years after birth (Pfau-Effinger 2004).

In the socialist GDR, in contrast, full integration of women into paid labor – as well as comprehensive control of children's socialization and education from an early age – were major national objectives, and the state provided full-time childcare institutions (Rosenfeld et al. 2004). This political and institutional framework entailed a re-shape of gender role beliefs in the population (Kreckel & Schenk 2000). Despite their full-time employment, however, women still bore the main responsibility for the family and household labor (Kolinsky 2003). Thus, differences between GDR and FRG – in particular the gendered division of homework and childcare – were less pronounced in the private sphere than at the workplaces. In sum, before the reunification Eastern and Western Germany could be characterized as manifesting different 'gender arrangements' (Pfau-Effinger 2004), which suggests that qualitatively different attitude patterns might exist.

The second reason is that German social policies have been subject to considerable changes following the reunification of the two German states in 1990. In the 1990s, family policies in reunified Germany were mainly based on the West German model, promoting a gendered division of labor in the family (Pfau-Effinger & Smidt 2011). However, the mid-2000s experienced a paradigm shift in family policies. Since then, public childcare institutions have been substantially expanded, and an updated parental leave law supports the financial autonomy of caring parents for one year, promoting the return to the workplace after that time and the participation of fathers in

childrearing (Kluve & Tamm 2013). The tax system, however, still benefits an unequal distribution of income in married couples.

In sum, different social contexts in the former states of FRG and GDR as well as substantial changes in welfare-state policies since the unification render the assessment of gender role change in Germany an interesting research question. Previous studies employing dimensional models show that while there is a trend towards more egalitarian attitudes in Germany since the reunification, the attitudinal gap between East and West persists or has even widened (Bauernschuster & Rainer 2012; Lee et al. 2007). In research on the class structure of gender role attitudes, however, only data from Western Germany has been assessed (Grunow et al. 2018) or East and West have been combined (Knight & Brinton 2017).

Drawing on our theoretical assumptions and previous research, we expect to find at least three different types of role expectations in the German sample. First, a class that represents almost unequivocally traditional attitudes, supporting a male breadwinner model and locating women's responsibilities mainly in the private sphere, in particular childcare. Second, the opposite type that has a completely egalitarian point of view, arguing in favor of women's labor force participation and strongly rejecting traditionally slanted items. Third, we expect at least one class of respondents differentiating between the roles of mother and wife, supporting gender equality but stressing the importance of mothers as primary caregivers for small children. Further differentiations of these three basic profiles at one or both time points are entirely possible. With regard to regional and temporal differences, we expect the egalitarian class(es) to be more prevalent in Eastern Germany and to increase over time in both parts of the country. The traditional and differentiated class(es) should be more pervasive in Western Germany, with the traditional class(es) decreasing from 1991 to 2012. Due to the assumption of increasingly complex role sets, we expect the differentiated class(es) to increase over time.

3. Data and Methods

We use data from the GGSS, a representative cross-sectional survey of the German adult population that is conducted every two years. The gender role questionnaire was fielded for the first time in 1982 in the GGSS (only FRG) and every four years as of 1991 (when a supplementary survey was fielded due to the re-unification of the FRG and GDR in 1990) onwards. Due to model complexity when testing for measurement invariance, we restrict our analyses to two points of time. In Germany, no reversals of the trend towards more egalitarianism have been reported (Blohm & Walter 2016), thus we use the longest time span for which data are available on both parts of the country, comparing German gender role expectations soon after the reunification

(1991) to those held at present (2012). The sample size was 3,049 in 1991 and, due to a sample split in terms of gender role items, it was 1,726 in 2012. The questions on gender role attitudes in the GGSS cover statements about the division of labor in the family and the consequences of women's labor force participation, measured on a four-point Likert scale from 'fully agree' to 'fully disagree' (see table 2 for item wording).

The typology of gender role attitude patterns is constructed by applying LCA to the items. LCA is a statistical method that can identify previously unobserved groups in a population. These groups – the latent classes – are represented by a categorical latent variable. The classification into latent classes is based on respondents' answer patterns to several categorical indicators. In the case at hand, the indicators are the six survey items on gender role attitudes, while the latent class variable represents different types of role expectations.

In contrast to variable-centered approaches such as factor analysis which assume that a linear structure of relations between variables will hold for all individuals, LCA has been termed a person-centered approach that identifies groups (clusters) of individuals who exhibit similar answer profiles (Bergman & Magnusson 1997; Collins & Lanza 2013). Thus, consistent with our theoretical considerations, the interplay of different gender role expectations in distinct groups can be investigated. Another criterion for the choice of LCA is that the gender role items we use were measured on a four-point scale and should thus be treated as categorical rather than metric variables.

Unlike other clustering methods such as k-means or hierarchical clustering, LCA derives clusters using a probabilistic model, which allows for model selection and the assessment of goodness of fit based on likelihood statistics. Indicators of absolute model fit used here are log-likelihood (LL) and likelihood-ratio chi-square test (L^2). The optimal number of classes is determined by several statistical information criteria. We use the Bayesian Information Criterion (BIC) and Akaike's information criteria (AIC and AIC3) as well as the Lo-Mendell-Rubin-test (LMR) and entropy (see, e.g., Kankaraš et al. 2010; Rudnev, Magun, & Schmidt 2016). BIC, AIC and AIC3 are minimized in the solution that provides the best relative fit to the data. LMR is a likelihood-based technique to compare the fit of a k-class model to the model with k-1 classes. It indicates whether there is a significant improvement in model fit when adding one class. Entropy values show the goodness of class separation, ideally approaching 1. However, information criteria may differ in their selection of the optimal model. While BIC has been reported to perform well, particularly in large samples, AIC and AIC3 often overestimate the number of classes (Morgan 2015; Nylund, Asparouhov & Muthén, 2007; Yang 2006)¹. Due to the possible inconsistencies

¹ In a simulation study, Nylund et al. (2007) find that a bootstrap likelihood ratio test (BLRT) is slightly preferred over the LMR. However, we use weighted data, for which the BLRT is not provided in MPlus. A run without the

associated with statistical criteria, interpretability and compliance with theoretical expectations are important criteria in model selection as well. The substantive interpretation of the classes is based on respondents' class-specific response propensities (in this case, the propensity to choose between four ordered response categories in each of the six indicator variables). As the allocation of respondents to classes is probabilistic, the possibility of measurement error is taken into account.

In the context of LCA, measurement equivalence is defined as 'equivalence in the class-specific conditional response probabilities' (Eid et al. 2003, p. 206). The comparison of the size of classes is only sensible when this condition of equal class profiles across groups or time points is satisfied. It is tested by comparing the fit of a multi-group LC model in which the item-response probabilities across groups are unrestricted (heterogeneous model), to the fit of a restricted (homogeneous) model. In the latter, structurally equivalent model, the item-response probabilities are equal across groups. The models are nested, therefore enabling a direct comparison of fit (McCutcheon 2002). If the fit of the homogeneous model is not significantly worse than the heterogeneous one, measurement equivalence in terms of the grouping variable can be assumed. Following the recommendation of Kankaraš et al. (2010) and Collins and Lanza (2013) for model comparison when sample size is large, we use the BIC as the main criterion in assessing measurement equivalence.

The analysis proceeded in several steps: First, we fitted separate LC models to the samples in 1991 and 2012 to explore the optimal number of classes separately for each time point. Then, we pooled the two samples and tested for measurement equivalence across time and region. To this end, we conducted several multi-group LCAs with a grouping variable reflecting the possible combinations of region and time point (4 groups). The relative fit of these models, in which the conditional response probabilities had the constraint of being equal across time, region, or both, was then compared to the fit of the unrestricted model. After having established the comparability of class profiles, we interpreted the role expectation types and studied attitude change in terms of differences in class prevalence from 1991 to 2012 and between Eastern and Western Germany. All models were estimated using MPLUS 8 (Muthén & Muthén 1998-2017).

4. Results

4.1 Model selection and equivalence testing

In the first step, we estimated separate models for the years 1991 and 2012 with one to seven classes each. Information criteria regarding the optimal number of classes were inconsistent: while AIC and AIC3 continually decreased, thus favoring seven classes or more, LMR suggested that in

weight variable showed that in 1991, the BLRT was consistently significant up to 9 classes, suggesting a model with 9 classes or more, whereas in 2012, it pointed to a 7-class solution.

1991, the solution had not significantly improved when adding a fifth class, while in 2012 a 2-class model was regarded as sufficient. In both years, the BIC was minimized in the five-class solution (see table A1 in the appendix). An inspection of the conditional response probabilities revealed that classes representing the expected profile of differentiated role expectations towards mother and wife emerged most clearly in the 5-class model in both years. We also found that when considering five classes, class profiles were comparable in 1991 and 2012: Correlations of corresponding class profiles were between .80 and .95, suggesting similar attitude patterns (Rudnev et al. 2016). We thus decided to proceed with the 5-class solution.

The next step was testing for measurement equivalence between Eastern and Western Germany and across time. This was done by fitting multi-group LC models in which years, regions, or both were constrained to be equal and compared to the unrestricted, fully heterogeneous model (model 1). In model two, the class profiles were allowed to differ between Eastern and Western Germany, but constrained to be equal over time. Model three required the conditional response probabilities to be equal in both Eastern and Western Germany, but allowed different profiles in 1991 and 2012. Model four was the most restrictive model, with class profiles being equal both across regions and over time. As can be seen in table 1, class separation (entropy) was sufficient in all models, while the model assuming full measurement invariance (model 4) provided the best fit in terms of BIC.

Table 1 Measurement equivalence: comparison of model fit, lowest BIC value in bold

Model	Npar	LL	BIC	entropy
(1) fully heterogeneous*	379	-36719	76649	0.85
(2) years invariant, region heterogeneous	199	-36977	75640	0.83
(3) region invariant, years heterogeneous	199	-36948	75581	0.84
(4) years and region invariant	109	-37122	75167	0.84

* best LL not replicated

Consequently, our final model was the fully invariant five-class multi-group LC model, with time by region (four groups) as the grouping variable. Allaying concerns about a possible lack of measurement equivalence, the statistical analysis in fact suggested that conditional response probabilities – and therefore both structure and interpretation of the five classes – were the same

across time and region. Accordingly, class sizes can be meaningfully compared: temporal change and East-West differences are captured in differential class prevalence (see chapter 4.3).

4.2 Class interpretation

The interpretation of the five classes is based on the conditional response probabilities in each class (see table 2). The prevalence of the respective classes in 1991 and 2012 in Eastern and Western Germany can be seen in table 3, while table A2 in the appendix gives illustrative information on class composition with regard to gender, age, marital status, educational qualification, income, and work status.

Table 2 Conditional response probabilities in the fully invariant five-class multi-group LC model (probabilities that guide class interpretation in bold)

		(1) "family-oriented traditional"	(2) "child-oriented traditional"	(3) "unconditionally egalitarian"	(4) "conflicted egalitarian"	(5) "moderately egalitarian"
<i>A working mother can establish just as loving and secure a relationship with her children as a mother who doesn't work.</i>	<i>fully agree</i>	38.4	36.0	97.7	64.6	75.2
	<i>agree</i>	17.2	43.4	1.8	26.0	23.6
	<i>disagree</i>	28.8	20.0	0.0	6.5	1.1
	<i>fully disagree</i>	15.7	0.6	0.4	2.8	0.1
<i>It's more important for a wife to help her husband with his career than to pursue her own career.</i>	<i>fully agree</i>	25.6	5.3	3.5	0.3	4.3
	<i>agree</i>	38.1	32.0	5.8	2.2	17.3
	<i>disagree</i>	22.7	50.2	20.5	19.5	66.5
	<i>fully disagree</i>	13.6	12.5	70.2	78.0	11.9
<i>A small child is bound to suffer if his or her mother goes out to work.</i>	<i>fully agree</i>	81.4	28.0	3.8	14.5	8.1
	<i>agree</i>	14.0	59.4	8.2	42.4	28.1
	<i>disagree</i>	3.4	9.5	20.9	37.9	53.1
	<i>fully disagree</i>	1.2	3.2	67.1	5.2	10.7
<i>It is much better for everyone concerned if the man goes out to work and the woman stays at home and looks after the house and children.</i>	<i>fully agree</i>	64.3	7.9	0.5	0.2	1.4
	<i>agree</i>	28.0	56.5	2.1	1.6	11.1
	<i>disagree</i>	6.8	33.4	5.4	46.4	68.6
	<i>fully disagree</i>	0.9	2.3	92.0	51.8	18.9
<i>A child actually benefits if his or her mother has a job rather than just concentrating on the home.</i>	<i>fully agree</i>	4.7	1.1	59.0	5.0	16.4
	<i>agree</i>	10.1	25.7	28.8	52.5	66.4
	<i>disagree</i>	35.3	60.7	6.1	30.1	15.8
	<i>fully disagree</i>	49.9	12.5	6.1	12.5	1.5
<i>A married woman should not work if there are not enough jobs to go round and her husband is also in a position to support the family.</i>	<i>fully agree</i>	56.8	8.9	4.5	0.9	9.0
	<i>agree</i>	26.0	52.2	6.7	7.8	22.7
	<i>disagree</i>	9.4	30.1	16.9	29.7	48.6
	<i>fully disagree</i>	7.9	8.7	72.0	61.6	19.7

Note: English translation of German items as suggested by Wasmer 2014; values important for interpretation in bold

Respondents in the first class mainly reject the notion that a mother's role should involve active labor market participation, strongly agreeing that 'a small child is bound to suffer if his or her mother goes out to work'. However, respondents' opinions as to whether a loving and secure relationship between children and a working mother can be established vary. Although the probability of fully disagreeing with this statement is higher compared to all other classes, it still accounts for only 15.7%, whereas the likelihood of agreeing or fully agreeing is more than 50%.

Regarding the division of labor between husband and wife, the male breadwinner model is embraced by a clear majority. Opinions towards the statement that ‘it is more important for a wife to help her husband with his career than to pursue her own career’ are more ambiguous: (Full) agreement is higher than in all other classes, but there is also a one in three chance of disagreeing. One might assume that pursuing a career is not part of the role expectations of a number of respondents in this class; this interpretation is supported by the observation that educational qualification and household net income are lowest among this type (see table A2), which suggests a high frequency of jobs with no or limited career opportunities. Further, mean age and the share of married respondents are highest compared to the other classes. On the whole, the response patterns of class one point to a comparatively strong endorsement of traditional gender roles, both in terms of parenting and partnership, as well as a certain disinterest in work-related matters. Therefore, we label this type of role expectation ‘family-oriented traditional’.

Respondents in the second class are also quite clear in their opinion that a mother’s place is with her children. Agreement with the statement concerning the quality of working mother’s relationship to her child is nevertheless high (almost 80%). Positions towards the division of labor in married couples are split: The probability of agreeing with the two statements advocating a male breadwinner model is about 60%, whereas the idea that ‘it is more important for a wife to help her husband with his career than to pursue her own career’ is rejected by more than half of the respondents. Respondents in this class exhibit an overall tendency to pick moderate rather than extreme values. In contrast to the ‘family-oriented’ class, respondents are not as fierce in their embracement of traditional values. However, they are skeptical towards women going out to work when (small) children are involved. Thus, role expectations of this type are referred to as ‘child-oriented traditional’.

In the third class, the notion that labor market participation is an integral part of the role of mother and wife is strongly embraced. Nearly all respondents are convinced that children will not suffer, but rather benefit if their mother has a job outside the home. Similarly, strong disagreement with items suggesting a traditional allocation of roles between husband and wife is very prevalent. This type of respondent thus has a firmly egalitarian point of view, regardless of whether a woman is referred to as being a mother or a wife. In terms of response behavior, this is expressed in a pronounced tendency to choose extreme values throughout the whole set of items. Consequently, the role expectation in this class is labelled ‘unconditionally egalitarian’. This class has the highest share of women (almost two-thirds), household income and educational qualification are high.

Respondents in the fourth class exhibit the highest relative probability of fully disagreeing with the notion that ‘it is more important for a wife to help her husband with his career than to

pursue her own career', and emphatically dismiss the statements about the reduced labor market participation of married women as well. Expectations towards mothers are, however, ambivalent. Although working mothers are defended by overall agreement with the statement that 'a working mother can establish just as loving and secure a relationship with her children as a mother who doesn't work', a majority fears that small children suffer when their mothers work, and opinions as to whether children might benefit vary. Thus, respondents in this class strongly support a female career and argue against any subordination of a wife to her husband's work obligations; at the same time, many seem to expect mothers to assume the main responsibility for childcare. This type's role expectation is referred to as 'conflicted egalitarian'; it is especially prevalent among young, well-educated respondents.

The fifth class consists of respondents who have a high probability of considering the quality of a working mother's relationship to her children as good. The majority also agrees that a mother's job can have benefits for children, whereas the likelihood of fearing negative consequences is less than 40%. The three statements in favor of a male breadwinner model are mainly rejected, albeit not as resolutely as in classes three and four. Compared to the conflicted egalitarian type, respondents in this class are more optimistic regarding the reconciliation of family and work for mothers, but at the same time less insistent on women's careers. As there is a visible tendency to choose moderate rather than extreme values, the role expectation in this class is called 'moderately egalitarian'. Here, respondents' qualification and income is lower than in the other egalitarian classes.

On the whole, the ambiguity of certain items, especially in the domain of the child- and family-oriented types, shows that some role expectations are individually varied even within segments of the population who otherwise share distinct positions towards the division of labor in the family. This also suggests that certain items might be more difficult to respond to for subgroups of the population, since certain aspects of the roles – such as support for a partner's career – do not relate to their daily life (Biddle 1979, p. 257f).

In general, there are few respondents who strongly disagree with the notion that 'a working mother can establish just as loving and secure a relationship with her children as a mother who doesn't work' (in sum 4.0% in 1991 and 1.7% in 2012); strong support for the idea that 'it is more important for a wife to help her husband with his career than to pursue her own career' is also rare (8.6% in 1991; 4.6% in 2012). These items are not only strongly skewed, but also differentiate less between latent classes compared to the other items. Regarding a possible revision of question formulation, they should be among the first to be considered.

4.3 Changing role expectations

In the final step of our analysis, we assess changes in gender role expectations by comparing latent class prevalence in 1991 and 2012 in Eastern and Western Germany. Table 3 shows the percentages of the five types in 1991 and 2012 based on estimated posterior probabilities (most likely class membership), separately for Eastern and Western Germany. It is clear that over two decades significant changes have taken place, while pronounced regional differences remain. In 1991, the share of family-oriented and child-oriented traditional types accounted for about 65% of the Western German population, whereas the unconditionally egalitarian mindset was prevalent only in 5%. 21% exhibited conflicting role expectations towards women. In Eastern Germany, the percentage of traditional (39%) and conflicted egalitarian (11.5%) role expectations was much lower than in the West, while more respondents expressed moderately or unconditionally egalitarian views.

In both parts of Germany, role expectations have shifted towards a more egalitarian perspective over time, albeit at a different pace. In the West, the proportion of traditional types was reduced by half, constituting 30% of the population in 2012. In the East, as few as 11% of the total population expressed family or child-oriented traditional role expectations in 2012, which is a reduction by two thirds compared to 1991. In contrast, about half of the Eastern German population had become unconditionally egalitarian, whereas this class covered not even a quarter of Western Germans. The proportion of ‘conflicted egalitarians’ remained stable at one fifth of the population in Western Germany. In comparison, only one percent of Eastern Germans still perceived a conflict between a women’s career and childcare in 2012.

Table 3 Class prevalence in 1991 and 2012 in percent, separately for Eastern and Western Germany

		family-oriented traditional	child-oriented traditional	unconditionally egalitarian	conflicted egalitarian	moderately egalitarian	N	East-West difference
1991	West	27.7	37.7	5.3	20.9	8.4	1510	X ² =473.2 df=4 CV=0.39 p<.001
	East	12.0	27.1	16.5	11.5	32.9	1539	
2012	West	11.8	20.7	22.5	19.5	25.4	1175	X ² =275.5 df=4 CV=0.40 p<.001
	East	3.5	7.2	50.5	1.1	37.8	550	

Although the composition of the population in terms of class prevalence underwent pronounced changes in both parts of Germany, the effect size of the difference between East and

West has, with a Cramér's V of .40, remained virtually the same². A similar effect size is observed in terms of the difference between 1991 and 2012 when Western ($X^2=418.8$ $df=4$ $CV=0.39$ $p<.001$) and Eastern ($X^2=339.8$ $df=4$ $CV=0.40$ $p<.001$) Germany are observed separately. Thus, one could say that the difference in terms of gender role expectations between Eastern and Western Germany accounts for a gap of about twenty years.

5. Discussion

The aim of this article was to assess the complexity of change and persistence in gender role attitudes, taking Germany as an example. From the perspective of sociological role theory, we argued for a classificatory operationalization of gendered role expectations. LCA was used to differentiate segments in the German population based on survey respondents' expectations with regard to the behavior of mothers, fathers, husbands, and wives. Five latent classes were distinguished that varied in attitude extremity, the desired division of labor in couples, and the extent to which respondents distinguished between expectations towards mothers and wives. We found that 'traditional' positions differed in terms of attitudinal extremity and the prioritization of the whole family or a mother's involvement with children, respectively. In the 'egalitarian' spectrum, a woman's right to a professional career was supported by most respondents, while role expectations differed strongly when children were involved. The hypothesis of different attitude profiles was therefore confirmed, and we were able to present a differentiated typology of gender role attitudes in Germany. In particular, class-specific combinations of role expectations towards women as part of a (married) couple and as mothers (especially prevalent in the 'conflicted egalitarian' and 'child-oriented traditional' types) suggest that dimensional approaches might overlook heterogeneity and potential role conflicts. In this respect, our results are consistent with other studies stressing the existence of gender role attitude types that do not fit into a linear continuum from traditional to egalitarian attitudes (Braun 2008; Grunow et al. 2018; Knight & Brinton 2017).

We further argued that, when assessing change in gender role attitudes, the possibilities of structural differences in the concept, shifts of meaning, or changes in response behavior over time should be taken into account. For these reasons, our comparative analysis of gender role attitudes

² These results are based on chi-square tests of posterior class probabilities (allocation of respondents to most likely class in the final model with year by region as grouping variable). As classification error is ignored when using most likely class membership, the standard errors are too small which leads to an overestimation of significance. However, the internal comparison of effect sizes from analyses based on the same latent class model is possible.

in Germany from 1991 to 2012 was preceded by testing for measurement equivalence in class-specific response probabilities. Due to persisting cultural, political, and institutional differences between Eastern and Western Germany, regional differences were also assessed. In contrary to our concerns, statistical testing for measurement invariance confirmed that the structure of the five classes could be considered equivalent between 1991 and 2012 and for Eastern and Western Germany, suggesting *structural stability* of gender role attitudes. The ensuing comparison of class prevalence revealed significant *quantitative differences* over time and between Eastern and Western Germany. Corroborating previous findings (Lee et al. 2007; Bauernschuster & Rainer 2012; Blohm & Walter 2016), the analysis showed that the share of the population with ‘egalitarian’ role expectations has increased in both parts of Germany since the early 1990s. This development can be interpreted as a reciprocal adaptation of role expectations and the social context (Dahrendorf 1964). Social changes – such as women’s increasing labor market participation, political measures to promote the reconciliation of work and family, and media discourse on the subject – provide possibilities for new role conceptions. Acting out these roles, in turn, not only affects the role owners, but also their social circles, which again creates the potential for social change. However, more complex role sets also entail the possibility of role conflict (Lopata 2006; Dahrendorf 1964), which is especially pronounced in the type of ‘conflicted egalitarian’ role expectations in our study. In this respect, regional differences are of particular significance: In Western Germany, the share of the population who supports women’s equal participation in paid labor while at the same time fearing negative consequences for children, has remained stable at one fifth over time. In contrast, the ‘conflicted egalitarian’ type accounted for only 11.5 percent of Eastern Germans in 1991 and vanished almost completely in 2012. While the persistence, or even widening, of the gap between more egalitarian attitudes in the post-socialist Eastern part of Germany and traditionalism in the West has been stated in several studies (Lee et al. 2007; Bauernschuster & Rainer 2012), we could show in addition that a major difference between East and West is constituted by the amount of perceived conflict between women’s careers and the demands of motherhood. The question as to whether this can mainly be traced back to socialization in different regimes, with respective norms persisting even after the downfall of the GDR (Bauernschuster & Rainer 2012), institutional factors such as the higher availability of childcare facilities in Eastern Germany, or an interaction of both (Pfau-Effinger & Smidt 2011), provides further research opportunities.

In sum, we can conclude that analyzing gender role expectations in a latent class framework sheds light on heterogeneous attitude profiles in the population, while at the same time enabling differentiated analyses of change and persistence. Thus, we were able to complement previous findings such as an overall trend towards more egalitarian attitudes and a persisting gender role gap between Eastern and Western Germany by distinguishing between several classes of role

expectations and their temporal development. Arguing for a differentiated analysis of role-conflicts and role-sets, one limitation of the study at hand is the small number of items used in class construction. While the goal of analyzing temporal changes in gender role expectations necessitated the restriction to items that were available at both time points, we are aware that a depiction of the full complexity of gender role expectations would require a larger set of statements. In particular, a stronger focus should be placed on role expectations towards fathers and husbands, and on gendered roles in non-familial contexts. Instruments covering these aspects do exist, but have scarcely been implemented in representative, longitudinal surveys, let alone sufficiently tested for cross-temporal and cross-cultural measurement equivalence (Constantin & Voicu 2015; McHugh & Frieze 1997). Moreover, both qualitative and quantitative research has shown that items promoting specific, non-traditional views, for example, a female breadwinner model or the equal sharing of all tasks, pose comprehension problems for a considerable number of respondents (Braun 2008; Behr et al. 2012). Similarly, in our analysis it became apparent that the differentiating power of established items is not equally distributed in the population, for example when there is no clear position towards a wife's support of her husband's career in distinct latent classes. It can be assumed that in such cases, the aspect in question does not constitute an integral part of respondents' role expectations, making the respective statement difficult to respond to. Therefore, more research on the adaptation of existing measurement instruments is needed. In this respect, it should be emphasized that considering the socio-historical embeddedness of role expectations, the development of an instrument that is time-invariant, equally valid in all population strata, and applicable to different cultural contexts, is unrealistic. Nevertheless, existing data still provide a rich source for comparative analyses given the application of adequate methods and the assessment of measurement equivalence.

6. References

- Barth, A. (2016). The changing nature of attitude constructs: an application of multiple correspondence analysis on gender role attitudes. *Quality & Quantity*, 50(4), 1507-1523.
- Bauernschuster, S., & Rainer, H. (2012). Political regimes and the family: how sex-role attitudes continue to differ in reunified Germany. *Journal of Population Economics*, 25(1), 5-27.
- Beere, C. A. (1990). *Gender roles: A handbook of tests and measures*. Westport, Connecticut: Greenwood Publishing Group.
- Behr, D., Braun, M., Kaczmirek, L., & Bandilla, W. (2012). Testing the validity of gender ideology items by implementing probing questions in web surveys. *Field Methods*, 25(2), 124-141.
- Bergman, L. R., & Magnusson, D. (1997). A person-oriented approach in research on developmental psychopathology. *Development and Psychopathology*, 9(02), 291-319.
- Biddle, B. J. (1979). *Role Theory: Expectations, Identities, and Behaviors*. New York: Academic.
- Blasius, J., & Thiessen, V. (2006). Assessing Data Quality and Construct Comparability in Cross-National Surveys. *European Sociological Review*, 22(3), 229-242.
- Blohm, M. & Walter, J. (2016). Einstellungen zur Rolle der Frau und der des Mannes. In: Statistisches Bundesamt & Wissenschaftszentrum Berlin für Sozialforschung (Hg.): Datenreport 2016. bpb, Bonn.
- Braun, M. (1998). Gender roles. In: J.W. van Deth (Ed.), *Comparative Politics. The Problem of Equivalence*. London/New York: Routledge.
- Braun, M. (2008). Using egalitarian items to measure men's and women's family roles. *Sex Roles*, 59(9-10), 644-656.
- Braun, M. & J. Scott (2009). Gender-Role Egalitarianism – Is the Trend Reversal Real. *International Journal of Public Opinion Research* 21(3), 362-367.
- Collins, L. M., & Lanza, S. T. (2013). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. Hoboken, NJ: John Wiley & Sons.
- Constantin, A., & Voicu, M. (2015). Attitudes Towards Gender Roles in Cross-Cultural Surveys: Content Validity and Cross-Cultural Measurement Invariance. *Social Indicators Research*, 123(3), 733-751.
- Cotter, D., Hermsen, J. M., & Vanneman, R. (2011). The End of the Gender Revolution? Gender Role Attitudes from 1977 to 2008. *American Journal of Sociology*, 117(1), 259-289.
- Dahrendorf, R. (1964). *Homo sociologicus*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Dirksmeier, P. (2015). The intricate geographies of gender ideologies in Germany. *Geoforum*, 64, 12-24.
- Dorius, S. F., & Alwin, D. F. (2012). The global reach of egalitarianism: Worldwide trends in gender beliefs. Unpublished document, Iowa State University. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.706.6361&rep=rep1&type=pdf>
- Eid, M., Langeheine, R., & Diener, E. (2003). Comparing typological structures across cultures by multigroup latent class analysis a primer. *Journal of Cross-Cultural Psychology*, 34(2), 195-210.
- Grunow, D., Begall, K., & Buchler, S. (2018). Gender Ideologies in Europe: A Multidimensional Framework. *Journal of Marriage and Family*, 80(1), 42-60.

- Gusfield, J. R. (1967). Tradition and modernity: Misplaced polarities in the study of social change. *American Journal of Sociology*, 351-362.
- Kankaraš, M., Moors, G., & Vermunt, J. K. (2010). Testing for measurement invariance with latent class analysis. In: Davidov, E., Schmidt, P., Billiet, J. (eds). *Cross-cultural analysis: Methods and Applications*, Routledge: New York/Hove, pp. 359-384.
- Kluve, J., & Tamm, M. (2013). Parental leave regulations, mothers' labor force attachment and fathers' childcare involvement: evidence from a natural experiment. *Journal of Population Economics*, 26(3), 983-1005.
- Knight, C. R. & Brinton, M.C. (2017). One Egalitarianism or Several? Two Decades of Gender-Role Attitude in Europe. *American Journal of Sociology*, 122(5), 1485-1532.
- Knudsen, K., & Wærness, K. (2001). National context, individual characteristics and attitudes on mothers' employment: A comparative analysis of Great Britain, Sweden and Norway. *Acta Sociologica*, 44(1), 67-79.
- Kolinsky, E. (2003). Gender and the Limits of Equality in East Germany. In: Kolinsky, E. & Nickel, H. M. (eds): *Reinventing Gender – Women in Eastern Germany since reunification*. Frank Cass Publishers, London/Portland, pp. 100-130.
- Kreckel, R., & Schenk, S. (2000). Full-Time or Part-Time? The Contradictory Integration of the East German Female Labor Force in Unified Germany. In: Marshall, V.W., Heinz, W.R., Krüger, H., and Verma, A. (Eds.), *Restructuring work and the life course*, University of Toronto Press; Toronto, 159-76.
- Lee, K. S., Alwin, D. F., & Tufiş, P. A. (2007). Beliefs about women's labour in the reunified Germany, 1991–2004. *European Sociological Review*, 23(4), 487-503.
- Lindsey, L. L. (2015). *Gender roles: A sociological perspective*. Abingdon/New York: Routledge.
- Linton, R. (1936). *The study of man: an introduction*. Oxford: Appleton-Century.
- Lopata, H.Z. (2006). Gender and Social Roles. In: Chafetz, J.S. (Ed.). *Handbook of the Sociology of Gender*. Springer Science & Business Media, New York.
- Lugtig, P. J., Boeijs, H. R., & Lensvelt-Mulders, G. J. L. M. (2011). Change? What change? An exploration of the use of Mixed-methods research to understand longitudinal measurement invariance. *Methodology*, 8(4), 1-9.
- Mason, K. O., Czajka, J. L., & Arber, S. (1976). Change in US women's sex-role attitudes, 1964-1974. *American Sociological Review*, 41(4), 573-596.
- Mason, K. O., & Lu, Y. H. (1988). Attitudes toward women's familial roles: Changes in the United States, 1977-1985. *Gender & Society*, 2(1), 39-57.
- Mays, A. (2012). Determinanten traditionell-sexistischer Einstellungen in Deutschland—eine Analyse mit ALLBUS-Daten. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 64(2), 277-302.
- McCutcheon, A. L. (2002). Basic concepts and procedures in single-and multiple-group latent class analysis. In: Hagenaars, J., McCutcheon, A.L. (Eds.): *Applied latent class analysis*, Cambridge University Press, 56-88.
- McHugh, M. C., & Frieze, I. H. (1997). The Measurement of Gender-Role Attitudes. A Review and Commentary. *Psychology of Women Quarterly*, 21(1), 1-16.
- Merton, R. K. (1957). The role-set: Problems in sociological theory. *The British Journal of Sociology*, 8(2), 106-120.

- Morgan, G.B. (2015). Mixed Mode Latent Class Analysis: An Examination of Fit Index Performance for Classification. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(1), 76-86.
- Muthén, L.K., Muthén, B.O. (1998-2017). *Mplus User's Guide*. Eighth Edition. Los Angeles, CA: Muthén & Muthén.
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling*, 14(4), 535-569.
- Parsons, T. (1951). *The Social System*. Routledge & Kegan Paul: London.
- Pfau-Effinger, B. (2004). Socio-historical paths of the male breadwinner model—an explanation of cross-national differences. *The British Journal of Sociology*, 55(3), 377-399.
- Pfau-Effinger, B., & Smidt, M. (2011). Differences in women's employment patterns and family policies: eastern and western Germany. *Community, Work & Family*, 14(2), 217-232.
- Rosenfeld, R. A., Trappe, H., & Gornick, J. C. (2004). Gender and work in Germany: Before and after reunification. *Annual Review of Sociology*, 30, 103-124.
- Rudnev, M., Magun, V. & Schmidt, P. (2016). Basic Human Values: Stability of Value Typology in Europe. In: Voicu, M., Mochmann, I.C. & Dülmer, H. (eds). *Values, Economic Crisis and Democracy*. Routledge, Abingdon/New York, 21-49.
- Scott, J. (2008). Changing gender role attitudes. In: Scott, J., Dex, S., Joshi, H. (eds). *Women and employment: Changing lives and new challenges*, pp. 156-176.
- Sjöberg, O. (2004). The role of family policy institutions in explaining gender-role attitudes: a comparative multilevel analysis of thirteen industrialized countries. *Journal of European Social Policy*, 14(2), 107-123.
- Smith, T.W. (2005). The Laws of Studying Societal Change. *Survey Research*, 36(2), 1-5.
- Smith, T.W., Marsden, P. V., Hout, M., & Kim, J. (2013). *General Social Survey Cumulative Codebook, 1972–2012 [MRDF]*. Chicago: National Opinion Research Center.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-70.
- Van de Vijver, F., & Leung, K. (1997). Methods and data analysis of comparative research. In: Berry, J. W., Poortinga, Y. H., Pandey, J. (Eds). *Handbook of cross-cultural psychology: Theory and method*, Vol. 1. Needham Heights, MA, US: Allyn & Bacon, 257-300.
- Walter, J.G. (2017). The adequacy of measures of gender role attitudes: a review of current measures in omnibus surveys. *Quality & Quantity* (online first publication), doi: 10.1007/s11135-017-0491-x.
- Wasmer, M. (2014). *German General Social Survey 2012. English Translation of the German "ALLBUS"-Questionnaire*. GESIS-Technical Reports 2014/01, Köln.
- Yang, C. (2006). Evaluating latent class analysis models in qualitative phenotype identification. *Computational Statistics & Data Analysis*, 50 (4), 1090-1104.

7. Appendix

Table A1: Fit statistics of the estimated 1 to 7-class solutions in 1991 and 2012 (values indicating best fit in bold)

	LL	AIC	BIC	AIC3	Npar	L ²	p (L ²)	df (L ²)	entropy	LMR(p)
1	-22741	45517	45626	45535	18	6439	0	4072	-	-
2	-21141	42356	42579	42393	37	4003	0.7	4049	0.76	0
3	-20630	41371	41708	41427	56	3177	1	4029	0.73	0
4	-20444	41039	41490	41114	75	2883	1	4010	0.70	0
5	-20353	40894	41460	40988	94	2757	1	3992	0.68	0.16
6	-20288	40802	41482	40915	113	2669	1	3973	0.70	0.78
7	-20234	40733	41528	40865	132	2581	1	3954	0.69	0.8

	LL	AIC	BIC	AIC3	Npar	L ²	p (L ²)	df (L ²)	entropy	LMR(p)
1	-12224	24483	24581	24501	18	3534	1	4032	-	-
2	-11249	22572	22774	22609	37	2568	1	4036	0.76	0
3	-10917	21946	22251	22002	56	2091	1	4020	0.76	0.67
4	-10816	21782	22192	21857	75	1906	1	4002	0.74	0.61
5	-10741	21670	22182	21764	94	1777	1	3982	0.72	0.76
6	-10704	21634	22250	21747	113	1722	1	3965	0.71	0.76
7	-10669	21603	22322	21735	132	1635	1	3944	0.74	0.77

Table A2: Class-specific distribution of socio-demographic variables (based on most likely class membership)

	(1) family-oriented traditional	(2) child-oriented traditional	(3) unconditionally egalitarian	(4) conflicted egalitarian	(5) moderately egalitarian
ratio: female	46.7	47.5	64.8	48.7	53.1
age (mean)	54.7	46.4	44.8	39.5	46.4
ratio: married	64.0	61.9	56.6	53.0	59.2
educational qualification					
<i>None/compulsory education</i>	70.1	51.3	20.8	25.4	36.6
<i>Secondary school</i>	21.0	28.1	39.0	30.3	40.9
<i>Higher educational entrance qualification</i>	8.9	20.5	40.2	44.3	22.4
Household net income (mean)	1578	1853	2554	2282	2084



What was I thinking? A theoretical framework for analysing panel conditioning in attitudes and (response) behaviour

Michael Bergmann & Alice Barth

To cite this article: Michael Bergmann & Alice Barth (2018) What was I thinking? A theoretical framework for analysing panel conditioning in attitudes and (response) behaviour, International Journal of Social Research Methodology, 21:3, 333-345, DOI: [10.1080/13645579.2017.1399622](https://doi.org/10.1080/13645579.2017.1399622)

To link to this article: <https://doi.org/10.1080/13645579.2017.1399622>



© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 08 Nov 2017.



[Submit your article to this journal](#)



Article views: 397



[View Crossmark data](#)

What was I thinking? A theoretical framework for analysing panel conditioning in attitudes and (response) behaviour

Michael Bergmann^a and Alice Barth^b

^aMunich Center for the Economics of Aging (MEA), Technical University of Munich (Chair for the Economics of Aging), Munich, Germany; ^bInstitute of Political Science and Sociology, University of Bonn, Bonn, Germany

ABSTRACT

Though panel data are increasingly used in the social sciences, the question whether repeatedly participating in a panel survey affects respondents' attitudes and (response) behaviour is still largely unsolved. Drawing on a model of associative networks that is extended by assumptions on survey satiscing, we present a theoretical framework that emphasizes the role of strength-related attributes of attitudes (accessibility, internal consistency, extremity) and motivation in respondents' information processing. In particular, we argue that – depending on respondents' predispositions – occupation with survey questions enhances attitude strength, which results in increasing attitude stability and influence on thoughts and behaviours. Against this background, we bring together hitherto unconnected results from previous research and thus contribute to a more thorough understanding of both the mechanisms and the multifaceted outcomes of panel conditioning.

ARTICLE HISTORY

Received 17 February 2017
Accepted 18 October 2017

KEYWORDS

Panel conditioning;
associative network model;
attitude strength; survey
satiscing; response
behaviour

Introduction

One of the major advantages of panel surveys is that repeated observations of the same respondents over time permit to analyse intra-individual changes. This enables researchers to approach causal relationships that cannot be modelled in an analogous manner on the basis of cross-sectional studies or time-series analyses. Consequently, data from panel surveys have been constituting a major source of scientific insights for many years. At the same time, panel data bear two problems that complicate the generalization of empirical results: non-random *attrition* (see, e.g. Schifeling, Cheng, Reiter, & Hillygus, 2015; Waterton & Lievesley, 1987) and *panel conditioning*. Although numerous studies explore the conditions and the extent of panel conditioning effects, the question 'whether repeated interviews are likely, in themselves, to influence a respondent's opinions' (Lazarsfeld, 1940, p. 128) and her (response) behaviour has not been sufficiently solved yet. In recent years, evidence for the existence of panel conditioning has accumulated (e.g. Bergmann, 2015; Halpern-Manners, Warren, & Torche, 2014; Kroh, Winter, & Schupp, 2016; Warren & Halpern-Manners, 2012). At the same time, a substantial number of studies has found no or only very small effects (e.g. Axinn, Jennings, & Couper, 2015; Barber, Gatny, Kusunoki, & Schulz, 2016; Mann, 2005; Smith, Gerber, & Orlich, 2003).

CONTACT Michael Bergmann  bergmann@mea.mpisoc.mpg.de

The contradictory nature of results is commonly ascribed to large differences in research designs (e.g. Warren & Halpern-Manners, 2012) and the difficulty of empirically separating panel conditioning from attrition and true change (e.g. Yan & Eckman, 2012).

In this paper, we argue that a thorough analysis of the occurrence, magnitude and direction of panel conditioning effects does not only require empirical but also theoretical rigour. While a substantial number of studies has suggested sophisticated measurement approaches for the separation of panel conditioning effects from attrition bias and true change (Crossley, de Bresser, Delaney, & Winter, 2017; Das, Toepoel, & van Soest, 2011; Struminskaya, 2016; Yan & Eckman, 2012), the phenomenon's underlying mechanisms and driving factors are still far from clear. Hence, this paper aims to integrate proposed mechanisms behind panel conditioning in a comprehensive theoretical framework by drawing on insights from research on cognitive information processing, attitude strength and respondents' survey response strategies. The framework shall serve as a basis for the formulation of hypotheses on the causes and consequences of panel conditioning for attitudes and (response) behaviour, thereby providing a broad theoretical foundation for future analyses.

Literature shows that most existing assumptions on the mechanisms of panel conditioning (at least implicitly) refer to cognitive processes, either focusing on the role of memory in a sense that survey responses depend on experiences in previous panel waves (e.g. Crespi, 1948; Struminskaya, 2016; Waterton & Lievesley, 1989) or postulating some kind of stimulation by answering survey questions (e.g. Jagodzinski, Kühnel, & Schmidt, 1987; Sturgis, Allum, & Brunton-Smith, 2009). This observation provides the rationale for connecting and expanding current explanations in a model of associative networks (Anderson, 1983). In this context, we make use of the concept of attitude strength (Krosnick & Petty, 1995) in a two-step procedure: first, we illustrate that strength-related attributes of attitudes, namely accessibility, internal consistency and extremity, are likely to be affected by repeated participation in a panel study. This is what we call the *mechanisms* of panel conditioning. Generally speaking, repeated processing of question-related information is supposed to simplify and accelerate the cognitive steps that are involved in answering a survey question: understanding the meaning of a question, retrieving relevant information from memory, constructing an answer from the available information, and reporting it (Tourangeau, Rips, & Rasinski, 2000). Second, we argue that substantial changes in attitudes and behaviour can be regarded as *consequences* of the distinctive features of strong attitudes.

Additionally, we refer to Krosnick's (1991) approach of survey satisficing to highlight the role of respondent characteristics, particularly motivation, in explaining observed changes in their response behaviour. Against this background, we demonstrate that our theoretical framework facilitates the integration of different, sometimes contradictory results from previous studies, derive a set of hypotheses, and give an empirical example of measuring panel conditioning with regard to political indecision while controlling for individual attitude strength.

Previous research on panel conditioning

In the 75 years that have passed since Lazarsfeld's (1940) first concerns on the potential influence of repeated interviews on respondents' opinions, numerous studies in various disciplines have investigated this topic. Due to the diversity of methodological approaches, data sources and study designs, we do not aim at a comprehensive meta-analysis of detected effects. Instead, we give an overview on popular streams of proposed theoretical explanations for panel conditioning effects in the realm of attitudes and (response) behaviour. It has repeatedly been remarked that the theoretical basis of research concerning panel conditioning effects is rather thin (Struminskaya, 2016; Sturgis et al., 2009; Warren & Halpern-Manners, 2012). While a number of sound assumptions have been formulated, most propositions are not linked to a broader theoretical framework. In trying to narrow this research gap, we hope to shed some light on the reasons why some studies detect considerable effects of repeated interviewing while others do not.

Memory effects

When respondents are asked questions several times, they are likely to remember the structure of the questionnaire as well as their previous answers. Basically, this is the rationale behind proposing recollection or learning as causal mechanisms of panel conditioning, which can have different consequences: first, respondents may remember that answering a certain question yields a number of follow-up questions, which leads them to choose less time-consuming alternatives in subsequent waves (e.g. Toepoel, Das, & van Soest, 2008). This mechanism has been suspected with regard to substantially inexplicable changes in later waves of a panel survey, such as less reported jobs (Warren & Halpern-Manners, 2012), a decline in the size of personal networks (Eagle & Proeschold-Bell, 2015), and even reduced use of toothpaste (Nancarrow & Cartwright, 2007).

In the absence of external factors, the avoidance of lengthy questions is clearly attributable to a desire to reduce burden and can thus be regarded as a problem of response behaviour (Struminskaya, 2016). However, more exact reporting has been proposed – sometimes even simultaneously – as a second memory effect (Silberstein & Jacobs, 1989; Waterton & Lievesley, 1989). In this respect, learning the rules that govern a survey interview might also lead to better informed respondents who give more exact answers. Additionally, familiarization with the survey procedure due to repeated participation is supposed to increase respondents' trust, which is why more truthful answers are expected in later waves (Struminskaya, 2016; Waterton & Lievesley, 1989). Finally, respondents remembering their answers from earlier waves may be disinclined to change their once stated opinion as they are reluctant to appear fickle, resulting in higher response consistency (Crespi, 1948). In sum, hypotheses based on memory effects address modifications in response behaviour, which are effectuated by participation in previous waves. However, due to the – often simultaneous – prediction of opposite effects, their value in the advancement of theory on mechanisms of panel conditioning is limited.

Stimulation effects

Next to remembering one's answers as well as the general survey procedure, many studies assume that being interviewed about certain topics leads to a sensitization of the respondent, making her think about the survey topics more thoroughly and, as a consequence, increase topical interest. For example, Waterton and Lievesley (1989) found evidence for an increase in partisanship and a decrease of the likelihood of 'don't know' responses. This is explained by an increased cognitive engagement with the survey topic. Sturgis and colleagues (2009) argue in the same way when they propose a 'cognitive stimulus' that serves to crystallize previously weak and inconsistent attitudes. Jagodzinski and colleagues (1987), who label the phenomenon of higher attitude consistency in later waves as 'socratic effect', and Crespi (1948), who terms it 'clarification effect', refer to an increased cognitive processing of survey topics, too. Particularly in the domain of attitudes, the idea of a cognitive stimulation that yields stronger opinions due to repeated interviewing is the most convincing theoretical approach to date. However, analyses are largely restricted to the outcomes of the assumed stimulation, while the exact mechanisms of individual cognitive processes and the role of respondents' characteristics have not been focused on.

Mere measurement effects

The stream of research that operates under the heading 'mere measurement' expands the idea of stimulation from surveys on political and social attitudes to consumer research and other fields. Frequently, evidence from external validation data is provided, which proves that being surveyed does not only affect attitudes but also behaviour. In this area, increased attitude accessibility is often proposed as the main mechanism accounting for panel conditioning (see Dholakia, 2010). Morwitz, Johnson, and Schmittlein (1993) revealed that being surveyed on the intention to buy a car or a computer leads to an increase in purchase behaviour, especially when participants had no prior experience with these

products. The same effect was shown for candy bars (Morwitz & Fitzsimons, 2004) as well as for online grocery purchases and automotive services (Chandon, Morwitz, & Reinartz, 2004; Morwitz et al., 1993). The authors hypothesize that repeated activation of attitudes leads to higher attitude accessibility that in turn increases attitude-behaviour consistency. Despite the robustness of the mere measurement effect and its replication in different fields, its drivers as well as its effect size are, however, still not fully obvious (Dholakia, 2010).

Self-prophecy effects

Originally referred to by Sherman (1980) as ‘self-erasing errors of prediction’, the term self-prophecy (coined by Greenwald, Carnot, Beach, & Young, 1987) describes the phenomenon that merely asking people to predict whether they are willing to perform a certain behaviour increases their probability of acting in line with their prediction. In a series of experiments, it was shown that previous self-assessments increase socially desirable behaviours, such as charity work (Sherman, 1980), voter registration and turnout (Greenwald et al., 1987) or health club visits (Spangenberg, 1997), whereas socially undesirable actions decrease (Spangenberg & Obermiller, 1996). The authors assume that due to the socially normative nature of the behaviours, respondents overestimate their adherence to perceived societal rules. However, in order to preserve a positive self-perception, they subsequently feel compelled to act according to their own prediction, which makes the normative bias a self-erasing one. This explanation fits in with the theory of cognitive dissonance (Festinger, 1957) and is currently considered the leading one, though other mechanisms such as heightened self-awareness or script evocation have been proposed as well (see Dholakia, 2010).

In the last years, there have been efforts to unite the streams of mere measurement and self-prophecy under the heading of a ‘question-behaviour-effect’. Several meta-analyses confirm a small, but significant positive effect of asking questions on the subsequent performance of related behaviour (Rodrigues, O’Brien, French, Glidewell, & Sniehotta, 2015; Spangenberg, Kareklas, Devezer, & Sprott, 2016; Wilding et al., 2016). However, the question whether the effect is mainly driven by attitude accessibility, processing fluency, behavioural simulation, motivation or the desire to reduce cognitive dissonance remains a matter of debate.

To sum up, most of the listed studies share the (implicit) assumption that the survey interview itself causes some kind of change in an individual’s information processing. Though all four explanations provide important clues for the causal mechanisms behind panel conditioning, unified discussions of the foundations of panel conditioning and its consequences have not initiated to date. In the following, this gap is minimized by elaborating a comprehensive theoretical framework that specifies mechanisms of cognitive information processing on the one hand and accentuates respondents’ individual preconditions as a main differentiating factor on the other. Against the background of associative networks, assumptions of mere measurement and self-prophecy can be generalized, whereas the stimulation of respondents and the conditions for differential outcomes of memory effects can be stated more precisely. Additionally, by extending the framework with Krosnick’s (1991) approach on survey satisficing, we are able to predict changes in respondents’ reporting behaviour that can manifest themselves in different directions.

A theoretical framework for analysing panel conditioning

Associative network models, originally based on works about semantic connections (Anderson, 1983), currently belong to the most popular models for the explanation of information processing in cognitive psychology, but have been adopted by other disciplines as well. They assume that information is stored in long-term memory in the form of objects representing individuals, issues, specific events, but also values or ideological principles. Those objects can be imagined as nodes that are connected with specific attributes as well as with other objects in a network structure (see, e.g. Lodge & McGraw, 1991). In such a mental network, the links between particular objects contain information on the nature of their

relationship. Moreover, objects are associated with positive or negative evaluations that vary in their intensity. Thus, initially knowledge-based cognitive models of information processing are extended by an affective component taking into account that socio-political concepts are affectively charged (e.g. Kim, Taber, & Lodge, 2010). Within this model, an attitude can be defined as a summarized or balanced reaction towards an object that can be positive or negative and differs in intensity.

Regarding the investigation of panel conditioning effects, this outline of the model architecture allows for several assumptions concerning changes in respondents' associative networks due to repeatedly responding to the same questions. We posit that repeated interviewing affects respondents' individual information processing. More precisely, we expect the automatic activation of question-related objects to increase attitude accessibility, internal consistency and extremity (see Figure 1). These underlying *mechanisms* of panel conditioning take place largely automatically due to the mere confrontation of respondents with survey questions; they are only dependent on the individual arrangement of a respondent's associative network (for example their initial experience with a certain topic). Based on this, further substantial *consequences* for respondents' attitudes and (response) behaviour can be derived. These consequences, which are mediated by respondents' predispositions, such as perceived social norms, topical interest or motivation, include effects on the stability of attitudes, attitude formation and related actual behaviours as well as the reporting of survey answers.

Mechanisms of panel conditioning

Firstly, we assume that repeated survey participation enhances the *accessibility* of object nodes. Accessibility can be defined as 'the speed and ease with which the attitude can be accessed from memory' (Fazio, Chen, McDonel, & Sherman, 1982, p. 340) and is a function of prior activation. This process happens automatically within a few hundred milliseconds after the perception of the object and is not consciously controlled by respondents (Fazio, Sanbonmatsu, Powell, & Kardes, 1986). The more one has come into contact with an object in the recent past, for example because one has heard or thought about it in a survey, the more mentally accessible this object, its associated evaluation, but also closely connected objects are. Moreover, frequent contacts with an object accelerate the process of attitude formation and the utterance of an evaluative judgment. Therefore, quick answers in surveys are regarded as an indicator for the accessibility of the object in question (e.g. Fazio et al., 1986). These considerations are in line with findings on heightened attitude accessibility in mere measurement research (e.g. Morwitz & Fitzsimons, 2004).

Additionally, it is very likely that panelists directly resort to a previously formed attitude in subsequent waves, as repeated survey participation increases the chance that a summary evaluation of a repeatedly activated object is stored in long-term memory next to singular object evaluations. This contrasts the situation in cross-sectional surveys where respondents frequently do not have a pre-formed attitude; hence, they have to newly construct an attitude from memory by retrieving, evaluating and

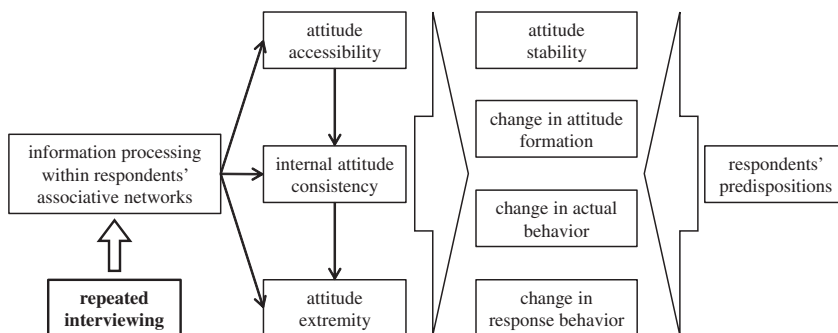


Figure 1. Framework for analysing panel conditioning.

putting together individual pieces of information when confronted with a new survey question (Bizer, Tormala, Rucker, & Petty, 2006). When a summary evaluation is available, attitude formation can be shortened or even skipped, which increases response speed. We thus assume that the time needed to answer a question decreases over the course of a panel survey and that the effect is most pronounced at the beginning, attenuating – also because of cognitive limits – with the number of repetitions.

Next to an increase in accessibility, we secondly expect that panel surveys enhance attitude *consistency*. The theory of spreading activation in associative networks poses that querying a specific object will also activate adjacent objects and their associated evaluations that are evaluatively consistent (Kim et al., 2010). A number of studies have demonstrated that objects exhibiting evaluative consistency are easier activated and transferred from long-term memory to working memory than objects that are associated with both positive and negative evaluations (e.g. Fazio et al., 1986; Judd & Brauer, 1995). Since the early work of Festinger (1957) on dissonance theory, it is known that individuals usually prefer consistent to inconsistent attitudes. In a panel survey, the chance is high that respondents subliminally perceive inconsistencies in their evaluations and try to eliminate them by reducing the relevance of inconsistent information, adding concordant elements, or reinterpreting dissonant elements. Therefore, repeated administration of the same questions enhances selective processing of evaluatively consistent objects and their associated evaluations, while inconsistent attributes become increasingly meaningless in attitude formation. Researchers adhering to variants of the stimulation hypothesis have presented empirical evidence for increased attitude consistency (Jagodzinski et al., 1987; Sturgis et al., 2009).

Thirdly, the repeated activation of consistent evaluations results in increased attitude *extremity*. This hypothesis is based on the notion that each activation of objects, for example when respondents are confronted with questions on certain topics, strengthens the connections between objects that are evaluatively consistent. This results in a selective processing of information and, in turn, a higher influence of consistent evaluations on attitude formation (Tesser, 1978). Consequently, the summary evaluation of an object is incrementally steered towards a certain direction and the extremity of attitudes is expected to increase (Judd & Brauer, 1995; Kim et al., 2010). In a panel study this would mean, for instance, that the evaluation of a politician is getting more and more extreme from wave to wave.

So far, we have argued that repeated interviews have the potential to influence the strength-related attributes of attitudes. While the automatic activation of objects due to answering questions in a survey can be considered as fairly universal, it is important to state that panel conditioning effects will be more pronounced the lesser respondents have engaged in thinking about the objects in question before. This is a logical consequence of the mechanisms described above: for respondents who already have accessible, consistent and extreme attitudes towards a certain object, being asked a question about it only brings about a marginal change, if any. In contrast, respondents whose attitudes are less elaborated and partly inconsistent are likely to experience significant changes in their associative network structure due to survey participation (Judd & Brauer, 1995; Tesser, 1978). The latter situation should be frequently the case at the beginning of a panel study. Changes in associative networks therefore provide a clue why panel conditioning effects are most pronounced between the first and the second measurement occasion. In addition, these considerations are helpful in explaining the finding that shorter intervals between waves increase the probability for panel conditioning (Warren & Halpern-Manners, 2012): the less time has passed between the last measurement and a renewed activation of an attitude object in the next wave, the more mentally accessible is this object as well as its associated evaluations and the more internally consistent and extreme attitudes will become over time. In such a situation, panel conditioning can be expected to produce particularly strong effects.

While we consider the associative network model very useful for the explanation of actual attitude changes, we extend our framework by Krosnick's (1991) approach of survey satisficing in order to account more precisely for changes in reporting. Krosnick states that a respondent's propensity to reduce the cognitive effort that is necessary to answer survey questions accurately is moderated by three factors: the more difficult it is for a respondent to provide the correct answer, and the less able and motivated to do so she is, the more likely it is that a satisficing strategy is employed. This can,

for instance, result in selecting the first answer that seems reasonable instead of carefully reflecting all alternatives, indiscriminately agreeing with assertions, or saying ‘don’t know’ despite having a substantive opinion.

Summing up, we assume that the cognitive mechanisms operating during a panel survey change the way information is processed, resulting in stronger attitudes towards repeatedly assessed objects. This central claim facilitates the connection of hitherto freestanding explanations of panel conditioning. It thus helps to understand in detail what exactly happens in the head of respondents when repeatedly confronted with a survey question instead of largely concentrating on manifestations of the underlying cognitive processes. Moreover, it enables systematic analyses of the underlying individual mechanisms of panel conditioning taking into account respondents’ previous experience with the survey objects by operative indicators of attitude strength (Bassili, 1996). An increase in attitude strength implies, in turn, that distinctive features of strong attitudes will become more pronounced. We therefore expect attitudes to become more stable and more influential towards thoughts and actual behaviour due to repeatedly answering the same questions (Krosnick & Petty, 1995). Further, being repeatedly confronted with identical questions in a panel survey may also affect response behaviour. These consequences are exposed in more detail in the following.

Consequences of panel conditioning

Stability is one of the most important characteristics associated with strong attitudes (Krosnick & Petty, 1995) and is consequentially expected to increase with repeated survey participation. Furthermore, the probability that respondents resort to a previously stored summary evaluation increases over the course of a panel survey, while the influence of situational factors in the construction of an attitude decreases (Bizer et al., 2006). This should contribute to less inter-wave fluctuation of individual responses, too.

Stability can be defined as rank-order stability that is testable with test–retest correlations or absolute stability which is quantified by changes in intra-individual response sequences (e.g. Prior, 2010). While the former variant has been assessed in the context of stimulation effects (Jagodzinski et al., 1987; Sturgis et al., 2009), the latter is often associated with positing that respondents remember their answers from previous interviews (e.g. Bridge et al., 1977). Applying the model of associative networks, these two streams can be integrated as attitude strength and the recourse to a summary evaluation suggest an increase in both forms of stability over the course of a panel study. Empirical findings regarding stability are mixed: while Bridge et al. (1977), Jagodzinski et al. (1987) and Sturgis et al. (2009) report a significant increase in stability, Waterton and Lievesley (1989) as well as Silberstein and Jacobs (1989) only find a negligible increase. Next to differences in operationalization and measurement, such ambiguous results can be attributed to the neglect of differences in respondent characteristics because a significant increase in stability is expected only for those respondents whose initial attitudes on the surveyed topics were rather weak and inconsistent.

In addition to more stable attitudes, we assume that repeated confrontation with identical survey questions leads to systematic and substantial *changes in the formation of attitudes*. Due to selective information processing, where certain (consistent) beliefs are activated and processed more frequently than others, attitudes are likely to be systematically altered depending on respondents’ existing preferences and intents (e.g. Morwitz et al., 1993). Such attitude changes may have consequences for closely connected objects within the respondents’ knowledge structure as well. In this context, empirical studies have shown that persons with strong attitudes exhibit an increase in evaluative differences between objects that are perceived as being contrary to each other (attitude polarization) when they process information that is in line with their previous attitudes towards these objects (Taber & Lodge, 2006). In the view of Tesser (1978, p. 298), there is no doubt that even mere thinking about an object, for example during a survey, is sufficient to cause such effects: ‘thought about some particular object in the absence of any new information will tend to produce attitude polarization’.

This mechanism also provides an explanation why respondents’ trust in the survey and hence their willingness to reveal sensitive information tends to increase during the course of a panel study

(Struminskaya, 2016; Waterton & Lievesley, 1989): as one can reasonably assume that respondents need a minimum amount of positive feelings towards both the survey organization and the asked topics as a prerequisite to initially participate, their evaluations towards the survey are likely to evolve in a positive direction, while inconsistent beliefs such as feeling uncomfortable during the interview become less and less important. This also means that respondents who initially agree to participate despite having negative feelings towards the survey, for example, because the requested information is seen as too sensitive, have a higher probability to drop out from the study.

Besides influences on thoughts, stronger attitudes can have consequences on the *actual behaviour* of respondents as well. It is well established that more consistent and accessible attitudes allow more accurate predictions of subsequent behaviour (e.g. Fazio et al., 1986). With respect to panel conditioning effects, it has been demonstrated that respondents' perceived social norms play a crucial role (e.g. Spangenberg & Greenwald, 1999). The perception of existing inconsistencies between attitudes and behaviour is more likely the more often one participates in a panel survey and thus intellectually engages with the survey topics. The desire to resolve or at least reduce unpleasant inconsistencies should be especially strong when the respective behaviour is perceived as socially (un)desirable. The repeated confrontation with discrepancies is therefore hypothesized to trigger cognitive dissonance that should lead to an adaptation of one's own behaviour to the social norm and subsequently more accordance between attitudes and behaviour.

Taking into account social norms that are perceived as important by the respondents, hence considering individual differences in respondents' cognitive structures and normative beliefs, gives researchers the opportunity to predict the magnitude as well as the direction of changes, thus moving beyond the assumption of mere behavioural stimulation caused by panel conditioning. In this sense, observed panel conditioning effects such as a higher turnout in elections (e.g. Clausen, 1968; Granberg & Holmberg, 1992) or a higher percentage of correct answers to knowledge questions (e.g. Toepoel, Das, & van Soest, 2009) can be interpreted as normatively charged behaviour changes.

Finally, participating in a panel may not only influence attitudes and actual behaviour, but also *response behaviour*. By applying Krosnick's (1991) survey satisficing model against the background of the presented mechanisms in associative networks, it can be predicted how the conditions that foster satisficing (or optimizing as its counterpart) evolve over the course of a panel study. On the one hand, an increase in attitude strength makes it easier to process information from an existing knowledge structure, thus reducing difficulty, potentially improving ability, and therefore increasing the chance of optimizing. On the other hand, decreasing motivation might outweigh possible gains in ability and losses in difficulty. It has been shown that fatigue – due to the lengthy or repetitive character of a survey – decreases motivation and can lead to a strong desire to reduce the burden of answering cognitively demanding survey questions, resulting in a systematic underreporting of events or symptoms (e.g. Duan, Alegria, Canino, McGuire, & Takeuchi, 2007), avoidance of follow-up questions (Mathiowetz & Lair, 1994) or speeding through the questionnaire (Roßmann, 2017).

The presence of such effects is most likely in questions that have already been experienced as burdensome in a preceding interview (Das et al., 2011; Toepoel et al., 2008). Positioning individual motivation as a key concept in analyses of panel conditioning thus substantially reduces the arbitrariness that characterizes most previous theoretical approaches towards changes in reporting survey answers, where 'better' and 'worse' response behaviour is predicted simultaneously (Warren & Halpern-Manners, 2012; Waterton & Lievesley, 1989; Yan & Eckman, 2012) and can help to distinguish different directions of changes in reporting.

An empirical example

To illustrate the analytical potential of our framework, we give an example of panel conditioning effects on respondents' political indecision, using data from the German Longitudinal Election Study. In particular, we use a campaign panel with six interviews conducted before the 2009 federal elections in Germany as well as several independent cross-sectional studies that serve as control groups (see

Bergmann, 2015; Steinbrecher, Roßmann, & Bergmann, 2013 for detailed information on the data). To separate panel conditioning from confounding effects such as panel attrition, propensity score weighting was applied that takes into account respondents' socio-demographic characteristics as well as their political interest, which is often correlated with panel attrition (e.g. Lazarsfeld, 1940). This procedure serves to guarantee that panel (treatment group) and cross-sectional (control group) respondents can be reasonably compared and observed effects can be correctly attributed to repeated interviewing. In addition, we separated panel conditioning effects from real changes over time by using the differences between the observed changes from one panel interview to the next and the aggregate change in two parallel cross-sections as a control variable in our analyses (see Waterton & Lievesley, 1989).

Our main focus is on intra-individual changes, therefore, we ran a fixed-effects model that accounts for person-specific heterogeneity (see, e.g. Allison, 2009). As dependent variable, we use the change in respondents' political indecision regarding party vote intention (saying 'don't know' which party one is going to vote for in the upcoming election). As explaining variables, we use the frequency of being interviewed as well as the indicators of attitude strength (accessibility, internal consistency and extremity¹) as suggested above. Following our theoretical model, it is hypothesized that panel respondents experience an increase in attitude strength which leads to a decrease in political indecision. The effect is expected to be stronger for respondents whose attitudes were weak at the beginning of the study.

Table 1 shows the influence of repeatedly participating in a panel study on the change of political indecision between the first and the sixth panel wave during the election campaign when individual attitude strength is accounted for. The consistent negative effect of interview frequency can be interpreted as a strong decrease in political indecisiveness of panel respondents that is significantly larger than the observed decrease in the general population (time trend) which we control for. In addition, model 2 shows that political indecisiveness is reduced significantly less for panel respondents whose party vote intention was highly accessible at the beginning. The same is true for attitude extremity towards the parties running for election (see model 3). Only the consistency in evaluating the personality of the candidates for chancellorship seems to have no independent influence on the change in political indecision (model 4).

Overall, this brief example mainly confirms the model's assumptions: The repeated confrontation of panel respondents with their party vote intentions led to substantial changes, surpassing those in the general population. As we accounted for confounding effects, panel attrition cannot serve as a possible explanation for these changes. Moreover, the initial strength with which an attitude is held at the beginning of a panel survey plays an important moderating role. Especially panel respondents with weak attitudes towards the issue in question show considerable effects, while respondents who already have crystallized and strong attitudes show significantly smaller effects.

Table 1. Panel conditioning effect on the change of respondents' political indecision.

	Change of political indecision			
	Model 1	Model 2	Model 3	Model 4
Time trend	-.17* (-2.16)	-.17* (-2.15)	-.17* (-2.15)	-.17* (-2.19)
Frequency of being interviewed	-1.43*** (-7.22)	-1.44*** (-7.30)	-1.44*** (-7.30)	-1.45*** (-7.35)
<i>Attitude strength at the beginning of the study</i>				
high accessibility * number of interviews		.86*** (4.76)	.86*** (4.81)	.85*** (4.72)
high internal consistency * number of interviews			.13 (.73)	.01 (.03)
high extremity * number of interviews				.81*** (4.12)
Constant	22.26***	22.27***	22.27***	22.26***
R ² (within)	.01	.02	.02	.02
N (observations)			13,341	
N (individuals)			3,468	

Notes: Fixed-effects regression coefficients with *t*-values based on panel-robust standard errors in parentheses. All coefficients have been multiplied with 100 to increase legibility.

Significance level: **p* < .05, ****p* < .001.

Discussion

Although the notion that participating in a panel *can* change respondents' attitudes and (reporting) behaviour has almost become a truism of social science research, questions of when, why, how, and to what extent such changes occur are far from being solved. In this paper, we argued that while recent works greatly advanced the methods to adequately measure effects of panel conditioning (Crossley et al., 2017; Das et al., 2011; Halpern-Manners et al., 2014; Kroh et al., 2016; Yan & Eckman, 2012), in order to thoroughly understand the mechanisms behind panel conditioning and to produce testable hypotheses on its occurrence, magnitude and direction, a theoretical framework is needed that is more comprehensive than existing singular assumptions.

Based on research in cognitive information processing, we proposed that repeatedly answering the same questions can be understood as a trigger for structural changes in associative networks, thus increasing (1) accessibility, (2) internal consistency and (3) extremity of attitudes. These processes happen largely automatically and are only moderated by respondents' previous experience with objects in the sense that (4) the effects of repeated questioning are stronger the less respondents have come into contact with the survey topics before. We further argued that the increase in attitude strength due to repeated interviewing effectuates (5) higher stability of attitudes, (6) systematic attitude formation and (7) higher attitude-behaviour correspondence conditional on existing preferences and social desirability of topics. Finally, we claimed that (8) respondents' motivation is a key element in predicting whether repeated surveying enhances satisficing or optimizing with respect to response behaviour.

At this point, it is important to state that the whole framework is not restricted to attitude questions, but can be regarded as applicable to any repeated question with an evaluative component, such as behavioural frequencies or knowledge questions. Changes in this respect can manifest themselves, for instance, in an active search for information in order to answer questions that previously could not be answered. The proposed framework connects existing streams of explanations and thus serves to improve the understanding of the cognitive foundations of panel conditioning. We also provided a short example, demonstrating how some of the model's most important assumptions can be tested. However, only further empirical applications will allow firm conclusions about the hypotheses' validity. Thus, we hope that the proposed framework can serve as a starting point for more comprehensive, theoretically informed future research on panel conditioning.

In this regard, several aspects are important in our view: first, empirical analyses should adopt a differentiated perspective on the consequences of panel conditioning by taking respondents' differences more strongly into account. This applies in particular to respondents' experience with the survey topic, which determines attitude strength, but also to predispositions in terms of social norms, topical interest and motivation that affect attitude stability and formation as well as (response) behaviour. In our view, accounting for these differences might help to understand some of the contradictory or null findings in previous research.

Second, panel conditioning needs to be carefully distinguished from other effects, such as panel attrition, but also real changes over time in the population of interest. The problem of correctly identifying panel conditioning in the presence of attrition has been extensively addressed in the literature (e.g. Struminskaya, 2016; Warren & Halpern-Manners, 2012). In order to separate panel conditioning from real change over time, researchers can exploit the potential of cross-sectional control groups by contrasting the observed changes from one interview to the next in a panel study with the aggregate change in two parallel cross-sections as shown in our example.

Third, we recommend complementing comparisons between repeatedly surveyed panelists on the one hand and cross-sectional respondents on the other by intra-individual analyses of change. Only the latter allows identifying causal effects of repeated interviewing by controlling unit-specific heterogeneity. Such an analysis strategy constitutes an important extension to previous research, which is frequently limited to univariate comparisons at the aggregate level, therefore comprising the danger of overlooking contrary effects that cancel each other out.

Finally, we completely agree with recently expressed demands of conducting methodological experiments (Halpern-Manners et al., 2014) that vary key conditions of panel conditioning, such as mode of data collection, time between waves, the number of repetitions as well as question wording. Such an experimental setting (that can also be integrated in classical surveys; see, e.g. Barber et al., 2016) has the advantage to carefully manipulate and simultaneously test how respondents process specific information and in what way this influences their attitudes and (response) behaviour.

We are aware that while some of our suggestions are fairly easy to implement, setting up experimental designs and integrating new variables into panel surveys is costly and time consuming. Further, identifying mechanisms and conditions that foster panel conditioning and correcting for bias are quite different matters. Nevertheless, we are convinced that the latter is not possible without the former. A theoretically informed understanding of panel conditioning would thus be a helpful starting point for measuring and correcting for panel conditioning effects in an accurate way.

Note

1. Accessibility was measured by individual response latencies regarding party vote intention. Internal attitude consistency was operationalized by comparing respondents' positive and negative evaluations of the two candidates for chancellorship. Using the evaluations of all parties, attitude extremity was calculated as the average of absolute deviations from the neutral midpoint of the scale (see Bergmann, 2015, pp. 164–176 for a detailed description).

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Michael Bergmann is a senior researcher at the Technical University of Munich (Chair for the Economics of Aging) and the Munich Center for the Economics of Aging (MEA). He received his PhD at the University of Mannheim. His main research interests are changing attitudes and behaviour in longitudinal studies as well as methodological issues, such as interviewer effects, panel attrition and panel conditioning.

Alice Barth is currently working as a research assistant at the department of Sociology in Bonn and completing her PhD in the field of survey methodology and applied statistics. Her research interests are geometric data analysis, survey data quality, panel conditioning, and the interplay between attitudes and social structure.

References

- Allison, P. D. (2009). *Fixed effects regression models*. Thousand Oaks, CA: Sage.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Axinn, W. G., Jennings, E. A., & Couper, M. P. (2015). Response of sensitive behaviors to frequent measurement. *Social Science Research*, 49, 1–15. doi:10.1016/j.ssresearch.2014.07.002
- Barber, J. S., Gatny, H. H., Kusunoki, Y., & Schulz, P. (2016). Effects of intensive longitudinal data collection on pregnancy and contraceptive use. *International Journal of Social Research Methodology*, 19(2), 205–222. doi:10.1080/13645579.2014.979717
- Bassili, J. N. (1996). Meta-judgmental versus operative indexes of psychological attributes: The case of measures of attitude strength. *Journal of Personality and Social Psychology*, 71(4), 637–653. doi:10.1037/0022-3514.71.4.637
- Bergmann, M. (2015). *Panel conditioning. Wirkungsmechanismen und Konsequenzen wiederholter Befragungen [Panel conditioning. Underlying mechanisms and consequences of repeated interviewing]*. Baden-Baden: Nomos.
- Bizer, G. Y., Tormala, Z. L., Rucker, D. D., & Petty, R. E. (2006). Memory-based versus on-line processing: Implications for attitude strength. *Journal of Experimental Social Psychology*, 42(5), 646–653. doi:10.1016/j.jesp.2005.09.002
- Bridge, R. G., Reeder, L. G., Kanouse, D., Kinder, D. R., Nagy, V. T., & Judd, C. M. (1977). Interviewing changes attitudes – Sometimes. *Public Opinion Quarterly*, 41(1), 56–64. doi:10.1086/268352
- Chandon, P., Morwitz, V. G., & Reinartz, W. J. (2004). The short- and long-term effects of measuring intent to repurchase. *Journal of Consumer Research*, 31(3), 566–572. doi:10.1086/425091
- Clausen, A. R. (1968). Response validity: Vote report. *Public Opinion Quarterly*, 32(4), 588–606. doi:10.1086/267648

- Crespi, L. P. (1948). The interview effect in polling. *Public Opinion Quarterly*, 12(1), 99–111. doi:10.1086/265924
- Crossley, T. F., de Bresser, J., Delaney, L., & Winter, J. (2017). Can survey participation alter household saving behaviour? *Economic Journal*. doi:10.1111/ecoj.12398
- Das, M., Toepoel, V., & van Soest, A. (2011). Nonparametric tests of panel conditioning and attrition bias in panel surveys. *Sociological Methods & Research*, 40(1), 32–56. doi:10.1177/0049124110390765
- Dholakia, U. M. (2010). A critical review of question-behavior effect research. *Review of Marketing Research*, 7(7), 145–197. doi:10.1108/S1548-6435(2010)0000007009
- Duan, N., Alegria, M., Canino, G., McGuire, T. G., & Takeuchi, D. (2007). Survey conditioning in self-reported mental health service use: Randomized comparison of alternative instrument formats. *Health Services Research*, 42(2), 890–907. doi:10.1111/j.1475-6773.2006.00618.x
- Eagle, D. E., & Proeschold-Bell, R. J. (2015). Methodological considerations in the use of name generators and interpreters. *Social Networks*, 40, 75–83. doi:10.1016/j.socnet.2014.07.005
- Fazio, R. H., Chen, J.-M., McDonel, E. C., & Sherman, S. J. (1982). Attitude accessibility, attitude-behavior consistency, and the strength of the object-evaluation association. *Journal of Experimental Social Psychology*, 18(4), 339–357. doi:10.1016/0022-1031(82)90058-0
- Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, 50(2), 229–238. doi:10.1037/0022-3514.50.2.229
- Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston: Row, Peterson and Company.
- Granberg, D., & Holmberg, S. (1992). The Hawthorne effect in election studies: The impact of survey participation on voting. *British Journal of Political Science*, 22(2), 240–247. doi:10.1017/S0007123400006359
- Greenwald, A. G., Carnot, C. G., Beach, R., & Young, B. (1987). Increasing voting behavior by asking people if they expect to vote. *Journal of Applied Psychology*, 72(2), 315–318. doi:10.1037/0021-9010.72.2.315
- Halpern-Manners, A., Warren, J. R., & Torche, F. (2014). Panel conditioning in a longitudinal study of illicit behaviors. *Public Opinion Quarterly*, 78(3), 565–590. doi:10.1093/poq/nfu029
- Jagodzynski, W., Kühnel, S. M., & Schmidt, P. (1987). Is there a “socratic effect” in nonexperimental panel studies? Consistency of an attitude toward guestworkers. *Sociological Methods & Research*, 15(3), 259–302. doi:10.1177/0049124187015003004
- Judd, C. M., & Brauer, M. (1995). Repetition and evaluative extremity. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 43–72). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kim, S.-Y., Taber, C. S., & Lodge, M. (2010). A computational model of the citizen as motivated reasoner: Modeling the dynamics of the 2000 presidential election. *Political Behavior*, 32(1), 1–28. doi:10.1007/s1109-009-9099-8
- Kroh, M., Winter, F., & Schupp, J. (2016). Using person-fit measures to assess the impact of panel conditioning on reliability. *Public Opinion Quarterly*, 80(4), 914–942. doi:10.1093/poq/nfw025
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3), 213–236. doi:08884080/91/030213-24
- Krosnick, J. A., & Petty, R. E. (1995). Attitude strength: An overview. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 1–24). Mahwah, NJ: Lawrence Erlbaum Associates.
- Lazarsfeld, P. F. (1940). “Panel” studies. *Public Opinion Quarterly*, 4(1), 122–128. doi:10.1086/265373
- Lodge, M., & McGraw, K. M. (1991). Where is the schema? Critiques. *American Political Science Review*, 85(4), 1357–1380.
- Mann, C. B. (2005). Unintentional voter mobilization: Does participation in preelection surveys increase voter turnout? *ANNALS of the American Academy of Political and Social Science*, 601(1), 155–168. doi:10.1177/0002716205278151
- Mathiowetz, N. A., & Lair, T. J. (1994). Getting better? Change or error in the measurement of functional limitations. *Journal of Economic and Social Measurement*, 20(3), 237–262. doi:10.3233/JEM-1994-20305
- Morwitz, V. G., & Fitzsimons, G. J. (2004). The mere-measurement effect: Why does measuring intentions change actual behavior? *Journal of Consumer Psychology*, 14(1), 64–74. doi:10.1207/s15327663jcp1401&2_8
- Morwitz, V. G., Johnson, E., & Schmittlein, D. (1993). Does measuring intent change behavior? *Journal of Consumer Research*, 20(1), 46–61. doi:10.1086/209332
- Nancarrow, C., & Cartwright, T. (2007). Online access panels and tracking research: The conditioning issue. *International Journal of Market Research*, 49(5), 573–594.
- Prior, M. (2010). You’ve either got it or you don’t? The stability of political interest over the life cycle. *Journal of Politics*, 72(3), 747–766. doi:10.1017/S0022381610000149
- Roßmann, J. (2017). *Satisficing in Befragungen. Theorie, Messung und Erklärung [Satisficing in surveys. Theory, measurement and explanation]*. Wiesbaden: Springer VS.
- Rodrigues, A. M., O’Brien, N., French, D. P., Glidewell, L., & Sniehotta, F. F. (2015). The question-behavior effect: Genuine effect or spurious phenomenon? A systematic review of randomized controlled trials with meta-analyses. *Health Psychology*, 34(1), 61–78. doi:10.1037/hea0000104
- Schifeling, T. A., Cheng, C., Reiter, J. P., & Hillygus, D. S. (2015). Accounting for nonignorable unit nonresponse and attrition in panel studies with refreshment samples. *Journal of Survey Statistics and Methodology*, 3(3), 265–295. doi:10.1093/jssam/smv007
- Sherman, S. J. (1980). On the self-erasing nature of errors of prediction. *Journal of Personality and Social Psychology*, 39(2), 211–221. doi:10.1037/0022-3514.39.2.211

- Silberstein, A. R., & Jacobs, C. A. (1989). Symptoms of repeated interview effects in the consumer expenditure interview survey. In D. Kasprzyk, G. J. Duncan, G. Kalton, & M. P. Singh (Eds.), *Panel surveys* (pp. 289–303). New York, NY: Wiley.
- Smith, J. K., Gerber, A. S., & Orlich, A. (2003). Self-prophecy effects and voter turnout: An experimental replication. *Political Psychology*, 24(3), 593–604. doi:10.1111/0162-895X.00342
- Spangenberg, E. (1997). Increasing health club attendance through self-prophecy. *Marketing Letters*, 8(1), 23–31. doi:10.1023/A:1007977025902
- Spangenberg, E., & Greenwald, A. G. (1999). Social influence by requesting self-prophecy. *Journal of Consumer Psychology*, 8(1), 61–89. doi:10.1207/s15327663jcp0801_03
- Spangenberg, E., Kareklas, I., Devezer, B., & Sprott, D. E. (2016). A meta-analytic synthesis of the question-behavior effect. *Journal of Consumer Psychology*, 26(3), 441–458. doi:10.1016/j.jcps.2015.12.004
- Spangenberg, E., & Obermiller, C. (1996). To cheat or not to cheat: Reducing cheating by requesting self-prophecy. *Marketing Education Review*, 6(3), 95–103. doi:10.1080/10528008.1996.11488565
- Steinbrecher, M., Roßmann, J., & Bergmann, M. (2013). *The short-term campaign panel of the German longitudinal election study 2009: Design, implementation, data preparation, and archiving*. GESIS – Technical Reports 2013/20. Köln: GESIS.
- Struminskaya, B. (2016). Respondent conditioning in online panel surveys: Results of two field experiments. *Social Science Computer Review*, 34(1), 95–115. doi:10.1177/0894439315574022
- Sturgis, P., Allum, N., & Brunton-Smith, I. (2009). Attitudes over time: The psychology of panel conditioning. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 113–126). Chichester: Wiley.
- Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755–769. doi:10.1111/j.1540-5907.2006.00214.x
- Tesser, A. (1978). Self-generated attitude change. *Advances in Experimental Social Psychology*, 11, 289–338. doi:10.1016/S0065-2601(08)60010-6
- Toepoel, V., Das, M., & van Soest, A. (2008). Effects of design in web surveys comparing trained and fresh respondents. *Public Opinion Quarterly*, 72(5), 985–1007. doi:10.1093/poq/nfn060
- Toepoel, V., Das, M., & van Soest, A. (2009). Relating question type to panel conditioning: Comparing trained and fresh respondents. *Survey Research Methods*, 3(2), 73–80.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.
- Warren, J. R., & Halpern-Manners, A. (2012). Panel conditioning in longitudinal social science surveys. *Sociological Methods & Research*, 41(4), 491–534. doi:10.1177/0049124112460374
- Waterton, J., & Lievesley, D. (1987). Attrition in a panel study of attitudes. *Journal of Official Statistics*, 3(3), 267–282.
- Waterton, J., & Lievesley, D. (1989). Evidence of conditioning effects in the British social attitudes panel. In D. Kasprzyk, G. J. Duncan, G. Kalton, & M. P. Singh (Eds.), *Panel surveys* (pp. 319–339). New York, NY: Wiley.
- Wilding, S., Conner, M., Sandberg, T., Prestwich, A., Lawton, R., Wood, C., ... Sheeran, P. (2016). The question-behaviour effect: A theoretical and methodological review and meta-analysis. *European Review of Social Psychology*, 27(1), 196–230. doi:10.1080/10463283.2016.1245940
- Yan, T., & Eckman, S. (2012). Panel conditioning: Change in true value versus change in self-report. In *Proceedings of the Survey Methods Research Section* (pp. 4726–4736). Alexandria, VA: ASA.