

Der Einfluss der Ausführlichkeit von Vignetten auf die Erfassung prosozialer Einstellungen

Ergebnisse zweier Split-Ballot Experimente

Von Stefanie Eifler und Knut Petzold*

Zusammenfassung: Szenario-Techniken oder Vignetten wurden häufig im Rahmen der Analyse heikler Themen eingesetzt. Bis heute ist unklar, ob Vignetten als Verfahren der indirekten Einstellungsmessung in einem solchen Forschungszusammenhang zu validieren Ergebnissen führen als klassische direkte Verfahren der Einstellungsmessung (z. B. Items). Es wird angenommen, dass die Ausführlichkeit der Darstellung von Urteilsgegenständen und die individuelle Tendenz zu sozial erwünschtem Antwortverhalten die Erfassung von Einstellungen zu prosozialem Verhalten beeinflussen. Die Annahme, dass insbesondere eine ausführliche Vignette zu ehrlicheren Antworten führt, wird anhand von zwei Split-Ballot Experimenten untersucht. Ein Experiment bezieht sich inhaltlich auf das Erleben einer Gelegenheit zu prosozialem Verhalten im Alltag, also auf einen konkreten Ausschnitt alltäglicher Erfahrung. Ein anderes Experiment bezieht sich inhaltlich auf die Normalität alltäglicher Gelegenheiten zu prosozialem Verhalten, also auf die im konkreten Ausschnitt der alltäglichen Erfahrung geltenden normativen Erwartungen. Die empirische Analyse erfolgt auf der Grundlage von Daten, die im Rahmen Computergestützter Telefoninterviews (CATI) an einer Auswahl von 560 Einwohnern einer deutschen Großstadt im Alter von 18 bis 65 Jahren erhoben wurden. Logistische und OLS-Regressionsanalysen führen zu dem Ergebnis, dass die Ausführlichkeit der Vignetten nur in Situationen hypothesenkonform von Bedeutung ist, die sich auf die Normalität alltäglicher Gelegenheiten zu prosozialem Verhalten beziehen, nicht aber in Situationen, die sich auf das Erleben konkreter Gelegenheiten im Alltag beziehen. Die Ergebnisse werden mit Bezug auf methodologische und methodische Fragen diskutiert.

1 Einleitung

In der sozialwissenschaftlichen Methodenforschung wird seit den 1980er Jahren häufig die *Methode der gegabelten Befragung (Split-Ballot Experiment)* eingesetzt. Diese wurde bereits von Benson (1941) beschrieben und vor allem mit den Arbeiten von Cantril (1944) bekannt, hat sich aber erst durchgesetzt, als systematisch die kognitiven Prozesse bei der Beantwortung von Fragen untersucht wurden (Schwarz / Sudman 1992; Sudman et al. 1996). Gegenstand gegabelter Befragungen sind Themen der Methodenforschung wie die Analyse von Kontext- und Reihenfolgeeffekten, die Analyse der Einflüsse unterschiedlicher Frageformulierungen und Einflüsse verschiedener Antwortformate (vgl. zusammenfassend Petersen 2002). Bei einer gegabelten Befragung wird die Untersuchungsstichprobe zufällig in mindestens zwei Gruppen geteilt, die im Rahmen einer Umfrage systematisch mit unterschiedlichen Stimuli konfrontiert werden. Insofern als prinzipiell eine Randomisierung der Probanden in Experimental- und Kontrollgruppe(n) erfolgt, sind gegabelte Befragungen „echte“ Experimente, die einen Test von Unterschiedshypothesen erlauben.

Im vorliegenden Zusammenhang wird die Methode der gegabelten Befragung eingesetzt, um methodische Aspekte des Einsatzes von Vignetten im Rahmen der Umfrageforschung empirisch zu analysieren. Vignetten sind kurze Beschreibungen von realen Situationen, die Probanden im Rahmen einer Umfrage dargeboten werden (Kerlinger 1986). Sie dienen der Erfassung normativer Einstellungen und Handlungsintentionen. Anders als bei der Erfassung von Einstellungen in der Tradition von Thurstone, Likert und Guttman, wo Items verwendet werden, die aus einem Satz bzw. einer Aussage bestehen, wird der zu beurteilende

* Wir danken den beiden anonymen Gutachtern sowie den Herausgebern für wertvolle Hinweise zu einer früheren Version dieses Beitrags.

Gegenstand in Vignetten in ausführlicherer Form dargestellt. Auf diese Weise wird eine größere Nähe der Inhalte von Vignetten zum alltäglichen Erleben hergestellt. Gerade diese Ausführlichkeit wird im Allgemeinen als Vorzug bei der Messung betrachtet, da die größere Genauigkeit der Beschreibung die Tendenz zu sozial erwünschtem Antwortverhalten mindern soll: Es wird angenommen, dass die Befragten einen geringeren Interpretationsspielraum haben und daher bei ihrer Antwort nicht auf normative Aspekte bei der Beurteilung der geschilderten Situationen oder Gegenstände zurückgreifen. Allerdings wurde genau diese Annahme bislang nicht eigens empirisch analysiert, sodass sich die folgende Frage ergibt: Hat die Ausführlichkeit von Vignetten tatsächlich Vorteile in dem Sinne, dass ausführliche Vignetten in geringerem Maße zu normativ geprägtem Antwortverhalten führen als nicht ausführliche Vignetten? Und sind entsprechend Antworten auf ausführlichere Vignetten valider als Antworten auf nicht ausführliche Vignetten?

Zur Beantwortung dieser Frage wird ein Beispiel aus dem Bereich prosozialen Verhaltens herangezogen, da hier eine besondere Anfälligkeit gegenüber sozial erwünschten bzw. normativ geprägten Antworten vermutet wird. Als „prosozial“ werden Handlungen bezeichnet, mit denen Akteure einer anderen Person absichtlich eine Annehmlichkeit bereiten möchten, ohne dass sie hierzu verpflichtet wären. In der Literatur wird ein organisiertes prosoziales Engagement – etwa im Rahmen eines zivilen Engagements – von spontanen Hilfleistungen in alltäglichen Situationen unterschieden. Dabei hat sich die experimentelle Forschung bislang auf die Analyse von spontanem prosozialem Verhaltens konzentriert und beispielsweise die Rückgabe vermeintlich verlorener Briefe, das Aufheben von heruntergefallenen Gegenständen oder das Vorlassen in Warteschlangen untersucht (vgl. Bierhoff 2002, 2010). Im vorliegenden Zusammenhang wird das Beispiel des Vorlassens in Warteschlangen aufgegriffen und zum Gegenstand der Analyse gemacht. Über die Anwendung zweier Split-Ballot Experimente werden zugleich die Stärken experimenteller Ansätze speziell im Zusammenhang mit methodischen Fragestellungen herausgearbeitet.

Nach einer kurzen Erörterung des Vignetten-Ansatzes und seiner Anwendungsfelder werden die grundlegenden kognitiven Prozesse bei der Beantwortung von Fragen behandelt. Die Konzeptualisierung der Studie beinhaltet die Ableitung von Hypothesen über die Auswirkungen der Ausführlichkeit von Vignetten auf das Ausmaß normativ geprägten Antwortverhaltens. Der Vorstellung der Stichprobe und der Ergebnisse folgt schließlich die Diskussion.

2 Problemstellung

Im Vergleich zu herkömmlichen Messungen von Einstellungen über direkte Fragen bzw. Items wird Vignetten als zentralem Vorteil eine Minderung sozial erwünschten Antwortverhaltens zugeschrieben. Aufgrund der Messung von Reaktionen auf (systematisch variierte) Situationsmerkmale kann indirekt auf Effekte sozialer Erwünschtheit geschlossen werden (Kerlinger 1986: 476; Armacost et al. 1991). Während Studien, die sich auf einen Vergleich von direkten mit indirekten Einstellungsmessungen beziehen, schon selten vorkommen, mangelt es besonders an Untersuchungen, die speziell den Einfluss der Ausführlichkeit von Vignetten auf das Ausmaß normativ geprägter Antworten thematisieren.

2.1 Vignetten und ihre Anwendungsfelder

Als *Vignetten* werden kurze, realitätsnahe Beschreibungen von Situationen bezeichnet, die im Rahmen einer Umfrage als Stimuli eingesetzt werden, wobei normative Einstellungen oder Handlungsintentionen als Reaktion auf diese Stimuli gemessen werden (Kerlinger 1986). Sie entsprechen damit einer Forderung, die Campbell (1950) im Rahmen des *indirect assessment movement* formuliert hat: „There is a very real need for instruments which do not destroy the natural form of the attitude in the process of describing it“ (Campbell 1950: 15).

Vignetten – von Campbell (1950) als „situations“ oder „episodes“ bezeichnet – werden eingesetzt, um soziale Einstellungen und/oder (vermutliches) soziales Handeln realitätsnah und (dennoch) standardisiert zu erfassen.

Die Anwendungsfelder von Vignetten sind vielfältig: Am bekanntesten ist der Einsatz von Vignetten im Rahmen des Faktoriellen Surveys (Rossi 1979). Dabei handelt es sich um ein surveybasiertes Experiment, in dem bei der Beschreibung fiktiver Objekte, Personen oder Situationen Merkmale (Dimensionen) über verschiedene Levels variiert werden. Die Methode erlaubt die Berücksichtigung einer Vielzahl von Merkmalen verschiedener Ausprägungen (Faktorielle Designs) und wurde bislang vor allem zur Messung von Normen und Einstellungen (Rossi / Anderson 1982; Beck / Opp 2001), mitunter aber auch zur Messung von Handlungsinentionen (Abraham / Meyer 2012) eingesetzt (vgl. zusammenfassend Wallander 2009). Darüber hinaus werden Vignetten auch im Rahmen von Szenario-Techniken – mit und ohne die Kopplung an experimentelle Designs – verwendet, um Einflüsse von Situationsmerkmalen auf soziale Einstellungen und (vermutliches) soziales Handeln zu untersuchen. Dabei liegen in aller Regel experimentelle between-subject Designs ohne Messwiederholung zugrunde. Typische Forschungsfelder sind üblicherweise: Situationen mit Dilemmacharakter, Verantwortungsdiffusion bei Zuschauern einer Notsituation, Hilfeverhalten oder Anwendungen im Kontext der Abschreckungstheorie (vgl. z. B. Finch 1987; Klepper / Nagin 1989; Nagin / Paternoster 1994). Demgegenüber weniger geläufig ist die Verwendung von Vignetten im Rahmen herkömmlicher Einstellungsskalen (vgl. z. B. Pedhazur 1969).

Gemeinsam ist den verschiedenen Anwendungen von Vignetten das Bemühen, die Stimuli den Situationen des alltäglichen Erlebens und Verhaltens möglichst anzugeleichen (vgl. z. B. Kerlinger 1986). Sofern eine solche Angleichung gelingt, werden Vignetten insbesondere im Kontext der Analyse heikler Themen als geeignete Stimuli angesehen: „They can [...] be unobtrusive approaches to sensitive information about the subjects (e.g., prejudiced attitudes, needs, and sexual preferences“ (Kerlinger 1986: 476).

Die bisherigen Ausführungen verdeutlichen, dass Vignetten für besser geeignet gehalten werden, insbesondere sensible Themen zu erfassen, weil sie eine größere Realitätsnähe aufweisen als direkte Messungen über Fragen und einfache Items. Diese Realitätsnähe wird im Vergleich zu direkten Messinstrumenten über eine stärkere *Ausführlichkeit* der Beschreibungen erreicht.

Im Kontext des Faktoriellen Survey Approach bezieht sich die Idee der Realitätsnähe auf die Komplexität der verwendeten Stimuli, die sich aus der Anzahl der Dimensionen ergibt, die im Rahmen des experimentellen Designs variiert werden. In diesem Zusammenhang sind erste Studien der Frage nachgegangen, ob und inwiefern die Gestaltung der Vignetten bezüglich der Anzahl der variierten Dimensionen und der Darstellung unrealistischer Fälle zu Lern-, Ermüdungs- und Reihenfolgeeffekten führen kann (Auspurg et al. 2009; Auspurg et al. 2010; Auspurg / Jäckle 2012). Es zeigte sich beispielsweise, dass die Komplexität von Vignetten vor allem Unterschiede im Antwortverhalten nach Alter und Bildung nach sich zieht (Sauer et al. 2011).

Im Kontext der vorliegenden Untersuchung bezieht sich die Idee der Realitätsnähe demgegenüber auf die Ausführlichkeit der Vignetten. Dabei wird angenommen, dass ein und dieselbe Information über eine Situation oder einen Urteilsgegenstand in einem oder in mehreren Sätzen beschrieben werden kann. Der Aspekt der Ausführlichkeit einer Situationsbeschreibung bezieht sich in dieser Untersuchung also auf die Formulierung oder den Wortlaut (wording) der Vignetten. Einflüsse der Ausführlichkeit von Vignetten wurden unseres Wissens bisher nicht gezielt untersucht. In bisherigen Studien blieb weiterhin häufig implizit,

wieso eine ausführlichere Situationsbeschreibung überhaupt einen Effekt auf das Antwortverhalten zeigen sollte. Bisherige Studien sind dieser Frage nicht in ausreichendem Maße nachgegangen, sodass hier sowohl theoretisch als auch empirisch untersucht werden soll, welche Bedeutung die Ausführlichkeit von Vignetten für die Validität der Antworten hat.

2.2 Soziale Erwünschtheit und die Qualität der Messung von Vignetten

Die Analyse von Vignetten unter methodischen Gesichtspunkten soll im Folgenden im Rahmen der *Total Survey Error Heuristik (TSE)* (Groves 1989) thematisiert werden. Unter dem Begriff des TSE werden alle Fehlerquellen aufgefasst, die Abweichungen zwischen einem gemessenen Wert und dem „wahren“ Wert verursachen können – im Rahmen einer Umfrage sind dies Merkmale des Interviewers, Merkmale des Befragten, Merkmale des Erhebungsinstruments und der Modus, in dem die Umfrage durchgeführt wird (Groves 1989). Im vorliegenden Zusammenhang sollen Einflüsse des Befragten und Merkmale des Erhebungsinstruments auf das Umfrageergebnis betrachtet werden.

Als eine Fehlerquelle in diesem Antwortprozess wird die Neigung von Befragten, sich selbst in einem guten Licht zu präsentieren, betrachtet, da sie die Ermittlung eines „wahren Wertes“ verzerrt. Seit Edwards (1957) wird die Tendenz von Befragten, sich im Rahmen einer empirischen Untersuchung möglichst günstig darzustellen, als „social desirability“ bezeichnet, allgemein übersetzt mit „soziale Erwünschtheit“. Als sozial erwünscht gelten Äußerungen, wenn sie den Werten und Normen innerhalb einer Gesellschaft entsprechen. Die Abgabe einer nicht dem wahren Wert entsprechenden Antwort wird als „faking“ bezeichnet, wobei zwischen „faking good“ und „faking bad“ unterschieden wird: „Faking good“ bezeichnet die Tendenz „to create a more favorable impression or, in general, the tendency to give socially more favorable impression or, in general, the tendency to give socially desirable responses to items in self-description“ (Edwards 1957: 53), „faking bad“ heißt entsprechend die Tendenz, einen eher unvorteilhaften Eindruck von der eigenen Person zu präsentieren.

Seit Crowne und Marlowe (1964) wird die Neigung, sich selbst in ein vorteilhaftes Licht zu rücken, weiterhin als Ausdruck eines individuellen Bedürfnisses nach sozialer Anerkennung (approval motive) betrachtet. Je stärker dieses Bedürfnis ausgeprägt ist, desto stärker ist die Tendenz, Antworten im Rahmen einer Befragung an geltenden Standards auszurichten und sich damit einen vorteilhaften Eindruck von der eigenen Person zu erzeugen.

Ausgehend von diesen klassischen Vorläufern hat sich Paulhus (1986, 1991) mit der Antworttendenz der sozialen Erwünschtheit beschäftigt: „A response bias is a systematic tendency to respond to a range of questionnaire items on some basis other than the specific item content (i. e., what the items were designed to measure)“ (Paulhus 1991: 17). Soziale Erwünschtheit wird von Paulhus folgendermaßen definiert: SDR (socially desirable responding) ist definiert als „the tendency to give answers that make the respondent look good“ (Paulhus 1991: 17). In einer systematischen Analyse von Skalen zur Erfassung der Tendenz zur sozialen Erwünschtheit hat Paulhus (1991) gezeigt, dass diese Skalen zwei Faktoren abbilden, einen Faktor, der inhaltlich den Bereich der „Angst“ repräsentiert, und einen Faktor, der inhaltlich mit dem Bereich der „Freundlichkeit“ korrespondiert. „Paulhus (1984, 1986) provided evidence that these two SDR factors represent (a) self-deceptive positivity (an honest but overly positive self-presentation) and (b) impression management (self-presentation tailored to an audience)“ (Paulhus 1991: 21).

Im vorliegenden Zusammenhang wird diese Unterscheidung von Paulhus aufgegriffen. Die Tendenz, in einer Weise zu antworten, die die eigene Person vor dem eigenen Selbst in einem guten Licht dastehen lässt, wird als *Self-Deception* (SD) bezeichnet; die Tendenz, in

einer Weise zu antworten, die die eigene Personen gegenüber anderen vorteilhaft erscheinen lässt, wird als *Other-Deception* (OD) bezeichnet. Insbesondere das Konstrukt der Other-Deception setzt erstens voraus, dass eine soziale Interaktion zwischen Interviewer und Befragtem stattfindet und dass der Interviewte zweitens die Erwünschtheit bezüglich seiner Antworten auch subjektiv wahrnimmt (vgl. hierzu auch Stocké 2004). Nach unserer Auffassung kann von einer sozialen Interaktion auch dann gesprochen werden, wenn Interviewer und Befragter sich nicht physisch gegenüberstehen, sondern beispielsweise – wie im Rahmen einer CATI-Umfrage – telefonieren.

Tourangeau und Yan (2007) erarbeiten in ihrem Überblicksartikel ebenfalls die Wechselwirkung zwischen dem Respondenten, dem Erhebungsinstrument und dem Interviewer. Über alle Studien hinweg sehen die Autoren einerseits Ursachen bei den Respondenten, die im Survey mit Fragen konfrontiert werden, auf die sie peinliche Antworten geben müssten. Andererseits entstehe eine solche Peinlichkeit nur in der Interaktion mit dem Interviewer oder im Zusammenhang mit Dritten. Sozial erwünschte Antworten seien damit insbesondere bei heiklen Themen wie Drogenmissbrauch oder Sexualität zu erwarten. Das Kernmerkmal eines heiklen Themas ist eine potentielle Rückwirkung der Antwort auf den Probanden, die eine gefällige und damit verzerrte Antwort fördert: „Socially sensitive research refers to studies in which there are potential social consequences or implications, either directly for the participants in the research or for the class of individuals represented by the research“ (Sieber / Stanley 1988: 49).

Neben der Neigung des Befragten, sozial erwünscht zu antworten, liegt eine weitere zentrale Fehlerquelle im Befragungsprozess im Erhebungsinstrument, das einen Bias durch soziale Erwünschtheit bedingen kann. Vorliegende Split-Ballot Experimente zeigen, dass Verzerrungen in den Antworten etwa vom Fragekontext und vom Wording, von der Anordnung der Fragen im Survey sowie vom Antwortformat abhängen (Schumann / Presser 1981; Schwarz / Sudman 1992, 1996; Sudman et al. 1996; Bradburn et al. 2004).

Für herkömmliche, direkte Formen der Einstellungsmessung kann die Validität der Antworten durch eine Reihe von Techniken, wie beispielsweise die Randomized Response Technique, die Unmatched Count Technique oder die Nominative Technique gesteigert werden (vgl. zusammenfassend Krumpal 2013). Darüber hinaus wurden indirekte Techniken der Einstellungsmessung entwickelt, wie etwa der Einsatz von Vignetten im Rahmen von Szenario-Techniken. Diese Methoden sind explizit mit der Absicht erarbeitet worden, die Einstellung in dem Kontext, in dem sie auch in der Realität der Befragten relevant ist, zu erfassen (vgl. Campbell 1950; Kerlinger 1986). Auf diese Weise sollten validere Einstellungsmessungen möglich sein (Alexander / Becker 1978; Finch 1987).

Es mangelt allerdings weitestgehend an Studien, die diesen methodischen Vorteil tatsächlich belegen. Bisherige Ansätze vergleichen zumeist verschiedene Vignetteninhalte, Vignetten mit direkten Techniken der Befragung oder Vignetten mit anderen Datenerhebungsformen. Neff (1979) verglich etwa Vignetten, die sich auf den Probanden selbst beziehen, mit Vignetten, die sich auf eine hypothetische Person beziehen, und konnte zeigen, dass das Ausmaß sozial erwünschten Antwortverhaltens geringer ist, wenn hypothetische Personen geschildert werden. Im Vergleich zu anderen Erhebungstechniken sind die Ergebnisse zur Validität von Vignettenmessungen unterschiedlich: Einige Autoren sehen keine eindeutige Über- oder Unterlegenheit einer direkten Messung gegenüber einer indirekten (Armacost et al. 1991; Nisic / Auspurg 2009); andere konnten zeigen, dass Vignetten im Vergleich zu direkten Techniken zu differenzierteren Ergebnissen führen. So konnten Unterschiede zwischen Gruppen, die im Falle von direkten Techniken nicht gefunden wurden, mit Vignetten gefunden werden (Burstin 1980). Weiterhin weisen Studien, in denen Vignetten mit nichtteilnehmenden Beobachtungen verglichen werden, darauf hin, dass Vignetten nicht zu vali-

deren Messungen führen. Es werden gerade bei heiklen Themen weniger sozial unerwünschte (Fundunterschlagung, kleine Normverstöße im Straßenverkehr) und mehr sozial erwünschte Verhaltensweisen (Rücksendung verlorener Briefe) berichtet als im Feld beobachtet – allerdings, durch die Anlage der Untersuchungen bedingt, an unterschiedlichen Stichproben (Eifler 2007, 2010; Groß / Börensen 2009).

Gleichwohl ist der Einfluss der Ausführlichkeit der Formulierung einer Vignette auf das Ausmaß sozial erwünschten Antwortverhaltens bislang nicht untersucht worden, was noch einmal die Bedeutung der Fragestellung der vorliegenden Studie verdeutlicht. Insgesamt halten wir daher die Untersuchung der Auswirkungen der Ausführlichkeit von Vignetten für notwendig. Wir gehen davon aus, dass gerade am Beispiel prosozialen Verhaltens die Anfälligkeit unterschiedlicher Arten der Formulierung von Vignetten für sozial erwünschte Antworten im Sinne der Ergebniskontrolle untersuchungsbedürftig ist.

2.3 Konzept und Hypothesen

Der zentrale Vorteil von Vignetten wurde stets darin gesehen, dass Vignetten als realitätsnahe Schilderungen von alltäglichen Situationen auch zu einem dem alltäglichen Erleben und Verhalten entsprechenden Antwortverhalten der Befragten führen. Eine theoretische Begründung für diese Annahme wurde in den meisten Studien jedoch nicht angeführt. Obwohl von Anfang an die Möglichkeit, mit Vignetten Simulationen von alltäglichen Handlungsvollzügen zu realisieren, kontrovers diskutiert wurde (vgl. z. B. Hughes 1998), haben zahlreichen Autoren Vignetten zur Erfassung des alltäglichen Erlebens eingesetzt (vgl. Hegtvedt 1988, 1990). Die Befunde verschiedener empirischer Untersuchungen nähren Zweifel an der Leistungsfähigkeit von Vignetten: So führten empirische Studien, in denen direkte Messungen und Vignetten systematisch miteinander verglichen wurden, zu unterschiedlichen Ergebnissen; in Studien, in denen Vignetten mit Feldbeobachtungen kontrastiert wurden, deutete sich an, dass gerade bei heiklen Themen mehr sozial erwünschte Verhaltensweisen (Rücksendung verlorener Briefe) angesichts einer Vignette berichtet als im Feld beobachtet werden (z. B. Eifler 2010).

Um sich dem Problem theoretisch anzunähern, erscheint eine Adaptation des *Modells des Antwortprozesses* nach Tourangeau (1984) an das Vignettendesign sinnvoll. Im Rahmen einer Umfrage ist der Befragte vor eine komplexe Aufgabe gestellt. Tourangeau (1984; Tourangeau et al. 2000) hat die kognitiven Aufgaben, die ein Befragter dabei bewältigen muss, mit dem Modell des Antwortprozesses beschrieben: Demnach muss ein Befragter, der eine Frage gestellt bekommt, zunächst den Inhalt der Frage interpretieren (interpretation), sodann Informationen aus seinem Gedächtnis abrufen (retrieval), ein Urteil bilden (judgement) und sodann seine Antwort in ein vorgegebenes Antwortformat einordnen (response selection). Überträgt man diese Überlegung auf die Messung mit Vignetten, bekommt ein Proband nach der Präsentation einer Vignette eine Frage zu deren Beurteilung gestellt und muss zunächst die Frage selbst interpretieren sowie mit Bezug darauf Informationen aus seinem Gedächtnis abrufen, bevor er sich ein Urteil bilden und eine Antwort geben kann.

Als theoretische Begründung für die Idee, dass Vignetten dem Erleben einer alltäglichen Situation sehr nahe kommen und in diesem Sinne extern valide Daten zu erheben erlauben, können daran anknüpfend kognitionspsychologische Schema- und Skripttheorien herangezogen werden (Bartlett 1932; Markus 1977; Abelson 1976). Insbesondere aus der Skripttheorie (Abelson 1976; Schank / Abelson 1977) kann die These abgeleitet werden, dass das Wissen über alltägliche Handlungsvollzüge in Form von Skripten kognitiv organisiert und verfügbar ist. Mit der Präsentation von Vignetten können diese Skripte evoziert werden, sodass das Antwortverhalten sich auf der Basis des jeweils evozierten Skripts gestaltet (vgl. hierzu ausführlicher Eifler 2007, 2010). Aus der Perspektive der Skripttheorie kann ange-

nommen werden, dass mit einer kurzen Schilderung einer alltäglichen Situation ein allgemeines Skript evoziert wird, während mit einer langen Schilderung einer alltäglichen Situation ein situationsspezifisches Skript aktiviert wird. Vorstellungen über Alltagssituationen im Allgemeinen führen dementsprechend zu eher allgemeinen Urteilen über die in den Vignetten geschilderten Situationen, während mit zunehmender Ausführlichkeit der Information in der Vignette bezogen auf die geschilderten Situationen passfähigere Informationen aus dem Gedächtnis abgerufen werden können. Entsprechend kann angenommen werden, dass die Antworten umso situationsspezifischer und damit umso weniger sozial erwünscht ausfallen, je ausführlicher ein und dieselbe Situation in einer Vignette beschrieben wird. Wird der Proband dagegen mit einer wenig ausführlichen Situationsbeschreibung konfrontiert, die er beurteilen soll, ist er gezwungen, auf sehr allgemeine Skripte zurückzugreifen. Es ist sehr plausibel, davon auszugehen, dass in diesem Fall eher Normen als Verhaltensregelmäßigkeiten und als Vorstellungen über typische Abläufe der geschilderten Ausschnitte der alltäglichen Erfahrung als Interpretationsgrundlage dienen. Bei einer weniger ausführlichen Beschreibung kann daher davon ausgegangen werden, dass die Antworten weniger situationsspezifisch und stärker normativ geprägt sind.

Ausgehend von den theoretischen Überlegungen werden im Folgenden die Untersuchungshypothesen hinsichtlich der Einflüsse der Ausführlichkeit von Vignetten zur Erfassung prosozialer Einstellungen abgeleitet. Dabei werden zwei Aspekte der prosozialen Einstellungen jeweils gesondert als abhängige Variablen betrachtet; zum einen das Vorhandensein eines Verpflichtungsgefühls zu prosozialem Verhalten – als binäre Variable mit den Ausprägungen ja oder nein, und zum anderen die Intensität eines Verpflichtungsgefühls zu prosozialem Verhalten – als metrische Variable.

Hypothese 1: Einfluss der Ausführlichkeit der Vignetten

(a) *Wenn die Information über eine Gelegenheit zu prosozialem Verhalten in einer nicht ausführlichen Form präsentiert wird, dann wird das Vorhandensein eines Verpflichtungsgefühls zu prosozialem Verhalten eher berichtet.*

(b) *Wenn die Information über eine Gelegenheit zu prosozialem Verhalten in einer nicht ausführlichen Form präsentiert wird, dann wird eine höhere Intensität des Verpflichtungsgefühls zu prosozialem Verhalten berichtet.*

Ferner ist hier auch der direkte Einfluss der individuellen Neigung zu sozial erwünschtem Antwortverhalten von Interesse. Diese Neigung wird in Anlehnung an Paulhus (1991) als Self-Deception und Other-Deception berücksichtigt, wobei ein genereller Einfluss erwartet wird.

Hypothese 2: Einfluss der Tendenz zu sozial erwünschtem Antwortverhalten

(a) *Je stärker die individuelle Tendenz zu sozial erwünschtem Antwortverhalten ausgeprägt ist, desto eher wird das Vorhandensein eines Verpflichtungsgefühls zu prosozialem Verhalten berichtet.*

(b) *Je stärker die individuelle Tendenz zu sozial erwünschtem Antwortverhalten ausgeprägt ist, desto höher ist die Intensität des Verpflichtungsgefühls zu prosozialem Verhalten.*

Schließlich kann eine Interaktion zwischen der Tendenz, sozial erwünscht zu antworten und der Ausführlichkeit der Vignetten angenommen werden. Dabei wird der negative Zusammenhang zwischen der Ausführlichkeit und dem berichteten Verpflichtungsgefühl bzw. dessen Intensität noch weiter gesteigert.

Hypothese 3: Interaktion zwischen der Tendenz zu sozial erwünschtem Antwortverhalten und der Ausführlichkeit der Vignetten

(a) Je stärker die Tendenz zu sozial erwünschtem Antwortverhalten ist, desto stärker führt eine nicht ausführliche Information über eine Gelegenheit zu prosozialem Verhalten dazu, dass das Vorhandensein eines Verpflichtungsgefühls zu prosozialem Verhalten berichtet wird.

(b) Je stärker die Tendenz zu sozial erwünschtem Antwortverhalten ist, desto stärker führt eine nicht ausführliche Information über eine Gelegenheit zu prosozialem Verhalten zu einer höheren Intensität des Verpflichtungsgefühls zu prosozialem Verhalten.

Zur empirischen Überprüfung dieser Hypothesen greifen wir auf die Methode der gegabelten Befragung (Split-Ballot Experiment) zurück. Darin verbinden sich die Vorteile des experimentellen Designs mit denen eines Surveys. Das experimentelle Design erlaubt über die Strategien der Variation der Formulierung der Vignetten und der Randomisierung zweier Untersuchungsgruppen aus einer Stichprobe über die beiden Treatments die Identifikation des isolierten Einflusses eines mehr oder weniger ausführlichen Wortlauts auf das Antwortverhalten. Die Einbindung in einen Survey unterstützt dabei die Umsetzung des experimentellen Designs im Rahmen einer repräsentativen Bevölkerungsumfrage in einer deutschen Großstadt.

Außerdem werden die vorgenannten Hypothesen in Verbindung mit unterschiedlichen Instruktionen getestet, um die Idee der allgemeinen und situationsspezifischen Skripte vertieft untersuchen zu können. Dabei wird angenommen, dass die in den Hypothesen 1 bis 3 vermuteten Einflussbeziehungen wirken, auch wenn die Vignetten mit unterschiedlichen Instruktionen zur Beantwortung vorgegeben werden. Es werden Instruktionen unterschieden, die sich zum einen auf das Erleben einer Gelegenheit zu prosozialem Verhalten und zum anderen auf den Alltag bzw. die Normalität des Alltags beziehen. Das Experiment zum Erleben im Alltag nimmt dabei einen starken Bezug auf die konkret erlebte Situation. Im Sinne der Skripttheorie würde der Bezug zum Erleben die situationsspezifischen Skripte gezielt evozieren, und der Vorteil einer ausführlichen Vignette könnte zum Tragen kommen. Das zweite Experiment fokussiert dagegen eine eher gewohnte Situation, die an sich schon einen stärkeren Normbezug aufweist. Eine Instruktion, die die Regelmäßigkeiten und die Normalität des Alltags anspricht, würde eher allgemeine Skripte aktivieren. Es wäre mit einer solchen Instruktion schwerer, den Vorteil einer ausführlichen gegenüber einer weniger ausführlichen Situationsbeschreibung herauszuarbeiten. In diesem Sinne stellt eine Instruktion, die sich auf die Normalität des Alltags bezieht, einen strengeren Test der Untersuchungshypothesen dar, da hier von einer stärkeren Tendenz zu sozial erwünschtem Antwortverhalten ausgegangen werden kann und Einflüsse der Ausführlichkeit der Vignetten schwerer nachweisbar sein müssten. Um die Aussagekraft der Ergebnisse dieser Studie weiter zu erhöhen, erfolgte die Variation der Ausführlichkeit der Vignetten also in zwei separaten Split-Ballot Experimenten mit jeweils unterschiedlichen Instruktionen.

3 Methode

Die empirische Untersuchung wurde im Rahmen der Teilstudie „Soziales Kapital in der Stadt“ des Projekts „Zusammenleben in der Stadt“ in der Zeit von April 2011 bis April 2013 durch das Zentrum für Sozialforschung Halle e.V. (ZSH) durchgeführt. Die Experimente zur Analyse der formalen Gestaltungsmerkmale von Vignetten waren in diesen Projektkontext eingebettet.

3.1 Verfahren der Datenerhebung

Die Studie wurde als computergestützte telefonische Befragung (CATI) zu einem Erhebungszeitpunkt (Querschnittstudie) in Leipzig durchgeführt. Als Erhebungszeitraum wurde die Zeit von Oktober bis November 2012 festgelegt. Das Erhebungsinstrument wurde im Vorfeld der Studie einem Pretest an wenigen Probanden ($n = 6$) unterzogen. Dieser führte zu dem Ergebnis, dass die Befragung inhaltlich und technisch fehlerfrei verlief. An der Datenerhebung waren insgesamt 39 Interviewer beteiligt, die zuvor eine Interviewerschulung durchlaufen hatten.

3.2 Stichprobe

Grundgesamtheit und Auswahlverfahren

Im Rahmen der Studie sollten insgesamt 630 Interviews geführt werden, wobei die Befragten sich gleichmäßig auf alle 63 Leipziger Ortsteile verteilen sollten. Auf der Basis des Einwohnermelderegisters der Stadt Leipzig wurde eine Zufallsauswahl vorgenommen. Als Grundgesamtheit wurden die im Meldewesen erfassten Bürgerinnen und Bürger bestimmt, die am Tag der Stichprobenziehung im Alter zwischen 18 und 65 Jahren waren und ihren Hauptwohnsitz in der Stadt Leipzig hatten. Um Sprach- und Verständnisprobleme während der Interviews zu vermeiden, wurden nur Personen mit deutscher Staatsangehörigkeit ausgewählt. Männliche und weibliche Personen sollten zu jeweils gleichen Anteilen in der Stichprobe enthalten sein. Aus jedem Ortsteil wurden rund 318 Personen, insgesamt also $n = 20.034$ Personen ausgewählt. Für die erhaltenen Adressen wurden anschließend Telefonnummern recherchiert, wobei für einige Ortsteile eine nachträgliche Recherche in Telefonverzeichnissen erforderlich wurde. Als Grundlage für die telefonische Befragung standen schließlich $n = 3.546$ Adressen mit Telefonnummern zur Verfügung. Insgesamt konnten 648 Interviews realisiert werden, was einem Rücklauf von ca. 18,27% entspricht.

Realisierte Stichprobe

Die soziodemographische Zusammensetzung der realisierten Stichprobe weist Abweichungen von der Grundgesamtheit der Studie auf. Insgesamt sind 11,65% der Befragten unter 18 Jahren oder über 65 Jahre alt. Die Daten dieser Befragten wurden aus der weiteren Analyse ausgeschlossen, da insbesondere für ältere Befragte eine Konfundierung des Alters mit den untersuchten unabhängigen Variablen zu erwarten gewesen wäre. Auf diese Weise verblieben $n = 560$ Probanden in der realisierten Stichprobe, wodurch sich der Rücklauf auf ca. 15,79% reduzierte.

Das mittlere Alter der Befragten betrug 50,72 Jahre (Standardabweichung $s = 10,84$). Alter und Geschlecht der Befragten wurden mit der Grundgesamtheit verglichen, die in der amtlichen Statistik repräsentiert ist. Hierzu wurde eine Vorversion des Statistischen Jahrbuches aus dem Jahr 2013 verwendet, welches Daten für das Jahr 2011 enthielt. Zum Zeitpunkt der Datenanalyse waren Daten zum Erhebungszeitraum noch nicht verfügbar. Es zeigte sich, dass unter 30-jährige Personen in der realisierten Stichprobe im Vergleich zur Grundgesamtheit sehr deutlich unterrepräsentiert sind, während umgekehrt über 30-jährige Personen in der realisierten Stichprobe im Vergleich zur Grundgesamtheit sehr deutlich überrepräsentiert sind. Dabei ist insbesondere der Anteil von Personen im Alter zwischen 54 und 65 Jahren in der realisierten Stichprobe mehr als doppelt so hoch wie in der Grundgesamtheit. Darüber hinaus zeigt sich, dass in der realisierten Stichprobe im Vergleich zur Grundgesamtheit überproportional viele Frauen vertreten sind (vgl. Tabelle A-1 im An-

hang). Aufgrund der im Vergleich zur Grundgesamtheit überproportionalen Anteile von älteren Befragten und Frauen in der realisierten Stichprobe werden Alter und Geschlecht in die statistischen Analysen zur Überprüfung der Untersuchungshypothesen einbezogen.

3.3 Operationalisierung

In der vorliegenden Studie wurden Einstellungen zu prosozialem Verhalten auf der Grundlage von Vignetten erhoben. Diese Vignetten bezogen sich inhaltlich auf eine alltägliche Situation, in der die Möglichkeit besteht, sich gegenüber einer anderen Person in prosozialer Weise zu verhalten. Konkret ging es um eine Situation, in der man im Supermarkt an der Kasse jemanden vorgehen lässt, der offenbar in Eile ist. Anhand dieser Situation sollten Einflüsse der Gestaltung von Vignetten untersucht werden, und zwar die Ausführlichkeit, mit der die Situation geschildert wird. Es wurden zwei Experimente durchgeführt. Im ersten Experiment wurden die Probanden gebeten, sich vorzustellen, dass sie die geschilderte Situation im Alltag erleben. Die Vignette wurde in einer nicht ausführlichen und einer ausführlichen Variante zur Beantwortung vorgegeben. Im zweiten Experiment wurden die Probanden gebeten, an ihren Alltag zu denken. Auch hier wurde die Vignette in einer nicht ausführlichen und einer ausführlichen Variante zur Beantwortung vorgegeben. Die Probanden wurden zufällig einem der beiden Experimente und einer der beiden Variationen zugewiesen. Jedem Experiment lag folglich ein einfaktorielles between-subjects-Design zugrunde. Im Folgenden werden die beiden Bedingungen der Experimente dargestellt. Die Abbildungen 1 und 2 zeigen die verwendeten Instruktionen und Vignetten im Wortlaut.

Abbildung 1: Experiment I mit Variationen

Bitte stellen Sie sich vor, Sie erleben im Alltag die folgende Situation:

Vignette – nicht ausführlich

Im Supermarkt bemerken Sie hinter sich in der Warteschlange an der Kasse jemanden, der es sehr eilig hat.

Vignette – ausführlich

Im Supermarkt stellen Sie sich in einer Warteschlange an der Kasse an. Hinter Ihnen stellen sich weitere Personen an. Sie bemerken jemanden, der es sehr eilig hat.

Abbildung 2: Experiment II mit Variationen

Bitte denken Sie an Ihren Alltag!

Vignette – nicht ausführlich

Im Alltag kommt es häufig vor, dass man in einer Warteschlange hinter sich jemanden bemerkt, der es sehr eilig hat.

Vignette – ausführlich

Im Alltag kommt es häufig vor, dass man sich in einer Warteschlange anstellt. Regelmäßig stellen sich hinter einem weitere Personen an. Oft bemerkt man jemanden, der es sehr eilig hat.

In einem weiteren Schritt wurde die Anzahl der Befragten in beiden Experimenten und beiden Bedingungen bestimmt. Wie bereits gesagt, wurden die Befragten zufällig einem der beiden Experimente zugewiesen und innerhalb des jeweiligen Experiments zufällig mit einer ausführlichen oder einer nicht ausführlichen Vignette konfrontiert. Im Folgenden wird die Anzahl der Befragten in beiden Experimenten und beiden Bedingungen nach Alter und Geschlecht differenziert dargestellt (vgl. Tabelle 1).

Tabelle 1: Verteilung der Befragten in Experiment I und II

Vignette	Experiment I: Erleben der Situation		Experiment II: Normalität des Alltags	
	nicht ausführlich	ausführlich	nicht ausführlich	ausführlich
absolute (relative) Häufigkeit	112 (42,75%)	150 (57,25%)	138 (46,46%)	159 (53,54%)
Altersdurchschnitt	51,51 Jahre	49,12 Jahre	52,41 Jahre	50,18 Jahre
Anteil Frauen	63,39%	62,00%	56,52%	62,89%

Die Verteilung der Probanden über die beiden Bedingungen der beiden Experimente zeigt ein relativ ausgeglichenes Gefüge. Die Altersunterschiede der Probanden zwischen den beiden Bedingungen der Experimente sind relativ gering und nicht signifikant (Experiment I: $t = 1,76$; n.s.; Experiment II: $t = 1,82$; n.s.). Auch der Anteil der Frauen variiert geringfügig aber nicht signifikant über die experimentellen Bedingungen in beiden Experimenten (Experiment I: $\phi = -0,01$; n.s.; Experiment II: $\phi = 0,06$; n.s.). Dennoch werden sowohl Alter als auch Geschlecht als Kontrollvariablen in alle nachfolgenden statistischen Analysen der Untersuchungshypothesen aufgenommen. Die Einbeziehung weiterer Variablen wie beispielsweise Bildung und/oder Beruf war im vorliegenden Zusammenhang nicht möglich, da diese Informationen über die Befragten nicht erhoben wurden.

Soziale Erwünschtheit

Die Analyse der individuellen Neigung, sich in vorteilhafter Weise zu präsentieren, wurde anhand der auf Paulhus (1986, 1991) zurückgehenden Unterscheidung zwischen Self-Deception und Other-Deception vorgenommen. Die Operationalisierung der beiden Aspekte von sozialer Erwünschtheit erfolgte anhand ausgewählter Items aus dem *Inventar zur Erfassung von zwei Faktoren sozialer Erwünschtheit* von Musch et al. (2002). Im Rahmen der vorliegenden Studie wurden aus der im Original umfassenderen Skala für beide Bereiche der sozialen Erwünschtheit jeweils vier Items ausgewählt, die die Konstrukte gut repräsentierten (vgl. Musch et al. 2002). Diese ausgewählten Items wurden den Befragten zur Beantwortung vorgegeben. Die Beantwortung der Items erfolgte auf einer siebenstufigen Skala von 0 (= völlige Ablehnung) bis 6 (= völlige Zustimmung). Abbildung 3 zeigt die Items zur Erfassung von sozialer Erwünschtheit im Wortlaut.

Abbildung 3: Items zur Erfassung von sozialer Erwünschtheit

Self-Deception	
SD-1	Ich bin mir meiner Urteile sehr sicher.
SD-2	Ich bin ein vollkommen rational denkender Mensch.
SD-3 (-)	Ich bin nicht immer mir selbst gegenüber ganz ehrlich gewesen.
SD-4 (-)	Ich weiß nicht immer die Gründe für meine Handlungen.
Other-Deception	
OD-1	Ich fluche niemals.
OD-2	Ich nehme niemals Dinge an mich, die mir nicht gehören.
OD-3 (-)	Ich bin schon einmal wegen einer angeblichen Krankheit nicht zur Arbeit/Schule gegangen.
OD-4 (-)	Manchmal lüge ich, wenn ich muss.

Die Items zur Erfassung von Self-Deception und Other-Deception wurden zunächst jeweils einer Hauptkomponentenanalyse mit schiefwinkliger Rotation unterzogen.¹ Für die

1 Die Interitemkorrelationen für Self-Deception und Other-Deception sind in Tabelle A-2 im Anhang wiedergegeben.

Items zur Erfassung von Self-Deception ergaben sich zwei Komponenten, bestehend aus Komponente 1: SD-3 und SD-4 sowie Komponente 2: SD-1 und SD-2 (vgl. Tabelle 2). Da der Aspekt der Selbsttäuschung eher durch die Items SD-3 und SD-4 erfasst wird, beide Items auf einer Hauptkomponente laden und miteinander korreliert sind ($r = 0,37; p \leq 0,05$), wurde nur aus den beiden Items „Ich bin nicht immer mir selbst gegenüber ganz ehrlich gewesen“ und „Ich weiß nicht immer die Gründe für meine Handlungen“ in rekodierter Form ein Summenscore gebildet (SD-Score, $M = 3,58, s = 1,65$) und in die nachfolgenden Analysen einbezogen.

Tabelle 2: Hauptkomponentenanalyse der Items zur Erfassung von Self-Deception

Items	Komponente 1*	Komponente 2*
SD-1	0,0195	0,6929
SD-2	-0,0200	0,7062
SD-3 umgepolt	0,7330	-0,0909
SD-4 umgepolt	0,6798	0,1137
Eigenwert	1,56	1,18
Erklärte Varianz	38,95%	29,38%

* Ladungen nach schiefwinkliger Rotation (Oblimin)

Für die Items zur Erfassung von Other-Deception ergaben sich ebenfalls zwei Komponenten, bestehend aus Komponente 1: OD-3 und OD-4 sowie Komponente 2: OD-1 und OD-2 (vgl. Tabelle 3). Da der Aspekt der Fremdtäuschung eher durch die Items OD-3 und OD-4 erfasst wird, beide Items auf einer Hauptkomponente laden und miteinander korreliert sind ($r = 0,25; p \leq 0,05$), wurde nur aus den beiden Items „Ich bin schon einmal wegen einer angeblichen Krankheit nicht zur Arbeit oder Schule gegangen“ und „Manchmal lüge ich, wenn ich muss“ in rekodierter Form ein Summenscore gebildet (OD-Score, $M = 3,83, s = 1,71$) und in den nachfolgenden Analysen verwendet.

Tabelle 3: Hauptkomponentenanalyse der Items zur Erfassung von Other-Deception

Items	Komponente 1*	Komponente 2*
OD-1	0,2721	0,6117
OD-2	-0,1335	0,7798
OD-3 umgepolt	0,6765	-0,1147
OD-4 umgepolt	0,6722	0,0768
Eigenwert	1,31	1,07
Erklärte Varianz	32,86%	26,66%

* Ladungen nach schiefwinkliger Rotation (Oblimin)

Abhängige Variablen

Im Anschluss an die Vignette, die den Befragten vom Interviewer vorgelesen wurde, wurden prosoziale Einstellungen zu der jeweils geschilderten Situation erhoben. Die Befragten wurden um eine Auskunft darüber gebeten, ob man sich in der geschilderten Situation verpflichtet fühlen kann, sich prosozial zu verhalten (im Folgenden: Vorhandensein Verpflichtungsgefühl bzw. Vorhandensein), und wenn ja, in welchem Ausmaß dies der Fall ist (im Folgenden: Intensität Verpflichtungsgefühl bzw. Intensität). Abbildung 4 zeigt die Items im Wortlaut:

Abbildung 4: Items zur Erfassung der prosozialen Einstellungen

Vorhandensein Verpflichtungsgefühl

In dieser Situation kann man sich verpflichtet fühlen, die Person, die es eilig hat, vorzulassen.

Intensität Verpflichtungsgefühl

In welchem Maße fühlen Sie sich in dieser Situation verpflichtet, die Person, die es eilig hat, vorzulassen?

Das Item zur Erfassung des Vorhandenseins des Verpflichtungsgefühls wurde mit dichotomem Antwortformat (0 = trifft zu; 1 = trifft nicht zu), das Item zur Erfassung der Intensität des Verpflichtungsgefühls wurde mit einer zehnstufigen Skala von 0 (= überhaupt nicht verpflichtet) bis 9 (= in höchstem Maße verpflichtet) vorgegeben. Tabelle 2 informiert über die Häufigkeitsverteilung der beiden Items zur Erfassung der prosozialen Einstellung in der realisierten Stichprobe.

Tabelle 4: Verteilung des Items zur Erfassung des Vorhandenseins des Verpflichtungsgefühls

Vorhandensein	Absolute Häufigkeit	Relative Häufigkeit
Trifft zu	487	88,22%
Trifft nicht zu	65	11,78%
Total	552	100%

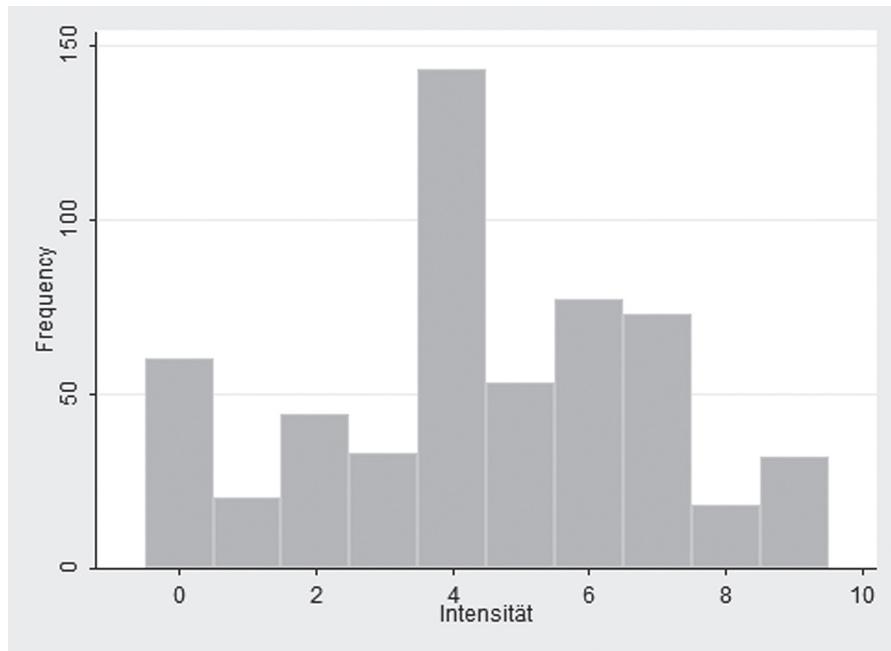
Aus Tabelle 4 geht hervor, dass mit 88,22% der überwiegende Teil der Probanden angesichts der in der Vignette geschilderten Situation zu der Einschätzung gelangt, dass man sich in einer solchen Situation verpflichtet fühlen kann, die Person, die es eilig hat, an der Kasse vorzulassen.

Wie Abbildung 5 zeigt, ist die Verteilung der Intensität des Verpflichtungsgefühls nur an nähernd unimodal ($M = 4,42$, $s = 2,47$). Sie entspricht aber hinsichtlich ihrer Schiefe (Schiefe = $-0,16$), nicht aber hinsichtlich ihrer Wölbung (Kurtosis = $2,34$) einer Normalverteilung, wie das Ergebnis eines Anpassungstest gezeigt hat. Angesichts dessen scheinen die Variablen insgesamt für die Anwendung multivariater Analyseverfahren geeignet zu sein.

3.4 Verfahren der Datenanalyse

In einem ersten Schritt der Datenanalyse wurden bivariate Analysen der untersuchten Zusammenhänge vorgenommen. Dabei wurden für beide Experimente zunächst Unterschiede in den abhängigen Variablen Vorhandensein und Intensität des Verpflichtungsgefühls nach der Ausführlichkeit der Vignetten und nach der Tendenz zu sozial erwünschtem Antwortverhalten berechnet. In einem zweiten Schritt der Datenanalyse wurden multivariate Analysen der Untersuchungshypothesen vorgenommen. Einflüsse der unabhängigen Variablen Ausführlichkeit der Vignetten und soziale Erwünschtheit auf das Vorhandensein und die Intensität des Verpflichtungsgefühls wurden anhand von Regressionstechniken bestimmt, nachdem die Anwendungsvoraussetzungen dieser Verfahren geklärt worden waren. Für die abhängige Variable Vorhandensein des Verpflichtungsgefühls war dies das Verfahren der logistischen Regression, für die abhängige Variable Intensität des Verpflichtungsgefühls war dies das Verfahren der OLS-Regression, nachdem sich die Robustheit dieses Verfahrens im Vergleich zur ordinalen Regression bestätigt hatte. Um den angenommenen Interaktionseffekt zu prüfen, wurden für jedes Experiment ein Haupteffektmodell und zusätzlich ein Modell berechnet, das zwei multiplikative Terme enthält. Die Spannweiten der Ausgangsvariablen wurden dafür zwischen 0 und 1 fixiert.

Abbildung 5: Verteilung des Items zur Erfassung der Intensität des Verpflichtungsgefühls



4 Ergebnisse

Im Folgenden werden die Ergebnisse der Datenanalysen für beide Experimente dargestellt. Dabei werden in einem ersten Schritt die Ergebnisse der bivariaten Analysen berichtet, in deren Rahmen Unterschiede im Vorhandensein und in der Intensität des Verpflichtungsgefühls nach der Ausführlichkeit der Vignette und nach der Tendenz zu sozial erwünschtem Antwortverhalten berechnet wurden. In einem zweiten Schritt werden die Ergebnisse der multivariaten Analysen berichtet, in deren Rahmen Einflüsse der Ausführlichkeit der Vignetten und der Tendenz zu sozial erwünschtem Antwortverhalten auf das Vorhandensein und die Intensität des Verpflichtungsgefühls berechnet wurden.

4.1 Bivariate Analysen der untersuchten Zusammenhänge

Unterschiede im Vorhandensein und in der Intensität des Verpflichtungsgefühls nach Ausführlichkeit der Vignetten

Im ersten Experiment, in dem die Vignetten mit einer Instruktion zum Erleben der geschilderten Situation im Alltag zur Beantwortung vorgegeben wurden, ergibt sich, dass keine Unterschiede der Häufigkeitsverteilungen des Vorhandenseins ($phi = 0,01$; n.s.) und der Intensität ($t = -0,38$; n.s.) des Verpflichtungsgefühls in Abhängigkeit von der Ausführlichkeit der Vignetten bestehen (Tabelle 5).

Tabelle 5: Verteilung des Vorhandenseins und der Intensität des Verpflichtungsgefühls nach Ausführlichkeit

Vignette	Experiment I: Erleben der Situation		Experiment II: Normalität des Alltags	
	nicht ausführlich	ausführlich	nicht ausführlich	ausführlich
<i>Vorhanden</i>				
absolute (relative) Häufigkeit	99 (89,19%)	132 (88,59%)	119 (88,15%)	137 (87,26%)
<i>Nicht vorhanden</i>				
absolute (relative) Häufigkeit	12 (10,81%)	17 (11,41%)	16 (11,85%)	20 (12,74%)
<i>Intensität</i>				
Mittelwert	4,323	4,458	4,744	4,180
Standardabweichung	2,614	2,263	2,389	2,246

Anders ist die Situation für das zweite Experiment, in dem die Vignetten mit einer Instruktion zum Alltag im Allgemeinen und damit zur Normalität des Alltags zur Beantwortung vorgegeben wurden. Hier sind zwar hinsichtlich des Vorhandenseins des Verpflichtungsgefühls die durch die Ausführlichkeit der Vignette bedingten und unbedingten Häufigkeitsverteilungen identisch ($phi = 0,01$; n.s.), jedoch zeigt sich für die Intensität des Verpflichtungsgefühls ($t = 1,91$, $p \leq 0,05$), dass der Mittelwert in der Gruppe der Probanden, die eine nicht ausführliche Vignette beantwortet haben, höher ist als in der Gruppe der Probanden, die eine ausführliche Vignette beantwortet haben. Unter der Bedingung, dass eine ausführliche Situation beschrieben wird, ist also durchschnittlich eine weniger positive Einstellung zu prosozialem Verhalten zu verzeichnen. Ein signifikanter Unterschied in der Intensität des Verpflichtungsgefühls zeigt sich – anders als erwartet – nicht unter der weniger strengen Bedingung, in der die Instruktion auf das Erleben der geschilderten Situation Bezug nimmt, sondern nur unter der strengeren Bedingung, in der sich die Instruktion auf die Normalität des Alltags bezieht.

Unterschiede im Vorhandensein und in der Intensität des Verpflichtungsgefühls nach sozialer Erwünschtheit

Auch die Analyse der bivariaten Zusammenhänge zwischen der Tendenz zu sozial erwünschtem Antwortverhalten und dem Vorhandensein und der Intensität des Verpflichtungsgefühls zeigt ein differenziertes Bild: Im ersten Experiment, dessen Instruktion auf das Erleben im Alltag bezogen war, sind die Zusammenhänge zwischen Other-Deception und dem Vorhandensein des Verpflichtungsgefühls (punktwise Korrelation) und der Intensität des Verpflichtungsgefühls (Rangkorrelation) nicht signifikant (Vorhandensein: $r = 0,043$; n.s.; Intensität: $rho = 0,143$; n.s.). Im zweiten Experiment, dessen Instruktion sich auf die Normalität des Alltags bezog, findet sich demgegenüber kein signifikanter Zusammenhang zwischen Other-Deception und dem Vorhandensein des Verpflichtungsgefühls, allerdings findet sich ein sehr schwacher und nur grenzwertig signifikanter Zusammenhang zwischen Other-Deception und der Intensität des Verpflichtungsgefühls (Vorhandensein: $r = -0,117$, n.s.; Intensität: $rho = 0,133$, $p \leq 0,05$).

4.2 Multivariate Analysen der Untersuchungshypothesen

Die folgenden Ausführungen berichten über die Ergebnisse der empirischen Analysen der Untersuchungshypothesen, und zwar für beide Experimente gesondert und nacheinander.

Experiment I: Instruktion zum Erleben im Alltag

Die Regressionsmodelle in Tabelle 6 zeigen, dass die Angaben der Befragten hinsichtlich des Vorhandenseins und der Intensität des Verpflichtungsgefühls nicht von deren Alter oder Geschlecht, aber auch nicht von der Ausführlichkeit der Vignettenformulierung abhängen. In dem Experiment, in dem sich die Vignetten auf das Erleben einer Situation im Alltag beziehen, ist es also unerheblich für die Erfassung der Einstellung zu prosozialem Verhalten, ob eine mehr oder weniger ausführliche Version der Vignette verwendet wird. Auch die angenommene Interaktion zwischen der Ausführlichkeit und der Tendenz, sozial erwünscht zu antworten, kann in diesen Modellen nicht bestätigt werden. Die Daten liefern demnach in diesem Experiment weder Hinweise für das Zutreffen der Hypothese I noch für das der Hypothese III.² Bemerkenswert ist aber der signifikante, positive Koeffizient der Other-Deception im Modell zur Intensität des Verpflichtungsgefühls. Das weist darauf hin, dass unter der Bedingung des Erlebens der Situation im Alltag eine umso stärkere Intensität des Verpflichtungsgefühls berichtet wird, je stärker die Neigung zur Other-Deception ist. Das Modell ist allerdings insgesamt nicht signifikant, und auch der Effekt ist sehr schwach und nur knapp signifikant. Für das Ausmaß des Verpflichtungsgefühls in einer erlebten Alltagssituation deutet sich demnach Unterstützung für die Annahmen aus Hypothese II an.

Tabelle 6: Einflüsse der Ausführlichkeit der Vignette und der Tendenz zu sozial erwünschtem Antwortverhalten auf das Vorhandensein und die Intensität des Verpflichtungsgefühls

Experiment I Erleben im Alltag	Vorhandensein Verpflichtungsgefühl (Logistische Regression)				Intensität Verpflichtungsgefühl (OLS-Regression)			
	<i>b</i>	<i>z</i>	<i>b</i>	<i>z</i>	<i>b</i>	<i>t</i>	<i>b</i>	<i>t</i>
Ausführlichkeit	0,100	0,25	1,087	0,83	0,179	0,50	-0,539	-0,50
Self-Deception	-0,024	-0,20	0,184	0,89	0,047	0,45	-0,095	-0,53
Other-Deception	0,077	0,62	0,023	0,11	0,258*	2,44	0,288 ⁺	1,70
SD*Ausführlichkeit			-1,935	-1,27			1,337	0,99
OD*Ausführlichkeit			-0,374	0,24			-0,188	-0,15
Alter	0,007	0,35	0,008	0,40	-0,023	-1,46	-0,023	-1,49
Frau	-0,008	-0,02	-0,071	-0,17	-0,025	-0,07	0,007	0,02
Konstante	-2,65		-3,280		4,361		4,809	
Modellanpassung	<i>n</i> = 255		<i>n</i> = 255		<i>n</i> = 223		<i>n</i> = 223	
	LL = -90,002		LL = -89,166		F(5,217) = 1,63		F(7,215) = 1,30	
	LR Chi ² = 0,65		LR Chi ² = 2,33		<i>p</i> = 0,153		<i>p</i> = 0,252	
	<i>p</i> = 0,986		<i>p</i> = 0,940		R ² = 0,03		R ² = 0,04	
	Pseudo R ² = 0,004		Pseudo R ² = 0,013					

Logistische und OLS-Regressionen. Abhängige Variablen: Vorhandensein Verpflichtungsgefühl (Modelle 1 und 2), Intensität Verpflichtungsgefühl (Modelle 3 und 4). * $p \leq 0,05$; ⁺ $p \leq 0,10$.

2 Interaktionen zwischen den Kontrollvariablen Alter und Geschlecht mit den Prädiktoren wurden im Rahmen zusätzlicher multivariater Analysen ausgeschlossen. Da solche Interaktionen nicht bestanden, werden diese Modelle im vorliegenden Zusammenhang nicht gesondert ausgewiesen. Konfundierungen zwischen dem Alter der Befragten und der Ausführlichkeit von Vignetten aufgrund von möglicherweise altersbedingt eingeschränkten Kompetenzen wurden neben der Randomisierung auch bereits dadurch vermieden, dass die Grundgesamtheit auf bis zu 65-jährige Befragte limitiert wurde (vgl. Abschnitt 3.2).

Experiment II: Instruktion zur Normalität im Alltag

Auch für das Experiment, das sich inhaltlich auf die Normalität im Alltag bezog, wurden direkte Einflüsse der unabhängigen Variablen der Ausführlichkeit der Vignettenformulierung, der Self- und Other-Deception, sowie der Kontrollvariablen Alter und Geschlecht auf das Vorhandensein und die Intensität des Verpflichtungsgefühls anhand von Regressionsanalysen berechnet. Die Ergebnisse dieser Analysen sind in Tabelle 7 berichtet, wobei erneut für das Vorhandensein des Verpflichtungsgefühls eine logistische Regression, für die Intensität des Verpflichtungsgefühls eine OLS-Regression berechnet wurde.

Tabelle 7: Einflüsse der Ausführlichkeit der Vignette und der Tendenz zu sozial erwünschtem Antwortverhalten auf das Vorhandensein und die Intensität des Verpflichtungsgefühls

Experiment II Normalität im Alltag	Vorhandensein Verpflichtungsgefühl (Logistische Regression)				Intensität Verpflichtungsgefühl (OLS-Regression)			
	b	z	b	z	b	t	b	t
Ausführlichkeit	-0,015	-0,04	1,712	1,49	-0,634*	-2,18	-0,974	-1,10
Self-Deception	0,100	-0,83	0,198	1,08	0,002	0,02	-0,036	0,25
Other-Deception	-0,063	-0,57	0,099	0,55	-0,132	-1,43	-0,205	-1,54
SD*Ausführlichkeit			-1,064	-0,73			-0,414	-0,35
OD*Ausführlichkeit			-1,690	-1,22			0,848	0,77
Alter	-0,010	-0,60	-0,007	-0,41	-0,004	-0,31	-0,005	-0,35
Frau	-0,079	-0,21	-0,062	-0,17	0,172	0,55	0,169	0,54
Konstante	-1,504		-2,714		5,444		5,662	
Modellanpassung	<i>n</i> = 283		<i>n</i> = 283		<i>n</i> = 242		<i>n</i> = 242	
	LL = -105,188		LL = -105,034		F(5,236) = 1,40		F(7,234) = 1,08	
	LR Chi ² = 1,41		LR Chi ² = 4,24		<i>p</i> = 0,224		<i>p</i> = 0,375	
	<i>p</i> = 0,923		<i>p</i> = 0,751		R ² = 0,029		R ² = 0,031	
	Pseudo R ² = 0,007		Pseudo R ² = 0,020					

Logistische und OLS-Regressionen. Abhängige Variablen: Vorhandensein Verpflichtungsgefühl (Modelle 1 und 2), Intensität Verpflichtungsgefühl (Modelle 3 und 4). * $p \leq 0,05$; + $p \leq 0,10$.

Die Ergebnisse der logistischen Regression des Vorhandenseins des Verpflichtungsgefühls auf das Alter oder Geschlecht der Befragten, die Tendenz zu sozial erwünschtem Antwortverhalten und die Ausführlichkeit der Vignette zeigen, dass keine Einflüsse der Prädiktoren auf das Vorhandensein des Verpflichtungsgefühls bestehen. Demgegenüber deutet sich ein negativer Einfluss der Ausführlichkeit der Vignettenformulierung auf die Intensität des Verpflichtungsgefühls an, wenn dieser unter Kontrolle der Tendenz zur Self- und Other-Deception und der Kovariaten Alter und Geschlecht berechnet wird: Ebenso wie bereits die bivariaten Analysen gezeigt hatten (vgl. Abschnitt 4.1), wird eine geringere Intensität des Verpflichtungsgefühls berichtet, wenn die Vignette in einer ausführlichen Form dargeboten wird. Allerdings bleiben auch in diesem Experiment die Interaktionsterme zwischen der Ausführlichkeit und der Self-Deception sowie zwischen der Ausführlichkeit und der Other-Deception statistisch unbedeutend.

Die Daten dieses Experiments lassen demnach Hinweise auf die bedingte Gültigkeit der Hypothese I erkennen. Hypothese II kann aber auch hier nicht gestützt werden. Anders als in dem Experiment, das sich auf das Erleben einer Situation im Alltag bezieht, ist es also in diesem Experiment, in dem es um die Normalität des Alltags geht, für die Antworten der Befragten tendenziell von Bedeutung, in welcher Form eine Vignette präsentiert wird. Dieses Ergebnis befindet sich im Einklang mit der Idee der Anwendung von Vignetten: Als Vorzug von Vignetten wurde angeführt, dass Vignetten aufgrund ihrer Ausführlichkeit weniger anfällig für normativ geprägte Antworten sind. Betrachtet man eine höhere Intensität

des Verpflichtungsgefühls im vorliegenden Zusammenhang als Hinweis auf ein normativ geprägtes Antwortverhalten, so scheint eine ausführliche Vignette weniger Anlass dazu zu bieten: Dort wird von den Befragten ein durchschnittlich geringeres Verpflichtungsgefühl berichtet. Das Ausmaß der individuellen Tendenz zu sozial erwünschtem Antwortverhalten ist gleichzeitig nicht von direktem Belang für die Intensität des Verpflichtungsgefühls. Jedoch verliert der Haupteffekt der Ausführlichkeit im Interaktionseffektmodell seine statistische Bedeutsamkeit. Das heißt, es liegt zwar ein gewisses Maß an gemeinsamer Varianz zwischen der Ausführlichkeit und den Skalen zur individuellen Tendenz, sozial erwünscht zu antworten, vor, diese ist jedoch nicht ausreichend, um einen eigenständigen Effekt der Interaktion zu begründen.

Die Ergebnisse der empirischen Analyse der Untersuchungshypothesen sind insofern interessant, als die hier berichteten Split-Ballot Experimente gezeigt haben, dass es weitestgehend unerheblich ist, ob Vignetten in einer mehr oder weniger ausführlichen Form präsentiert werden. Allenfalls in einem Experiment, in dem die Vignetten selbst einen Normbezug aufweisen, kommen die vielfach behaupteten Vorzüge von Vignetten andeutungsweise zum Tragen. Interessant ist außerdem, dass im vorliegenden Zusammenhang ein schwacher Einfluss der Neigung, sich in vorteilhafter Weise gegenüber anderen darzustellen, nur im Experiment zum Erleben im Alltag beobachtet werden konnte.

5 Diskussion

In der vorliegenden Studie wurde der Frage nachgegangen, ob Vignetten als Technik der indirekten Erfassung von Einstellungen, insbesondere zu heiklen Themen, aufgrund ihrer Ausführlichkeit weniger normativ geprägte Antworten generieren. Hintergrund ist die Argumentation, dass Vignetten gerade aufgrund der größeren Realitätsnähe weniger sozial erwünschtes Antwortverhalten stimulieren. Diese Vermutung ist jedoch bislang nur sehr unzureichend empirisch belegt und theoretisch kaum untermauert.

Den Ausgangspunkt der vorliegenden Studie bildete die Überlegung, dass die größere Realitätsnähe von Vignetten durch deren Ausführlichkeit erzielt wird. Es wird vermutet, dass eine ausführlichere Vignette besser geeignet ist, die geschilderte Situation im Rahmen eines Surveys erfahrbar bzw. erlebbar zu machen, als eine nicht ausführliche Vignette. Daraus sollten Urteile, die auf der Basis einer ausführlichen Vignette gebildet werden, näher am Erleben der geschilderten Situation in alltäglichen Handlungsvollzügen und in diesem Sinne realitätsnäher bzw. weniger durch soziale Erwünschtheit und normativ geprägt sein. Eine Antwort gerade bei sensiblen Fragen, wie der bei prosozialem Verhalten, sollte durchschnittlich weniger zustimmend sein. Ist die Vignette weniger ausführlich formuliert, muss der Proband die Situation selbst stärker ausdeuten und greift hierfür möglicherweise eher auf bekannte Interpretationsmuster, also etwa auf Normen, zurück. Die Folge könnte eine stärkere Zustimmung zu prosozialem Verhalten sein. Zur Kontrolle wurde weiterhin vermutet, dass die Zustimmung zusätzlich von der individuellen Neigung, sozial erwünscht zu antworten, abhängig ist. Hier wurde ein eigenständiger Haupteffekt ebenso angenommen, wie eine Wechselwirkung mit der Ausführlichkeit der Formulierung. Unter der Bedingung einer starken Neigung zu sozial erwünschtem Antwortverhalten müsste der negative Zusammenhang zwischen der Ausführlichkeit der Vignette und dem berichteten Verpflichtungsgefühl sowie dessen Intensität noch stärker ausfallen und das Verpflichtungsgefühl bei einer geringen Ausführlichkeit noch höher sein, so die Annahme.

Diese Annahmen wurden mithilfe zweier Split-Ballot Experimente am Beispiel einer Situation an der Supermarktkasse und auf der Basis in Leipzig erhobener Daten vorläufig geprüft. Dabei wurden die Probanden zufällig einer ausführlicheren und einer weniger ausführlichen Vignette zugewiesen. Darüber hinaus fokussierte ein Experiment das Erleben im All-

tag und das andere die Normalität des Alltags, um die vermuteten Zusammenhänge stärker abzusichern. Hier stellt die Situation mit der Instruktion zur Normalität eine strengere Testsituation dar, weil normative Antworttendenzen bereits implizit in der Instruktion enthalten sind. Es zeigte sich, dass die Ausführlichkeit der Vignettenformulierung weitestgehend unbedeutend für das Ausmaß der Zustimmung zu prosozialem Verhalten ist. Eine hypothesenkonforme Tendenz zeigt sich nur für die Intensität des Verpflichtungsgefühls in dem Experiment, das auf die Normalität des Alltags zielte. Dort fand eine ausführliche Vignettenformulierung weniger Zustimmung, was als weniger normativ geprägte Antwort interpretiert werden kann. Es hat sich also gezeigt, dass ein Einfluss der Ausführlichkeit der Vignettenformulierung in dem Experiment sichtbar wird, in dem aufgrund der Instruktion die strengere Testsituation hergestellt wird, nicht jedoch in dem Experiment, dessen Instruktion unmittelbar das Erleben im Alltag fokussiert. Mit einer höheren Neigung, sozial erwünscht zu antworten, geht zudem ein stärker empfundenes Verpflichtungsgefühl im Experiment zum Erleben der Situation im Alltag einher. Hier ist die Skala zur sozialen Erwünschtheit sogar der stärkste Prädiktor der berichteten Intensität, was die Hypothese stützt. Die Ergebnisse sprechen jedoch nicht für die vermutete Interaktion zwischen der Neigung zu sozial erwünschtem Antwortverhalten und der Ausführlichkeit der Vignetten. Die Unterschiede im Zusammenhang der Ausführlichkeit und der Zustimmung zwischen Personen mit einer starken und einer schwachen Neigung zu sozialer Erwünschtheit sind weder statistisch noch substanzial bedeutsam. Die Hypothese kann demnach mit den vorliegenden Daten nicht bestätigt werden.

Für die Anwendung von Vignetten scheint sich daraus ableiten zu lassen, dass die Ausführlichkeit der Vignettenformulierung allein noch kein ausschlaggebendes Kriterium für die Validität einer indirekten Einstellungsmessung sein kann. Vielmehr sprechen die Ergebnisse für die Überlegung, dass weniger die Ausführlichkeit als vielmehr die Instruktion ausschlaggebend für die Validität der Daten ist. Insbesondere der Befund, dass sich der Einfluss der Ausführlichkeit nur in der strengeren Testsituation findet, in der die Instruktion auf die Normalität des Alltags bezogen ist, nicht aber in der Testsituation, die sich auf das Erleben im Alltag bezieht, ist hier von Bedeutung: Durch eine Instruktion, das Geschilderte zu *erleben* (Experiment I), wird ein Abrufen überwiegend normativer Interpretationsmuster scheinbar verhindert, während es durch den normalen *Alltagsbezug* (Experiment II) bei gleichzeitig wenig ausführlicher Vignettenformulierung gefördert wird. Eine rein informationstheoretische Interpretation dieser Ergebnisse führt zu dem Schluss, dass die Ausführlichkeit der Vignetten nur in Situationen mit Normbezug bedeutsam wird, nicht jedoch in Situationen, die ein Erleben simulieren sollen. Für diese Überlegung spricht auch, dass im Experiment mit der Instruktion zum Erleben die Neigung zu sozial erwünschtem Antwortverhalten am stärksten mit der Intensität des berichteten Verpflichtungsgefühls zusammenhängt. Gerade in einer aktiven Erlebensinstruktion, die weniger normative Interpretationsmuster aktivieren soll, ist demnach die individuelle Neigung zu sozialer Erwünschtheit deutlich wichtiger als die Ausführlichkeit der Vignettenformulierung. Hinsichtlich der Frage, ob die Ausführlichkeit einer Vignette oder aber die Instruktion, mit der sie zur Beantwortung vorgegeben wird, für die Realitätsnähe der Antworten maßgeblich ist, kann die vorliegende Studie keine zufriedenstellenden Antworten geben; vielmehr sind hier weitere Untersuchungen erforderlich.

Damit scheint die indirekte Messung mithilfe von Vignetten weitestgehend robust gegenüber der Ausführlichkeit in der Vignettenformulierung zu sein. Da es sich um die Messung normativer Einstellungen anhand von fiktiven Situationsbeschreibungen handelt, sollte jedoch der Formulierung der Instruktion, wie die Situation wahrgenommen werden soll, ebenfalls sorgfältig überlegt sein. Schließlich lässt sich die Robustheit und damit die Validität der Messung weiterhin absichern, wenn zusätzlich die individuelle Neigung, sozial erwünscht zu antworten, kontrolliert wird. Techniken der Formulierung von Szenarien in Vignetten er-

weisen sich damit erneut als grundlegend fruchtbar für die indirekte Erfassung von Einstellungen und Normen.

Es ist aber auch denkbar, dass die Vermutung schon theoretisch nicht erschöpfend ist. Rekurriert man nochmals auf das Modell des Antwortprozesses, ist beispielsweise auch denkbar, dass die Differenz in der Information gar nicht so deutlich durch den Probanden wahrgenommen wird, oder dass tatsächlich doch noch andere als normativ eingefärbte Informationen aktiviert werden, wenn die Vignette kurz formuliert ist. Darüber hinaus liegen Studien vor, die Hinweise darauf liefern, dass ausführlichere Informationen nicht notwendigerweise in die Antwortentscheidung einfließen (Krosnick / Presser 2010). Gerade vor dem Hintergrund der Erhebung im Rahmen von CATI-Interviews in dieser Untersuchung ist auch plausibel, dass zu hohe Anforderungen an die Konzentrations- und Memorierungsfähigkeit der Probanden gestellt worden sind. So wäre denkbar, dass die Zuordnung der Antwort auf der Skala vom „wahren“ Urteil abweicht. Hierin liegt weiterer Forschungsbedarf, und es sind zukünftig differenziertere Untersuchungen notwendig.

Die Reichweite der Ergebnisse ist zudem im Hinblick auf methodische Aspekte gewissen Einschränkungen unterworfen, die bei der Interpretation berücksichtigt werden sollten. Zwar kann aufgrund des Experimentalcharakters und der Zufallsstichprobe davon ausgegangen werden, dass die Ergebnisse nicht durch übermäßige Selbstselektionsprozesse verzerrt sind, sodass die Befunde weitestgehend gesichert sind. Allerdings sind weder das Alter noch das Geschlecht über die Gruppen gleich verteilt, was eine nicht perfekte Randomisierung nahelegt. Auch eine Kontrolle nach der Bildung ist aufgrund fehlender Daten nicht möglich. Weiterhin kann ein systematischer Ausfall in der Antwortrate der kontaktierten Haushalte nicht ausgeschlossen werden. Zudem mussten in einigen Stadtteilen Telefonnummern nachträglich recherchiert werden. Allerdings ist sichergestellt worden, dass die Probanden gleichmäßig in allen Leipziger Stadtteilen rekrutiert worden sind, sodass unter Vorbehalt auf die Bewohner der Stadt Leipzig geschlossen werden kann. Gleichwohl schränkt die Stichprobe den Geltungsbereich streng genommen weiter auf die Stadt Leipzig ein. Ein zukünftiges Desiderat liegt demnach in der Anwendung des Designs auf andere regionale Stichproben bzw. deren Vergleich. Das würde auch das Problem lösen, dass die statistische Power in dieser Untersuchung vergleichsweise gering ist.

Des Weiteren liegen mögliche Fehlerquellen in der Konstruktion des Erhebungsinstruments selbst. Hier sollten zukünftige Studien anknüpfen. Erstens ist es möglich, dass die Variation der Ausführlichkeit nicht ausreichend war. Möglicherweise führen deutlichere Differenzen in der Ausführlichkeit der Formulierung auch zu deutlicheren Ergebnissen im Antwortverhalten. Zweitens könnte neben der Ausführlichkeit der Modus der Präsentation entscheidend sein. Die Daten wurden im Rahmen einer CATI-Befragung erhoben, die vielleicht weniger geeignet zur Präsentation von Vignetten ist, da die Vignetten den Probanden vorgelesen werden, was einen vergleichsweise hohen kognitiven Aufwand für diese bedeutet. Eine systematische Untersuchung des Präsentationsmodus wäre zukünftig wünschenswert. Drittens ist möglicherweise das gewählte Beispiel des prosozialen Verhaltens im Supermarkt nicht heikel bzw. sensibel genug. Zu prüfen wäre, ob sich das Ergebnis etwa auch bei heikleren Themen wie Drogenmissbrauch oder Sexualverhalten replizieren lässt. Schließlich wäre es viertens möglicherweise sinnvoll, eigenständige Haupteffekte der Instruktionsvariation nach Normalität und Erleben auf das berichtete Verpflichtungsgefühl und dessen Intensität anzunehmen. In der vorliegenden Untersuchung sind die Instruktionen als Kontexte der Antwortentscheidung aufgefasst worden. An dieser Stelle wäre dann die Konzeption einer weiteren experimentellen Variation zusätzlich zur Ausführlichkeit möglich.

Gerade die Instruktion hinsichtlich der Wahrnehmung der im Experiment formulierten Situation scheint demnach eine erhebliche Bedeutung für die Validität der Messung zu haben.

Dieser Aspekt sollte demnach nicht nur gewissenhaft in entsprechenden experimentellen Untersuchungen berücksichtigt werden, sondern auch in zukünftigen Methodenstudien, um weitere, vertiefte Einblicke in die Wirkungen der Gestaltung von Vignetten auf die Antwortvalidität zu erhalten.

Literatur

- Abelson, R.P. (1976): Script Processing in Attitude Formation and Decision Making, in: J. S. Carroll / J. W. Payne (Hrsg.), *Cognition and Social Behavior*, Hillsdale / MI, S. 33-67.
- Abraham, M. / C. Meyer (2012): Bedingungen des Erfahrungsaustausches in wirtschaftlichen Transaktionen: Das Beispiel der Reputation von Weiterbildungsanbietern, in: *Zeitschrift für Soziologie* 41, S. 166-181.
- Alexander, C.S. / H.I. Becker (1978): The Use of Vignettes in Survey Research, in: *Public Opinion Quarterly* 42, S. 93-104.
- Auspurg, K. / A. Jäckle (2012): First Equals Most Important? Order Effects in Vignette-Based Measurement, ISER Working Paper 2012-01.
- Auspurg, K. / T. Hinz / S. Liebig / C. Sauer (2010): Wie unplausibel darf es sein? Der Einfluss von Designmerkmalen auf das Antwortverhalten in Faktoriellen Surveys, in: H.-G. Soeffner (Hrsg.), *Unsichere Zeiten. Herausforderungen gesellschaftlicher Transformation*, Verhandlungen des 34. Kongresses der Deutschen Gesellschaft für Soziologie in Jena 2008, Wiesbaden, CD-ROM.
- Auspurg, K. / T. Hinz / S. Liebig (2009): Komplexität von Vignetten, Lerneffekte und Plausibilität im Faktoriellen Survey, in: *Methoden, Daten, Analysen* 3, S. 59-96.
- Armacost, R.L. / J.C. Hosseini / S.A. Morris / K.A. Rehbein (1991): An Empirical Comparison of Direct Questioning, Scenario, and Randomized Response Methods for Obtaining Sensitive Business Information, in: *Decision Sciences* 22, S. 1073-1090.
- Bartlett, F.C. (1932): *Remembering*, Cambridge.
- Beck, M. / K.-D. Opp (2001): Der faktorielle Survey und die Messung von Normen, in: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 53: S. 283-306.
- Benson, L.E. (1941): Studies in Secret-Ballot Technique, in: *Public Opinion Quarterly* 5: S. 97-82.
- Bierhoff, H.-W. (2002): *Prosocial Behavior*, London.
- Bierhoff, H.-W. (2010): *Psychologie prosozialen Verhaltens*, Stuttgart.
- Bradburn, N.M. / S. Sudman / B. Wansink (2004): *Asking Questions: The Definitive Guide to Questionnaire Design – for Market Research, Political Polls, and Social and Health Questionnaires*, San Francisco / CA.
- Burstin, K. / E. Doughtie / A. Raphaeli (1980): Contrastive Vignette Technique: An Indirect Methodology Designed to Address Reactive Social Attitude Measurement, in: *Journal of Applied Social Psychology* 10, S. 147-165.
- Cantril, H. (1944): *Gauging Public Opinion*, Princeton / NJ.
- Campbell, D.T. (1950): The Indirect Assessment of Social Attitude, in: *Psychological Bulletin* 47, S. 15-38.
- Crowne, D.P. / D. Marlowe (1964): The Approval Motive. Studies in Evaluative Dependence, New York / NY.
- Edwards, A.E. (1957): *Techniques of Attitude Scale Construction*, New York / NY.
- Edwards, A.E. (1957): The Social Desirability Variable in Personality Assessment and Research, New York / NY.
- Eifler, S. (2007): Evaluating the Validity of Self-Reported Deviant Behavior Using Vignette Analyses, in: *Quality & Quantity* 41, S. 303-318.

- Eifler, S. (2010): Validity of a Factorial Survey Approach to the Analysis of Criminal Behavior, in: *Methodology* 6, S. 139-146.
- Finch, J. (1987): The Vignette Technique in Survey Research, in: *Sociology* 21, S. 105-114.
- Groß, J. / C. Börensen (2009): Wie valide sind Verhaltensmessungen mittels Vignetten? Ein methodischer Vergleich von faktoriellem Survey und Verhaltensbeobachtung, in: P. Kriwy / C. Gross (Hrsg.), *Klein aber Fein! Quantitative empirische Sozialforschung mit kleinen Fallzahlen*, Wiesbaden, S. 149-178.
- Groves, R.M. (1989): *Survey Errors and Survey Costs. An Introduction to Survey Error*, New York / NY.
- Hegtvedt, K.A. (1988): Social Determinants of Perception: Power, Equity, and Status Effects in an Exchange Situation, in: *Social Psychology Quarterly* 51, S. 141-153.
- Hegtvedt, K. A. (1990): The Effects of Relationship Structure on Emotional Responses to Inequity, in: *Social Psychology Quarterly* 53, S. 214-228.
- Hughes, R. (1998): Considering the Vignette Technique and its Application to a Study of Drug-Injecting and HIV Risk and Safer Behavior, in: *Sociology of Health and Illness* 20, S. 381-400.
- Kerlinger, F.N. (1986): *Foundations of Behavioral Research*, 3. Aufl., New York / NY.
- Klepper, S. / D. Nagin (1989): Tax Compliance and Perceptions of the Risks of Detection and Criminal Prosecution, in: *Law & Society Review* 23, S. 209-240.
- Krosnick, J.A. / S. Presser (2010): Question and Questionnaire Design, in: P. V. Marsden / J. D. Wright (Hrsg.), *Handbook of Survey Research*, West Yorkshire, S. 263-313.
- Krumpal, I. (2013): Determinants of Social Desirability Bias in Sensitive Surveys: A Literature Review, in: *Quality and Quantity* 47, S. 2025-2047.
- Lee, R.M. (1993): *Doing Research on Sensitive Topics*, London.
- Markus, H.R. (1977): Self-Schemata and Processing Information about the Self, in: *Journal of Personality and Social Psychology* 35, S. 63-78.
- Musch, J. / R. Brockhaus / A. Bröder (2002): Ein Inventar zur Erfassung von zwei Faktoren sozialer Erwünschtheit, in: *Diagnostica* 48, S. 121-129.
- Nagin, D.S. / R. Paternoster (1994): Personal Capital and Social Control: The Deterrence Implications of a Theory of Individual Differences in Criminal Offending, in: *Criminology* 32, S. 581-606.
- Neff, J. Alan (1979) Interactional versus Hypothetical Others: The Use of Vignettes in Attitude Research, in: *Sociology & Social Research* 64, S. 105-125.
- Nisic, N. / K. Auspurg (2009): Faktorieller Survey und klassische Bevölkerungsumfrage im Vergleich – Validität, Grenzen und Möglichkeiten beider Ansätze, in: P. Kriwy / C. Gross (Hrsg.), *Klein aber fein! Quantitative Sozialforschung mit kleinen Fallzahlen. Forschung und Entwicklung in der Analytischen Soziologie*, Wiesbaden, S. 211-245.
- Paulhus, D.L. (1986): Self-Deception and Impression Management in Test Responses, in: A. Angleitner / J. S. Wiggins (Hrsg.), *Personality Assessment via Questionnaires. Current Issues in Theory and Measurement*, Berlin, S. 143-165.
- Paulhus, D. L. (1991): Measurement and Control of Response Bias, in: J. P. Robinson / P. R. Shaver / L. S. Wrightsman (Hrsg.), *Measures of Personality and Social Psychological Attitudes*, San Diego / CA, S. 17-69.
- Pedhazur, E.J. (1969): Pseudoprogressivism and Assessment of Teacher Behavior, in: *Educational and Psychological Measurement* 29, S. 377-386.
- Petersen, T. (2002): *Das Feldexperiment in der Umfrageforschung*, Frankfurt / Main.
- Rossi, P.H. (1979): Vignette Analysis: Uncovering the Normative Structure of Complex Judgements, in: R.K. Merton / J. S. Coleman / P. H. Rossi (Hrsg.), *Qualitative and Quantitative Social Research. Papers in Honor of Paul F. Lazarsfeld*, New York / NY.

- Rossi, P.H. / A.B. Anderson (1982): The Factorial Survey Approach: An Introduction, in: P. H. Rossi / S. L. Nock (Hrsg.), *Measuring Social Judgments. The Factorial Survey Approach*, Beverly Hills / CA, S. 15-67.
- Sauer, C. / K. Auspurg / T. Hinz / S. Liebig (2011): The Application of Factorial Survey in General Population Samples: The Effect of Respondent Age and Education on Response Times and Response Consistency, in: *Survey Research Methods* 5, S. 89-102.
- Schank, R. / R.P. Abelson (1977): *Scripts, Plans, Goals and Understanding*, Hillsdale / MI.
- Schuman, H. / S. Presser / J. Ludwig (1981): Context Effects on Survey Responses to Questions about Abortion, in: *Public Opinion Quarterly* 45, S. 216-223.
- Schwarz, N. / S. Sudman (1992): Context Effects in Social and Psychological Research, New York / NY.
- Sieber, J.E. / B. Stanley (1988): Ethical and Professional Dimensions of Socially Sensitive Research, in: *American Psychologist* 43, S. 49-55.
- Stadt Leipzig (2013): *Statistisches Jahrbuch 2013*. <http://statistik.leipzig.de/statpubl/index.aspx?cat=1&rub=1>, letztes Abrufdatum: 15.1.2014.
- Stocké, V. (2004): Entstehungsbedingungen von Antwortverzerrungen durch soziale Erwünschtheit. Ein Vergleich der Prognosen der Rational-Choice-Theorie und des Modells der Frame-Selektion, in: *Zeitschrift für Soziologie* 33, S. 303-320.
- Sudman, S. / N.M. Bradburn / N. Schwarz (1996): *Thinking About Answers. The Application of Cognitive Processes to Survey Methodology*, New York / NY.
- Tourangeau, R. (1984): Cognitive Science and Survey Methods, in: T. B. Jabine / M. L. Straf / J. M. Tanur / R. Tourangeau (Hrsg.), *Cognitive Aspects of Survey Methodology. Building a Bridge between Disciplines*, Washington / DC, S. 73-100.
- Tourangeau, R. / L.J. Rips / K. Rasinski (2000): *The Psychology of Survey Response*, Cambridge.
- Tourangeau, R. / T. Yan (2007): Sensitive Questions in Surveys, in: *Psychological Bulletin* 133, S. 859-883.
- Wallander, L. (2009): 25 Years of Factorial Survey in Sociology: A Review, in: *Social Science Research* 38, S. 505-520.

Prof. Dr. Stefanie Eifler
 Dr. Knut Petzold
 Katholische Universität Eichstätt Ingolstadt
 Lehrstuhl für Soziologie und empirische Sozialforschung
 Ostenstr. 26
 85072 Eichstätt
 stefanie.eifler@ku.de
 knut.petzold@ku.de

Anhang

Tabelle A-1: Vergleich von realisierter Stichprobe und Grundgesamtheit nach Geschlecht und Alter

	Realisierte Stichprobe		Amtliche Statistik*
	n	%	%
18-29 Jahre	31	5,54	27,41
30-41 Jahre	67	11,96	25,09
42-53 Jahre	195	34,82	26,28
54-65 Jahre	267	47,68	21,22
Gesamt	560	100	100,00
männlich	217	38,75	50,28
weiblich	343	61,25	49,72
Gesamt	560	100	100,00

* Quelle: Stadt Leipzig (2013)

Tabelle A-2: Bivariate Zusammenhänge zwischen den Items zur Erfassung von Self-Deception und Other-Deception

Self-Deception	SD-1	SD-2	SD-3	SD-4
SD-1	1			
SD-2	0,347	1		
SD-3	0,041	0,029	1	
SD-4	0,172	0,168	0,366	1
Other-Deception	OD-1	OD-2	OD-3	OD-4
OD-1	1			
OD-2	0,087	1		
OD-3	0,064	0,005	1	
OD-4	0,133	0,087	0,249	1

Alle Korrelationen sind Spearmans Rangkorrelationen.