

Informationsverfügbarkeit und Response-Effects: die Prognose von Einflüssen unterschiedlich kategorisierter Antwortskalen durch Antwortsicherheiten und Antwortlatenzen

Stocké, Volker

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Stocké, V. (2003). Informationsverfügbarkeit und Response-Effects: die Prognose von Einflüssen unterschiedlich kategorisierter Antwortskalen durch Antwortsicherheiten und Antwortlatenzen. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 52, 6-36. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-198938>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Informationsverfügbarkeit und Response-Effects: Die Prognose von Einflüssen unterschiedlich kategorisierter Antwortskalen durch Antwortssicherheiten und Antwortlatenzen

von Volker Stocké¹

Zusammenfassung

Die Angaben von Befragten über die Häufigkeit oder Dauer bestimmter Verhaltensweisen werden in vielen Fällen durch die Verwendung unterschiedlich kategorisierter, objektiv aber identischer Antwortskalen beeinflusst. Als zentrale Entstehungsbedingung dieses Response-Effects wird die kognitive Verfügbarkeit der jeweils abgefragten Informationen vermutet. In der vorliegenden Untersuchung wird diese Hypothese getestet, indem die individuellen Unterschiede in der subjektiven Antwortssicherheit und der jeweils zur Antwortgenerierung benötigten Zeit analysiert werden. Die Studie vergleicht die relative Prognosekraft dieser beiden Indikatoren der Informationsverfügbarkeit für die Entstehung des untersuchten Response-Effects. Ein weiterer Untersuchungsgegenstand betrifft die bislang ungeklärte Frage, durch welche der Datenanalyse vorangehende Transformationen der Antwortlatenzen deren Validität und Vorhersagekraft für das Vorliegen von Response-Effects maximiert werden können. Es werden drei Vorschläge hierzu aufgegriffen und deren Validität mit der von untransformierten Rohreaktionszeiten verglichen. Die Daten eines Feldexperimentes mit einer lokalen Zufallsstichprobe zeigen, dass der Einfluss unterschiedlich kategorisierter Antwortskalen auf den von den Befragten berichteten Fernsehkonsum sowohl durch die Antwortssicherheiten als auch durch die Antwortlatenzen vorhergesagt werden kann. Dabei kommt der Antwortgeschwindigkeit jedoch eine deutlich stärkere Prognosekraft zu. Beim Vergleich der

¹ Dr. **Volker Stocké** ist Hochschulassistent am Sonderforschungsbereich 504 „Rationalitätskonzepte, Entscheidungsverhalten und ökonomische Modellierung“ der Universität Mannheim, L13, 15; D-68131 Mannheim. E-Mail: vstocke@rumms.uni-mannheim.de

unterschiedlich transformierten Versionen der Antwortlatenzen erweist sich die am Mittelwert der individuellen Antwortgeschwindigkeiten zentrierte Variante als bester Prädiktor der Skaleneinflüsse.

Abstract

Respondents' self reports about the frequency or duration of certain activities are often found to be influenced by differences in the categorisation of substantially identical response scales. The cognitive availability of the information to be recalled is supposed to be the pivotal determinant of this type of response-effect. In the following article this hypothesis is tested, first of all, by analysing individual differences in the interviewees' response certainty and the response latencies needed to answer the focal question. Secondly, the predictive power of these two indicators of information availability for the existence of the response-effects under investigation is compared. Another open question addressed in this paper is how response latencies should be transformed prior to data analysis in order to maximise their accuracy in predicting response-effects. For this purpose, the validity of three transformations proposed in the literature with those of raw response-latencies is compared. Data from a local survey based on a random probability sample is used to answer these questions. The results show that both, response certainty and response latencies are able to predict response scale effects on the reported television consumption. However, response latencies proved to be the much stronger predictor. A comparison of differently transformed versions of the response latencies reveals that difference-scores between the target response latencies and the individual interviewees' mean response speed during the rest of the interview is the best predictor for their susceptibility to the investigated response-effect.

1 Einleitung²

Empirische Untersuchungsergebnisse haben gezeigt, dass sich scheinbar unbedeutende Unterschiede in der Präsentation von Antwortoptionen teilweise dramatisch

2 Der Autor dankt den Teilnehmern des 56th Annual Meeting of the American Association for Public Opinion Research 2001 (Montreal, Kanada, 17.-20.05.2002), einem anonymen Gutachter dieser Zeitschrift und speziell **Bürbel Knäuper** für hilfreiche Kommentare zu einer früheren Version dieser Arbeit. **Hartmut Esser** gilt ein ganz besonderer Dank für anregende Diskussionen und zielsichere Kommentare. **Birgit Becker** war eine große Unterstützung bei der Datenaufbereitung und **Eva Feuerbach** sowie **Tobias Stark** bei der Erstellung des Manuskriptes. Der Beitrag wurde durch finanzielle Förderung der Deutschen Forschungsgemeinschaft an den Sonderforschungsbereich 504 der Universität Mannheim ermöglicht.

auf das Antwortverhalten von Umfrageteilnehmern auswirken können. So werden etwa unterschiedliche Angaben der Befragten beobachtet, wenn die Antwortmöglichkeiten bei Items vom Likert-Typ entweder mit Zahlen von 0 bis 10 oder von 1 bis 11 markiert werden (*Schwarz* et al. 1998). Auch die Verwendung von Antwortskalen mit einer ungeraden statt einer geraden Anzahl an Antwortoptionen und damit die Bereitstellung einer mittleren Antwortmöglichkeit wirkt sich in vielen Fällen auf die inhaltlichen Angaben der Befragten aus (*Pecher* und *Raaijmakers* 1999).

Ein für die Praxis der Umfrageforschung besonders schwerwiegendes Problem besteht darin, dass die Angaben von Befragten über die Häufigkeit bestimmter Verhaltensweisen oder die zeitliche Länge alltäglicher Erlebnisse durch eine unterschiedliche Kategorisierung der Antwortskalen beeinflusst werden. Statt einer offenen Abfrage der anvisierten Informationen wird das Antwortkontinuum oft zu Intervallen zusammengefasst und den Befragten dann als distinkte Antwortmöglichkeiten vorgelegt. Durch diese Vorgehensweise sollen die zur Antwortgenerierung notwendigen kognitiven Anforderungen reduziert und gleichzeitig das Ausmaß des Item-Nonresponse minimiert werden. Ein unerwünschter Nebeneffekt dieser Maßnahme zur Verbesserung der Datenqualität besteht darin, dass sich die Art der Kategorisierung des Antwortkontinuums bei unterschiedlichsten Befragungsthemen auf die inhaltlichen Angaben der Befragten auswirkt. So unterscheiden sich die Antworten teilweise dramatisch, wenn entweder im hohen oder niedrigen Bereich des Häufigkeitskontinuums eine differenziertere Kategorisierung vorgenommen und den Befragten damit eine größere oder kleinere Anzahl an Antwortoptionen vorgelegt wird. So berichten die Befragten bei Verwendung einer hoch-frequenten, im Vergleich zu einer nieder-frequenten Antwortskala mehr Konsumausgaben (*Menon* et al. 1997, *Winter* 2002), ein größeres Ausmaß an sexueller Aktivität (*Schwarz* und *Scheuring* 1988), häufigere Bedrohungsgefühle (*Gaskell* et al. 1994) sowie eine größere Anzahl an Albträumen (*Ji* et al. 2000) und kulturellen Aktivitäten (*Bless* et al. 1992).

Theoretische Erklärungen für Einflüsse unterschiedlich kategorisierter Häufigkeitsskalen gehen davon aus, dass die mangelhafte kognitive Verfügbarkeit der jeweils abgefragten Informationen im Gedächtnis der Befragten die zentrale Entstehungsbedingung für diese Art von Response-Effects ist (*Schwarz* und *Hippler* 1987)³.

3 Mit „kognitiver Informationsverfügbarkeit“ ist im vorliegenden Beitrag die Leichtigkeit gemeint, mit der eine abgefragte Information von den Probanden aus dem Gedächtnis abgerufen werden kann. Dabei ist nicht relevant, ob eine hohe Informationsverfügbarkeit auf einer „chronischen“ Zugänglichkeit der Information im Gedächtnis beruht oder darauf, dass deren Abruf durch Kontexteffekte erleichtert wurde.

Diese Annahme wird dadurch gestützt, dass sich Subgruppen von Befragten, die vermutlich die betreffende Information mit geringerer Wahrscheinlichkeit im Gedächtnis verfügbar haben, als stärker beeinflussbar erweisen (*Schwarz* 1999). Auch die Beobachtung, wonach bei Befragungsthemen mit wahrscheinlich geringerer Salienz für die Befragten stärkere Skaleneinflüsse vorliegen, kann ebenfalls als Hinweis auf die Bedeutung der Informationsverfügbarkeit interpretiert werden (*Rockwood* et al. 1997). Trotz Plausibilität dieser Argumentation liegt beim derzeitigen Forschungsstand keine direkte Evidenz für die Prognosekraft unterschiedlicher Grade der Informationsverfügbarkeit vor.

Ein deutlich zwingenderer Nachweis der Rolle mangelnder Informationsverfügbarkeit ist dann möglich, wenn individuelle Unterschiede in dieser Hinsicht direkt bei den Befragten erfasst werden und dann deren Erklärungskraft für die relative Beeinflussbarkeit durch Skaleneinflüsse gezeigt wird. Eine solche Vorgehensweise macht jedoch die valide Messung unterschiedlicher Grade der Informationsverfügbarkeit notwendig. Im Bereich faktischer Befragungsgegenstände kommen hierfür zwei unterschiedliche Verfahren in Betracht. Dabei handelt es sich *erstens* um das von den Befragten selbstberichtete Ausmaß ihrer Antwortsicherheit als subjektiver Indikator der Informationsverfügbarkeit. *Zweitens* ist die für die Antwortgenerierung benötigte Zeit ein objektiver Indikator für die Verfügbarkeitsdimension (*Wegener* et al. 1995). Was die Validität dieser beiden Messkonzepte angeht, so liegen im Bereich der Einstellungsmessung teilweise inkonsistente und bei der Abfrage faktischer Informationen praktisch keine Untersuchungsergebnisse vor.

Die sehr heterogene Praxis bei der vorbereitenden Behandlung von Antwortlatenzen ist eine mögliche Ursache für die inkonsistente Ergebnislage über deren Validität als Indikator der Informationsverfügbarkeit. So werden Latenzdaten vor der Datenanalyse in unterschiedlicher Weise transformiert, um eine Reduktion ihrer charakteristischen Schiefe zu erreichen. Es herrscht außerdem kein Konsens darüber, wie mit dispositionalen und somit über verschiedene Befragungsthemen stabilen Unterschieden in der Antwortgeschwindigkeit zwischen den Befragten umgegangen werden soll. Solche Unterschiede werden teilweise als Störgröße betrachtet und durch eine Standardisierung an den mittleren Antwortgeschwindigkeiten der Befragten eliminiert. Andere Forscher sehen in diesem Varianzbestandteil eine relevante Erklärungsgröße und belassen diese in den Latenzdaten (*Bargh* und *Chartrand* 2000, *Fazio* 1990). Durch welche der Vorgehensweisen die Konstruktvalidität und damit Prognosekraft der Antwortlatenzen maximiert werden kann, wurde bisher empirisch nicht systematisch untersucht.

Die vorliegende Untersuchung verfolgt insgesamt drei, eng miteinander verbundene Zielsetzungen. So soll *erstens* durch eine direkte Messung individueller Unterschiede

in der Informationsverfügbarkeit die Annahme untermauert werden, dass diese die Beeinflussbarkeit der Befragten durch unterschiedlich kategorisierte Antwortskalen vorhersagen. Diese, auf die inhaltlichen Bestimmungsfaktoren des Befragtenverhaltens abzielende Fragestellung bildet den Ausgangspunkt für zwei stärker messtheoretisch orientierte Zielsetzungen. So soll in einer vergleichenden Perspektive *zweitens* überprüft werden, ob den von den Befragten selbstberichteten Antwortsicherheiten als meta-kognitive Operationalisierung oder den Antwortlatenzen als operationalem Index der Informationsverfügbarkeit ein höheres Maß an Validität zukommt. Als Validierungskriterium wird hierbei die Fähigkeit der beiden Indikatoren zur Vorhersage der Einflüsse unterschiedlich formatierter Antwortskalen herangezogen. Die *dritte* Fragestellung der vorliegenden Arbeit bezieht sich auf den Einfluss verschiedener Transformationen von Reaktionszeitdaten auf deren Gültigkeit und damit Fähigkeit zur Prognose der Beeinflussbarkeit der Befragten durch Response-Effects. Die Datengrundlage der vorliegenden Untersuchung besteht in computergestützten Interviews mit einer Zufallsstichprobe von Befragten. Die Ergebnisse lassen somit auch praxisrelevante Aussagen darüber zu, ob das innovative, bisher aber hauptsächlich im Kontext von Laborexperimenten eingesetzte Instrument der Antwortlatenzen auch in einer „normalen“ Bevölkerungsumfrage gewinnbringend verwendet werden kann.

2 Forschungsstand

Der folgende Überblick über den relevanten Forschungsstand gliedert sich in drei Abschnitte. So werden im *ersten* Schritt die vorliegenden Ergebnisse über die Ursachen der Beeinflussbarkeit von Befragten durch unterschiedlich kategorisierte Antwortskalen dargestellt. Der *zweite* Abschnitt befasst sich dann mit dem Forschungsstand über die Validität subjektiver Antwortsicherheiten und Antwortlatenzen als alternative Indikatoren für die kognitive Verfügbarkeit der abgefragten Informationen. Im *dritten* Abschnitt wird schließlich die Angemessenheit unterschiedlicher Möglichkeiten der Transformation von Antwortlatenzen diskutiert.

2.1 Informationsverfügbarkeit als Ursache für Skaleneinflüsse

Ein besonders intensiv erforschter Befragungsgegenstand, bei dem regelmäßig Einflüsse unterschiedlich kategorisierter Antwortskalen beobachtet werden, ist die Erfassung des täglichen Fernsehkonsums der Befragten (*Bless* et al. 1992, *Menon* et al. 1995, *Rockwood* et al. 1997, *Schwarz* 1988, *Schwarz* und *Bienias* 1990, *Schwarz* et al. 1985, *Schwarz* und *Hippler* 1987, *Winter* 2002). Die Probanden beantworten hierbei die Frage nach der Länge ihres täglichen Fernsehkonsums im

Rahmen eines „Split-Ballot“-Experimentes entweder mithilfe einer hoch- oder nieder-frequenten Antwortskala. Dabei steigt die hoch-frequente Skalenversion, ausgehend von der Option „bis zu 2.5 Stunden“, in Schritten von einer halben Stunde bis zur höchsten Antwortmöglichkeit „mehr als 4.5 Stunden“ an. Dagegen variiert die nieder-frequente Skala, ebenfalls in halbstündigen Intervallen, zwischen den Antwortoptionen „bis zu 0.5 Stunden“ und „mehr als 2.5 Stunden“. Durch eine Dichotomisierung der Angaben in die beiden Kategorien „2.5 Stunden oder weniger Fernsehkonsum“ und „mehr als 2.5 Stunden Fernsehkonsum“ lassen sich die Antwortinhalte bei den beiden Skalenversionen vergleichbar machen. Die Ergebnisse beim ersten Experiment mit dieser Versuchsanordnung zeigen, dass bei Verwendung des nieder-frequenten Skalentyps 16.2 Prozent und bei einer hoch-frequenten Art der Kategorisierung 37.5 Prozent der Probanden in die Kategorie der „Vielseher“ fallen (*Schwarz et al. 1985*).

Neben anderen Faktoren, wie etwa der Regelmäßigkeit und Beobachtbarkeit des abgefragten Verhaltens (*Menon et al. 1995, Ji et al. 2000*), der Schwierigkeit der Fragestellung sowie der Motivation der Befragten (*Stocké 2001*), wird das Ausmaß der kognitiven Verfügbarkeit der jeweils abgefragten Information als zentrale Entstehungsbedingung der Skaleneinflüsse angesehen (*Schwarz und Hippler 1987*). Dabei wird angenommen, dass die Befragten, auch wenn ihnen die jeweils abgefragten, meist alltäglichen Ereignisse nicht als episodisch organisierte Informationen im Gedächtnis verfügbar sind, dennoch zu einer sinnvollen Beantwortung der betreffenden Frage motiviert sind. Daher ziehen sie die Formatierung der jeweils vorgelegten Häufigkeitsskala als Heuristik bei der Auswahl der angemessenen Antwortoption heran. Die Befragten gehen davon aus, dass sich in der Art der Kategorisierung des Antwortkontinuums das Wissen des Sponsors der Befragung über die wahre Verteilung des betreffenden Merkmals in der Gesamtpopulation widerspiegelt. Aus der Sicht der Probanden repräsentiert somit die jeweils mittlere Antwortoption der Skalen die durchschnittliche Prävalenz des Merkmals in der Gesellschaft. Ausgehend von diesem Referenzpunkt wählen die Befragten dann jene Antwortoption, die im Vergleich zu ihrer Überzeugung darüber, wie stark über- oder unterdurchschnittlich ihre eigene Ausprägung auf der betreffenden Dimension ist, mehr oder weniger weit über oder unter dem jeweiligen Skalenmittelpunkt liegt. Da sich der Inhalt der mittleren Antwortoption zwischen den Skalentypen deutlich unterscheidet, gelangen die Probanden bei Verwendung der gleichen Antwortgenerierungsstrategie zu unterschiedlichen Antwortinhalten.

Einige Studien finden Hinweise auf die empirische Angemessenheit der skizzierten Erklärung für die beobachteten Skaleneinflüsse. So konnte etwa gezeigt werden, dass die Befragten den durchschnittlichen Fernsehkonsum in der Gesamtgesell-

schaft höher einschätzen, wenn sie zuvor Fragen über ihr eigenes Fernsehverhalten bei Verwendung einer hoch-frequenten statt einer nieder-frequenten Antwortskala beantwortet haben (*Schwarz* und *Hippler* 1987). Dieses Ergebnis kann als Beleg für die informative Funktion unterschiedlich kategorisierter Antwortskalen verstanden werden. Andere Forschungsergebnisse finden stärkere Einflüsse des verwendeten Skalentyps auf das Antwortverhalten, wenn die Befragten statt über ihr eigenes Fernsehverhalten, Angaben über den Fernsehkonsum ihrer Freunde oder über den von Fremden machen sollen (*Schwarz* und *Bienias* 1990). Dieses Ergebnis wird mit Hinweis darauf interpretiert, dass die bei den verschiedenen Fragestellungen abgefragten Informationen den Befragten immer weniger verfügbar sind und diese daher immer stärker auf die informative Funktion der Antwortskalen angewiesen sind. Die Angaben von Studenten über die Länge ihres täglichen Fernsehkonsums werden stärker durch die Kategorisierung der Antwortskala beeinflusst als ihre Angaben über die Häufigkeit positiver Leistungsbewertungen in der Vergangenheit. Auch dieser Unterschied lässt sich als Ergebnis einer unterschiedlichen kognitiven Verfügbarkeit der beiden Informationen deuten (*Rockwood* et al. 1997). In einer weiteren Studie erweisen sich ältere im Vergleich zu jüngeren Befragten stärker durch die Kategorisierung der verwendeten Antwortskalen beeinflusst, wenn nach der Häufigkeit des Fleischkonsums gefragt wird (*Schwarz* 1999). Dieser Unterschied wird mit einem altersbedingten Rückgang der Gedächtnisleistung und der damit verbundenen geringeren Verfügbarkeit der betreffenden Information erklärt. Allerdings finden sich umgekehrte Unterschiede in der Beeinflussbarkeit der Altersgruppen bei Fragen nach der Häufigkeit von Krankheitssymptomen. Dies kann darauf zurückgeführt werden, dass ältere Menschen gesundheitsbezogenen Ereignissen mehr Aufmerksamkeit schenken, so dass diese besser im Gedächtnis verfügbar sind (*Schwarz* 1999).

2.2 Die Validität von Antwortlatenzen und der subjektiven Antwortsicherheit als Indikatoren der Informationsverfügbarkeit

Bei der direkten Erfassung unterschiedlicher Grade der Informationsverfügbarkeit lassen sich meta-kognitive und operationale Messverfahren unterscheiden (vgl. *Bassili* 1996b). Bei meta-kognitiven Verfahren werden die Befragten zusätzlich zu ihren inhaltlichen Angaben um eine qualifizierende Einschätzung ihrer Antworten gebeten. Dabei geben die Befragten beispielsweise an, für wie reflektiert sie ihre Angaben halten oder wie ambivalent sie zwischen verschiedenen Antwortmöglichkeiten sind. Der von den Befragten selbstberichteten Sicherheit ihrer Angaben kommt wegen der häufigen Verwendung dieses Indikators eine besonders hervorgehobene Stellung zu (vgl. als Überblick: *Wegener* et al. 1995). Operationale Indi-

katoren der Informationsverfügbarkeit beruhen dagegen auf Merkmalen des Antwortgenerierungsprozesses oder formalen Eigenschaften der resultierenden Angaben, die nicht auf den Angaben der Befragten beruhen. Neben der Extremität der Angaben im Bereich der Einstellungsmessung kann die Geschwindigkeit der Antwortgenerierung als wichtigster operationaler Indikator der Informationsverfügbarkeit angesehen werden (*Bassili* 1996b, *Bassili* und *Fletcher* 1991, *Fazio* 1990). Operationalen Messverfahren kommt möglicherweise daher ein höheres Ausmaß an Validität zu, weil diese nicht die anspruchsvolle Fähigkeit der Befragten zu einer Selbstreflexion ihrer Angaben voraussetzen (*Bassili* 1996b).

Bei der Verwendung von Antwortlatenzen liegt allerdings das Problem vor, dass die Validität dieses Indikators der Informationsverfügbarkeit durch miterfasste Fremddimensionen eingeschränkt sein kann. So werden durch Antwortlatenzen neben der für die Gedächtnissuche nach antwortrelevanten Informationen benötigten Zeit auch weiterer Teilaspekte des gesamten Antwortgenerierungsprozesses erfasst. Es handelt sich hierbei um die Zeit, die für die Interpretation der Frage, für die Selektion der passenden Antwortvorgabe und die Korrektur hinsichtlich potentiell bestehender Anforderungen durch soziale Erwünschtheit benötigt wird (*Schwarz* und *Strack* 1991). Demnach erhöhen beispielsweise schwere und unverständlich formulierte Fragen sowie die Verwendung komplexer Messinstrumente die Länge des Antwortgenerierungsprozesses (*Bassili* und *Scott* 1996). Bei bestehenden Anreizen durch soziale Erwünschtheit verlängern die kognitiven Aktivitäten bei der Ausbildung eines Kompromisses zwischen den sozialen Anforderungen an das Antwortverhalten und den „wahren“ Werten der Befragten die benötigte Antwortzeit. Es kann jedoch argumentiert werden, dass sich diese Messstörungen gleichförmig auf alle Befragten auswirken und somit die Antwortlatenzen nur um eine Konstante erhöhen. Unter diesen Umständen bleiben die relativen Unterschiede in der Antwortgeschwindigkeit zwischen Befragten mit hoher und geringer Informationszugänglichkeit unberührt.

Praktisch alle Untersuchungen der Validität subjektiver Antwortsicherheiten sowie Antwortlatenzen stammen aus dem Bereich der Einstellungsmessung und beziehen sich damit auf die Operationalisierung des Konzepts der Einstellungsstärke. Ähnlich wie bei der Erfassung von faktischen Informationen soll hier die kognitive Verfügbarkeit der zugrunde liegenden Urteile und Überzeugungen der Befragten erfasst werden. Die Stärke einer Einstellung lässt sich durch die Stabilität der Bewertungsinhalte und danach definieren, in welchem Umfang diese die Wahrnehmungen und Handlungen der Akteure vorhersagen können (*Krosnick* und *Petty* 1995). Demnach wäre bei starken Einstellungen auch mit einer hohen Resistenz gegenüber Einflüssen von Response-Effects auf das Antwortverhalten zu rechnen. Für die Einschät-

zung der Validität von Antwortsicherheiten und Antwortlatenzen sind daher Studien aus drei Bereichen relevant. Dabei handelt es sich *erstens* um Untersuchungen über die Prognosekraft der beiden Indikatoren für das Ausmaß der Einstellungs-Verhaltenskonsistenz, *zweitens* um Studien über die Vorhersagbarkeit der Einstellungsstabilität sowie *drittens* um deren Eignung zur Vorhersage der Beeinflussbarkeit unterschiedlicher Befragten durch Response-Effects.

Die Validität der subjektiv geäußerten Antwortsicherheit kann in einigen, jedoch keineswegs allen Untersuchungen belegt werden. Demnach steigt das Ausmaß der beobachteten Einstellungs-Verhaltenskonsistenz mit der von den Probanden selbstberichteten Sicherheit ihrer Bewertungsurteile an. So konnte festgestellt werden, dass die Prognosekraft der Einstellung gegenüber politischen Wahlen für die tatsächliche Wahlbeteiligung signifikant mit der selbstberichteten Sicherheit dieser Einstellungsinhalte ansteigt (*Warland* und *Sample* 1973). Auch finden sich zunehmend stärkere Zusammenhänge zwischen der Bewertung einer Liberalisierung von Schwangerschaftsabbrüchen und den Verhaltensabsichten der Probanden, wenn diese ihre Bewertungen in zunehmendem Ausmaß als sicherer bezeichnen (*Renata* 1999). Dagegen lässt sich in einer Untersuchung von *Mielke* (1985) nicht nachweisen, dass die Stärke des Zusammenhangs zwischen den Einstellungen der Befragten zum Umweltschutz und ihren Angaben über früheres Umweltschutzverhalten durch die subjektive Antwortsicherheit moderiert wird.

Die Ergebnisse über die Bedeutung der subjektiven Antwortsicherheit als Moderatorvariable bei der Entstehung von Response-Effects sind besonders inkonsistent. So konnte in drei Untersuchungen gezeigt werden, dass Befragte, die ihre Einstellung gegenüber einer Liberalisierung von Schwangerschaftsabbrüchen als sicher bezeichnet haben, signifikant weniger durch Fragereihenfolgeeffekte beeinflusst werden. Dagegen erweist sich die subjektive Antwortsicherheit bei vier weiteren Experimenten mit dem gleichen Befragungsgegenstand nicht als valider Prädiktor für die Beeinflussbarkeit der Probanden durch vorangegangene Fragen (*Krosnick* und *Schuman* 1988, *Schuman* et al. 1981, *Stocké* 2002). In einer Untersuchung von Lavine und Mitarbeitern (*Lavine* et al. 1998) wird der Einfluss von Kontextitems mit liberalem oder konservativem Inhalt auf die Zustimmung zu erhöhten Sozialausgaben und einer Verbesserung der Rechte von Angeklagten untersucht. Dabei finden sich zwar reduzierte Einflüsse der Kontextfragen bei Befragten mit höherer Einstellungssicherheit, dieser Moderatoreffekt kann jedoch nicht als statistisch abgesichert angesehen werden.

Auch die Evidenz für die Validität von Antwortlatenzen als Indikator für die Informationsverfügbarkeit muss beim derzeitigen Forschungsstand als teilweise inkon-

sistent bezeichnet werden. Dabei liegen aus dem Bereich der Prognose der Einstellungs-Verhaltenskonsistenz die meisten bestätigenden Ergebnisse vor. So konnte in einer ganzen Serie von Untersuchungen gezeigt werden, dass der Zusammenhang zwischen den Einstellungen gegenüber politischen Kandidaten und dem tatsächlichen Wahlverhalten der Befragten durch die Geschwindigkeit der Beantwortung der Einstellungsfragen vermittelt wird. Die vor der Wahl geäußerten Bewertungen stimmen mit steigender Geschwindigkeit dieser Angaben immer mehr mit dem nach der Abstimmung erhobenen Wahlverhalten überein (*Bassili* 1993, *Bassili* 1995, *Bassili* und *Bors* 1997, *Fazio* und *Williams* 1986, *Fletcher* 2000). Die Prognosekraft der Antwortlatenzen lässt sich auch dadurch belegen, dass die Einstellungen gegenüber einer Reihe unterschiedlicher Konsumprodukte in zunehmendem Ausmaß mit der berichteten Konsumabsicht der Befragten übereinstimmen, wenn die Einstellungsfragen zunehmend schneller beantwortet werden (*Kokkinaki* und *Lunt* 1997). In einer weiteren Untersuchung hat sich gezeigt, dass die inhaltlichen Angaben der Befragten über ihre Einstellungen gegenüber Umfragen dann stärker ihre Unterstützung der Befragung durch korrektes und weniger sozial erwünschtes Antwortverhalten vorhersagen können, wenn die Einstellungsantworten zunehmend schneller geäußert wurden (*Stocké* 2003).

Forschungsergebnisse belegen auch, dass schnell geäußerte Einstellungsurteile relativ änderungsresistent sind. So erweisen sich Einstellungen gegenüber Geschlechterquoten bei der Einstellung von Beschäftigten oder gegenüber Pornographie dann zunehmend stabiler gegenüber Überzeugungsversuchen, wenn die Geschwindigkeit der Urteile ansteigt (*Bassili* 1996b). Die Antwortlatenzen bei der Beantwortung von Fragen über die Einstellung zu der Freigabe von Schwangerschaftsabbrüchen kann auch als Indikator dafür herangezogen werden, für wie wahrscheinlich die Umfrageteilnehmer einen Wandel dieser Einstellung bei sich selbst ansehen. So berichten Befragte mit langsamerer Urteils geschwindigkeit mehr Zweifel an ihrer Einstellungsstabilität (*Huckfeldt* und *Sprague* 2000).

Was die Prognosekraft der Antwortlatenzen für die Beeinflussbarkeit der Befragten durch Response-Effects angeht, so liegt bisher nur wenig und inkonsistente Evidenz vor. In einer Untersuchung wurde die Fähigkeit von Antwortlatenzen für die Prognose von Antwortreihenfolgeeffekten, Einflüssen durch Akquieszenz und die Bedeutung einer mittleren Antwortoption für das Antwortverhalten überprüft (*Bassili* und *Krosnick* 2000). Dabei hat sich die Antwortgeschwindigkeit in keinem Fall als signifikanter Prädiktor für die Beeinflussbarkeit der Befragten erwiesen. Dagegen konnte in einer anderen Untersuchung gezeigt werden, dass der üblicherweise bei der sequenziellen Abfrage der Einstellung gegenüber der generellen Freigabe von Schwangerschaftsabbrüchen und jener im Falle einer Vergewaltigung beobachtete

Fragereihenfolgeeffekt zuverlässig durch die Antwortgeschwindigkeit der Befragten vorhergesagt werden kann (*Stocké* 2002). In dieser Studie wurden auch die subjektiven Antwortsicherheiten der Befragten einbezogen und deren Vorhersagefähigkeit für die Entstehung der Fragereihenfolgeeffekte mit der von Antwortlatenzen verglichen. Dabei hat sich gezeigt, dass die anfänglich bestehende Vorhersagekraft der Antwortsicherheiten bei gleichzeitiger Kontrolle der Erklärungskraft der Antwortgeschwindigkeit vollständig absorbiert wird. Dies führt zur Schlussfolgerung des Autors, dass den Antwortlatenzen als operationalem Indikator für die Informationsverfügbarkeit, zumindest bei der Vorhersage von Fragereihenfolgeeffekten, eine größere Validität zukommt.

2.3 Die unterschiedliche Transformation von Antwortlatenzen als Validitätsfaktor

In der Literatur finden sich unterschiedliche Vorschläge, wie Antwortlatenzen vor deren Verwendung als Erklärungsfaktor transformiert werden sollen. Dabei beziehen sich die ersten beiden Vorschläge auf mathematische Transformationen, durch die die charakteristische Schiefe von Reaktionszeitverteilungen reduziert und stärker an eine, bei einer Reihe statistischer Auswertungsverfahren angenommene Normalverteilung angenähert werden können. Zu diesem Zweck werden Rohreaktionszeiten häufig logarithmiert oder einer Wurzeltransformation unterzogen (*Bargh* und *Chartrand* 2000). Eine dritte, weniger statistisch als inhaltlich begründete Art der vorbereitenden Behandlung von Antwortlatenzen besteht in deren Standardisierung am individuellen Mittelwert der Befragten. Dabei wird von jeder Zielreaktionszeit die bei einer möglichst großen Anzahl an Standardisierungsfragen beobachtete mittlere Antwortgeschwindigkeit jedes individuellen Befragten abgezogen. Die resultierende Ergebnisvariable spiegelt dann ausschließlich itemspezifische Unterschiede in den Latenzen zwischen den Befragten und damit das vorliegende Ausmaß der Informationsverfügbarkeit wider. Dadurch werden, so das Argument, dispositional stabile und von der eigentlich zu erfassenden Zugänglichkeitsdimension unabhängige Unterschiede zwischen den Befragten und damit Fehlervarianz der Messung beseitigt (*Fazio* 1990). Gegen diese Vorgehensweise wurde allerdings der Einwand vorgebracht, dass hierbei auch individuelle Unterschiede in der vom Befragungsthema unabhängigen, globalen Informationszugänglichkeit zwischen den Befragten eliminiert werden. Dadurch wird möglicherweise die Validität der resultierenden Messergebnisse geschwächt (*Bassili* und *Krosnick* 2000).

Bei einer Durchsicht der oben dargestellten Untersuchungen über die Validität von Antwortlatenzen findet sich eine große Heterogenität bezüglich der jeweils durchge-

fürten Transformationen der Reaktionszeiten. So verwendet **Bassili** (1996) erfolgreich Rohreaktionszeiten zur Prognose der Einstellungsstabilität bei verschiedenen Befragungsthemen. Auch in einer Reihe von Untersuchungen über das Ausmaß der Einstellungs-Verhaltenskonsistenz werden erfolgreich untransformierte und unstandardisierte Antwortlatenzen herangezogen (**Bassili** und **Bors** 1997, **Bassili** 1995, **Fletcher** 2000). Hinweise auf die relative Validität dieser Basisversion der Antwortgeschwindigkeit und durch Logarithmierung transformierter Latenzdaten findet sich in einer Untersuchung, in der beide Varianten zur Prognose der von den Befragten berichteten Wichtigkeit bestimmter Einstellungsthemen herangezogen werden. Dabei erweisen sich die bei der Beantwortung der Einstellungsfragen beobachteten Rohreaktionszeiten als vorhersagekräftig, und eine Logarithmierung hat keinen Einfluss auf die Prognoseleistung (**Krosnick** 1989). In einer Studie von **Bassili** (1993) werden erfolgreich standardisierte Antwortlatenzen als Moderatorvariable zwischen der Einstellung zu politischen Kandidaten und dem Wahlverhalten der Befragungsteilnehmer herangezogen. Hier wird von einer, allerdings nicht näher quantifizierten, schwächeren Vorhersagekraft logarithmierter Antwortlatenzen berichtet. Bei Verwendung standardisierter Antwortlatenzen finden sich auch in anderen Studien durchgängig positive Ergebnisse. In den bereits oben erwähnten Untersuchungen von **Kokkinaki** und **Lunt** (1997) sowie der von **Stocké** (2002) haben sich standardisierte Rohreaktionszeiten als signifikanter Prädiktor für den Grad der Übereinstimmung zwischen Einstellungen und Verhalten sowie für die Beeinflussbarkeit der Befragten durch Fragereihenfolgeeffekte erwiesen. Im Kontrast zu der insgesamt positiven Bilanz zugunsten aller anderen Reaktionszeitenversionen erbringt die Studie von **Bassili** und **Krosnick** (2000) ausschließlich negative Evidenz für die Prognosekraft wurzeltransformierter Latenzverteilungen. Keine der vier untersuchten Arten von Response-Effects lässt sich durch diese Operationalisierung der Informationsverfügbarkeit vorhersagen.

Zusammenfassend erweist sich die Praxis der Transformation von Antwortlatenzen als genau so heterogen wie die in der Literatur vorgeschlagenen Vorgehensweisen. Obwohl einige Ergebnisse die Schlussfolgerung nahe legen, dass sich die Validität der verschiedenen Operationalisierungen der Informationsverfügbarkeit unterscheiden, liegen hierzu bisher keine systematischen Vergleiche vor.

3 Empirische Untersuchung

Die Fragestellung der vorliegenden Untersuchung wird durch ein „Split-Ballot“-Experiment im Rahmen einer allgemeinen Bevölkerungsumfrage analysiert. Dabei beantworten zufällig ausgewählte Teilgruppen der Befragten die Frage nach ihrem

täglichen Fernsehkonsum entweder auf einer nieder- oder hoch-frequenten Häufigkeitsskala. Es wird analysiert, ob die Stärke der hierbei in anderen Studien beobachteten Einflüsse der unterschiedlich kategorisierten Häufigkeitsskalen durch die gleichzeitig erfassten Indikatoren der Informationsverankerung prognostiziert werden können. Zu diesem Zweck werden die Befragten einerseits über das Ausmaß ihrer subjektiven Antwortsicherheit befragt. Andererseits wird im Rahmen der computergestützten Interviews die von den Befragten zur Antwortgenerierung benötigte Zeit gemessen.

3.1 Stichprobe

Die Studienteilnehmer entstammen einer lokalen, mehrstufigen Zufallsstichprobe der Wohnbevölkerung einer süddeutschen Großstadt. Die Grundgesamtheit bildeten Bürger mit Lebensmittelpunkt im Stadtgebiet, die zum Befragungszeitpunkt mindestens 18 Jahre alt waren und die deutsche Staatsbürgerschaft besaßen. Es wurde eine „Random-Walk“-Prozedur als Auswahlverfahren verwendet, wobei zunächst nach dem Zufallsprinzip im Stadtgebiet Startpunkte bestimmt wurden, von denen ausgehend eine Haushaltsstichprobe generiert werden konnte. Die Auswahl der Zielperson innerhalb der Haushalte erfolgte mittels der „Geburtstagsmethode“. Hinsichtlich der sozialstrukturellen Merkmale, die auch bei der Datenanalyse als Kontrollvariablen verwendet werden, setzt sich die Stichprobe wie folgt zusammen: An den 110 Interviews haben 53.6 Prozent Frauen und 46.4 Prozent Männer teilgenommen, die am Erhebungszeitpunkt im Durchschnitt 46.8 Jahre alt waren und 10.9 Jahre Schulausbildung absolviert hatten. Die Stichprobe besteht weiterhin aus 13.6 Prozent Arbeitern, 65.5 Prozent Angestellten und Beamten, 9.1 Prozent selbstständig Beschäftigten sowie 11.2 Prozent Befragten, die noch nie erwerbstätig waren. Das durchschnittliche Netto-Haushaltseinkommen der Stichprobe beträgt 3784.- DM. Die Ausschöpfungsquote liegt bei 34.0 Prozent.

3.2 Vorgehensweise

Die Datenerfassung wurde in der Form computergestützter persönlicher Interviews bei den Befragten zu Hause durchgeführt, wobei die Probanden nach dem Zufallsprinzip einer der beiden Experimentalbedingungen mit unterschiedlich kategorisierten Antwortskalen zugewiesen wurden. Die jeweilige Version der Antwortskala wurde auf „show cards“ visuell präsentiert und den Befragten vorgelegt. Das Experiment wurde als Teil eines durchschnittlich 58-minütigen Interviews durchgeführt. Die Angaben der Befragten über die Länge ihres täglichen Fernsehkonsums und deren subjektive Antwortsicherheit wurden im letzten Drittel der Befragung erfasst.

Die Umfrage wurde im Anschreiben sowie in der Kontaktphase als Befragung über „gesellschaftliche Probleme“ und „alltägliche Gewohnheiten“ angekündigt. Weder der experimentelle Charakter noch die Tatsache der Erfassung von Antwortgeschwindigkeiten wurden den Teilnehmern vor Ende des Interviews mitgeteilt.

3.3 Operationalisierung

Die vorliegende Fragestellung macht die Operationalisierung der unterschiedlich kategorisierten Antwortskalen zur Erfassung der Länge des Fernsehkonsums, der hierbei subjektiv von den Befragten erlebten Antwortsicherheit sowie die Messung und unterschiedliche Transformation der Antwortgeschwindigkeiten notwendig.

- *Unterschiedlich kategorisierte Antwortskalen:* Das Antwortkontinuum der möglichen Länge des täglichen Fernsehkonsums wurde bei jeder der beiden Versionen der Antwortskala zu 7 Antwortoptionen zusammengefasst. Dabei waren beim nieder-frequenten Skalentyp die extremen Antwortmöglichkeiten durch die Optionen „überhaupt nicht“ und „mehr als 2 Stunden“ definiert, wohingegen diese bei der hoch-frequenten Version der Skala aus den Kategorien „bis zu 2 Stunden“ und „mehr als 4.5 Stunden“ bestanden. Der Wertebereich zwischen den jeweils extremen Antwortoptionen wurde bei beiden Antwortskalen in fünf Kategorien unterteilt, die jeweils ein Zeitintervall von einer halben Stunde umfassten (vgl. Tabelle 1 in Abschnitt 3.4.1 für eine Repräsentation der beiden Skalenversionen).⁴ Die inhaltlichen Antworten wurden bei den beiden Skalentypen vergleichbar gemacht, indem die Befragten jeweils in eine Gruppe mit einem geringen Fernsehkonsum von „bis zu 2.5 Stunden“ und einem hohen Konsum von „mehr als 2.5 Stunden“ klassifiziert wurden.

- *Subjektive Antwortsicherheit:* Die subjektive Sicherheit beziehungsweise Unsicherheit der Befragten bei der Beantwortung der Frage über ihr tägliches Fernsehverhalten wurde durch eine siebenstufige Antwortskala vom Likert-Typ erfasst.⁵ Die Endpunkte dieser Antwortskala waren mit den Aussagen „ich bin mir völlig sicher“ und „ich bin mir überhaupt nicht sicher“ verankert.

- *Rohreaktionszeiten:* Die von den Befragten benötigte Zeit zur Beantwortung der Frage über ihr Fernsehverhalten wurde zusammen mit den inhaltlichen Angaben

4 Die genaue Frageformulierung lautet wie folgt: „Können Sie mir bitte sagen, wie lange Sie an einem durchschnittlichen Werktag Fernsehen schauen. Bitte nennen Sie mir die Nummer der zutreffenden Antwortkategorie auf dieser Liste, die Ihren Fernsehgewohnheiten am ehesten entspricht“.

5 Die Frage ist wie folgt formuliert: „Können Sie mir bitte sagen, wie sicher Sie sich bei der Schätzung über die Häufigkeit Ihres Fernsehkonsums sind?“.

durch eine Sequenz von insgesamt vier Bildschirmen des computeradministrierten Fragebogens erfasst. Der Interviewer hat den Fragetext vom ersten Bildschirm abgelesen und sofort nach dessen Beendigung die Zeitmessung eingeschaltet. Der dadurch aktivierte zweite Bildschirm signalisierte die aktive Latenzmessung und blieb solange sichtbar, bis der Interviewer ihn im Moment der Antwortgenerierung durch den nächsten Tastendruck deaktivierte. Auf dem nächsten, automatisch aufgerufenen dritten Bildschirm wurden dann die Antworten der Befragten in den Laptop-Computer eingegeben. Auf dem die Befragungssequenz abschließenden vierten Bildschirm wurde dann dokumentiert, ob die zuvor erfasste Antwortlatenz valide gemessen wurde.

Das zentrale Problem bei dieser „manuellen“ Art der Zeiterfassung besteht darin, den genauen Zeitpunkt für die Beendigung der Messung zu treffen. Eine „konservative“, weil das Risiko von zu langen Zeitmessungen minimierende Vorgehensweise besteht darin, die Messung sofort bei der ersten Äußerung der Befragten abzuschalten. Immer, wenn es sich bei dieser Äußerung der Befragten nicht um eine fertige Antwort handelt, ergeben sich bei dieser Vorgehensweise invalide Zeitmessungen. Dies ist etwa dann unvermeidlich der Fall, wenn die Befragten Verständnisfragen stellen oder spontan in eine Kommunikation mit dem Interviewer eintreten (**Bassili** 1996a). Auch Ablenkungen der Befragten durch Kontextfaktoren und die damit verbundene Invalidierung der Messung lassen sich kaum verhindern. Allerdings begleiten die Befragten ihren Antwortgenerierungsprozess häufig durch „lautes Nachdenken“, was in eine korrekt gemessene Antwortgeschwindigkeit einbezogen werden muss. In diesem Fall führt die „konservative“ Strategie der Beendigung der Zeitmessung zu Fehlmessungen, die durch eine flexiblere Handhabung vermieden werden können. So haben die Ergebnisse von Voruntersuchungen gezeigt, dass Interviewer mit zunehmender Erfahrung mit der Messprozedur schnell lernen, den Antwortgenerierungsprozess begleitende Selbstgespräche von der Äußerung fertiger Antworten zu unterscheiden. Diese Fähigkeit wurde daher im Rahmen einer praktischen Interviewerschulung intensiv geübt. Der dennoch unvermeidliche Anteil der nicht validen Reaktionszeitmessungen liegt in der vorliegenden Studie bei 21.8 Prozent der Fälle. Die durch die Software bedingte, technische Obergrenze der Messgenauigkeit liegt bei einer Hundertstelsekunde.

- *Transformierte Versionen der Rohreaktionszeiten:* In der Literatur wird vorgeschlagen, die charakteristische Schiefe der Verteilung von Reaktionszeiten durch Logarithmierung oder durch Anwendung der Quadratwurzel zu beseitigen (**Bargh** und **Chartrand** 2000). Es wurden daher zwei entsprechend transformierte Versionen der Rohreaktionszeiten erstellt und in die Validierungsstudie einbezogen.

- *Standardisierte Rohreaktionszeiten*: Bei der Erstellung der standardisierten Version der Antwortlatenzen wurde wie folgt vorgegangen: In einem ersten Schritt wurde für jeden Befragungsteilnehmer dessen durchschnittliche Antwortgeschwindigkeit bei allen anderen, im Interview beantworteten Fragestellungen berechnet. Bei diesen Standardisierungsfragen handelt es sich um insgesamt 126 Fragestellungen aus dem Bereich politischer Überzeugungen, Einstellungen zu ethnischen Fremdgruppen und gegenüber Umfragen. Es wurde weiterhin nach verschiedenen Lebensgewohnheiten der Befragten und deren sozialstrukturellen Merkmalen gefragt. Die resultierende Verteilung der Mittelwerte spiegelt die individuellen Unterschiede in der generellen Antwortgeschwindigkeit zwischen den Befragten wider. Die standardisierte Antwortlatenz wurde berechnet, indem die Basisgeschwindigkeit jedes Befragten von der bei der Beantwortung der Frage nach dem täglichen Fernsehkonsum beobachteten Antwortlatenz abgezogen wurde (*Fazio* 1990). Haben die Befragten bei der Zielfragestellung, im Vergleich zu der ansonsten im Interview beobachteten durchschnittlichen Antwortlatenz, überdurchschnittlich schnell geantwortet, so nimmt die Ergebnisvariable negative Werte an. Eine positive Ausprägung auf dieser Dimension zeigt dagegen an, dass die Befragten überdurchschnittlich lange für die Generierung der Antwort brauchten.

3.4 Ergebnisse

Die Darstellung der empirischen Ergebnisse gliedert sich in vier Abschnitte. *Zuerst* werden die Angaben der Befragten über ihr Fernsehverhalten bei Verwendung der unterschiedlich kategorisierten Antwortskalen dargestellt und die dabei benötigte Zeit sowie das erlebte Gefühl der Antwortsicherheit beschrieben. Es wird auch analysiert, ob den unterschiedlichen Indikatoren für die Informationsverfügbarkeit konvergente Validität zugesprochen werden kann. Im *zweiten* Abschnitt soll dann die Prognosekraft der subjektiven Antwortsicherheit für die Beeinflussbarkeit der Befragten durch die hier untersuchte Art von Response-Effects untersucht werden. Diese wird dann im *dritten* Teil der Datenauswertung mit jeder der Antwortgeschwindigkeiten verglichen. Der abschließende *vierte* Analyseteil untersucht dann die relative Fähigkeit der verschiedenen Versionen der Antwortlatenzen zur Identifikation jener Befragten, deren Angaben in besonderem Ausmaß durch die unterschiedlich kategorisierten Antwortskalen beeinflusst werden.

3.4.1 Deskriptive Ergebnisse

Die in Tabelle 1 dargestellten Angaben über die Länge des täglichen Fernsehkonsums weisen deutliche Einflüsse der unterschiedlich kategorisierten Antwortskalen

auf. So ist erkennbar, dass 80.4 Prozent der Befragten bei Verwendung der niederfrequenten Antwortskala einen täglichen Fernsehkonsum zwischen 0 und 2.5 Stunden angeben. Wenn dagegen die hoch-frequente Skalenversion herangezogen wird, so liegt dieser Anteil jedoch nur bei 57.4 Prozent. Demnach lassen sich die üblicherweise beobachteten Skaleneinflüsse auch in der vorliegenden Studie replizieren.

Tabelle 1 Das Antwortverhalten bei den unterschiedlich kategorisierten Antwortskalen

Nieder-frequente Antwortskala			Hoch-frequente Antwortskala		
	N	%		N	%
(1) überhaupt nicht	1	1.8			
(2) bis zu einer halben Stunde	4	7.1			
(3) zwischen 0.5 und 1 Stunde	7	12.5			
(4) zwischen 1 und 1.5 Stunden	17	30.4			
(5) zwischen 1.5 und 2 Stunden	10	17.9	(1) bis zu 2 Stunden	24	44.4
(6) zwischen 2 und 2.5 Stunden	6	10.7	(2) zwischen 2 und 2.5 Stunden	7	13.0
(7) mehr als 2.5 Stunden	11	19.6	(3) zwischen 2.5 und 3 Stunden	13	24.1
			(4) zwischen 3 und 3.5 Stunden	4	7.4
			(5) zwischen 3.5 und 4 Stunden	5	9.3
			(6) zwischen 4 und 4.5 Stunden	0	0.0
			(7) mehr als 4.5 Stunden	1	1.9
Dichotomisiertes Antwortverhalten					
0 bis 2.5 Std. Fernsehkonsum	45	80.4	0 bis 2.5 Std. Fernsehkonsum	31	57.4
mehr als 2.5 Stunden Fernsehkonsum	11	19.6	mehr als 2.5 Stunden Fernsehkonsum	23	42.6
Total	56	100	Total	54	100

Auf der Grundlage der von den Befragten selbst berichteten Aussagen lässt sich ein insgesamt hohes Ausmaß an Antwortssicherheit über die Länge ihres täglichen Fernsehkonsums ableiten. Die durchschnittlichen Angaben liegen bei einem Wert von 1.9 und damit weniger als einen Skalenpunkt über der maximalen Sicherheit bei 1.0 (vgl. Tabelle 2). Dagegen verweist die mittlere Rohreaktionszeit von rund 6.6 Sekunden auf die Notwendigkeit eines doch eher aufwendigen Antwortgenerierungsprozesses. Dieser Eindruck wird durch den Vergleich mit der durchschnittlich bei der Beantwortung der 126 Standardisierungsfragen beobachteten Antwortlatenz verstärkt. Bei diesen Fragestellungen liegt der Mittelwert der benötigten Zeit nur bei 4.3 Sekunden. Dieser Unterschied spiegelt sich im Durchschnittswert der standardisierten Antwortlatenzen von plus 2.4 Sekunden wider und besagt, dass die Beantwortung der Frage nach der Länge des täglichen Fernsehkonsums überdurchschnittlich aufwendig ist.

Eine Analyse der Rohreaktionszeiten und der unterschiedlich transformierten Antwortlatenzen zeigt, dass diese in verschieden starkem Umfang von einer Normalverteilung abweichen. Die auf dem Kolmogorov-Smirnov-Test beruhende Lilliefors-Teststatistik beträgt bei der Rohreaktionszeit 0.12 ($p \leq 0.01$), bei der logarithmierten Variante 0.14 ($p \leq 0.01$) und bei der wurzeltransformierten Antwortgeschwindigkeit 0.06 ($p \geq 0.1$). Es kann somit festgestellt werden, dass sich die Antwortlatenzen nur bei einer Wurzeltransformation nicht mehr statistisch signifikant von einer Normalverteilung unterscheiden.

Tabelle 2 Beschreibung der unterschiedlichen Operationalisierungen der Informationsverfügbarkeit

	Min/Max	Mittelwert	Median	Std.
Subjektive Sicherheit ¹	1.0 / 7.0	1.9	1.0	1.3
Rohreaktionszeit ²	27.0 / 3263.0	662.3	566.0	529.1
Logarithmierte Rohreaktionszeit	3.3 / 8.1	6.2	6.3	0.9
Wurzeltransformierte Rohreaktionszeit	5.2 / 57.1	23.8	23.8	9.8
Standardisierte Rohreaktionszeit	-666.4 / 1906.5	231.0	163.6	436.5
Mittlere Reaktionszeit bei den 126 Standardisierungsfragen	80.1 / 1506.4	431.3	355.2	249.9

¹ Skalenwert 1=“völlig sicher“, Skalenwert 7=“überhaupt nicht sicher“; ² Die dargestellte Maßeinheit sind Hundertstelsekunden. Datengrundlage: Eigene Erhebung; N=110.

Die Struktur der Korrelationen zwischen den unterschiedlichen Indikatoren der Informationsverfügbarkeit zeigt die, bei bestehender konvergenter Validität der Messungen zu erwartenden Zusammenhänge (vgl. Tabelle 3). So finden sich positive Zusammenhänge zwischen der von den Befragten selbstberichteten Antwortunsicherheit und allen Versionen der Antwortlatenzen. Demnach berichten die Probanden ein zunehmend stärkeres Gefühl der Antwortsicherheit, wenn die Geschwindigkeit ihres Antwortgenerierungsprozesses zunimmt. Allerdings ist nur der Zusammenhang zwischen den subjektiven Antwortsicherheiten und der standardisierten Version der Antwortgeschwindigkeiten nach konventionellen Kriterien statistisch signifikant. Dieser Art der Antwortlatenzen kann demnach das höchste Ausmaß an konvergenter Validität zugesprochen werden. Die insgesamt eher schwachen Korrelationskoeffizienten zwischen 0.11 und 0.19 belegen für alle Versionen der Antwortlatenzen, dass diese und die Sicherheitsangaben keineswegs identische Sach-

verhalte erfassen. Zwischen den unterschiedlich transformierten Antwortlatenzen und der Rohreaktionszeit finden sich dagegen naturgemäß starke Korrelationen mit einem Wertebereich zwischen 0.77 und 0.96.

Tabelle 3 Korrelationen zwischen den unterschiedlichen Indikatoren der Informationsverfügbarkeit (Pearsons r)

	Subjektive Sicherheit	Rohreaktionszeit (R-RZ)	Logarithmierte R-RZ	Wurzeltransformierte R-RZ	Standardisierte R-RZ
Subjektive Sicherheit	1.00	--	--	--	--
Rohreaktionszeit (R-RZ)	0.15	1.00	--	--	--
Logarithmierte R-RZ	0.11	0.85**	1.00	--	--
Wurzeltransformierte R-RZ	0.14	0.96**	0.96**	1.00	--
Standardisierte R-RZ	0.19*	0.88**	0.77**	0.86**	1.00

Datengrundlage: Eigene Erhebung; N = 110; Signifikanz: ** $p \leq 0.01$; * $p \leq 0.05$

3.4.2 Erklärungskraft der Antwortsicherheit für die Skaleneinflüsse

Im folgenden Abschnitt wird untersucht, ob die Beeinflussbarkeit der Befragten durch die unterschiedlich kategorisierten Antwortskalen mithilfe der selbstberichteten Antwortsicherheit vorhergesagt werden kann.⁶ Dabei wird das in Tabelle 1 dargestellte, dichotomisierte und daher zwischen den Skalenversionen vergleichbare Antwortverhalten als abhängige Variable der Analyse herangezogen. Die in Tabelle 4 dargestellten Ergebnisse des ersten logistischen Regressionsmodells zeigen zuerst einmal, dass der Einfluss der unterschiedlich kategorisierten Antwortskalen als sta-

6 Bei der Schätzung multiplikativer Interaktionsparameter tritt typischerweise das Problem hoher Multikollinearität zwischen den Ausgangsvariablen und den multiplikativen Faktoren auf. Dies trifft auch im vorliegenden Fall zu. Ohne Gegenmaßnahmen liegt bei den folgenden Regressionsanalysen die Toleranz der Interaktionsterme in vielen Fällen bei einem Wert von 0.02 und damit deutlich unter einem akzeptablen Niveau. Diesem Problem und der damit verbundenen Inflationierung der für die Regressionsparameter geschätzten Standardfehler kann durch die Verwendung z-standardisierter Variablen begegnet werden (Cronbach 1987). Durch diese Transformation können zwar die Schätzergebnisse von Parametern niedrigerer Ordnung beeinflusst werden, die für unser Analyseziel relevanten Parameter der jeweils höchsten Interaktionsstufe bleiben hiervon jedoch unberührt (Aiken und West 1991: 28ff.). Bei den folgenden Regressionsmodellen 1 bis 10 gehen daher alle an den Interaktionseffekten beteiligten Variablen als z-standardisierte Versionen in die Analyse ein. Bei dieser Vorgehensweise liegt nur noch ein geringes Ausmaß an Multikollinearität vor, und die Toleranz fällt für keine der Parameterschätzungen unter einen Wert von 0.48.

tistisch abgesicherte Determinante des Antwortverhaltens der Befragten angesehen werden kann. Von den gleichzeitig in das Regressionsmodell aufgenommenen sozialstrukturellen Kontrollvariablen erweist sich die Länge der Schulbildung als signifikanter Prädiktor für die Angaben. Demnach geht der berichtete Fernsehkonsum mit steigender Schulbildung deutlich zurück.⁷ Die Analyse zeigt auch, dass sowohl die subjektive Sicherheit wie auch die Rohreaktionszeiten positiv mit der Wahrscheinlichkeit der Nennung eines hohen Fernsehkonsums zusammenhängen: diese Wahrscheinlichkeit steigt an, wenn beide (Un-)Sicherheitsindikatoren ein geringeres Ausmaß an Informationsverfügbarkeit anzeigen. Dieser unerwartete, allerdings nur für die Antwortlatenzen überhaupt marginal signifikante Effekt, kann so interpretiert werden, dass Befragte mit einem hohen Fernsehkonsum am wenigsten zu einer sicheren Quantifizierung ihres täglichen Fernsehkonsums in der Lage sind.

Tabelle 4 Skaleneinflüsse auf die Angaben über den täglichen Fernsehkonsum und die moderierende Wirkung der Antwortsicherheit für die Stärke der Effekte (Ergebnisse logistischer Regressionsanalysen)

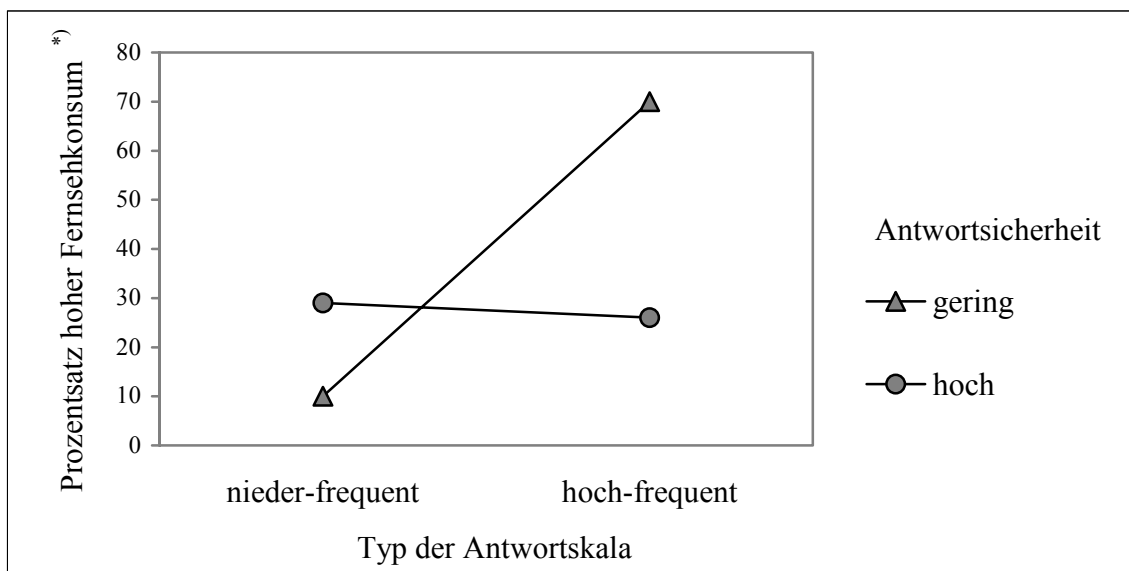
	Modell 1 B (Wald-Statistik)	Modell 2 B (Wald-Statistik)
Kontroll-Variablen		
(1) SCHULBILDUNG (Jahre)	-0.41 (6.51)**	-0.26 (2.91)*
(2) ALTER (Jahre)	0.02 (2.36)	0.03 (4.35)**
Erklärende Variablen		
(3) SKALENTYP (hoch-frequent) ¹	1.67 (10.15)**	1.55 (9.11)**
(4) SUBJEKTIVE SICHERHEIT (z-Werte)	0.23 (1.09)	-0.06 (0.05)
(5) ROHREAKTIONSZEIT (z-Werte)	0.46 (3.54)*	--
(6) SKALENTYP • SICHERHEIT	--	1.20 (4.25)**
Konstante	1.48 (0.54)	-0.30 (0.02)
Pseudo R ² Gesamtmodell	0.19	0.20
N	110	110

¹ Referenzkategorie: „nieder-frequent“; abhängige Variable: 0=„0 bis 2.5 Stunden Fernsehkonsum“, 1=„mehr als 2.5 Stunden Fernsehkonsum“; die Variable „Skalentyp“ geht in Dummy-Kodierung (0,1) in die Analyse ein; Signifikanz: * p ≤ 0.1; ** p ≤ 0.05

⁷ Abgesehen von den dargestellten Variablen wurde außerdem überprüft, ob sich das Geschlecht, der Erwerbsstatus, das Haushaltseinkommen sowie die Konfessionszugehörigkeit und der Familienstand der Befragten auf deren Antwortverhalten auswirken (Ergebnisse nicht dargestellt). Da dies bei keiner der angeführten Dimensionen der Fall ist, werden diese Faktoren in die weiteren Analysen nicht einbezogen.

In Regressionsmodell 2 wird überprüft, ob die Beeinflussbarkeit der Befragten durch die unterschiedlich kategorisierten Antwortskalen durch die subjektive Antwortsicherheit prognostiziert werden kann (vgl. Tabelle 4). Es wird erwartet, dass eine hoch-frequente Antwortskala mit steigender Antwortunsicherheit der Befragten einen zunehmend positiven Einfluss auf die Nennung eines hohen Fernsehkonsums ausübt. Diese Erwartung wird durch den positiven und statistisch signifikanten Interaktionsparameter zwischen der Art der Antwortskala und den Sicherheitsangaben bestätigt.

Abbildung 1 Prognosekraft der subjektiven Antwortsicherheit für die Beeinflussbarkeit der Befragten durch unterschiedlich kategorisierte Antwortskalen⁸



^{*)} Prozentsatz der Befragten mit Angaben von mehr als 2.5 Stunden täglichen Fernsehkonsums

In Abbildung 1 wird der Inhalt dieses Interaktionseffektes graphisch dargestellt. Dabei erweist sich die Subgruppe der Befragten mit hoher Antwortsicherheit als praktisch immun gegenüber Einflüssen durch die unterschiedlichen Skalenversionen. Bei Verwendung der nieder-frequenten Antwortskala geben 30 Prozent und bei der hoch-frequenten Version 27 Prozent der Probanden einen Fernsehkonsum von mehr als 2.5 Stunden pro Tag an. Dagegen werden Befragte mit einer relativ geringen Antwortsicherheit sehr stark durch die Darstellung der Antwortoptionen beeinflusst. In dieser Gruppe geben bei Verwendung der nieder-frequenten Skala 10 Pro-

⁸ In der vorliegenden Darstellung wurde die Befragtenpopulation hinsichtlich ihrer Antwortsicherheit in zwei möglichst gleich große Gruppen unterteilt. Die Gruppe mit geringer Antwortsicherheit enthält 46 Prozent und die mit hoher Antwortsicherheit 54 Prozent der Befragten.

zent, bei der hoch-frequenten Skala aber 70 Prozent an, täglich mehr als 2.5 Stunden vor dem Fernseher zu verbringen. Die subjektive Antwortsicherheit kann somit als valider Indikator für die Informationsverfügbarkeit und guter Prädiktor für die untersuchte Art von Response-Effects angesehen werden.

3.4.3 Relative Erklärungskraft der Antwortsicherheit und Antwortlatenzen

Im folgenden Abschnitt wird die Prognosekraft der subjektiven Antwortsicherheit für die Stärke der Skaleneinflüsse mit der jeder der hier untersuchten Versionen der Antwortlatenzen verglichen. Zu diesem Zweck werden vier logistische Regressionsanalysen durchgeführt. Es wird genau wie bei der Antwortunsicherheit prognostiziert, dass eine hoch-frequente Antwortskala mit steigender Antwortgenerierungszeit einen zunehmend positiven Einfluss auf die Nennung eines hohen Fernsehkonsums ausübt. Demnach ist ein positiver Interaktionsparameter zwischen den Skalenversionen und den Antwortlatenzen der Befragten zu erwarten. Da bei der Analyse jeder der Versionen der Antwortlatenzen gleichzeitig die bereits oben festgestellte Erklärungskraft der Antwortsicherheiten statistisch kontrolliert wird, können Schlussfolgerungen über die relative Erklärungskraft der beiden Typen von Indikatoren der Informationsverfügbarkeit abgeleitet werden.

Die in Tabelle 5 dargestellten Ergebnisse von Regressionsmodell 3 belegen erstens, dass sich die Beeinflussbarkeit der Befragten durch Skaleneinflüsse auch dann in marginal signifikantem Umfang durch die Rohreaktionszeiten vorhersagen lässt, wenn zugleich die Bedeutung der subjektiven Antwortsicherheit kontrolliert wird. Dagegen verliert die selbstberichtete Antwortsicherheit bei gleichzeitiger Berücksichtigung der Antwortlatenzen jegliche diagnostische Kraft in dieser Hinsicht: der zuvor bedeutsame Interaktionseffekt zwischen dem Typ der Antwortskala und den Sicherheitsangaben erweist sich nun nicht mehr als signifikant.

Die Ergebnisse des vierten Regressionsmodells zeigen bei Verwendung der logarithmierten Rohreaktionszeiten das gleiche Ergebnis, wobei sich diese nun allerdings als nach konventionellen Kriterien statistisch abgesicherter Prädiktor der Skaleneinflüsse erweisen (vgl. Tabelle 5). Auch hier wird die Erklärungskraft der subjektiven Antwortsicherheit vollständig durch die Bedeutsamkeit der Antwortlatenzen absorbiert. Die gleiche Aussage trifft auch für die am Mittelwert zentrierte Version der Antwortlatenzen zu (vgl. Tabelle 5: Modell 6). Die Prognosekraft der wurzeltransformierten Version der Antwortlatenz (Modell 5) verfehlt bei gleichzeitiger Kontrolle der Antwortsicherheiten nur knapp das Kriterium der 5-prozentigen Irrtumswahrscheinlichkeit und erweist sich somit als marginal signifikant. Die vorliegende Analyse zeigt somit, dass jeder der untersuchten Versionen der Antwortla-

tenzen eine im Vergleich zur subjektiven Antwortsicherheit stärkere Fähigkeit zur Prognose der untersuchten Art von Response-Effects und damit eine höhere Validität als Indikator für die Informationsverfügbarkeit zukommt.

Tabelle 5 Relative Erklärungskraft der subjektiven Antwortsicherheit und unterschiedlicher Versionen der Antwortlatenzen für die Beeinflussbarkeit der Angaben über die Länge des täglichen Fernsehkonsums durch die Kategorisierung der Antwortskala
(Ergebnisse logistischer Regressionsanalysen)

	Modell 3 B (Wald-Statistik)	Modell 4 B (Wald-Statistik)	Modell 5 B (Wald-Statistik)	Modell 6 B (Wald-Statistik)
	Rohreaktionszeit	Logarithmierte Rohreaktionszeit	Wurzeltransformierte Rohreaktionszeit	Standardisierte Rohreaktionszeit
Kontroll-Variablen				
(1) BILDUNG (Jahre)	-0.38 (4.72)**	-0.40(5.22)**	-0.40(5.07)**	-0.34(3.95)**
(2) ALTER (Jahre)	0.03 (2.61)	0.03(3.04)*	0.03(2.72)*	0.03(3.00)*
Subjektive Sicherheit				
(3) SKALENTYP (hoch-frequent) ¹	1.77(10.52)**	1.66(9.08)**	1.73(10.12)**	1.86(10.82)**
(4) SUBJEKTIVE <u>SICHERHEIT</u> (z-Werte)	-0.06 (0.05)	-0.06(0.05)	-0.06(0.05)	-0.07(0.05)
(5) SKALENTYP • SICHERHEIT	0.91 (2.40)	0.83 (1.95)	0.86 (2.12)	0.90(2.01)
Antwortlatenzen				
(6) ANTWORT- <u>LATENZEN</u> (z-Werte) ²	0.09 (0.07)	0.09(0.05)	0.10(0.07)	0.05 (0.02)
(7) SKALENTYP • LATENZEN	1.06 (3.09)*	1.17 (3.80)**	1.07 (3.58)*	1.16 (3.71)**
(8) Konstante	1.24 (0.32)	1.33 (0.38)	1.38 (0.40)	0.73 (0.12)
Pseudo-R ² - Gesamtmodell	0.246	0.276	0.261	0.253
N	110	110	110	110

¹ Referenzkategorie: „nieder-frequent“; ² Es handelt sich bei den verschiedenen Modellen um den jeweils in der Kopfzeile der Tabelle beschriebenen Typ an Antwortlatenz; abhängige Variable: 0=„0 bis 2.5 Stunden Fernsehkonsum“, 1=„mehr als 2.5 Stunden Fernsehkonsum“; die Variable „Skalentyp“ geht in Dummy-Kodierung (0,1) in die Analyse ein; Signifikanz: * p ≤ 0.1; ** p ≤ 0.05

3.4.4 Relative Erklärungskraft verschiedener Versionen der Antwortlatenzen

Im abschließenden Teil der empirischen Untersuchung soll nun untersucht werden, ob und wie stark sich die unterschiedlichen Transformationen der Antwortlatenzen auf deren Fähigkeit zur Prognose der Skaleneinflüsse auswirken. Die Korrelationsanalysen in Abschnitt 3.4.1 haben gezeigt, dass die verschiedenen Antwortlatenzen unterschiedlich große Varianzbestandteile mit der subjektiven Antwortsicherheit teilen. Aus diesem Grund muss bei gleichzeitiger Berücksichtigung der Sicherheitsdimension mit einer bei den verschiedenen Latenzversionen unterschiedlich starken Absorption deren Erklärungskraft gerechnet werden. Eine solche Verzerrung der Untersuchungsergebnisse wird dadurch ausgeschlossen, indem die relative Validität der verschiedenen Latenztypen unter Ausschluss der Antwortsicherheit untersucht wird.

Die in Tabelle 6 dargestellten Ergebnisse entsprechender logistischer Regressionsanalysen belegen zuerst einmal, dass alle Versionen der Antwortlatenzen die Beeinflussbarkeit der Befragten durch unterschiedlich kategorisierte Antwortskalen in signifikantem Ausmaß vorhersagen können. Weiterhin lässt sich feststellen, dass den Rohreaktionszeiten (Modell 7), den logarithmierten Rohreaktionszeiten (Modell 8) sowie der wurzeltransformierten Variante der Antwortzeiten (Modell 9) ein praktisch identisches Ausmaß an Erklärungskraft zukommt. Durch die Berücksichtigung des Interaktionseffektes zwischen der Art der Antwortskala und der jeweiligen Version der Antwortlatenzen steigt der Pseudo- R^2 -Wert und damit die Erklärungskraft der Modelle um 0.037 bzw. 0.038 Einheiten an. Damit lässt sich zumindest bei der vorliegenden Art der Response-Effects keine Validitätsverbesserung durch eine größere Annäherung der Reaktionszeiten an eine Normalverteilung feststellen. Ausschließlich bei Verwendung der standardisierten Antwortgeschwindigkeiten liegt eine gewisse, allerdings schwache Verbesserung der Vorhersagefähigkeit vor. Ausgehend von den Rohreaktionszeiten erhöht sich die Prognosekraft um 0.006 auf einen Pseudo- R^2 -Wert von 0.044. Die Beseitigung der individuell stabilen Unterschiede in der Antwortgeschwindigkeit zwischen den Befragten reduziert somit die beobachtete Devianz zwischen dem prognostizierten und tatsächlichen Antwortverhalten um 0.6 Prozentpunkte.

Tabelle 6 Relative Erklärungskraft der Rohreaktionszeiten und unterschiedlich transformierter Versionen der Antwortlatenzen für die Beeinflussbarkeit der Angaben über die Länge des täglichen Fernsehkonsums durch die Kategorisierung der Antwortskala (Ergebnisse logistischer Regressionsanalysen)

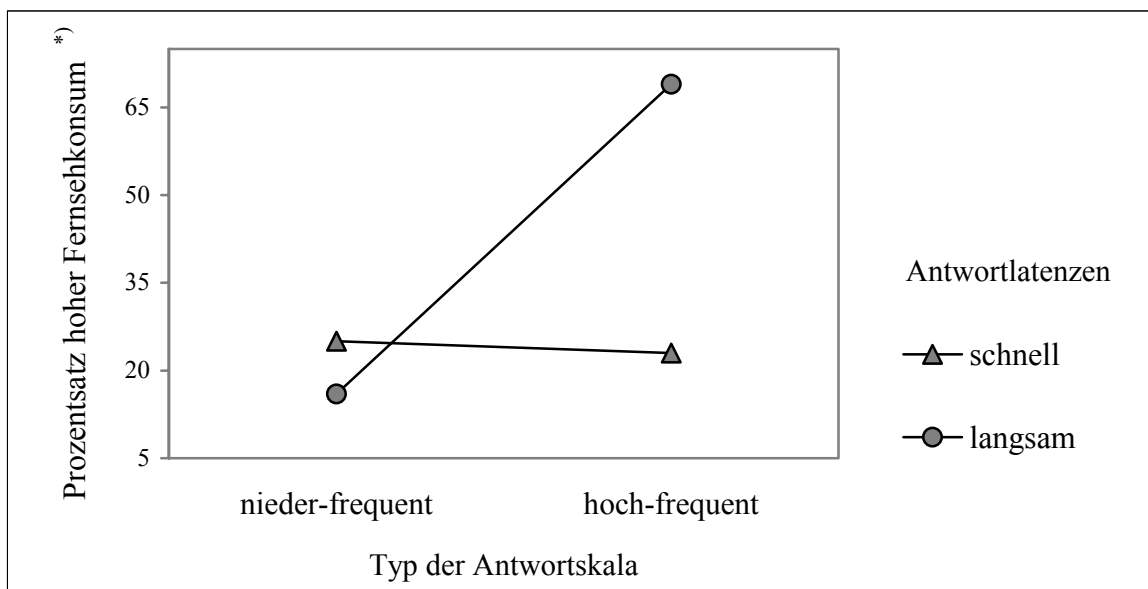
	Modell 7 B (Wald-Statistik)	Modell 8 B (Wald-Statistik)	Modell 9 B (Wald-Statistik)	Modell 10 B (Wald-Statistik)
	Rohreaktionszeit	Logarithmierte Rohreaktionszeit	Wurzeltransformierte Rohreaktionszeit	Standardisierte Rohreaktionszeit
Kontroll-Variablen				
(1) BILDUNG (Jahre)	-0.46 (7.57)**	-0.48 (7.89)**	-0.48 (7.82)**	-0.41 (6.41)**
(2) ALTER (Jahre)	0.02 (1.66)	0.02 (2.20)	0.02 (1.84)	0.02 (2.22)
Erklärende Variablen				
(3) SKALENTYP (hoch-frequent) ¹	1.67 (10.14)**	1.53 (8.51)**	1.63 (9.68)**	1.74 (10.76)**
(4) ANTWORT- LATENZ (z-Werte) ²	0.12 (0.14)	0.12 (0.09)	0.13 (0.14)	0.08 (0.06)
(5) SKALENTYP • LATENZ	1.25 (4.28)**	1.30 (4.77)**	1.20 (4.64)**	1.36 (5.29)**
(6) Konstante	2.38 (1.34)	2.34 (1.31)	2.45 (1.41)	1.71 (0.72)
Verbesserung Pseudo- R ² bei Einführung der SKALENTYP • LATENZ Interaktion	0.038	0.037	0.038	0.044
Pseudo-R ² Gesamtmodell	0.224	0.256	0.239	0.233
N	110	110	110	110

¹ Referenzkategorie: „nieder-frequent“; ² Es handelt sich bei den verschiedenen Modellen um den jeweils in der Kopfzeile der Tabelle beschriebenen Typ an Antwortlatenz; abhängige Variable: 0=„0 bis 2.5 Stunden Fernsehkonsum“, 1=„mehr als 2.5 Stunden Fernsehkonsum“; die Variable „Skalentyp“ geht in Dummy-Kodierung (0,1) in die Analyse ein; Signifikanz: * p ≤ 0.1; ** p ≤ 0.05

In Abbildung 2 wird der Inhalt des Interaktionseffektes zwischen den verschiedenen kategorisierten Antwortskalen und den standardisierten Antwortlatenzen dargestellt. Die Ergebnisse stimmen vollständig mit den bei der Untersuchung der subjektiven Antwortsicherheiten festgestellten überein. So berichten Probanden, die die Frage nach ihrem Fernsehverhalten schnell beantwortet haben, bei der nieder-frequenten zu 25 Prozent und bei der hoch-frequenten Antwortskala zu 23 Prozent einen tägli-

chen Fernsehkonsum von über 2.5 Stunden. Das Antwortverhalten dieser Gruppe von Befragten mit hoher Informationsverfügbarkeit bleibt somit durch die Art der Antwortskala praktisch unbeeinflusst. Dagegen finden sich bei Befragten mit relativ zu ihrer normalen Antwortgeschwindigkeit langen Latenzzeiten sehr starke Skaleneinflüsse. Hier steigt der Anteil der Probanden, die auf der Grundlage der eigenen Angaben als starke Fernsehkonsumenten klassifiziert werden müssen, ausgehend von einem Anteil von 16 Prozent bei der nieder-frequenten Skala auf 70 Prozent bei der hoch-frequenten Skalenversion.

Abbildung 2 Prognosekraft der standardisierten Antwortlatenzen für die Beeinflussbarkeit der Befragten durch unterschiedlich kategorisierte Antwortskalen⁹



^{*)} Prozentsatz der Befragten mit Angaben von mehr als 2.5 Stunden täglichen Fernsehkonsums

4 Zusammenfassung und Diskussion

Die vorliegende Untersuchung hat im ersten Schritt gezeigt, dass sich die unterschiedliche Kategorisierung der verwendeten Antwortskalen in dramatischem Ausmaß auf die inhaltlichen Angaben der Befragten über die Länge ihres täglichen Fernsehkonsums auswirkt. So haben in unserer Untersuchung bei Verwendung einer

⁹ In der vorliegenden Darstellung wurde die Befragtenpopulation hinsichtlich ihrer standardisierten Antwortgeschwindigkeit in zwei identisch große Teilgruppen klassifiziert. Die dargestellten Befragtengruppen mit langen und kurzen Antwortlatenzen repräsentieren jeweils 50 Prozent aller Umfrageteilnehmer.

nieder-frequenten Antwortskala 19.6 Prozent, beim Einsatz einer hoch-frequenten Skala aber 42.5 Prozent der Befragten einen täglichen Fernsehkonsum von mehr als 2.5 Stunden angegeben. Die Stärke dieses Response-Effects von 22.9 Prozentpunkten stimmt beinahe vollständig mit der in der ersten Studie mit dieser Fragestellung beobachteten Effektstärke von 21.3 Prozentpunkten überein (*Schwarz et al. 1985*). Somit wurde erneut gezeigt, dass sich scheinbar irrelevante Veränderungen des Befragungsinstrumentes in sehr starkem Umfang auf die Ergebnisse von Umfragen auswirken können.

Mit Hinblick auf die *erste* der drei zentralen Fragestellungen unserer Studie konnte gezeigt werden, dass sich die Beeinflussbarkeit der in unserer Stichprobe repräsentierten Befragten durch die Darstellung der Antwortskala in sehr starkem Umfang unterscheidet. Dabei hat sich die von den Probanden selbst berichtete Antwortsicherheit als guter Prädiktor für diese Unterschiede erwiesen. Demnach weichen die inhaltlichen Angaben bei den verschiedenen kategorisierten Antwortskalen dann in steigendem Ausmaß voneinander ab, wenn sich die Befragten zunehmend über ihre Angaben unsicher fühlen. Durch den Einsatz dieses meta-kognitiven Indikators der Informationsverfügbarkeit kann somit die in der Literatur bereits indirekt belegte Bedeutung dieses Faktors als zentrale Entstehungsbedingung von Skaleneinflüssen durch direkte Evidenz bestätigt werden (*Schwarz und Hippler 1987*).

Die Ergebnisse der vorliegenden Untersuchung belegen *zweitens* aber auch, dass bei Verwendung der von den Probanden zur Beantwortung der Frage über die Länge ihres Fernsehkonsums benötigten Zeit, deren Empfänglichkeit für Skaleneinflüsse besser als durch die Antwortsicherheiten prognostiziert werden kann. Multivariate Untersuchungsergebnisse haben gezeigt, dass die anfänglich bestehende gute Vorhersagekraft der subjektiven Antwortsicherheit vollständig durch die gleichzeitige Berücksichtigung der Antwortlatenzen absorbiert wird. Damit kann die Annahme bestätigt werden, dass die Antwortlatenzen als operationaler Indikator der Informationsverfügbarkeit, im Vergleich zur meta-kognitiven Operationalisierung mittels der selbstberichteten Antwortsicherheit, als validere Messung angesehen werden kann (*Bassili 1996b*).

Die Untersuchung hat hinsichtlich ihrer *dritten* Fragestellung gezeigt, dass die unterschiedlichen in der Literatur vorgeschlagenen und auch im forschungspraktischen Kontext oft eingesetzten Transformationen der Antwortlatenzen entweder nicht oder nur in recht moderatem Ausmaß zu einer verbesserten Validität der Messergebnisse führen. Ein Vergleich der Prognosekraft der Rohreaktionszeiten für die Beobachtete Stärke der Skaleneinflüsse mit der von logarithmierten und wurzeltransformierten Varianten der Latenzdaten zeigt keine Unterschiede in dieser Hin-

sicht. Ausschließlich die am Mittelwert der „normalen“ Antwortgeschwindigkeit jedes Befragten zentrierte Variante der Antwortlatenzen führt zu einer im Vergleich zu den anderen Latenzvarianten leichten Verbesserung der Vorhersagekraft. Damit kann zwar die Vermutung bestätigt werden, dass die Beseitigung dispositionaler und nicht mit dem Ausmaß der Informationsverfügbarkeit verbundener Unterschiede in der Antwortgeschwindigkeit die Validität dieses Indikators stärkt (*Stocké* 2002). Es finden sich aber keine Belege für unsere Annahme, wonach die uneinheitliche Praxis bei der Transformation von Latenzdaten die Ursache für die teilweise inkonsistenten Ergebnisse über die Prognosekraft von Antwortlatenzen ist.

Eine alternative, bisher aber nicht systematisch untersuchte Erklärung der widersprüchlichen Ergebnislage kann darin gesehen werden, dass Antwortlatenzen außer der zum Abruf der notwendigen Informationen aus dem Gedächtnis benötigten Zeit auch andere Aspekte des Antwortgenerierungsprozesses erfassen. So werden Latenzdaten außerdem durch die von der Befragtenmotivation abhängigen Elaboriertheit der Antworten, durch die Stärke von Anreizen durch soziale Erwünschtheit sowie durch die Komplexität des Befragungsinstrumentes beeinflusst. Mit zunehmendem Gewicht dieser Fremddimensionen der Messung muss mit einem immer stärkeren Rückgang der Validität von Antwortlatenzen als Indikator für die Informationsverfügbarkeit gerechnet werden. Zukünftige Untersuchungen müssen daher zeigen, welche Randbedingungen sich auf die relative Bedeutsamkeit der unterschiedlichen Bestimmungsfaktoren von Latenzdaten und damit auf deren Vorhersagekraft für die Stärke von Response-Effects auswirken.

Was die Schlussfolgerungen hinsichtlich der Umfragepraxis angeht, so kann festgestellt werden, dass die Antwortlatenzen in allen Transformationsvarianten sowie die subjektive Antwortsicherheit die Anfälligkeit bestimmter Befragtengruppen gegenüber Response-Effects und damit die erwartbare Qualität der Umfragedaten vorher sagen können. Es erscheint uns in dieser Hinsicht ratsam, derartige Indikatoren, zumindest bei „gefährdeten“ Fragestellungen, regelmäßig als Zusatzinformation zu den eigentlichen Befragtenangaben mitzuerheben. Antwortlatenzen haben hierbei nicht nur den Vorteil einer besseren Prognosekraft, sondern stehen bei computergestützten Befragungen auch als Nebenprodukt des normalen Interviewprozesses und damit ohne zusätzliche Kosten zur Verfügung. Die Möglichkeit eines fruchtbaren Einsatzes von Antwortlatenzen konnte bereits in Telefonumfragen mit allgemeinen Bevölkerungsstichproben belegt werden (*Bassili* 1993, 1996b, *Bassili* und *Bors* 1997). Die vorliegende Untersuchung zeigt dies auch für den Kontext persönlich-mündlicher Befragungen bei den Umfrageteilnehmern zu Hause. Ob dies auch bei größeren, beispielsweise bundesweiten Umfragen, und den hierbei normalerweise

zum Einsatz kommenden Interviewerstäben möglich ist, muss allerdings noch gezeigt werden.

Literatur

Aiken, Leona S.; West, Stephen G. 1991: Multiple Regression: Testing and Interpreting Interactions. Newbury Park: Sage.

Bargh, John A.; Chartrand, Tanya L. 2000: The Mind in the Middle. A Practical Guide to Priming and Automaticity Research. S. 253-285 in: **Reis, Harry T.; Judd, Charles M.** (Eds.), Handbook of Research Methods in Social and Personality Psychology. Cambridge, UK: Cambridge University Press.

Bassili, John N. 1993: Response Latency versus Certainty as Indexes of the Strength of Voting Intentions in a CATI Survey. In: Public Opinion Quarterly: 54-61.

Bassili, John N. 1995: Response Latency and the Accessibility of Voting Intentions: What Contributes to Accessibility and How it Affects Vote Choice. In: Personality and Social Psychology Bulletin: 686-695.

Bassili, John N. 1996a: The How and Why of Response Latency Measurement in Telephone Surveys. S. 319-346 in: **Schwarz, Norbert; Sudman, Seymour** (Eds.), Answering Questions. Methodology for Determining Cognitive and Communicative Processes in Survey Research. San Francisco: Jossey-Bass.

Bassili, John N. 1996b: Meta-Judgmental versus Operative Indexes of Psychological Attributes: The Case of Measures of Attitude Strength. In: Journal of Personality and Social Psychology: 637-653.

Bassili, John N.; Bors, Douglas A. 1997: Using Response-Latency to Increase Lead Time in Election Forecasting. In: Canadian Journal of Behavioural Science: 231-238.

Bassili, John N.; Fletcher, Joseph F. 1991: Response-Time Measurement in Survey Research. A Method for CATI and a New Look at Nonattitudes. In: Public Opinion Quarterly: 331-346.

Bassili, John N.; Krosnick, John A. 2000: Do Strength-Related Attitude Properties Determine Susceptibility to Response Effects? New Evidence From Response Latency, Attitude Extremity, and Aggregate Indices. In: Political Psychology: 107-132.

Bassili, John N.; Scott, B. S. 1996: Response Latency as a Signal to Question Problems in Survey Research. In: Public Opinion Quarterly: 390-399.

Bless, Herbert; Bohner, Gerd; Hild, Traudel; Schwarz; Norbert 1992: Asking Difficult Questions: Task Complexity Increases the Impact of Response Alternatives. In: European Journal of Social Psychology: 309-312.

Cronbach, Lee J. 1987: Statistical Tests for Moderator Variables: Flaws in Analyses Recently Proposed. In: Psychological Bulletin: 414-417.

Fazio, Russell H. 1990: Multiple Processes by which Attitudes Guide Behavior: The Mode Model as an Integrative Framework. S. 75-104 in: **Zanna, Mark P.** (Ed.), Advances in Experimental Social Psychology. San Diego: Academic Press.

Fazio, Russell H. 1990: A Practical Guide to the Use of Response Latency in Social Psychological Research. S. 74-97 in: **Hendrick, Clyde; Clark, Margaret S.** (Eds.), Review of Personality and Social Psychology: Vol. 11. Research Methods in Personality and Social Psychology. Newbury Park, CA: Sage.

Fazio, Russell H.; Williams, Carol J. 1986: Attitude Accessibility as a Moderator of the Attitude-Perception and Attitude-Behavior Relations: An Investigation of the 1984 Presidential Election. In: Journal of Personality and Social Psychology: 505-514.

- Fletcher, Joseph F.** 2000: Two Timing: Politics and Response Latencies in a Bilingual Survey. In: *Political Psychology*: 27-55.
- Gaskell, George D.; O'Muircheartaigh, Colm A.; Wright, Daniel B.** 1994: Survey Questions about the Frequency of Vaguely Defined Events. In: *Public Opinion Quarterly*: 241-254.
- Huckfeldt, Robert; Sprague, John** 2000: Political Consequences of Inconsistency: The Accessibility and Stability of Abortion Attitudes. In: *Political Psychology*: 57-79.
- Ji, Li-Jun; Schwarz, Norbert; Nisbett, Richard E.** 2000: Culture, Autobiographical Memory, and Behavioral Frequency Reports: Measurement Issues in Cross-Cultural Studies. In: *Personality and Social Psychology Bulletin*: 585-593.
- Kokkinaki, Flora; Lunt, Peter** 1997: The relationship between involvement, attitude accessibility and attitude-behaviour consistency. In: *British Journal of Social Psychology*: 497-509.
- Krosnick, Jon A.** 1989: Attitude Importance and Attitude Accessibility. In: *Personality and Social Psychology Bulletin*: 297-308.
- Krosnick, Jon A.; Petty, Richard E.** 1995: Attitude Strength: An Overview. S. 1-24 in: **Petty, Richard E.; Krosnick, Jon A.** (Eds.), *Attitude Strength. Antecedents and Consequences*. Mahwah, N.J.: Erlbaum.
- Krosnick, Jon A.; Schuman, Howard** 1988: Attitude Intensity, Importance, and Certainty and Susceptibility to Response Effects. In: *Journal of Personality and Social Psychology*: 940-952.
- Lavine, Howard; Huff, Joseph W.; Wagner, Stephen H.; Sweeney, Donna** 1998: The Moderating Influence of Attitude Strength on Susceptibility to Context Effects in Attitude Surveys. In: *Journal of Personality and Social Psychology*: 359-373.
- Menon, Geeta; Raghurir, Priya; Schwarz, Norbert** 1995: Behavioral Frequency Judgements: An Accessibility-Diagnosticity Framework. In: *Journal of Consumer Research*: 212-228.
- Menon, Geeta; Raghurir, Priya; Schwarz, Norbert** 1997: How Much Will I Spend? Factors Affecting Consumer's Estimates of Future Expense. In: *Journal of Consumer Psychology*: 141-164.
- Mielke, Rosemarie** 1985: Eine Untersuchung zum Umweltschutz- Verhalten (Wegwerf- Verhalten): Einstellung, Einstellungs- Verfügbarkeit und soziale Normen als Verhaltensprädiktoren. In: *Zeitschrift für Sozialpsychologie*: 196-205.
- Pecher, Diane; Raaijmakers, Jeroen G. W.** 1999: Automatic Priming Effects for New Associations in Lexical Decision and Perceptual Identification. In: *The Quarterly Journal of Experimental Psychology*: 593-614.
- Renata, Franc** 1999: Attitude strength and the attitude-behavior domain: magnitude and independence of moderating effects of different strength indices. In: *Journal of Social Behavior & Personality*: 177-197.
- Rockwood, Todd H.; Sangster, Roberta L.; Dillman, Don A.** 1997: The Effect of Response Categories on Questionnaire Answers. In: *Sociological Methods & Research*: 118-140.
- Schuman, Howard; Presser, Stanley; Ludwig, Jacob** 1981: Context Effects on Survey Responses to Questions About Abortion. In: *Public Opinion Quarterly*: 216-223.
- Schwarz, Norbert** 1988: Was Befragte aus Antwortvorgaben lernen: Zur informativen Funktion von Antwortvorgaben bei Verhaltensfragen. In: *Planung und Analyse*: 103-107.
- Schwarz, Norbert** 1999: Frequency Reports of Physical Symptoms and Health Behaviors: How the Questionnaire Determines the Results. S. 93-108 in: **Park, Denise C.; Morell, R. W.; Shifren, K.** (Eds.), *Processing in Medical Information in Aging Patients: Cognitive and Human Factors Perspectives*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Schwarz, Norbert; Bienias, Julia** 1990: What Mediates the Impact of Response Alternatives on Frequency Reports of Mundane Behaviors. In: *Applied Cognitive Psychology*: 61-72.

- Schwarz, Norbert; Grayson, Carla E.; Knäuper, Bärbel** 1998: Formal Features of Rating Scales and the Interpretation of Question Meaning. In: *International Journal of Public Opinion Research*: 177-183.
- Schwarz, Norbert; Hippler, Hans J.** 1987: What Response Scales May Tell your Respondents: Informative Functions of Response Alternatives. S. 163-178 In: **Hippler, Hans J.; Schwarz, Norbert; Sudman, Seymour** (Eds.), *Social Information Processing and Survey Methodology*. New York: Springer.
- Schwarz, Norbert; Hippler, Hans-J.; Deutsch, Brigitte; Strack, Fritz** 1985: Response Scales: Effects of Category Range on Reported Behavior and Comparative Judgments. In: *Public Opinion Quarterly*: 388-395.
- Schwarz, Norbert; Scheuring, Bettina** 1988: Judgements of Relationship Satisfaction: Inter- and Intraindividual Comparisons as a Function of Questionnaire Structure. In: *European Journal of Social Psychology*: 485-496.
- Schwarz, Norbert; Strack, Fritz** 1991: Context Effects in Attitude Surveys: Applying Cognitive Theory to Social Research. In: *European Review of Social Psychology*: 31-50.
- Stocké, Volker** 2001: The Influence of Frequency Scales on the Response Behavior. A Theoretical Model and its Empirical Examination. Arbeitspapier No. 02-35: Sonderforschungsbereich 504, Universität Mannheim.
- Stocké, Volker** 2002: Die Vorhersage von Fragenreihenfolgeeffekten durch Antwortlatenzen: Eine Validierungsstudie. In: *ZUMA-Nachrichten*: 26-53.
- Stocké, Volker** 2003: Einstellungen zu Umfragen. in: **Glöckner-Rist, Angelika** (Hrsg.), *ZUMA- Informationssystem. Elektronisches Handbuch sozialwissenschaftlicher Erhebungsinstrumente*. Mannheim: Zentrum für Umfragen, Methoden und Analysen.
- Warland, Rex H.; Sample, John** 1973: Response Certainty as a Moderator Variable in Attitude Measurement. In: *Rural Sociology*: 174-186.
- Wegener, Duane T.; Downing, John; Krosnick, Jon A.** 1995: Measures and Manipulations of Strength-Related Properties of Attitudes: Current Practice and Future Directions. S. 455-487 in: **Petty, Richard E.; Krosnick, Jon A.** (Eds.), *Attitude Strength. Antecedents and Consequences*. Mahwah, N.J.: Erlbaum.
- Winter, Joachim** 2002: Bracketing Effects in Categorized Survey Questions and the Measurement of Economic Quantities. Arbeitspapier No. 01-25: Sonderforschungsbereich 504, Universität Mannheim.