

# Leugnen oder gestehen? Bestimmungsfaktoren wahrer Antworten in Befragungen

## To Deny or to Confess? Determinants of Truthful Answers in Surveys

Anna Skarbek-Kozietulska, Peter Preisendörfer, Felix Wolter\*

Johannes Gutenberg-Universität Mainz, Institut für Soziologie, Colonel-Kleinmann-Weg 2, 55099 Mainz, Germany  
skarbak@uni-mainz.de; preisendoerfer@uni-mainz.de; felix.wolter@uni-mainz.de

**Zusammenfassung:** Der Artikel berichtet die Ergebnisse einer Validierungsstudie, in der es um das Problem wahrer Antworten in Umfragen geht. Auf der Grundlage einer schriftlichen Befragung von rund 300 Personen, die in den zurückliegenden Jahren eine oder mehrere kleinere Straftaten begangen hatten und dafür strafrechtlich verurteilt wurden, wird untersucht, wie häufig die Delinquenten in der Umfrage ihre strafrechtliche Verurteilung zugeben und von welchen Faktoren die Tendenz zum Leugnen bzw. Gestehen abhängt. Der Satz der im Einzelnen überprüften Einflussfaktoren wird aus verschiedenen Varianten der Rational-Choice-Theorie hergeleitet. Ausgewählte Ergebnisse: Insgesamt gestehen knapp zwei Drittel der Befragten ihre strafrechtliche Verurteilung. Frauen leugnen tendenziell häufiger als Männer, ältere Personen häufiger als jüngere und höher gebildete Befragte häufiger als niedriger gebildete. Die theoretisch bedeutsamen Variablen der Neigung zu sozial erwünschten Antworten (SD-Neigung) und der Erwünschtheitswahrnehmungen (SD-Belief) zeigen signifikante Effekte auf die Wahrscheinlichkeit einer wahren Antwort. Besondere Beachtung verdient der Befund, dass diejenigen, die den Fragebogen relativ spät bzw. erst nach wiederholter Aufforderung zurückschicken, eindeutig seltener wahrheitsgemäß antworten.

**Schlagnote:** Rational-Choice-Theorie; SEU-Theorie; Framing-Modelle; Validierungsstudie; Validität von Umfragen; schriftliche Befragung; soziale Erwünschtheit; Umfrageeinstellung; Umfrageerfahrung; Antwortlatenz.

**Summary:** This article presents results of a validation study dealing with the problem of truthful answers in surveys. Based on a mail survey of about 300 respondents who had committed one or more minor criminal offences in recent years and had been convicted of them in court we examine how many respondents include a confession in the questionnaire and what factors influence the tendency to give truthful answers. The set of influential factors which is analyzed is derived from different versions of rational choice theory. Selected findings: About two thirds of the total sample contain elements of confession. Women tend to deny their deviant behavior more often than men, older people more often than younger ones, and more highly educated respondents more frequently than less educated ones. The theoretically relevant variables "need for social approval" and "trait desirability" show significant effects on the probability of telling the truth. Particular attention is owed to the finding that those respondents who answered with a time delay, that is, send back the questionnaire late, clearly gave less valid answers.

**Keywords:** Rational Choice-Theory; SEU-Theory; Framing Models; Validation Study; Validity of Surveys; Mail Survey; Social Desirability; Survey Attitude; Survey Experience; Response Latency.

### 1. Einleitung

Ob und inwieweit Personen im Rahmen von Befragungen wahre Antworten geben, ist ein oder viel-

leicht sogar *das* Kernproblem der Survey-Methodologie. In zugespitzter Form stellt sich dieses Problem bei so genannten heiklen Fragen, also bei Fragen, die mehr oder weniger tabuisierte persönliche Sachverhalte ansprechen. Aus der Perspektive der Methodenforschung eignen sich „sensitive questions“ deshalb in besonderem Maße dazu, die Möglichkeiten und Grenzen der Befragungsmethode auszuloten, die Gangbarkeit neuer Wege zu erproben und im günstigen Fall innovative Lösungen hervorzubringen (Lee 1993: Kap. 1).

Angesichts dieser Ausgangslage verwundert es nicht, dass sich die Umfrageforschung seit ihren Anfängen

---

\* Der Beitrag entstand im Rahmen des von der Deutschen Forschungsgemeinschaft DFG geförderten Forschungsprojekts „Asking Sensitive Questions“ (Förderkennzeichen PR 237/6, DFG-Schwerpunktprogramm „Survey Methodologie“). Zusätzliche Unterstützung kam von der inneruniversitären Forschungsförderung der Johannes Gutenberg-Universität Mainz. Für wertvolle Hinweise und Anregungen danken wir Andreas Diekmann, Ulf Liebe, zwei anonymen Gutachtern und den Herausgebern der ZfS.

in den 1930er und 1940er Jahren mit dem Problem heikler Fragen „herumschlägt“ (vgl. Benson 1941; Hyman 1944; Barton 1958). Und es verwundert weiterhin nicht, dass auch neuere Review-Artikel (wie z. B. Tourangeau & Yan 2007; Lensvelt-Mulders 2008) nach wie vor diagnostizieren, es sei schwierig, auf heikle Fragen wahre bzw. valide Antworten zu erhalten. Immerhin kann man inzwischen halbwegs verlässlich sagen, dass es bei heiklen Themen Modus-Effekte in der Form gibt, dass selbstadministrierte Befragungen im Unterschied zu Befragungen, bei denen ein Interviewer beteiligt ist, vergleichsweise validere Ergebnisse liefern (Richman et al. 1999; Tourangeau et al. 2000: Kap. 10; Bradburn et al. 2004: Kap. 3; vgl. aber auch Stocké 2004). Weitgehend unabhängig von den Befragungsmodi gibt es zudem ein beachtliches Arsenal an speziellen Techniken, die mit dem Anspruch antreten, Antwortbarrieren bei heiklen Fragen zu reduzieren, z. B. die Technik des vertraulichen Kuverts, die Randomized-Response-Technik oder Bogus-Pipeline-Techniken (für Übersichten Lee 1993: Kap. 5; Barnett 1998; Tourangeau et al. 2000: Kap. 9). Auf diese Techniken soll hier nicht eingegangen werden, aber wir schließen uns der Einschätzung der zitierten Übersichtsarbeiten an, dass das Potenzial dieser Techniken zur Hervorbringung wahrer Antworten oft überschätzt wird.

Der entscheidende Engpass in der Forschung zum Wahrheitsgehalt von Befragungen und speziell von Befragungen zu heiklen Themen sind nach wie vor so genannte Validierungsstudien. In solchen Studien ist der wahre Wert eines Merkmals auf der Aggregatebene oder – besser – auf der individuellen Ebene bekannt und wird mit dem per Befragung ermittelten Wert abgeglichen. Der vorliegende Beitrag berichtet die Ergebnisse einer solchen Validierungsstudie auf Individualebene. Er stützt sich auf eine postalische Befragung von 301 Personen, die in den zurückliegenden Jahren eine oder mehrere kleinere Straftaten begangen hatten und dafür strafrechtlich verurteilt wurden. Die spannende Ausgangsfrage lautet dann, wie häufig die Delinquenten in der Umfrage „gestehen“, d. h. ihre strafrechtliche Verurteilung tatsächlich zugeben. So ergeben sich Hinweise darauf, in welchem Ausmaß in Umfragen zu selbstberichteter Delinquenz die gewonnenen Prävalenzschätzungen verzerrt sind. Weiterhin kann und soll untersucht werden, von welchen Faktoren es abhängt, ob delinquentes Verhalten in der Befragungssituation wahrheitsgemäß angegeben wird. Diese Frage ist zum einen vor dem Hintergrund des Problems bedeutsam, dass auch Zusammenhangsanalysen zwischen einem interessierenden heiklen

Verhalten und unabhängigen Variablen verzerrt sein können, wenn das Misreporting mit den untersuchten Prädiktoren korreliert. Zum anderen geht es darum, im Idealfall zu einer theoretischen Erklärung des Befragtenverhaltens zu gelangen.

Der Beitrag beginnt mit etwas ausführlicheren theoretischen Überlegungen, die im Ergebnis auf einen Satz von Faktoren hinführen, deren Einfluss auf das Antwortverhalten untersucht werden soll (2.). Anschließend werden die empirische Datenbasis und die zentralen Variablen vorgestellt (3.). Die Ergebnisse werden im ersten Schritt in der Form bivariater Aufgliederungen präsentiert; im zweiten Schritt folgen multivariate Modelle in der Form binärer logistischer Regressionen (4.). Der Beitrag endet mit ein paar knappen Schlussbemerkungen, die nicht zuletzt auch Hinweise auf die allgemeinere Relevanz der Ergebnisse für die Theorie und Praxis der Umfrageforschung geben (5.).

## 2. Theoretische Überlegungen und Hypothesen

Generell kann man davon ausgehen, dass Personen im Setting wissenschaftlicher Umfragen in der Regel nicht erratisch antworten oder gleichsam aus Spaß an der Freude lügen. Sie geben vielmehr nur dann falsche, d. h. nicht der Wahrheit entsprechende Antworten, wenn sie dafür einen oder mehrere Gründe haben. Fragt man beispielsweise jemanden danach, ob er/sie aktuell eine gültige Bahncard hat, gibt es kaum Anlass zu vermuten, dass es Gründe für eine Falschauskunft gibt. Gewiss mag es Leute geben, die – aus welchen Gründen auch immer – das Gegenteil dessen sagen, was zutrifft, aber man kann unterstellen, dass es wenige sind und dass sich derartige individuelle Idiosynkrasien im Aggregat kompensieren. Der Erfahrungstatbestand des begründeten Lügens legt es nahe, das Antwortverhalten in Befragungen theoretisch als rationales Entscheidungsverhalten zu konzeptualisieren und mithin einer Rational-Choice-Theorie des Befragtenverhaltens zu folgen (grundlegend dazu im deutschen Kontext v. a. Esser 1986, 1991).

Bei heiklen Fragen wie z. B. nach kleineren kriminellen Delikten, um die es empirisch in diesem Beitrag geht, gibt es klare Belege dafür, dass Befragte lügen, und gleichzeitig gute Gründe für ein solches Antwortverhalten.<sup>1</sup> Das Standardproblem bei heik-

<sup>1</sup> Vgl. neben den bereits genannten Überblicksartikeln von Tourangeau & Yan 2007 und Lensvelt-Mulders 2008 zu Antwortverzerrungen speziell bei „Delinquenz-Fragen“

len Fragen ist das „underreporting“, was dazu führt, dass auf Befragungen basierende Prävalenzschätzungen in der Regel eindeutig zu niedrig liegen.<sup>2</sup> Ist das Ausmaß des Underreporting für verschiedene Subgruppen unterschiedlich, werden auch Zusammenhangsanalysen zwischen dem heiklen Sachverhalt und mutmaßlichen Kovariaten in ihrer Gültigkeit beeinträchtigt.

Die Frage nach Gründen und möglichen Bestimmungsfaktoren für das Underreporting bei heiklen Themen sei im Folgenden vor dem generellen Hintergrund einer RC-Theorie des Befragtenverhaltens in drei Schritten angegangen: 1) Antwortverhalten im Lichte einer rudimentären RC-Theorie, 2) Antwortverhalten im Lichte der SEU-Theorie, 3) Antwortverhalten im Lichte von Framing-Modellen. Dieses Vorgehen in drei Schritten bedeutet, dass zunehmend komplexere Varianten der RC-Theorie in die Überlegungen Eingang finden.

## 2.1 Antwortverhalten im Lichte einer rudimentären RC-Theorie

Grundlegend für alle Varianten und Anwendungen der RC-Theorie ist die Sichtweise, dass Akteure auf der Basis von Nutzen-Kosten-Erwägungen handeln, wobei dieses Handeln als Wahl zwischen zur Verfügung stehenden Alternativen betrachtet wird. Die Nutzen und Kosten können – je nach RC-Variante – objektiv oder subjektiv bestimmt und inhaltlich mehr oder weniger weit gespannt werden. Zum Teil wird versucht, die Nutzen und Kosten direkt zu erheben, zum Teil wird mittels Brückenhypothesen indirekt an vermutete Nutzen und Kosten angeknüpft. Weiterhin „boomen“ seit mehreren Jahren so genannte Framing-Modelle, die sich in der Regel als Ergänzungen und Fortführungen, also als elabo-

rierte RC-Modelle verstehen. In der einfachsten Form verwendet man die RC-Theorie lediglich als eine rudimentäre Heuristik, die zur Strukturierung der Überlegungen im jeweiligen Anwendungsfall beiträgt und im günstigen Fall zur Herleitung von Hypothesen führt.

Bezogen auf die Beantwortung heikler Fragen erfolgt demnach eine (in der Regel unbewusst und schnell ablaufende) Nutzen-Kosten-Abwägung einer wahren gegen eine falsche Antwort, d. h. „respondents weigh the risks and benefits of responding truthfully“ (Tourangeau et al. 2000: 14). Für Personen, die das heikle Verhalten bzw. die heikle Eigenschaft nicht teilen, erscheint es naheliegend, dass sie eine wahre Antwort geben, denn es gibt im Normalfall keine guten Gründe, ein unerwünschtes Verhalten, das man nicht gezeigt hat, zu gestehen. Mithin ist der Terminus „heikle Fragen“ ein wenig missverständlich, da es eher um heikle Antworten geht. Ein echtes Entscheidungsproblem liegt eigentlich nur für diejenigen vor, deren Antwort heikel wäre, wenn sie wahrheitsgemäß ausfiele. Für diese Befragten besteht ein Anreiz, das Verhalten zu leugnen. Im Fall unserer Validierungsstudie haben alle Befragten das heikle Attribut, eine Straftat begangen zu haben, denn alle wurden strafrechtlich verurteilt. Mithin haben alle Befragten ein „echtes“ Entscheidungsproblem und einen Anreiz zu lügen.<sup>3</sup>

Was aber sind bei dem gegebenen Entscheidungsproblem die relevanten Nutzen und Kosten einer wahren im Vergleich zu denen einer falschen Antwort? Eine Antwort auf diese Frage fällt, wie Diekmann (2007: 445) bemerkt, schwer und im Detail bleibt die Frage zumeist ungeklärt. Es erscheint plausibel zu argumentieren, dass man in der Situation einer schriftlichen Befragung und im Anwendungsfall kleinerer krimineller Delikte mit einer wahren Antwort, also mit dem Eingeständnis einer strafrechtlichen Verurteilung, wenig gewinnen kann. Es verschafft vielen (oder vielleicht sogar den meisten) Menschen zwar eine innere Befriedigung, die Wahrheit zu sagen, aber sehr viel mehr an Nutzen ist hier nicht erkennbar. Mit den Kosten ist es noch schwieriger: In der Interviewsituation ist kein Interviewer anwesend, vor dem man einen guten Eindruck machen will. Das sollte der Wahrheit förderlich sein. Andererseits bleibt zu beachten, dass

die Validierungsstudien von Junger 1989, Elffers et al. 1992, van der Heijden et al. 2000 sowie Köllisch & Oberwittler 2004.

<sup>2</sup> Es sei angemerkt, dass sich die Forschung zu heiklen Fragen – ausgehend vom klassischen Fall von Verhaltensfragen mit einer Tendenz zu Underreporting – im Zuge der Zeit in zwei Richtungen erweitert hat: zum einen in Richtung Verhaltensfragen auch mit einer Tendenz zu Overreporting (Wahlbeteiligung, Kirchgangshäufigkeit usw.), zum anderen in Richtung heikle Einstellungsfragen (Rechtsextremismus, Fremdenfeindlichkeit usw.). Während bei Verhaltensfragen im Prinzip eine Validierung möglich ist, ist dies bei Einstellungsfragen nicht mehr gegeben, so dass das Konzept des „wahren Wertes“ problematisch wird. Für die genuine Methodenforschung erscheinen aus diesem Grund Verhaltensfragen in der Regel besser geeignet als Einstellungsfragen.

<sup>3</sup> Stillschweigend wird bei dieser Argumentation und generell bei der Qualifizierung einer Frage als „heikel“ unterstellt, dass es über alle Befragten hinweg einheitliche Vorstellungen und Wahrnehmungen hinsichtlich dessen gibt, was heikel bzw. unerwünscht ist (mehr dazu unter dem Stichwort SD-Belief in Abschnitt 2.2).

es sich anonym womöglich leichter lügen lässt als gegenüber einer realen Person. Der kritische Kostenfaktor ist wohl, dass dann, wenn man gesteht, das strafrechtliche Vergehen bekannt werden könnte. Dies ist zwar unwahrscheinlich und es ist unklar, wer genau eventuell davon erfährt oder erfahren könnte, aber risikoaverse Akteure können mit einer „Minimax Regret“-Antwort auf Nummer sicher gehen. Ignoriert man die Feinheiten der Nutzen-Kosten-Analyse für die Einzelpersonen und nutzt die RC-Theorie nur als allgemeine Arbeitsgrundlage, kann man auf der Kostenseite die allgemeine Hypothese formulieren, dass diejenigen Befragten seltener die Wahrheit sagen, die „viel zu verlieren“ haben (Verlust-Hypothese).<sup>4</sup>

Dabei dürfte die Höhe des potenziellen Verlusts nicht zuletzt von der Art des begangenen Delikts abhängen. Ein Befragter, der wegen einer eher relativ gravierenden Straftat wie einer Körperverletzung belangt wurde, hat durch eine mögliche Aufdeckung seiner Antwort mehr zu verlieren als einer, der lediglich ein Allerwelts- oder „Kavaliersdelikt“ wie etwa Schwarzfahren mit öffentlichen Verkehrsmitteln begangen hat. Wenngleich nicht ausgeschlossen werden kann, dass ein sich eventuell zeigender Einfluss der Schwere des Vergehens das Ergebnis eines Selektionseffekts ist und mithin nicht kausal interpretiert werden kann, verdient dieser Faktor (zumindest als statistische Kontrollvariable) auf jeden Fall eine Berücksichtigung in unseren empirischen Analysen.

Die Neigung zum Leugnen/Lügen dürfte weiterhin in Abhängigkeit vom Bildungsniveau der Befragten variieren. Eine strafrechtliche Verurteilung kann gebildeten und gemeinhin gesellschaftlich besser etablierten Personen mehr schaden als Personen, die wenig zu verlieren haben. Wir werden diese Hypothese in unseren empirischen Analysen dergestalt testen, dass wir die Erwartung überprüfen, dass Personen mit höherer Bildung seltener die Wahrheit sagen. Für sie stellt das Bekennen delinquenten Verhaltens eine besondere Gefahr dar, sie können im Fall des Bekanntwerdens einer strafbaren Handlung „tiefer fallen“, und sie werden ihre Tat deshalb häufiger leugnen.

## 2.2 Antwortverhalten im Lichte der SEU-Theorie

Ein Großteil der Forschung zum Wahrheitsgehalt von Umfragen und speziell auch zur Validität von Antworten auf heikle Fragen dreht sich um das Konzept der sozialen Erwünschtheit und den so genannten SD-Bias (SD = social desirability). In der Diskussion darum, was überhaupt eine heikle Frage ist, sehen einige Autoren solche Fragen lediglich als einen Spezialfall des Problems sozialer Erwünschtheit (z. B. DeMaio 1984; Esser 1991), andere Autoren schließen sich dieser verengenden Sichtweise nicht an (z. B. Lee 1993: 3ff.; Tourangeau et al. 2000: 257ff.). Aber wie auch immer man das Verhältnis von sozialer Erwünschtheit und heiklen Fragen sieht, klar ist, dass es eine beträchtliche Überschneidung in den damit verknüpften Problemlagen gibt.

Für die Charakterisierung und Erklärung des SD-Bias hat Esser (1986, 1991) in zwei vielzitierten Artikeln die „subjective expected utility theory“ (SEU-Theorie) als theoretische Basis vorgeschlagen. Nach Esser (1991: 64f.) unterscheidet diese Theorie drei Schritte im Prozess der Handlungswahl: die Kognition der Situation, die Bewertung der Konsequenzen bestimmter Handlungen und die Selektion einer spezifischen Handlung nach einer bestimmten Regel. Die Evaluation der Handlungsalternativen erfolgt auf der Basis der eigenen Präferenzen und der geschätzten subjektiven Wahrscheinlichkeiten über die Folgen einer bestimmten Handlung. Es handelt sich dabei um die Kalkulation der SEU-Werte für die jeweiligen Handlungsalternativen. Im Ergebnis wird die Handlungsalternative mit dem höchsten SEU-Wert gewählt und so die subjektive Nutzenerwartung maximiert.

Antwortverzerrungen in Umfragen führt Esser (1991: 62) dann auf Effekte sozialer Erwünschtheit zurück, die aus einer Kombination von Motiven, Bedürfnissen und Bewertungen mit Erwartungen darüber erwachsen, welche Antwort mit den vorliegenden Motiven übereinstimmen wird. Ursachen des SD-Bias können zum einen internalisierte Normen und Rollenerwartungen sein, zum anderen Erwartungen gegenüber dem Interviewer, wobei die befragte Person zentral an einem guten Eindruck gegenüber dem Interviewer interessiert ist („impression management“). Den Anreiz, eine Antwort zu geben, die gesellschaftlichen Normen und Rollenerwartungen entspricht, bezeichnet Esser als „kulturelle soziale Erwünschtheit“, den Anreiz, gegenüber dem Interviewer ein gutes Bild abzugeben, als „situationale soziale Erwünschtheit“.

<sup>4</sup> Ein der Verlust-Hypothese weitgehend analoges Argument findet man in „Voting“-Studien, bei denen es um Overreporting der Teilnahme an Wahlen geht, z. B. bei Bernstein et al. (2001: 41; vgl. auch Silver et al. 1986): „The key to understanding why people overreport is that those who are under the most pressure to vote are the ones most likely to misrepresent their behavior when they fail to do so.“

Geleitet durch die theoretische Perspektive der SEU-Theorie, postuliert Esser drei zentrale Determinanten des SD-Bias (vgl. Stocké auch 2004): 1) die Anonymität der Antwortsituation, 2) die SD-Neigung der befragten Person, d. h. ihr Bedürfnis nach sozialer Anerkennung, und 3) den SD-Belief der befragten Person, d. h. ihre Vorstellungen darüber, ob und in welchem Ausmaß ein bestimmtes Verhalten gesellschaftlich erwünscht bzw. unerwünscht ist. Das Auftreten eines SD-Bias ist nach Essers Vorstellungen normalerweise ein eher unwahrscheinliches Ereignis, denn die drei genannten Determinanten wirken gemäß Theorie interaktiv, d. h. ein SD-Bias tritt nur dann auf, wenn keine Anonymität der Antwortsituation gegeben ist, wenn die befragte Person ein Bedürfnis nach sozialer Anerkennung hat und wenn sie gleichzeitig das fragliche Verhalten (hier das Begehen eines geringfügigen kriminellen Delikts) für gesellschaftlich unerwünscht bzw. verwerflich hält. Für das Ausbleiben eines SD-Bias genügt es schon, wenn einer der drei kritischen Faktoren nicht gegeben ist. Sieht man die drei genannten Bestimmungsfaktoren als kontinuierliche Variablen, erscheint die These vom ausschließlich interaktiven Zusammenwirken zweifellos relativ strikt. Zudem liegen empirische Befunde aufgrund von Validierungsdaten vor, die dafür sprechen, dass die Dimensionen SD-Neigung und SD-Belief unabhängig voneinander bzw. additiv auf das Misreporting wirken (Phillips & Clancy 1972).

In unserem Anwendungsfall dürfte der erstgenannte Faktor für einen SD-Bias, fehlende Anonymität, weitgehend entfallen, denn wir haben eine selbstadministrierte schriftliche Befragung, bei der eine relativ hohe Anonymität gegeben ist. Die Nichtpräsenz eines Interviewers und damit das Entfallen der Komponente der situationalen Erwünschtheit wird in der Regel als Hauptgrund dafür angegeben, weshalb (wie oben angesprochen) selbstadministrierte Befragungen validere Antworten auf heikle Frage liefern (sollen).

Die SD-Neigung, die also im Wesentlichen ein Streben nach sozialer Anerkennung zum Ausdruck bringt, ist ein Faktor, der in der Diskussion um sozial erwünschtes Antwortverhalten seit langem im Mittelpunkt steht (vgl. Boe & Kogan 1964; DeMaio 1984; Hartmann 1991; Johnson & Fendrich 2002). Wenn kein Interviewer präsent ist, dem gegenüber man einen guten Eindruck machen möchte, resultiert die Neigung von Personen, Antworten zu geben, die ein normenkonformes Verhalten anzeigen, zum einen daraus, ein positives Selbstbild zu erhalten, zum anderen aus dem Bestreben nach sozialer Anerkennung durch eine zwar nicht unmittel-

telbar präsente, aber subjektiv vorgestellte Öffentlichkeit (Esser 1986). Das Streben nach sozialer Anerkennung wird in der einschlägigen Forschung häufig als ein Persönlichkeitsmerkmal gesehen, und auf der Basis der Annahme einer personenspezifischen Varianz der SD-Neigung wurden Skalen entwickelt, die dieses Merkmal messen sollen. Die nach wie vor bekannteste Skala zur Messung der Stärke des Bedürfnisses nach sozialer Anerkennung stammt von Crowne & Marlowe (1960, 1964). Diese Skala wurde auch ins Deutsche übertragen und dabei so weit gekürzt, dass sie auch in regulären Umfragen einsetzbar ist (Stocké 2009a). Trotz Kritik an der Konzeption der SD-Neigung als Persönlichkeitsmerkmal (z. B. Paulhus 2002) wird die Crowne/Marlowe-Skala in der praktischen Forschung noch immer gerne verwendet. Gestützt auf diese Skala werden wir in unseren Analysen die Hypothese überprüfen, dass eine höhere SD-Neigung die Wahrscheinlichkeit einer wahren Antwort senkt.

Zusätzlich zur SD-Neigung muss gemäß SEU-Theorie auch der sogenannte SD-Belief bzw. die Trait Desirability berücksichtigt werden (z. B. Esser 1986; Stocké 2004; Stocké & Hunkler 2004). Mit SD-Belief bzw. Trait Desirability sind die Erwünschtheitswahrnehmungen der Befragten gemeint. Sie beziehen sich auf die Einschätzung des Befragten, wie die soziale Umwelt eine bestimmte Verhaltensweise und entsprechende Antwort auf eine heikle Frage wahrnehmen würde. Allzu oft und bereitwillig geht man in der Umfrageforschung noch immer von gesellschaftlich einheitlichen SD-Beliefs aus, also z. B. davon, dass die Befragten einen Ladendiebstahl, illegale Drogen, Steuerhinterziehung, Plagiate, fremdenfeindliche Einstellungen usw. für gesellschaftlich unerwünscht halten. Dies ist aber nachweislich nicht der Fall (für besonders anschauliche Belege vgl. Köllisch & Oberwittler 2004; Stocké 2004), so dass man für sinnvolle Analysen des SD-Bias eigentlich immer noch miterheben muss, ob überhaupt und in welchem Ausmaß die Respondenten ein bestimmtes Verhalten für unerwünscht oder verwerflich halten. Auch dies haben wir in unserer empirischen Studie umgesetzt und können damit überprüfen, ob ein hoher SD-Belief die Tendenz zu wahren Antworten senkt. Je stärker kleinere kriminelle Vergehen von einer Person für verwerflich gehalten werden, desto eher kann man vermuten, dass sie leugnet, so etwas getan zu haben. Weiterhin sollte aufgrund des von der SEU-Theorie angenommenen Interaktionseffekts gelten, dass die SD-Neigung nur oder bevorzugt dann die Tendenz zum Lügen erhöht, wenn bei einer Person gleichzeitig ein hoher SD-Belief gegeben ist.

### 2.3 Antwortverhalten im Lichte von Framing-Modellen

Am SEU-Modell des Befragtenverhaltens kann man kritisieren, dass es Schwierigkeiten mit dem Tatbestand bzw. der empirischen Beobachtung hat, dass viele Personen dazu neigen, ihre Aufgabe als Befragungsperson relativ schematisch und schablonenhaft abzarbeiten und möglichst schnell hinter sich zu bringen. Ausgehend von einem instrumentalistischen Theorieverständnis kann man aus SEU-Sicht zwar argumentieren, dass ja nicht unterstellt wird, die befragten Personen würden (stets) bewusst kalkulieren und für jede Handlungsalternative den exakten SEU-Wert berechnen; vielmehr wird nur mit einer „als ob“-Prämisse gearbeitet. Trotzdem verbleibt gegenüber diesem „instrumentalistischen Rettungsanker“ ein gewisses Unbehagen. Es ist nicht unplausibel anzunehmen, dass Entscheidungen in einigen Fällen mehr, in anderen weniger einem rationalen Kalkulationsprozess unterliegen. Mit Blick auf das Verhalten in Umfragen führen Tourangeau et al. (2000: 14) z. B. aus: „Some respondents may decide in a rational way whether to respond truthfully or to give evasive answers; others may decide in an unconscious and automatic way, in accord with rules that originally evolved for dealing with other situations.“ Ein schablonenartiges Antwortverhalten wird auch dadurch gefördert, dass Personen in standardisierten Befragungen wenig Zeit zum Nachdenken haben (bei mündlichen und telefonischen Befragungen) bzw. sich wenig Zeit zum Nachdenken nehmen (bei schriftlichen Befragungen). Zudem steht bei der Beantwortung von Umfragen in der Regel wenig auf dem Spiel („low cost“-Situation), so dass auch die Motivation zu einer elaborierten Informationsverarbeitung gering sein dürfte. Die Befragten neigen zu spontanen und reflexartigen Antworten und ersparen sich eine genauere Auseinandersetzung mit den einzelnen Fragen.

Die Ausgangsprämisse von – inhaltlich mehr oder weniger weitreichenden und theoretisch mehr oder weniger bedrohlichen – Bereichen eingeschränkter Rationalität hat dazu geführt, dass RC-Modelle seit nunmehr etlichen Jahren durch so genannte Framing-Modelle ergänzt und erweitert werden (als Überblick Mayerl 2009). Das im deutschen Sprachraum einflussreichste Framing-Modell stammt von Esser (2001: 259ff., 2010; vgl. auch Kroneberg 2005). Zwei wichtige Grundideen der verschiedenen Varianten von Framing-Modellen sind: 1) Akteure greifen in ihrem Alltagsleben oft auf gesellschaftlich vorgefertigte inhaltliche Schablonen/Schemata/Skrip-

te/Rahmungen zurück, um bestimmte Situationen zu bewältigen. Diese Situationsdefinition entspricht dem Prozess des Framings. 2) Bei diesem Framingprozess „entscheiden“ sich Akteure in einem vorgelagerten Schritt, ob sie Framing und Situationsbewältigung in einem automatisch-spontanen Modus mit einem geringen Grad an Informationsverarbeitung angehen oder aber in einem reflexiv-kalkulierenden Modus, der mehr oder weniger der herkömmlichen RC-Theorie entspricht.

Mit Bezug auf Essers Modell der Frame-Selektion hat Stocké (2004) eine theoretische und empirische Umsetzung der erstgenannten Grundidee vorgelegt. Demnach stehen für die Absolvierung einer Befragung im Wesentlichen zwei gesellschaftliche Rollenschemata zur Verfügung, die aus einem Framingprozess resultieren können: zum einen eine kooperative, zum anderen eine konformistische Befragtenrolle. Befragte mit einer kooperativen Rollenorientierung streben „eine bedingungslose Unterstützung der aktuellen Umfrage durch möglichst korrekte Antworten“ (ebd.: 306) an. Eine solche Rahmung „bewirkt die Ausschaltung auch starker Anreize durch soziale Erwünschtheit“ (ebd.), d. h. die Befragten sagen – mögen die Fragen auch noch so heikel oder bedrohlich sein – stets die Wahrheit. Demgegenüber bedeutet eine konformistische Rollenorientierung, dass sich die Befragten gemäß SEU-Theorie verhalten und mithin der SD-Bias mit seinen drei Determinanten (siehe Abschnitt 2.2) zum Zuge kommt.

Ob das Framing nun auf eine kooperative oder eine konformistische Befragtenrolle hinausläuft, hängt nach den theoretischen Überlegungen von Stocké im Wesentlichen von ihrer generalisierten Umfrageeinstellung ab. Genauer: Personen mit einer stark verankerten positiven Umfrageeinstellung tendieren in der Regel zur kooperativen Befragtenrolle, was zur Ausblendung des SD-Bias führt, während die übrigen Befragten dem SD-Bias unterliegen. Mit einer positiven allgemeinen Umfrageeinstellung ist gemeint, dass Personen Umfragen als ein wichtiges, nützliches und der wissenschaftlichen Wahrheitsfindung dienendes Instrument sehen, mithin ihre Teilnahme an der Befragung als einen Beitrag zur Wissenschaft werten und sich deshalb zu wahren Antworten verpflichtet fühlen.

In der empirischen Umsetzung misst Stocké die generalisierte Umfrageeinstellung mit einem Index aus 16 Einzelitems („survey attitude scale“), die sich auf die wahrgenommene Nützlichkeit und Verlässlichkeit von Umfragen beziehen (vgl. auch Stocké 2009b). Für die Messung der Stärke der kogni-

tiven Verankerung verwendet er zwei Indikatoren: erstens die „Umfrageerfahrung“, d. h. die vor der aktuellen Befragung früher absolvierte Anzahl an Befragungen, zweitens die „Antwortlatenz“, d. h. die durchschnittliche Geschwindigkeit bei der Beantwortung der 16 Einzelfragen nach der allgemeinen Umfrageeinstellung. Vorhandene Umfrageerfahrung und eine niedrige Antwortlatenz sollen dabei eine starke Verankerung der Umfrageeinstellung anzeigen. Die empirischen Ergebnisse seiner Studie wertet Stocké dahingehend, dass die Umfrageerfahrung den besseren und zu bevorzugenden Indikator zur Messung der Verankerungstiefe darstellt.

Obwohl die theoretische Argumentation von Stocké von vorneherein etwas fraglich erscheint (bewirkt eine positiv verankerte Umfrageeinstellung tatsächlich ein komplettes Ausschalten des SD-Bias, z. B. auch bei sehr heiklen Fragen wie etwa nach der Selbstbefriedigungshäufigkeit?) und man zudem über die Operationalisierung der Stärke der kognitiven Verankerung streiten kann (sowohl die Umfrageerfahrung also auch die Antwortlatenz sind nicht mehr als Näherungsmaße), werden wir in unseren empirischen Analysen die generalisierte Umfrageeinstellung und die Umfrageerfahrung aufgreifen und deren Effekte auf die Wahrscheinlichkeit einer wahren Antwort überprüfen. Eine Erhebung der Antwortlatenz im Stockéschen Sinne war nicht möglich, da es sich bei unserer Studie um eine schriftliche Erhebung handelte, aber wir werden eine andere Variante von Antwortlatenz zum Einsatz bringen (mehr dazu gleich im Folgenden).

Nur am Rande erwähnt Stocké (2004: 311, Fußnote) die oben an zweiter Stelle genannte Grundidee von Framing-Modellen, nämlich die „dual process“-These, die im Framing-Modell von Esser begrifflich mit der Differenzierung zwischen einem automatisch-spontanen und einem reflexiv-kalkulierenden Modus angesprochen wird. Angewandt auf das Verhalten in Umfragen, erscheint es durchaus naheliegend, Probanden danach zu charakterisieren, ob sie bei der Erledigung ihrer Aufgabe (beim Ausfüllen des Fragebogens) auf einen automatisch-spontanen oder reflexiv-kalkulierenden Antwortmodus schalten. Befragte im reflexiv-kalkulierenden Modus sollten, fast per definitionem, sich nach der SEU-Theorie verhalten, d. h. bei fehlender Anonymität, hoher SD-Neigung und hohem SD-Belief mit einem SD-Bias antworten und mithin ihr delinquentes Verhalten leugnen. Für Befragte, die automatisch-spontan antworten, könnte man hingegen argumentieren, dass sie keinem (oder einem geringeren) SD-Bias unterliegen, denn Lügen

ist in der Regel schwierig und erfordert ein kalkulierendes Abwägen von damit vermutlich verbundenen Vorteilen und Risiken. Für diese Überlegung sprechen z. B. empirische Befunde von Holtgraves (2004), die zeigen, dass Befragte auf Fragen mit einer sehr hohen SD-Saliens langsamer antworten als auf Fragen mit niedriger SD-Saliens. Gewiss gibt es auch „routinierte Lügner“, die spontan und unverzüglich erst einmal alles leugnen, was mit Unannehmlichkeiten verknüpft sein könnte, aber zumindest als spekulative Hypothese verdient die Vermutung, dass im spontanen Antwortmodus seltener gelogen wird, eine empirische Überprüfung.

Ob eine Person eine Frage (oder auch einen ganzen Fragebogen) automatisch-spontan oder reflexiv-kalkulierend beantwortet, könnte man empirisch mit dem Indikator der Antwortlatenz messen. Eine rasche und unverzügliche Reaktion verweist auf den erstgenannten, eine zaudernde und zögerliche auf den zweitgenannten Antwortmodus. Im Ergebnis bedeuten diese zugestanden etwas gewagten Überlegungen, dass – wie bei Stocké – die Antwortlatenz als ein Bestimmungsfaktor für wahre bzw. falsche Antworten postuliert wird. Die nun folgenden empirischen Analysen werden zeigen, dass wir in unserer Studie zwar nur eine unbefriedigende Messung für die Antwortlatenz haben, diese Messung aber trotzdem ein substanziell wichtiges Ergebnis hervorbringt.

### 3. Datenbasis und Variablen

In einem ersten Abschnitt werden die notwendigen Informationen zur verwendeten Datenbasis gegeben. Ein zweiter Abschnitt stellt die abhängige Variable und die verschiedenen Gruppen von unabhängigen Variablen vor. Dabei ist natürlich die Grundauszählung für die abhängige Variable, d. h. welcher Anteil der Befragten die begangene Straftat bekennt, im ersten Schritt von besonderer Bedeutung.

#### 3.1 Empirische Datengrundlage

Unsere empirischen Analysen stützen sich auf eine im zweiten Halbjahr 2010 durchgeführte postalische Befragung von 352 Personen in einem westdeutschen Ballungsraum. Von 301 Befragten war aus Akten der Justizbehörden bekannt, dass sie im Zeitraum 2005–2008 für ein oder auch mehrere kleinere Vergehen (Schwarzfahren, Ladendiebstahl, Sozialversicherungsbetrug, Steuerhinterziehung, Körperverletzung u.ä.) strafrechtlich verurteilt worden

waren. Die restlichen 51 Befragten waren „saubere Fälle“, die lediglich aus Gründen des Datenschutzes in die Erhebung einbezogen wurden. Wir klammern diese Fälle aus den Analysen aus, so dass 301 „Delinquenten“ verbleiben.

Die 352 ausgefüllten Fragebögen resultieren aus einer Bruttostichprobe von insgesamt 2471 angeschriebenen Adressen. Von den angeschriebenen Adressen, die überwiegend aus den schon etwas „angestaubten“ Akten der Justizbehörden stammten, waren allerdings rund 1500 falsch, d. h. die Empfänger waren unter der angegebenen Adresse auf dem Postweg nicht mehr zu erreichen. Zusammen mit einigen anderen Ausfällen (Empfänger verstorben, Zielperson in Haft u.ä.) waren insgesamt 1553 „stichprobenneutrale Ausfälle“ (Diekmann 2007: 421) zu verzeichnen. Mit der dann verbleibenden Nettostichprobe im Umfang von 918 Fällen wurde eine Ausschöpfung von 38 Prozent erreicht, was angesichts der tendenziell „sperrigen“ Zielgruppe annehmbar erscheint.

Die Zielpersonen wurden in Anlehnung an die Total/Tailored-Design-Methode angegangen (Dillman 1978; Dillman et al. 2009): mit einem ersten Anschreiben plus Fragebogen und Antwortkuvert, einer Erinnerungspostkarte nach einer Woche und einem dritten Anschreiben plus Fragebogen und Antwortkuvert nach drei Wochen, aber unter Verzicht auf ein viertes Anschreiben nach sieben Wochen. Die Studie firmierte in der Außendarstellung unter einem Titel, der den Befragten eine Studie zum Thema Lebensqualität im Stadt/Land-Vergleich signalisierte. Das Anliegen der Studie machte diese Verschleierung erforderlich, und selbstverständlich durften die Zielpersonen auch nicht wissen, dass die Adresse aus Justizakten mit einer dort verzeichneten und per Gerichtsurteil bestätigten Straftat stammte.<sup>5</sup>

Das Rahmenthema der Lebensqualität machte es möglich, auch auf den Lebensbereich „Kriminalität“ einzugehen, da es diesbezüglich bekanntlich Stadt/Land-Unterschiede gibt. Erst nachdem einige Fragen zu Kriminalitätsfurcht im eigenen Lebensumfeld, zu persönlichen Viktimisierungserfahrungen

u.ä. gestellt worden waren, kam der Fragebogen zur „Gretchenfrage“ auch eigener Vergehen, sprich selbst begangener strafrechtlicher Delikte und eventuell nachfolgenden Verurteilungen.

### 3.2 Zentrale Variablen

Diese Vergehensfrage ist die zentrale abhängige Variable unserer Analysen. Sie hatte den Wortlaut: „Man wird nicht nur Opfer krimineller Handlungen, sondern verstößt manchmal auch selbst gegen Gesetze. Wurden Sie – per Strafbefehl oder Gerichtsverhandlung – jemals wegen eines kleineren oder auch größeren Delikts strafrechtlich verurteilt? Strafrechtlich bedeutet, dass die Staatsanwaltschaft eingeschaltet wurde.“ Es konnte nur mit ja oder nein geantwortet werden. Knapp zwei Drittel der aktenkundigen Delinquenten, genau 64,7 Prozent, gestanden ein, dass sie schon einmal strafrechtlich verurteilt wurden, und sagten damit die Wahrheit. Umgekehrt haben 35,3 Prozent die Tat lieber verschwiegen und konnten sich nicht zur Wahrheit durchringen (oder haben eventuell den Vorfall vergessen, was wir aber als unwahrscheinlich ansehen). Bei diesen Prozentwerten sind 18 Befragte ausgeklammert, die keine Antwort auf die Frage gegeben haben.<sup>6</sup> Im Vergleich zu anderen Validierungsstudien im Bereich delinquenten Verhaltens (z. B. Hessing et al. 1988; Junger 1989; Elffers et al. 1992; van der Heijden et al. 2000) ist unsere Quote wahrer Antworten relativ hoch. Dies unterstützt auch die einleitend angesprochene Vermutung validerer Prävalenzschätzungen bei Rückgriff auf selbstadministrierte Befragungen.

Die unabhängigen Variablen für die weiteren Auswertungen können – im Wesentlichen basierend auf den theoretischen Ausführungen von Abschnitt 2 – in vier Gruppen eingeteilt werden: 1) Kontrollvariablen: Personenübereinstimmung, Geschlecht und Alter, 2) elementare RC-Variablen: Deliktsschwere und Bildung, 3) SEU-Variablen: SD-Neigung und SD-Belief, 4) Framing-Variablen: Umfrageeinstellung, Umfrageerfahrung und Antwortlatenz. Diese insgesamt zehn Einzelvariablen werden im Folgenden erläutert.

<sup>5</sup> Forschungsethische und datenschutzrechtliche Fragen wurden im Vorfeld der Studie natürlich geprüft. Forschungsethisch ist die Studie auch insofern abgesichert, als diese im Rahmen eines von der Deutschen Forschungsgemeinschaft (DFG) geförderten Projekts durchgeführt wurde, somit den Begutachtungs- und Genehmigungsprozess durchlaufen hat und den DFG-Standards genügt. Zudem wurde der zuständige Landesdatenschutzbeauftragte bei der Projektplanung hinzugezogen.

<sup>6</sup> Aus der Forschung zu heiklen Fragen ist bekannt, dass Item-Nonresponse bei den meisten heiklen Fragen (Ausnahme: Einkommen) kein besonderes Problem ist, dass die Befürchtung, viele Probanden würden solche Fragen nicht beantworten, unbegründet ist und dass das Ausmaß von Item-Nonresponse kein guter Gradmesser für das Ausmaß der „Heikelheit“ einer Frage ist (mehr dazu z. B. Tourangeau et al. 2000: Kap. 9).

**Kontrollvariablen:** Die „Personenübereinstimmung“ ist eine 0/1-Variable und misst, ob das von den Probanden in der Befragung angegebene Geburtsjahr mit dem in den Akten verzeichneten Geburtsjahr übereinstimmt. Akzeptiert man eine Abweichung von plus/minus 1 Jahr, konnte bei 276 der 301 Fälle eine Übereinstimmung festgestellt werden. Keine Übereinstimmung gab es bei 19 Fällen und weitere 6 Fälle verweigerten in der Befragung die Angabe ihres Geburtsjahres; die 19 Fälle haben wir als „Nichtübereinstimmung“ codiert und die 6 Fälle ohne Angabe als fehlende Werte behandelt.<sup>7</sup> Nur 17 Prozent unserer Befragten sind Frauen, was mit dem niedrigen Anteil tatverdächtiger Frauen in der polizeilichen Kriminalstatistik korrespondiert (dort liegt – über alle Delikte hinweg und bezogen auf Deutschland insgesamt – der Frauenanteil bei knapp 25 Prozent; vgl. Bundeskriminalamt 2010). Das aus den Befragungsdaten berechnete Durchschnittsalter der Befragten liegt bei etwa 41 Jahren, mit jeweils rund einem Drittel der Befragten bis 30, 31–47 und über 47 Jahre.

**Elementare RC-Variablen:** Die metrische Variable zur Messung der „Deliktsschwere“ ist das Ergebnis eines externen Ratings von neun zusammengefassten Deliktategorien. Die vielen einzelnen Deliktarten wurden in einem ersten Schritt zu neun Kategorien zusammengefasst: Diebstahl (jeglicher Art), Schwarzfahren (Erschleichen von Leistungen), Alkohol/Drogen im Straßenverkehr, Unfallflucht, Fahren ohne Fahrerlaubnis, Drogenvergehen (Besitz, Erwerb, Handel mit Betäubungsmitteln), Körperverletzung (jeglicher Art), Beleidigung/Bedrohung/Nötigung/Erpressung und Betrug (jeglicher Art). 13 ausgewählte Personen wurden dann gebeten, die neun Straftaten nach ihrer Schwere zu ordnen, mit Nummern von eins (= schwerstes Delikt) bis neun (= harmlosestes Delikt). Als schwerstes Vergehen wurde die Körperverletzung eingestuft (Mittelwert = 1,1), als harmlosestes Schwarzfahren (Mittelwert = 8,6). Jedem Befragten wurde auf der Basis seiner Straftat der Mittelwert aus dem Rating zugewiesen, bei mehreren Delikten wurde auf das schwerste Vergehen abgestellt (ausführlicher zu dieser Messung der Deliktsschwere vgl. Wolter 2012: Kap. 4). Anschließend wurde die Skala umgepolt; empirisch bewegen sich ihre Werte von 1,4 bis 8,9. Der Mittelwert liegt bei 5,3. Die Bildung der Befragten wurde mit der Frage nach dem höchsten Bildungsabschluss erfasst (kein Schulabschluss bis hin zu einem Universitätsabschluss). Diese Variable wurde dann in eine metrische transformiert, indem auf die Zahl der Bildungsjahre übergangen wurde. Im Mittel belief sich die Zahl der Bildungsjahre auf 11 Jahre.

**SEU-Variablen:** Zur Messung der SD-Neigung wurden neun dichotome 0/1-Items verwendet, die in der Anhangtabelle A1 aufgeführt sind. Acht der neun Items entstammen der Kurzskaala zur Messung des Bedürfnisses nach sozialer Anerkennung von Stocké (2009a), ein zusätzliches Item kommt aus dem „Balanced Inventory of Desirable Responding“ von Paulhus (1991; vgl. auch Winkler et al. 2006). Aus den neun Items wurde ein Mittelwertindex (additiver Index dividiert durch die Zahl der Items) gebildet, der Werte zwischen 0 und 1 annehmen kann und einen Mittelwert von 0,63 erreichte. Je höher der Wert, desto ausgeprägter das Bedürfnis eines Befragten nach sozialer Anerkennung. In einer Faktorenanalyse laden die neun Items auf zwei Faktoren: Der eine Faktor umfasst die Aussagen zu erwünschtem Verhalten (Items 2, 5, 6, 7, 8), der zweite die Items zu unerwünschtem Verhalten (Items 1, 3, 4, 9). Trotz der Zweidimensionalität, die offenbar durch die unterschiedliche Polung der Items hervorgerufen wurde, haben wir einen einheitlichen Index gebildet. Cronbachs Alpha für die Skala liegt bei 0,62. Die Erwünschtheitswahrnehmung der Befragten, also der SD-Belief mit Bezug auf die wahrgenommene Verwerflichkeit delinquenten Verhaltens, wird ebenfalls mit einem additiven Index gemessen. Der Index besteht aus vier Items, die in der Anhangtabelle A2 dokumentiert sind. Die vier Items konnten auf einer Skala von 1–5 beantwortet werden, so dass der Mittelwertindex ebenfalls einen Wertebereich von 1–5 hat. Je höher der Indexwert, desto stärker die Wahrnehmung eines Befragten, dass kriminelles Verhalten sozial unerwünscht ist. Der Indexmittelwert beläuft sich auf 3,9, was anzeigt, dass die Mehrheit der Befragten durchaus der Ansicht ist, dass kriminelles Verhalten im engeren und weiteren sozialen Umfeld abgelehnt wird. Die Items des SD-Belief laden auf einen Faktor. Cronbachs Alpha erreicht 0,65.

**Framing-Variablen:** Die generalisierte Umfrageeinstellung wurde mit den vier in der Anhangtabelle A3 stehenden Items in Anlehnung an Stocké (2009b) erhoben. Die Items konnten auf einer Skala von 1–5 beantwortet werden. Diejenigen, die den Items zustimmen, haben eine „hohe Meinung“ von Umfragen. Die vier Items liefern in einer Faktorenanalyse eine einfaktorielle Lösung. Es wurde ein Mittelwertindex mit einem Wertebereich von 1 bis 5 gebildet, wobei höhere Werte eine positivere Umfrageeinstellung anzeigen. Die Reliabilität der Skala beläuft sich auf Cronbachs Alpha = 0,71. Der Index-Mittelwert in der Stichprobe liegt bei 3,4. Die Messung der Umfragerfahrung stützt sich auf die Antworten der Probanden auf die Frage, ob sie schon öfter befragt worden sind bzw. ob sie in der Vergangenheit schon einen oder mehrere Fragebögen ausgefüllt haben. Aus der ursprünglich ordinalen Variablen wurde eine quasi-metrische konstruiert, die auf die Anzahl der absolvierten Umfragen abstellt und Ausprägungen von 0 bis 5 hat. Sie erreicht einen Mittelwert von 2,3. Die oben als „Antwortlatenz“ bezeichnete Variable ist in unserer Operationalisierung nicht die Antwortgeschwindigkeit auf eine oder mehrere Einzelfragen (im Idealfall würde man wohl auf die Antwortgeschwindigkeit bei der heiklen Frage selbst abstellen), sondern die in Tagen gemessene Zeit, die ein Befragter gebraucht hat, um

<sup>7</sup> Am Rande sei vermerkt, dass – nicht zuletzt unter Beachtung der besonderen Stichprobe – eine so hohe Übereinstimmung der Geburtsjahre in der Befragung und den Akten ein starker Beleg für die Validität von Befragungen ist, genauer: ein Beleg für unsere Eingangsbemerkung zu Abschnitt 2, dass Befragte bei nicht oder wenig heiklen Themen in der Regel nicht lügen, sondern zutreffende Antworten geben.

**Tabelle 1** Selbstberichtete Delinquenz in Abhängigkeit von den mutmaßlichen Einflussfaktoren (Prozentangaben)

<b>Personenübereinstimmung***</b>		<b>Alter</b>	
nein	26,3	bis 30 Jahre	70,2
ja	67,2	31 bis 47 Jahre	63,7
<b>Geschlecht*</b>		48 Jahre und älter	59,4
Frau	50,0		
Mann	68,0		
<b>Deliktschwere<sup>+</sup></b>		<b>Bildung</b>	
niedrig	67,0	Hauptschule oder weniger	71,1
hoch	54,0	Mittlere Reife/Realschule	66,7
		Fachhochschulreife/Abi	57,0
<b>SD-Neigung**</b>		<b>SD-Belief***</b>	
schwach	84,8	schwach	80,0
mittel	64,5	mittel	79,5
stark	49,2	stark	53,5
<b>Umfrageeinstellung</b>		<b>Antwortlatenz**</b>	
negativ	63,2	Bis 14 Tage	77,3
mittel	63,7	15 bis 28 Tage	59,8
positiv	67,5	29 Tage und mehr	53,8
<b>Umfrageerfahrung</b>			
keine/wenig (bis 2x)	66,1		
viel (3+)	62,7		

*Erläuterung:* Signifikanzniveaus \* $p < 0,1$ , \* $p < 0,05$ , \*\* $p < 0,01$ , \*\*\* $p < 0,001$ . Die metrischen Variablen und Indizes wurden für die bivariate Analyse in ordinalskalierte Variablen umkodiert, wobei bei den Indizes die Kategorien in der Regel dergestalt gebildet wurden, dass jede Kategorie mit etwa der Hälfte bzw. einem Drittel besetzt war.

den Fragebogen ausgefüllt zurückzuschicken. Dies ist gewiss eine unbefriedigende Messung für unsere theoretisch postulierte Einflussgröße der Antwortlatenz im Rahmen des Framing-Modells (Abschnitt 2.3), aber in einer schriftlichen postalischen Befragung ist die Erfassung der Reaktionszeit auf einzelne Fragen nicht anders möglich. Die Zahl der Tage, bis der schriftliche Fragebogen ausgefüllt zu uns zurückkam, erstreckte sich von 2 bis 119 – dies mit einer erwartbar stark rechtsschiefen Verteilung und einem Median von 20 Tagen.<sup>8</sup>

#### 4. Empirische Ergebnisse

Die empirischen Befunde werden im Folgenden in zwei Schritten vorgestellt und erläutert: Im ersten Schritt werden die Ergebnisse einfacher bivariater Aufgliederungen der abhängigen Variable („Wer sagt die Wahrheit?“) nach den zehn unabhängigen Variablen präsentiert. Im zweiten Schritt geht es um eine Vertiefung und Absicherung der Resultate mittels multipler logistischer Regressionsmodelle.

<sup>8</sup> Deskriptive Statistiken zu allen Variablen der Analysen finden sich im Anhang in Tabelle A4.

#### 4.1 Bivariate Befunde

Betrachtet man auf der bivariaten Analyseebene in Tabelle 1 die Häufigkeit selbstberichteter Delinquenz und damit die Häufigkeit wahrer Antworten in Abhängigkeit von den vier Variablengruppen (Kontroll-, elementare RC-, SEU- und Framing-Variablen), kommt man zu Ergebnissen, die zum Teil unbedingt mit den theoretischen Ausgangserwartungen harmonisieren.

Erwartungsgemäß zeigt sich ein sehr deutlicher Zusammenhang zwischen der Personenübereinstimmung und der Wahrscheinlichkeit, dass sich eine befragte Person zu einer strafrechtlichen Verurteilung wegen eines begangenen Delikts bekennt. Dies ist aber kein substanzielles Ergebnis. Es unterstützt vielmehr nur die Vorabvermutung, dass es bei einer Nichtübereinstimmung der Geburtsjahre in der Befragung und den Akten (Verweigerungen der Altersangabe in der Befragung wurden als fehlende Werte gezählt) tatsächlich wohl öfter vorgekommen ist, dass nicht die eigentliche Zielperson den Fragebogen ausgefüllt hat. Kurz: Es handelt sich um eine

methodische Kontrollvariable, wobei aber nochmals darauf verwiesen sei, dass nur bei gut 6 Prozent aller gültigen Fälle eine Nichtübereinstimmung vorlag. Frauen geben signifikant seltener als Männer an, dass sie bereits delinquent waren. Während 68 Prozent der Männer die Wahrheit sagen, sind es 50 Prozent der Frauen. Bei der Verteilung wahrer Antworten nach Altersgruppen ergibt sich, dass mit steigendem Alter der Anteil der Personen sinkt, die zugeben, dass sie bereits strafrechtlich verurteilt wurden. In der jungen Altersgruppe gestehen 70 Prozent der Befragten ihre Delinquenz, in der älteren Gruppe sind es 59 Prozent (der Zusammenhang ist hier allerdings laut Chi-Quadrat-Test nicht signifikant).

Die bivariaten Ergebnisse bestätigen weitgehend die aus der elementaren RC-Theorie hergeleitete Verlust-Hypothese, dass diejenigen seltener die Wahrheit sagen, die dadurch „viel zu verlieren“ hätten. Es zeigt sich ein (nur auf dem 10 %-Niveau) signifikanter Zusammenhang zwischen der Deliktsschwere und der selbstberichteten Delinquenz. Von denjenigen, die ein relativ schweres Delikt begangen haben (Indexwerte 7 und mehr), bekennen sich 54 Prozent dazu, bei den Befragten mit einem eher geringfügigen Vergehen sind es 67 Prozent. In Abhängigkeit von der Bildung differieren die Wahrheitsquoten bivariat zwar nicht signifikant, aber von der Tendenz her folgen sie der Erwartung aus der Verlust-Hypothese (71 Prozent wahre Antworten bei niedriger, 67 bei mittlerer und 57 Prozent bei hoher Bildung).

Überraschend deutlich macht sich der im Rahmen der SEU-Theorie betonte SD-Bias bemerkbar. Beide Dimensionen sozialer Erwünschtheit – sowohl die SD-Neigung als auch der SD-Belief – zeigen einen drastischen Einfluss auf die wahrheitsgemäße Beantwortung der Validierungsfrage. Personen mit schwacher SD-Neigung (Indexwerte bis 0,44) antworten mit 85 Prozent am ehrlichsten; dieser Prozentsatz wahrer Antworten sinkt bei Personen mit mittlerer SD-Neigung (Indexwerte größer als 0,44 und kleiner als 0,78) auf 64 Prozent und bei Befragten mit einer starken SD-Neigung (Indexwerte ab 0,78) sogar auf 49 Prozent. Ähnlich gestaltet sich die Beziehung zwischen den Erwünschtheitswahrnehmungen und der wahren Antwort. Während Personen mit einem schwachen SD-Belief (Indexwerte bis 2), also Personen, die eine strafrechtliche Verurteilung nicht für besonders verwerflich halten, zu 80 Prozent ihre Delinquenz zugeben, bekennen nur 53 Prozent der Befragten mit starken Erwünschtheitswahrnehmungen (Indexwerte zwischen 4 und 5) ihr heikles Verhalten.

Für die aus den Framing-Modellen gewonnenen mutmaßlichen Einflussfaktoren auf die Wahrscheinlichkeit, eine wahre Antwort zu geben, ergibt sich bivariat ein gemischtes Bild. Die generalisierte Umfrageeinstellung und die Umfrageerfahrung zeigen keine nennenswerten Effekte. Aber der Zusammenhang zwischen der Antwortlatenz und der wahren Antwort ist statistisch hochsignifikant. Befragte, die den Fragebogen innerhalb von 14 Tagen seit dem ersten Versanddatum zurückgeschickt haben, geben zu 77 Prozent eine wahre Antwort; diejenigen, die erst nach 29 Tagen oder mehr reagiert haben, sagen nur zu 54 Prozent die Wahrheit. Dieser Unterschied im Umfang von 23 Prozentpunkten ist in der Tat sehr deutlich und (trotz theoretischer Einbettung) in seiner Größenordnung zweifellos überraschend.

## 4.2 Multivariate Befunde

Da die postulierten Einflussfaktoren auf die Tendenz zu wahren Antworten zum Teil miteinander korrelieren und da sowohl die SEU-Theorie als auch die Framing-Modelle bei strenger Auslegung keine signifikanten Haupteffekte, sondern signifikante Interaktionseffekte postulieren, ist es unverzichtbar, die voranstehenden bivariaten Aufgliederungen durch multiple Regressionsmodelle, d. h. im vorliegenden Fall durch binäre logistische Regressionen fortzuführen.

Im Vorspann dazu liefert die Tabelle A5 im Anhang zunächst die Matrix der bivariaten Korrelationen der insgesamt elf beteiligten Variablen. Ein Blick auf die erste Zahlenspalte in dieser Tabelle bekräftigt im Wesentlichen noch einmal das Bild aus Tabelle 1. Anders als in Tabelle 1 zeigen – bedingt durch den Übergang zu metrischen Variablen – jetzt aber auch das Alter (in Jahren) und die Bildung (in Jahren) einen signifikanten Zusammenhang mit der Wahrscheinlichkeit einer wahrheitsgemäßen Auskunft. Gemessen an den bivariaten Korrelationen sind in der Reihenfolge ihrer Nennung der SD-Belief, die SD-Neigung, die Personenübereinstimmung und die Antwortlatenz die besten Prädiktoren für das Eingeständnis, strafrechtlich verurteilt worden zu sein. Weitere nennenswerte Korrelationen ( $r > 0,2$ ) in Tabelle A5 verweisen auf positive Beziehungen zwischen SD-Belief und SD-Neigung, zwischen Bildung und dem Ausmaß an Umfrageerfahrung (= der Häufigkeit einer Beteiligung an Umfragen in der Vergangenheit) sowie zwischen SD-Belief und Deliktsschwere. Die letztgenannte positive Korrelation besagt, dass Personen, die Straftaten in höherem Maße für verwerflich halten,

selbst überdurchschnittlich oft eher schwere Delikte begangen haben – ein Ergebnis, das sich gewiss nicht ohne weiteres erklären lässt.

Die insgesamt vier logistischen Regressionsmodelle in Tabelle 2 berücksichtigen stets die Kontrollvariablen (Personenübereinstimmung, Geschlecht, Alter). Modell 1 konzentriert sich auf die aus der rudimentären RC-Theorie gewonnenen Einflussfaktoren (Deliktsschwere, Bildung), Modell 2 auf die im Rahmen der SEU-Theorie zentralen Determinanten (SD-Neigung, SD-Belief), Modell 3 auf die Variablen der Framing-Modelle (Umfrageeinstellung, Umfrageerfahrung, Antwortlatenz), und in dem Gesamtmodell 4 ist schließlich der komplette Satz der Kovariaten enthalten. Neben den unstandardisierten logistischen Regressionskoeffizienten und deren Standardfehlern sind in der Tabelle die durchschnittlichen marginalen Effekte (AME=average marginal effects) angegeben. Die AME informieren im Fall von Dummy-Variablen darüber, um wie viel im Durchschnitt die Wahrscheinlichkeit einer wahren Antwort bei Vorliegen der Eigenschaft (Code 1) im Vergleich zur Referenzkategorie (Code 0) höher liegt; im Fall von metrischen Variablen geht es um die durchschnittliche Erhöhung der Wahrscheinlichkeit einer wahren Antwort, wenn die unabhängige Variable um eine Einheit steigt (dazu z. B. Best & Wolf 2010). Vorab sei auch noch darauf verwiesen, dass für die Schätzung der Modelle alle metrischen Variablen (Alter bis Antwortlatenz) zentriert wurden und die Antwortlatenz wegen ihrer rechtsschiefen Verteilung zudem noch logarithmiert.

Wirft man in Tabelle 2 zunächst einen Blick auf die Kontrollvariablen, zeigt sich ein durchgehend signifikanter Effekt der Personenübereinstimmung. Geschlecht und Alter verlieren im Gesamtmodell (Modell 4) ihre Bedeutung und werden mithin von den theoretisch relevanten Variablen „übernommen“. Tabelle A5 macht in diesem Zusammenhang deutlich, dass Frauen einen höheren SD-Belief haben, d. h. Straftaten in höherem Maße für verwerflich halten, und ältere Befragte sowohl einen höheren SD-Belief als auch eine höhere SD-Neigung.

Modell 1 prüft die Verlusthypothese, dass diejenigen Befragten seltener die Wahrheit sagen, die etwas zu verlieren haben – also Befragte, die ein vergleichsweise schweres Delikt begangen haben, und Befragte, die einen höheren Bildungsabschluss vorzuweisen haben. Beide Variablen liefern – den theoretischen Erwartungen entsprechend – einen negativen Einfluss auf die Antwortvalidität, aber nur der Bildungseffekt ist dabei signifikant. Dieses Effektmuster bleibt so auch im Gesamtmodell erhalten.

Die Regressionskoeffizienten in Modell 2 und auch noch in Modell 4 verweisen auf einen sehr deutlichen Einfluss der SD-Neigung und des SD-Belief. Dies ist allerdings keine Bestätigung der SEU-Theorie. Diese behauptet nämlich, dass ein SD-Bias (eine Verzerrung aufgrund sozialer Erwünschtheit) nur dann eintritt, wenn bei einer Person zugleich eine hohe SD-Neigung und ein hoher SD-Belief vorliegen (Esser 1986, 1991; Stocké 2004). Technisch gesprochen bedeutet dies, dass es einen signifikanten Interaktionseffekt aus SD-Neigung und SD-Belief geben müsste. Dies ist aber in den Daten unserer Studie nicht der Fall: Ein erweitertes Modell 2, das zusätzlich auch noch den Interaktionseffekt aufnimmt, bringt für diesen keinen signifikanten Einfluss (dieses Modell ist in Tabelle 2 nicht eigens ausgewiesen).<sup>9</sup>

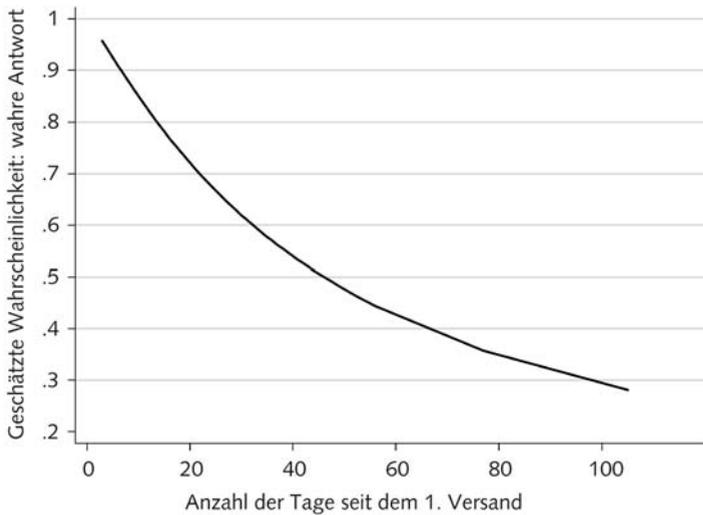
Modell 3 fokussiert die Framing-Variablen. Dabei ergibt sich das, was wir schon in den bivariaten Aufgliederungen gesehen haben: keine Effekte der generalisierten Umfrageeinstellung und der Umfrageerfahrung, aber ein starker Effekt der Antwortlatenz, d. h. der Zeitdauer bis der Fragebogen zurückgeschickt wurde. Diese Parameterkonstellation ändert sich auch im Gesamtmodell 4 nicht. Ähnlich wie bei der SEU-Theorie beziehen sich die Vorhersagen der Framing-Modelle aber nicht bzw. nicht in erster Linie auf Haupteffekte der beteiligten Variablen, sondern auf Interaktionseffekte. Bei Stocké (2004) werden aus den Variablen von SEU- und Framing-Theorie (SD-Neigung, SD-Belief, Umfrageeinstellung, Umfrageerfahrung, Antwortlatenz) eine Fülle von Interaktionstermen gebildet (Zweifach-, Dreifach- und noch höherstufige Interaktionen) und hinsichtlich ihrer Effekte auf bekundete Einstellungen gegenüber Ausländern überprüft. Dort finden sich erstaunlich viele empirische Bestätigungen der vergleichsweise diffizilen theoretischen Argumente. Von dieser Sachlage ausgehend, haben auch wir in unseren Daten nach signifikan-

<sup>9</sup> Bei der Berechnung von Interaktionseffekten in logistischen Modellen stellt sich das Problem, dass sich Richtung und Signifikanz der Interaktionseffekte nur auf die logarithmierten Odds und nicht auf die Wahrscheinlichkeiten, über die aber eigentlich Aussagen getroffen werden sollen, beziehen. Somit bedeutet ein mit Standardverfahren berechneter Interaktionseffekt nicht zwangsläufig, dass dieser für alle Merkmale der Stichprobe gilt, und auch nicht, dass er für alle Merkmale (nicht) signifikant ist (Norton et al. 2004; Best & Wolf 2010). Sogar das Vorzeichen kann variieren. Um diesem Problem zu begegnen, haben wir unsere Effekte mit dem von Norton et al. vorgeschlagenen Verfahren geprüft – mit dem Ergebnis, dass keine Interaktion vorliegt.

**Tabelle 2** Logistische Regressionen zu den mutmaßlichen Einflussfaktoren auf die selbstberichtete Delinquenz

	Modell 1			Modell 2			Modell 3			Modell 4		
	b	SE	AME	b	SE	AME	b	SE	AME	b	SE	AME
Personenüber. (1 = Ja)	1,787**	(0,559)	0,250	2,305**	(0,664)	0,281	2,001**	(0,622)	0,266	2,595***	(0,745)	0,280
Geschlecht (1 = Mann)	0,624 <sup>+</sup>	(0,338)	0,114	0,506	(0,352)	0,090	0,765*	(0,346)	0,133	0,495	(0,392)	0,079
Alter/10	-0,145	(0,094)	-0,030	-0,137	(0,096)	-0,026	-0,253**	(0,094)	-0,050	-0,099	(0,108)	-0,017
Deliktschwere	-0,139	(0,088)	-0,028							-0,095	(0,098)	-0,016
Bildung in Jahren	-0,100*	(0,047)	-0,020							-0,143*	(0,059)	-0,024
SD-Neigung				-1,350*	(0,636)	-0,259				-1,407*	(0,700)	-0,240
SD-Belief				-0,659**	(0,191)	-0,126				-0,815***	(0,215)	-0,139
Umfrageeinstellung							0,119	(0,168)	0,023	0,193	(0,186)	0,033
Umfrageerfahrung							0,000	(0,075)	0,000	0,048	(0,085)	0,008
Antwortlatenz (ln)							-0,865***	(0,223)	-0,170	-1,139***	(0,262)	-0,194
Konstante	-1,503*	(0,591)		-1,873**	(0,691)		-1,844**	(0,665)		-2,025**	(0,781)	
Pseudo R <sup>2</sup> (McFadden)	0,08			0,13			0,11			0,21		
N	274			275			278			269		

*Erläuterung:* Signifikanzniveaus <sup>+</sup> p < 0,1, \* p < 0,05, \*\* p < 0,01, \*\*\* p < 0,001; b = unstandardisierte Regressionskoeffizienten; SE = Standardfehler; AME = durchschnittliche marginale Effekte.



**Abb. 1** Wahrscheinlichkeit einer wahren Antwort in Abhängigkeit von der Antwortlatenz (d. h. von der Geschwindigkeit der Rücksendung des Fragebogens)

*Erläuterung:* Die Grafik basiert auf Modell 4, Tabelle 2. Dabei wurden die anderen Kovariaten des Gesamtmodells konstant gehalten, indem Dummy-Variablen auf 1 (Personenübereinstimmung und männliches Geschlecht) und metrische Variablen auf ihren Mittelwert gesetzt wurden.

ten Interaktionstermen der angesprochenen Variablen gesucht. Diese Suche erbrachte allerdings das ernüchternde Resultat, dass robuste, d. h. bei unterschiedlichen Modellspezifikationen stabil bleibende und konsistent den theoretischen Erwartungen entsprechende Interaktionseffekte nicht nachweisbar waren. Zudem änderte sich auch bei Berücksichtigung diverser Interaktionseffekte praktisch nie etwas an dem Ergebnis signifikanter Haupteffekte von SD-Neigung, SD-Belief und Antwortlatenz.

Das nur mit Haupteffekten operierende Gesamtmodell 4 erreicht mit Pseudo- $R^2 = 0,21$  einen recht guten Fit. Und das substantiell wichtigste Ergebnis ist dabei unseres Erachtens der markant negative Effekt der Antwortlatenz, d. h. der verstrichenen Zeit bis zur Rücksendung des Fragebogens. Gemessen am z-Wert ( $=b/SE$ ), handelt es sich um die Variable mit dem relativ stärksten Einfluss. Die spät Antwortenden sagen eindeutig seltener die Wahrheit als die rasch Antwortenden. Wie ausgeprägt dieser Zusammenhang zwischen Reaktionszeit und Antwortvalidität ist, illustriert noch einmal Abbildung 1, die auf der Grundlage von Modell 4 erstellt wurde.

Nun drängt sich natürlich die Frage auf, woran es liegt, dass die „Frühstarter“ tendenziell ehrlich antworten und die „Spätzünder“ es mit der Wahrheit weniger genau nehmen. Eine nochmalige Inspektion von Tabelle A5 (letzte Zeile) ergibt, dass die spät Antwortenden eine geringere Bildung haben, sich schon häufiger an Umfragen beteiligt haben und (überraschend) nicht relativ schwere, sondern eher leichte Delikte begangen haben – aber das hilft

bei der Erklärung auch nicht weiter. Vielleicht wäre es sinnvoll, zunächst einmal grundsätzlich zu überlegen, was sich hinter dem festgestellten Effekt verbergen könnte. Eher unplausibel erscheint, einen direkten Kausaleffekt der Antwortlatenz zu vermuten; bestenfalls wäre hier noch anzunehmen, dass unbeobachtete Variablen – wie z. B. nicht gemessene SD-Anreize – die Antwortlatenz beeinflussen und die unbeobachteten Merkmale vermittelt über die Antwortlatenz wirken (das ist im Grunde genommen auch die Argumentation der Framing-Theorie). Realistischer erscheint uns aber, hinter dem Effekt unbeobachtete Variablen zu vermuten, die sowohl die Antwortlatenz, als auch das Antwortverhalten beeinflussen. Bei Kontrolle dieser unbeobachteten Variablen würde der Latenzeffekt dann wegfallen. Diesbezüglich könnten zwei Mechanismen vermutet werden: Zum einen erscheint es möglich, dass durch die Tailored Design Method (TDM) tendenziell widerspenstige Befragte rekrutiert werden, die per se zu unehrlichem Antwortverhalten neigen (also auch lügen würden, wenn sie schnell antworteten). Zum anderen könnte speziell der Prozess der TDM (mehrmaliges Erinnern und Auffordern per Postkarte) Verärgerung, gefühlten Druck oder andere Effekte auslösen, die dann zu einem verzerrten Antwortverhalten führen. Aber wir wollen an dieser Stelle auf weitere ad hoc-Mutmaßungen verzichten und stattdessen zunächst einmal dazu auffordern, die Belastbarkeit unseres Antwortlatenzbefundes in weiteren Validierungsstudien zu überprüfen.

## 5. Schlussbemerkungen

Der vorliegende Beitrag hat auf der Basis einer schriftlichen Befragung zentrale Determinanten wahrer Antworten in Umfragen untersucht. Dies geschah im Rahmen einer Validierungsstudie, die sich auf eine Gruppe von Personen bezog, von denen wir aus Akten der Justizbehörden wussten, dass sie im Zeitraum 2005–2008 wegen eines kriminellen Delikts verurteilt worden waren. So konnten die Antworten auf die Deliktfrage gegen den tatsächlichen Wert abgeglichen werden.

Neben einer Reihe bemerkenswerter Einzelbefunde, z. B. dass insgesamt knapp zwei Drittel ihre strafrechtliche Verurteilung in der Befragung zugeben oder dass Frauen, ältere Personen und Befragte mit höherer Bildung bei der Deliktfrage tendenziell häufiger lügen, erscheinen zwei Ergebnisse für die Theorie und Praxis der Umfrageforschung von besonderer Relevanz:

Erstens hat sich gezeigt, dass die Variablen, die im Rahmen unterschiedlicher Varianten einer RC-Theorie der Befragung als Determinanten der Antwortvalidität benannt werden, tatsächlich eine wichtige Rolle spielen. Dies gilt vor allem für die SD-Neigung und den SD-Belief. Allerdings wird von Vertretern der SEU-Theorie (v. a. Esser 1986, 1991) zum einen das Konzept der SD-Neigung als Persönlichkeitsmerkmal kritisiert und zum anderen betont, dass nur bei einem gleichzeitigen Vorliegen einer hohen SD-Neigung und eines hohen SD-Belief mit einer Antwortverzerrung zu rechnen sei. Unsere Befunde widersprechen beiden Punkten, denn wir finden signifikante Haupteffekte für die SD-Neigung und den SD-Belief, aber keinen signifikanten Interaktionseffekt. Mithin können unsere Ergebnisse zu einer gewissen Rehabilitierung des vielfach kritisierten klassischen Konzepts der SD-Neigung (in der Tradition von Crowne & Marlowe) beitragen. Von den drei Variablen der Framing-Modelle, die in Anlehnung an Stocké (2004) spezifiziert wurden, zeigen die generalisierte Umfrageeinstellung und die Umfrageerfahrung keine nennenswerten Einflüsse, während die Antwortlatenz (die hier aber anders als bei Stocké definiert und gemessen wurde) voll durchschlägt. Weiterhin konnten wir keine einem konsistenten Muster folgende Effekte von Interaktionstermen finden, an denen Framing-Variablen beteiligt waren, was im Ergebnis gegen die Framing-Modelle spricht.

Der zweite wichtige Befund unserer Studie ist der gerade schon angesprochene hochsignifikante Effekt der Antwortlatenz, definiert als verstrichene Zeit bis der Fragebogen zurückgeschickt wurde.

Diejenigen, die auf die Aufforderung zur Teilnahme an einer postalischen Befragung spät reagiert haben, sagen eindeutig seltener die Wahrheit. Dieses Ergebnis, das nach unserem Wissen in der bisherigen Literatur so nicht auftaucht (was aber auch daran liegt, dass Validierungsstudien noch immer selten sind), erscheint nicht zuletzt von Bedeutung für die seit einiger Zeit geführte allgemeinere Diskussion darüber, ob für Umfragen und die dabei erzielten Rücklaufquoten die „the more, the better“-Annahme gilt (z. B. Koch 1998; Stoop 2005). In der Regel wird in der Survey-Methodologie angenommen, dass Umfragen mit einer höheren Ausschöpfung der angezielten Stichprobe eine höhere Datenqualität haben, und deshalb wird empfohlen, im Rahmen des Vertretbaren alle Möglichkeiten zu nutzen, den Rücklauf zu steigern. Gestützt wird diese Position durch empirische Studien, die speziell etwa für postalische Befragungen zeigen, dass es durch Erinnerungs- und Mahnschreiben gelingt, auch tendenziell sperrige Respondenten in die Studie einzubeziehen. In einer neueren Studie von Kunz (2010) zum Thema „Kriminalität im höheren Lebensalter“ wird berichtet, dass spät Antwortende u. a. eine geringere Bildung und allgemein einen niedrigeren sozialen Status haben. Aber da taucht dann die laut Kunz überraschende Besonderheit auf, „dass mit zunehmenden Kontaktierungen deutlich weniger delinquente Personen in die Befragung einbezogen werden“ (ebd.: 151). Die Autorin schließt daraus, kriminelles Verhalten der Bevölkerung würde überschätzt, wenn man in schriftlichen Umfragen zu selbstberichteter Delinquenz die Zielpersonen nur einmal anschreibe. Folgt man allerdings den Ergebnissen unserer Studie, spricht einiges dafür, dass ihr Befund schlicht durch vermehrte Leugnen der spät Antwortenden zustande gekommen ist. Allgemeiner bedeutet dies: Wenn die spät Antwortenden in schriftlichen (und eventuell auch anderen) Umfragen es mit der Wahrheit tatsächlich weniger genau nehmen, dann vermindert deren Teilnahme die Datenqualität und die „the more, the better“-Annahme wäre in dieser Hinsicht kontraproduktiv. Es dürfte sich lohnen, in weiteren Validierungsstudien genauer zu untersuchen, in welchem Ausmaß und in welchen inhaltlichen Bereichen unser Ergebnis eines Wahrheitsschwundes in Abhängigkeit von der „time-to-response“ die Brauchbarkeit von Umfragedaten gefährdet.

## Anhang

**Tabelle A1** Skala zur Messung des Bedürfnisses nach sozialer Anerkennung (SD-Neigung)

	„trifft zu“ in %
1. Manchmal bin ich beleidigt, wenn es nicht nach meinem Willen geht.	58,8
2. Ich bin stets ein guter Zuhörer, gleichgültig, wer mein Gesprächspartner ist.	69,2
3. Ich kann mich daran erinnern, dass ich schon mal krank gespielt habe, um eine Pflicht zu umgehen.	40,5
4. Bei Gelegenheit habe ich schon einmal jemanden ausgenutzt.	35,3
5. Wenn ich einen Fehler gemacht habe, bin ich stets bereit, das zuzugeben.	82,4
6. Ich halte mich immer selbst an Grundsätze, deren Befolgung ich von anderen erwarte.	74,3
7. Ich bin stets höflich, selbst zu Leuten, die ich abstoßend finde.	64,3
8. Ich habe noch nie absichtlich etwas gesagt, um die Gefühle anderer zu verletzen.	46,4
9. Zur Not lüge ich auch schon mal.	33,3

*Erläuterung:* Die Items 1–8 stammen von Stocké (2009a); die Personen wurden gefragt, welche Aussagen auf sie zutreffen; mögliche Antworten waren 0 („trifft nicht zu“) und 1 („trifft zu“). Item 9 entstammt der BIDR-Skala von Paulhus (1991, hier in deutscher Fassung von Winkler et al. 2006); dabei konnte auf einer fünfstufigen Skala von 1 („trifft überhaupt nicht zu“) bis 5 („trifft voll und ganz zu“) geantwortet werden; zur Bildung des additiven Indexes wurde das Item in eine Dummy-Variable umgewandelt mit 0 („trifft nicht zu“; 1,2,3=0) und 1 („trifft zu“; 4,5=1). Für den Index wurden die Items 1, 3, 4 und 9 umgepolt.

**Tabelle A2** Skala zur Messung der Erwünschtheitswahrnehmungen (SD-Belief)

	„verwerflich“ in %
1. Was meinen Sie: Wie verwerflich oder wie ungern gesehen ist es, in unserer Gesellschaft strafrechtlich verurteilt zu sein?	70,2
	„stimme zu“ in %
2. In den heutigen schwierigen Zeiten kann man auch mal gegen Gesetze verstoßen, wenn es nicht anders geht.	17,2
3. Kleinere Delikte sind nicht so schlimm; jeder schlägt mal über die Stränge.	25,1
4. Meine Freunde und Bekannten fänden es schlimm, wenn ich selbst kriminell würde.	83,8

*Erläuterung:* Die Antwortskala bei Item 1 war siebenstufig von 1 („überhaupt nicht verwerflich“) bis 7 („sehr verwerflich“). Der oben ausgewiesene Anteil „verwerflich“ fasst die Ausprägungen 4–7 zusammen. Die Antwortskala bei den Items 2–4 war fünfstufig mit 1 („stimme überhaupt nicht zu“) bis 5 („stimme voll und ganz zu“). Die oben ausgewiesenen Anteile zustimmender Antworten fassen die Ausprägungen 4 und 5 zusammen. Für die Indexbildung wurde Item 1 in eine fünfstufige Skala rekodiert, außerdem wurden die Items 2 und 3 umgepolt.

**Tabelle A3** Skala zur Messung der generalisierten Umfrageeinstellung

	„stimme zu“ in %
1. Umfragen sind sehr wichtig für Wissenschaft, Politik und Wirtschaft.	61,6
2. Umfragen bringen Abwechslung und sind interessant.	40,3
3. Die Ergebnisse von Umfragen sind in den meisten Fällen richtig.	34,3
4. Die Teilnehmer an Umfragen sagen meistens ihre wahre Meinung.	43,6

*Erläuterung:* Die vier Items stammen von Stocké (2009b). Die Antwortskala war fünfstufig von 1 („stimme überhaupt nicht zu“) bis 5 („stimme voll und ganz zu“). Die oben ausgewiesenen Anteile zustimmender Antworten fassen die Ausprägungen 4 und 5 zusammen.

**Tabelle A4** Deskriptive Statistiken der Analysevariablen

	Mittelwert	Standardabweichung	N
Strafrechtlich verurteilt (1=Ja)	0,647		283
Personenübereinstimmung (1=Ja)	0,936		295
Geschlecht (1=Mann)	0,831		295
Alter (Jahre)	40,634	14,599	295
Deliktschwere	5,336	1,632	301
Bildungsjahre	11,093	2,754	289
SD-Neigung	0,632	0,231	295
SD-Belief	3,939	0,879	290
Umfrageeinstellung	3,356	0,840	291
Umfrageerfahrung	2,343	1,841	293
Antwortlatenz (Tage)	22,326	1,841	293

**Tabelle A5** Matrix der bivariaten Korrelationen

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Wahrheit (1 = Ja)										
2. Personenüb. (1 = Ja)	0,214***									
3. Geschlecht (1 = M)	0,144*	0,138*								
4. Alter	-0,143*	-0,044	-0,028							
5. Deliktschwere	-0,120*	0,040	-0,063	0,146*						
6. Bildung	-0,141*	0,008	0,015	0,086	0,073					
7. SD-Neigung	-0,222***	-0,036	-0,095	0,126*	0,135*	0,020				
8. SD-Belief	-0,246***	0,048	0,130*	0,155**	0,224***	0,126*	0,297***			
9. Umfrage- einstellung	0,051	0,067	0,036	-0,001	-0,035	-0,116 <sup>+</sup>	0,115 <sup>+</sup>	0,103 <sup>+</sup>		
10. Umfrage- erfahrung	0,004	-0,039	-0,044	0,045	0,050	0,248***	-0,008	0,107 <sup>+</sup>	0,050	
11. Antwortlatenz	-0,163**	-0,045	0,049	-0,053	-0,121*	-0,133*	-0,017	-0,075	-0,029	0,141*

Erläuterung: Signifikanzniveaus <sup>+</sup>p<=0,1, \*p<=0,05, \*\*p<=0,01, \*\*\*p<=0,001.

## Literatur

- Barnett, J., 1998: Sensitive Questions and Response Effects. An Evaluation. *Journal of Managerial Psychology* 13: 63–76.
- Barton, A.H., 1958: Asking the Embarrassing Question. *Public Opinion Quarterly* 22: 67–68.
- Benson, L.E., 1941: Studies in Secret-Ballot Technique. *Public Opinion Quarterly* 5: 79–82.
- Bernstein, R., A. Chadha & R. Montjoy, 2001: Overreporting Voting. Why it Happens and Why it Matters. *Public Opinion Quarterly* 65: 22–44.
- Best, H. & C. Wolf, 2010: Logistische Regression. S. 827–854 in: C. Wolf & H. Best (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse*. Wiesbaden: VS.
- Boe, E.E. & W.S. Kogan, 1964: An Analysis of Various Methods for Deriving the Social Desirability Score, *Psychological Reports* 14: 23–29.
- Bradburn, N.M., S. Sudman & B. Wansink, 2004: *Asking Questions*. San Francisco, CA: Wiley.
- Bundeskriminalamt (Hrsg.), 2010: *Polizeiliche Kriminalstatistik 2009*. Bundesrepublik Deutschland. Wiesbaden: Bundeskriminalamt.
- Crowne, D.P. & D. Marlowe, 1960: A New Scale of Social Desirability Independent of Psychopathology. *Journal of Consulting Psychology* 24: 349–354.
- Crowne, D.P. & D. Marlowe, 1964: *The Approval Motive. Studies in Evaluative Dependence*. New York: Wiley.
- DeMaio, T.J., 1984: Social Desirability and Survey Meas-

- urement: A Review. S. 257–282 in: C.F. Turner & E. Martin (Hrsg.), *Surveying Subjective Phenomena*. New York: Sage Foundation.
- Diekmann, A., 2007: *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*. 18. Auflage. Reinbek: Rowohlt.
- Dillman, D.A., 1978: *Mail and Telephone Surveys. The Total Design Method*. New York: Wiley.
- Dillman, D.A., J.D. Smyth & L.M. Christian, 2009: *Internet, Mail, and Mixed-Mode Surveys. The Tailored Design Method*. Hoboken: Wiley.
- Elffers, H., H.S.J. Robben & D.J. Hessing., 1992: On Measuring Tax Evasion. *Journal of Economic Psychology* 13: 546–567.
- Esser, H., 1986: Können Befragte lügen? Zum Konzept des „wahren“ Wertes im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 38: 314–336.
- Esser, H., 1991: Die Erklärung systematischer Fehler in Interviews: Befragtenverhalten als „rational choice“. S. 59–78 in: R. Wittenberg (Hrsg.), *Person – Situation – Institution – Kultur*. Berlin: Duncker & Humblot.
- Esser, H., 2001: *Soziologie. Spezielle Grundlagen. Band 6: Sinn und Kultur*. Frankfurt: Campus.
- Esser, H., 2010: Das Modell der Frame-Selektion. Eine allgemeine Handlungstheorie für die Sozialwissenschaften?. S. 45–62 in: G. Albert & S. Sigmund (Hrsg.), *Soziologische Theorie kontrovers*. Sonderheft 50 der *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Wiesbaden: VS.
- Hartmann, P. 1991: *Wunsch und Wirklichkeit. Theorie und Empirie sozialer Erwünschtheit*. Wiesbaden: Deutscher Universitäts-Verlag.
- Hessing, D.J., H. Elffers & R.H. Weigel, 1988: Exploring the Limits of Self-Reports and Reasoned Action. An Investigation of the Psychology of Tax Evasion Behavior. *Journal of Personality and Social Psychology* 54: 405–413.
- Holtgraves, T., 2004: Social Desirability and Self-Reports: Testing Models of Socially Desirable Responding. *Personality and Social Psychology Bulletin* 30: 161–172.
- Hyman, H., 1944: Do They Tell the Truth?. *Public Opinion Quarterly* 8: 557–559.
- Johnson, T.P. & M. Fendrich, 2002: *A Validation of the Crowne-Marlowe Social Desirability Scale*, Chicago: University of Illinois, Survey Research Laboratory.
- Junger, M., 1989: Discrepancies between Police and Self-Report Data for Dutch Racial Minorities. *British Journal of Criminology* 29: 273–283.
- Koch, A., 1998: Wenn „mehr“ nicht gleichbedeutend mit „besser“ ist: Ausschöpfungsquoten und Stichprobenverzerrungen in allgemeinen Bevölkerungsumfragen. *ZUMA-Nachrichten* 42: 66–93.
- Köllisch, T. & D. Oberwittler, 2004: Wie ehrlich berichten männliche Jugendliche über ihr delinquentes Verhalten? Ergebnisse einer externen Validierung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 56: 708–735.
- Kroneberg, C., 2005: *Die Definition der Situation und die variable Rationalität der Akteure. Ein allgemeines Modell des Handelns*. *Zeitschrift für Soziologie* 34: 344–363.
- Kunz, F., 2010: Mahnaktionen in postalischen Befragungen. Empirische Befunde zu Auswirkungen auf den Rücklauf, das Antwortverhalten und die Stichprobenszusammensetzung. *Methoden – Daten – Analysen* 4: 127–155.
- Lee, R.M., 1993: *Doing Research on Sensitive Topics*. London: Sage.
- Lensvelt-Mulders, G., 2008: Surveying Sensitive Topics. S. 461–478 in: E.D. de Leeuw, J.J. Hox & D.A. Dillman (Hrsg.), *International Handbook of Survey Methodology*. New York: Erlbaum.
- Mayerl, J., 2009: *Kognitive Grundlagen sozialen Verhaltens: Framing, Einstellungen und Rationalität*. Wiesbaden: VS.
- Norton, E.C., H. Wang & C. Ai, 2004: Computing Interaction Effects and Standard Errors in Logit and Probit Models. *The Stata Journal* 4: 154–167.
- Paulhus, D.L., 1991: Measurement and Control of Response Bias. S. 17–59 in: J.P. Robinson, P.R. Shaver & L.S. Wrightsman (Hrsg.), *Measures of Personality and Social Psychological Attitudes*. San Diego: Academic Press.
- Paulhus, D.L., 2002: Socially Desirable Responding: The Evaluation of a Construct. S. 49–69 in: H.I. Braun, D.N. Jackson & D.E. Wiley (Hrsg.), *The Role of Constructs in Psychological and Educational Measurement*. Mahwah: Erlbaum.
- Phillips, D.L. & K.J. Clancy., 1972: Some Effects of “Social Desirability” in Survey Studies. *American Journal of Sociology* 77: 921–940.
- Richman, W.L., S. Kiesler, S. Weisband & F. Drasgow, 1999: A Meta-Analytic Study of Social Desirability Distortion in Computer-Administered Questionnaires, Traditional Questionnaires, and Interviews. *Journal of Applied Psychology* 84: 754–775.
- Silver, B.D., B.A. Anderson & P.R. Abramson, 1986: Who Overreports Voting? *American Political Science Review* 80: 613–624.
- Stocké, V., 2004: Entstehungsbedingungen von Antwortverzerrungen durch soziale Erwünschtheit. Ein Vergleich der Prognosen der Rational-Choice-Theorie und des Modells der Frame-Selektion. *Zeitschrift für Soziologie* 33: 303–320.
- Stocké, V., 2009a: Deutsche Kurzskaala zur Erfassung des Bedürfnisses nach sozialer Anerkennung. In: A. Glöckner-Rist (Hrsg.), *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. ZIS Version 13.00. Bonn: GESIS.
- Stocké, V., 2009b: Einstellungen zu Umfragen. In: A. Glöckner-Rist (Hrsg.), *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. ZIS Version 13.00. Bonn: GESIS.
- Stocké, V. & C. Hunkler, 2004: Die angemessene Erfassung der Stärke und Richtung von Anreizen durch soziale Erwünschtheit. *ZA-Information* 54: 53–88.
- Stoop, I., 2005: *The Hunt for the Last Respondent. Non-response in Sample Surveys*. The Hague: Social and Cultural Planning Office.

- Tourangeau, R., L.J. Rips & K.A. Rasinski, 2000: *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Tourangeau, R. & T. Yan, 2007: Sensitive Questions in Surveys. *Psychological Bulletin* 133: 859–883.
- van der Heijden, P.G.M., G. van Gils, J. Bouts & J.J. Hox, 2000: A Comparison of Randomized Response, Computer-Assisted Self-Interview, and Face-to-Face Direct Questioning. Eliciting Sensitive Information in the Context of Welfare and Unemployment Benefit. *Sociological Methods & Research* 28: 505–536.
- Winkler, N., M. Kroh & M. Spiess, 2006: Entwicklung einer deutschen Kurzskala zur zweidimensionalen Messung von sozialer Erwünschtheit. *Discussion Papers* 579. Berlin: DIW.
- Wolter, F., 2012: Heikle Fragen in Interviews: Eine theoretische und empirische Validierung der Randomized Response-Technik. Wiesbaden: VS (im Erscheinen).

### Autorenvorstellung

Anna Skarbak-Kozietulska, geb. 1983 in Ilmenau. Studium der Soziologie, Publizistik und der Deutschen Philologie in Warschau, Frankfurt am Main und Mainz (Abschluss 2009 in Mainz). Von 2009 bis 2011 wissenschaftliche Mitarbeiterin am Institut für Soziologie der Universität Mainz. Seit 2011 Analystin in einem Versicherungsunternehmen.

Peter Preisendörfer, geb. 1953 in Obersinn. Studium der Soziologie in Nürnberg, Wien, Chicago. Promotion 1984 in Nürnberg. Habilitation 1994 in München. Seit 1995 Professor für Soziologie zuerst in Rostock, jetzt in Mainz. Forschungsschwerpunkte: Organisationssoziologie, Entrepreneurship-Forschung, Umweltsoziologie, Methoden der quantitativen Sozialforschung.  
Ausgewählte Publikationen: *Umweltsoziologie* (mit A. Diekmann), Reinbek 2001; *Haushalte ohne Auto* (mit M. Rinn), Opladen 2003; *Der Erfolg neugegründeter Betriebe* (mit J. Brüderl und R. Ziegler), Berlin 2007; *Organisationssoziologie*, Wiesbaden 2011; zuletzt in dieser Zeitschrift: *Zahlungsbereitschaft für kollektive Umweltgüter* (mit U. Liebe), *ZfS* 36, 2007: 326–345.

Felix Wolter, geb. 1979 in Berlin. Studium der Soziologie, Politikwissenschaft und BWL in Mainz und Lyon (Frankreich). Dissertationsprojekt „Heikle Fragen in Interviews: Eine theoretische und empirische Validierung der Randomized Response-Technik“ (angenommen vom Fachbereich 02 der Universität Mainz). Seit 2007 wissenschaftlicher Mitarbeiter am Institut für Soziologie der Universität Mainz.  
Forschungsschwerpunkte: Methoden der empirischen Sozialforschung, Bildung und Weiterbildung, Arbeitsmarktsoziologie.  
Ausgewählte Publikationen: *Einkommenseffekte beruflicher Weiterbildung. Empirische Analysen auf Basis des Mikrozensus-Panels* (mit J. Schiener), *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 61, 2009; *Struktur und Dynamik der Ungleichheit von Erwerbslosigkeitsrisiken. Empirische Analysen auf Basis der Mikrozensus-Erhebungen 1996 bis 2004*, *Wirtschaft und Statistik* 2009; *Bildungschancen von Grundschulern: Die Bedeutung des Klassen- und Schulkontextes am Übergang auf die Sekundarstufe I* (mit A. Schulze und R. Unger), *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 61, 2009; *Heikle Fragen in Interviews: Eine theoretische und empirische Validierung der Randomized Response-Technik*, Wiesbaden 2012 (im Erscheinen).