

Institut für Agrarpolitik und Marktforschung  
der Justus-Liebig-Universität Giessen

Arbeitsbericht

**Nr. 51**

TOBIAS HENKEL

**Die Wirkung der Befragungsform  
auf das Antwortverhalten:  
Eine vergleichende Untersuchung am Beispiel  
des Carbon Footprints bei Lebensmitteln**

Gießen 2010

Bestell-Nr. 10/02

Anschrift des Instituts:

Senckenbergstr. 3  
35390 GIESSEN

Tel. Nr. 0641/99-37020; Fax: 0641/99-37029  
email: Sekretariat.Marktlehre@agrار.uni-giessen.de



## **Vorwort**

Das Internet ist aus vielen Bereichen des alltäglichen Lebens nicht mehr wegzudenken und hat in vielen Teilen des Alltags für deutliche Veränderungen gesorgt. Dieser Sachverhalt gilt auch für die Wissenschaft. Während über viele Jahre hinweg die mündliche Befragung als das Standardverfahren der Sozialwissenschaften galt, zeigen sich seit der Jahrtausendwende und der zunehmenden Verbreitung des Internets Tendenzen dahingehend, dass die mündliche Befragung zunehmend von online durchgeführten Befragungen verdrängt wird. Relevant hierbei ist jedoch, ob durch eine derartige Veränderung der Befragungsmethode dabei Veränderungen im Antwortverhalten der Befragten zu erwarten sind. Dieser Fragestellung widmet sich die vorliegende Studie am Beispiel des aktuell diskutierten „Product Carbon Footprints“ bei Lebensmitteln, welcher den Ausstoß an Treibhausgasen bei der Produktion eines Lebensmittels kennzeichnet. Anhand einer mündlichen Befragung sowie einer zeitnah durchgeführten Online-Befragung mit jeweils dem gleichen Fragebogen zu diesem Thema werden Unterschiede im Antwortverhalten mittels verschiedener statistischer Verfahren herausgearbeitet. Hierbei zeigt sich, dass nicht grundsätzlich von einem systematischen Einfluss der Befragungsmethode auf das Antwortverhalten ausgegangen werden kann.

Der vorliegende Arbeitsbericht stellt dabei eine überarbeitete Fassung einer empirischen Studie dar, die als Master-Thesis am Institut für Agrarpolitik und Marktforschung verfasst wurde. Ein besonderer Dank gilt dabei den, im Rahmen der empirischen Untersuchung eingesetzten, studentischen Interviewern, die durch ihre Tätigkeit maßgeblich zum Erfolg der Studie beigetragen haben. Sie haben es ermöglicht, dass innerhalb eines relativ kurzen Zeitraumes eine große Anzahl an Personen sachgerecht befragt werden konnte.

Ein weiterer Dank geht an die MitarbeiterInnen der Professur für Marktlehre der Agrar- und Ernährungswirtschaft. Sie haben mit verschiedenen hilfreichen Vorschlägen und Hinweisen sowie der Diskussionsmöglichkeit ebenfalls zum Erfolg der Studie beigetragen.

Gießen, im Juli 2010

Tobias Henkel

## **Kurzfassung**

Die vorliegende Arbeit führt einen empirischen Methodenvergleich zwischen einer mündlichen Befragung und einer Online-Befragung durch. Während F2F-Befragungen lange als „Königsweg“ der empirischen Sozialforschung galten, zeigen sich seit der Jahrtausendwende Tendenzen dahingehend, dass zunehmend mündliche Befragungen durch online durchgeführte Befragungen substituiert werden. Dieser Sachverhalt ist nicht verwunderlich, versprechen Online-Befragungen im Vergleich Vorteile wie niedrigere Durchführungskosten, die Möglichkeit der Integration multimedialer Inhalte oder auch eine höhere Durchführungsgeschwindigkeit. Bisher nur wenig erforscht ist allerdings die Fragestellung, ob bei der Substitution einer F2F-Befragung durch eine Online-Befragung mit Effekten im Antwortverhalten zu rechnen ist, die unter dem Stichwort der Methodeneffekte beschrieben werden können. Zur Beantwortung dieser Fragestellung wird ein Methodenvergleich unter Zuhilfenahme der sogenannten Split-Ballot-Technik durchgeführt, bei der das Antwortverhalten einer mündlich befragten Teilstichprobe (n=186) mit dem einer online befragten Teilstichprobe (n=514) über Signifikanztests sowie Regressionsanalysen verglichen wird. Um Ergebnisverzerrungen durch unterschiedliche Verteilungen der soziodemographischen Variablen zu vermeiden, wird eine Angleichung der Teilstichproben über Gewichtungsfaktoren vorgenommen. Als Ergebnis kann festgehalten werden, dass sich keine generellen Hinweise auf das Vorliegen von Methodeneffekten ergeben.

## **Abstract**

This paper conducts a comparison between the survey method “Face-to-Face” and an Online-Survey. For a long time, Face-to-Face-Interviews were the most used survey method in the empirical social research. But since the turn of the Millenium tendencies can be found, that interviews are substituted by online surveys. This fact is not surprising. Online surveys promise a lot of advantages like lower costs, advantages in time and the potential to integrate multimedial contents into the survey. The aim of the study is to point out differences in the answers of the participants, described in the literature as “mode effects”. The literature review shows that only a few studies focussed this case in retrospective for these two survey methods. To address this research topic a comparison of survey methods is done. Applying the split-ballot technique the answers of the Face-to-Face survey (n=186) and the online survey (n=514) are compared by statistical tests and a multiple regression analysis. To eliminate biases determined by unequal distributions of the socio-demographic variables in the surveys, a weighting procedure is used. The results indicate no general evidence for mode effects in the study.

## Inhaltsverzeichnis

<b>Abbildungsverzeichnis .....</b>	<b>IV</b>
<b>Tabellenverzeichnis .....</b>	<b>V</b>
<b>Verzeichnis der Tabellen im Anhang .....</b>	<b>VII</b>
<b>Abkürzungsverzeichnis .....</b>	<b>IX</b>
<b>1 EINLEITUNG.....</b>	<b>1</b>
1.1 Problemstellung .....	1
1.2 Zielsetzung .....	2
1.3 Vorgehensweise.....	3
<b>2 KLIMAWANDEL UND CARBON LABELLING-STRATEGIEN.....</b>	<b>5</b>
2.1 Klimawandel.....	5
2.1.1 Natürlicher Klimawandel .....	5
2.1.2 Anthropogener Klimawandel .....	6
2.1.3 Folgen des Klimawandels.....	8
2.2 Carbon-Labeling-Strategien .....	12
2.2.1 Carbon Labelling aus Konsumenten- und Unternehmenssicht .....	12
2.2.2 Definition des Product Carbon Footprints.....	14
2.2.3 Methodischer Hintergrund.....	16
<b>3 RELEVANTE METHODEN DER EMPIRISCHEN SOZIALFORSCHUNG.....</b>	<b>21</b>
3.1 Face-to-Face-Interviews .....	21
3.2 Online-Befragungen .....	30
3.3 Fehlerarten in der Befragung.....	38
3.3.1 Zufällige und systematische Fehler .....	39
3.3.2 Stichprobenfehler.....	40
3.3.3 Messfehler .....	47
<b>4 LITERATURÜBERBLICK: EMPIRISCHE ERGEBNISSE BISHERIGER STUDIEN ZU METHODENEFFEKTEN .....</b>	<b>57</b>
<b>5 EMPIRISCHE ANALYSE .....</b>	<b>65</b>
5.1 Arbeitshypothesen .....	65
5.2 Darstellung des Untersuchungsdesigns .....	66
5.2.1 Split-Ballot-Technik .....	66
5.2.2 Auswahl der Zielpopulation .....	67

---

5.2.3	Fragebogen .....	70
5.2.4	Durchführung der Befragungen.....	75
5.2.5	Aufbereitung und Auswertung der Daten.....	81
5.3	Ergebnisse zu den soziodemographischen Variablen.....	82
5.3.1	Soziodemographische Daten .....	82
5.3.2	Dauer der Fragenbeantwortung .....	89
5.3.3	Gewichtung der Fälle .....	92
5.3.4	Beantwortung der zurückgestellten Frage zum Einkommen.....	97
5.4	Ergebnisse der inhaltlichen Fragestellungen des Fragebogens .....	98
5.4.1	Fragen zur Relevanz der Stärkung verschiedener politischer Aufgaben- bereiche.....	98
5.4.2	Fragestellungen zum Themenkomplex „Klimaschutz“ .....	100
5.4.3	Verhaltensfrage zum Themenkomplex „Grüner“ Strom .....	104
5.4.4	Resultate des Quiz zum Themenkomplex „Klimaschutz“ .....	105
5.4.5	Überzeugungsfrage zu verschiedenen Gütezeichen .....	108
5.4.6	Fragestellungen zum Themenkomplex „Carbon Labelling“ .....	110
5.4.7	Zahlungsbereitschaft für Milch mit einer reduzierten CO <sub>2</sub> -Emission.....	116
5.5	Zwischenfazit .....	117
5.6	Einstellungsdimensionen der Befragten .....	120
5.7	Regressionsanalytische Auswertung .....	127
<b>6</b>	<b>DISKUSSION .....</b>	<b>133</b>
<b>7</b>	<b>ZUSAMMENFASSUNG UND AUSBLICK .....</b>	<b>138</b>
<b>8</b>	<b>LITERATURVERZEICHNIS .....</b>	<b>143</b>
<b>Anhang</b>	<b>.....</b>	<b>163</b>
	Fragebogen zur Thematik Carbon Labelling.....	165
	Verwendete Skalen sowie Labels .....	179
	Tabellen im Anhang .....	182

**Abbildungsverzeichnis**

Abbildung 1:	Schematische Darstellung des Treibhauseffektes .....	6
Abbildung 2:	Durch Verbrennung fossiler Brennstoffe verursachte CO <sub>2</sub> -Emissionen, Zeitraum 1751-2006 .....	7
Abbildung 3:	Erwartete Veränderungen der Durchschnittstemperatur im Zeitverlauf nach Szenarien.....	10
Abbildung 4:	Entwicklung der Prozentanteile der Befragungsmethoden im Zeitraum 1990 bis 2008 .....	21
Abbildung 5:	Die Interviewsituation als Reaktionssystem.....	25
Abbildung 6:	Morphologie der Fehlerarten im Rahmen der Befragung .....	39
Abbildung 7:	Undercoverage und Overcoverage .....	41
Abbildung 8:	(Non-)Response-Typen in Online-Befragungen .....	46
Abbildung 9:	Verzerrung durch soziale Erwünschtheit.....	52
Abbildung 10:	Darstellung einer Frage in LimeSurvey .....	74
Abbildung 11:	Histogramm zur Verteilung der Variablen "Alter" in der Gesamtstichprobe (n=697) .....	84
Abbildung 12:	Histogramm zur Verteilung der Variablen "Alter" in der F2F-Stichprobe (n=186) .....	85
Abbildung 13:	Histogramm zur Verteilung der Variablen "Alter" in der Online-Stichprobe (n=511) .....	85
Abbildung 14:	Graphische Darstellung der Häufigkeitsverteilungen der Variablen „Verfügbares Nettoeinkommen“ der beiden Teilstichproben .....	89
Abbildung 15:	Anzahl der korrekt beantworteten Items dieser Fragestellung .....	107
Abbildung 16:	Prozentanteile der korrekt gegebenen Teilantworten nach Stichproben ....	111
Abbildung 17:	Semantische Differentiale der beiden Teilstichproben zum Product Carbon Footprint-Zeichen .....	113
Abbildung 18:	Häufigkeitsverteilungen zur Fragestellung, ob ein CO <sub>2</sub> -Zeichen freiwillig oder gesetzlich vorgeschrieben sein sollte (F2F n=186; Online n=514).....	115
Abbildung 19:	Häufigkeitsverteilungen der Fragestellung zur Gestaltung eines CO <sub>2</sub> -Zeichens (F2F n=186; Online n=514) .....	115

## Tabellenverzeichnis

Tabelle 1:	Vergleich verschiedener europäischer PCF-Ansätze .....	19
Tabelle 2:	Verschiedene Formen der mündlichen Befragung .....	28
Tabelle 3:	Arten von Online-Befragungen .....	35
Tabelle 4:	Zusammenfassender Vergleich der beiden Befragungsformen F2F-Interview und Online-Befragung hinsichtlich ausgewählter Kriterien .....	38
Tabelle 5:	Häufigkeitsverteilung der Variablen "Geschlecht" in den Teilstichproben sowie in der Grundgesamtheit der JLU .....	83
Tabelle 6:	Deskriptive Statistik der Variablen "Alter", unterteilt nach Befragungsmodus.....	84
Tabelle 7:	Häufigkeitsverteilungen der Variablen "Fachbereich" in den Teilstichproben sowie in der Grundgesamtheit der JLU .....	86
Tabelle 8:	Häufigkeitsverteilungen der Variablen "Momentan angestrebter Abschluss" in den Teilstichproben sowie in der Grundgesamtheit der JLU .....	87
Tabelle 9:	Häufigkeitsverteilungen der Variablen "Verfügbares Nettoeinkommen" in den Teilstichproben sowie in der Gesamtstichprobe .....	88
Tabelle 10:	Deskriptive Statistik der Interviewdauer, unterteilt nach Befragungsmodus .....	90
Tabelle 11:	Zusammensetzung der Fachbereichs-Gruppen.....	96
Tabelle 12:	Änderung der Häufigkeitsverteilung der Variablen "Geschlecht" nach Durchführung der Gewichtung.....	96
Tabelle 13:	Tabelle 8: Änderung der Häufigkeitsverteilung der Variablen "Klassifizierter Fachbereich" nach Durchführung der Gewichtung.....	97
Tabelle 14:	Prozentanteile der Einkommensklassen in den beiden Befragungssamples vor und nach der Gewichtung.....	98
Tabelle 15:	Mittelwerte, Mittelwertdifferenzen sowie Ergebnisse des Mann-Whitney U-Tests für die Items der Fragestellung zur Relevanz einer zusätzlichen Stärkung verschiedener politischer Aufgabenbereiche in Deutschland.....	99
Tabelle 16:	Mittelwerte, Mittelwertdifferenzen sowie Ergebnisse des Mann-Whitney U-Tests für die Items der Einstellungsfrage zum Themenkomplex „Klimaschutz“ .....	101
Tabelle 17:	Mittelwerte, Mittelwertdifferenzen sowie Ergebnisse des Mann-Whitney U-Tests für die Items der Verhaltensfrage zum Themenkomplex „Klimaschutz“ .....	103
Tabelle 18:	Häufigkeiten sowie Prozentanteile der Fragestellung "Beziehst Du momentan 'Grünen' Strom oder hast Du schon mal 'Grünen' Strom bezogen?" .....	104

---

Tabelle 19: Resultate des Quiz zum Klimaschutz, aufgeteilt nach Teilstichproben. ....	106
Tabelle 20: Kenntnis der Befragten über verschiedene Zeichen aus der Bereich der Ernährungindustrie bzw. des Lebensmittelhandels .....	108
Tabelle 21: Deskriptive Statistik der Variablen "Zahlungsbereitschaftsdifferenz" .....	117
Tabelle 22: Zusammenstellung der Items bzw. Fragestellungen, bei denen signifikante Unterschiede zwischen den Teilstichproben ermittelt werden konnten .....	118
Tabelle 23: Rotierte Faktorladungsmatrix .....	121
Tabelle 24: Arithmetische Mittel der Einstellungsdimensionen in den Teilstichproben .....	125
Tabelle 25: Ergebnisse der Regression.....	129

## Verzeichnis der Tabellen im Anhang

Tabelle A 1:	Kolmogorov-Smirnov-Test auf Normalverteilung der Variablen "Alter" in der Gesamtstichprobe .....	182
Tabelle A 2:	Kolmogorov-Smirnov-Test auf Normalverteilung der Variablen "Alter" in den beiden Teilstichproben.....	182
Tabelle A 3:	Kolmogorov-Smirnov Z-Test auf unterschiedliche Verteilungen der Variablen "Alter" in den beiden Teilstichproben „F2F“ sowie „Online“ .....	182
Tabelle A 4:	Genauere Bezeichnungen der Fachbereiche (Legende zu den Abkürzungen der Tabelle 7).....	182
Tabelle A 5:	Ergebnisse des Mann-Whitney U-Tests zur klassifizierten Einkommensvariablen (ungewichtet).....	183
Tabelle A 6:	Kolmogorov-Smirnov-Test auf Normalverteilung der Variablen "Interviewdauer" in der Gesamtstichprobe.....	183
Tabelle A 7:	Ergebnisse des Mann-Whitney U-Tests zur Interviewdauer.....	183
Tabelle A 8:	Ergebnisse der Kontingenzanalyse zur Ermittlung des Zusammenhangs zwischen der Befragungsmethode und demographischen Merkmalen .....	183
Tabelle A 9:	Mann-Whitney U-Test zur Altersvariablen.....	183
Tabelle A 10:	Ergebnisse des Mann-Whitney U-Tests zur klassifizierten Einkommensvariablen (gewichtet).....	183
Tabelle A 11:	Ergebnisse des Kolmogorov-Smirnov-Tests zur Überprüfung der Items der Fragestellung zur Relevanz der Stärkung verschiedener politischer Aufgabenbereiche auf Normalverteilung .....	184
Tabelle A 12:	Ergebnisse des Mann-Whitney U-Tests zu den Items der Fragestellung zur Relevanz der Stärkung verschiedener politischer Aufgabenbereiche .....	184
Tabelle A 13:	Ergebnisse des Mann-Whitney U-Tests zu den Items der Einstellungsfrage zum Themenkomplex „Klimaschutz“ .....	184
Tabelle A 14:	Ergebnisse des Mann-Whitney U-Tests zu den Items der Verhaltensfrage zum Themenkomplex „Klimaschutz“. Genaue Fragestellung: „Wie oft hast Du bereits eine der folgenden Maßnahmen unternommen?“.....	185
Tabelle A 15:	Ergebnisse der Kontingenzanalyse zur Ermittlung des Zusammenhangs zwischen dem Bezug von „Grünem“ Strom und der Befragungsmethode .....	185
Tabelle A 16:	Ergebnisse der Kontingenzanalyse zur Ermittlung des Zusammenhangs zwischen der Beantwortung der einzelnen Quiz-Items und der Befragungsmethode.....	185
Tabelle A 17:	Ergebnisse des Kolmogorov-Smirnov-Tests zur Überprüfung der korrekt gegebenen Antworten des Klimaschutz-Quiz auf Normalverteilung.....	185

Tabelle A 18: Ergebnisse des Mann-Whitney U-Tests zu der Anzahl der korrekt gegebenen Antworten des Klimaschutz-Quiz.....	186
Tabelle A 19: Ergebnisse des Mann-Whitney U-Tests für die verschiedenen Zeichen. Genauer Wortlaut der Fragestellung: „Hast Du dieses Zeichen (bzw. Siegel) schon mal gesehen? Kennst Du diese Zeichen?“ .....	186
Tabelle A 20: Ergebnisse des Mann-Whitney U-Tests zu den korrekt gegebenen Teilerantworten hinsichtlich des Product Carbon Footprints (PCF).....	186
Tabelle A 21: Ergebnisse des Mann-Whitney U-Tests zum semantischen Differential des Product Carbon Footprints. Genauer Wortlaut der Fragestellung „Hältst Du ein solches ‚CO2-Zeichen‘ für...“, gefolgt von den Eigenschaftspaaren .....	186
Tabelle A 22: Ergebnisse der Kontingenzanalyse zur Ermittlung des Zusammenhangs zwischen der Frage nach der rechtlichen Regelung eines solchen PCFs und der Befragungsmethode.....	187
Tabelle A 23: Ergebnisse der Kontingenzanalyse zur Ermittlung des Zusammenhangs zwischen der Frage nach Ausgestaltung eines solchen PCFs und der Befragungsmethode .....	187
Tabelle A 24: Ergebnisse des Kolmogorov-Smirnov-Tests zur Überprüfung der gebildeten Variablen „Zahlungsbereitschaftsdifferenz“ auf Normalverteilung (Gesamtstichprobe).....	187
Tabelle A 25: Ergebnisse des Kolmogorov-Smirnov-Tests zur Überprüfung der gebildeten Variablen „Zahlungsbereitschaftsdifferenz“ auf Normalverteilung (Teilstichproben) .....	187
Tabelle A 26: Ergebnisse des Mann-Whitney U-Tests zu der Variablen „Zahlungsbereitschaftsdifferenz“ .....	187
Tabelle A 27: Erklärte Gesamtvarianz durch die Faktorenanalyse .....	188
Tabelle A 28: Ergebnisse des t-Tests zum Mittelwertvergleich der Einstellungsdimensionen aus der Faktorenanalyse (F2F n=176; Online n=477) .....	188

**Abkürzungsverzeichnis**

ADM	Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute
AGOF	Arbeitsgemeinschaft Online Forschung e. V.
BIP	Bruttoinlandsprodukt
CATI	Computer Assisted Telephone Interview
CQ	Chi-Quadrat-Test nach Pearson
CSAQ	Computerized Self-Administered Questionnaires
EF	Ecological Footprint
EIA	Energy Information Administration
F2F	Face-to-Face
HDE	Hauptverband des deutschen Einzelhandels
IPCC	Intergovernmental Panel on Climate Change
ISO	International Organization for Standardization
JLU	Justus-Liebig-Universität
KMO	Kaiser-Meyer-Olkin-Kriterium
LCA	Life Cycle Analysis
MW	Mann-Whitney U-Test
NSSE	National Survey of Student Engagement
n.s.	nicht signifikant
OECD	Organisation for Economic Co-Operation and Development
PAS	Public Available Standard
PASW	Predictive Analytics Software
PCF	Product Carbon Footprint
SPSS	Statistical Package for Social Sciences
THG	Treibhausgas
UNEP	United Nations Environment Programme
WTO	World Trade Organization

## 1 Einleitung

### 1.1 Problemstellung

Über viele Jahre hinweg war die mündliche Befragung, bei der eine Befragungsperson von einem Interviewer zu einer bestimmten Thematik befragt wird, in der empirischen Sozialforschung die am häufigsten anzutreffende Befragungsmethode. Aufgrund der häufigen Verwendung und verschiedener Vorteile, die mit ihr verbunden werden, galt sie über viele Jahre als „das“ Standardverfahren und als der „Königsweg“ in den Sozialwissenschaften.

Seit der Jahrtausendwende und der stetig zunehmenden Verbreitung des Internets zeigen sich jedoch Tendenzen dahingehend, dass die mündliche Befragung von online durchgeführten Befragungen mehr und mehr verdrängt wird. Der immer noch andauernde Siegeszug des Internets in vielen Bereichen des alltäglichen Lebens hat nicht vor der Wissenschaft und vor wissenschaftlichen Untersuchungsmethoden Halt gemacht. Dies lässt sich auch anhand von Zahlen des ARBEITSKREISES DEUTSCHER MARKT- UND SOZIALFORSCHUNGSINSTITUTE belegen: So lag im Jahre 1990, vor der weltweiten Expansion des Internets, der Anteil der mündlichen Befragungen noch bei 65 %. Alternativ herangezogene Befragungsmethoden wie telefonische oder schriftliche Befragungen wurden entsprechend in der Summe nur bei 35 % der durchgeführten Befragungsprojekte verwendet. Im Jahre 2008 hingegen zeigt sich ein deutlich anderes Bild: Während nur noch 21 % der Befragungen mündlich durchgeführt werden, ist der Anteil der online durchgeführten Befragungen bereits auf 31 % angestiegen (ADM 2008: 12).

Aufgrund verschiedener, mit Online-Befragungen verbundener Vorteile wie einer schnelleren Verfügbarkeit der Daten, Kostenvorteilen oder auch der möglichen Implementierung multimedialer Inhalte ist mit einem weiteren Anstieg des Anteils von Online-Befragungen in der Zukunft zu rechnen. An dieser Stelle stellt sich jedoch die Frage, ob die den vorangegangenen Zahlen entnehmbare Substitution von mündlichen Befragungen durch Online-Befragungen ausschließlich mit vermeintlich positiven Aspekten verbunden ist oder ob auch negative Auswirkungen zu erwarten sind. Eine wichtige Fragestellung diesbezüglich lautet, ob mündliche Befragungen und online durchgeführte Befragungen letztlich zu gleichen und somit validen Ergebnissen führen. Nur wenn sichergestellt werden kann, dass beide Befragungsmethoden zu gleichlautenden Ergebnissen führen, ist eine Substitution von mündlichen Befragungen durch Online-Befragungen im Bereich der empirischen Sozialforschung überhaupt vertretbar. Andernfalls stellt sich die Frage, ob ein Wechsel der Befragungsmethode in Längsschnittunters-

chungen, wie er teilweise durchgeführt wird, überhaupt akzeptabel sein kann, oder ob eine derartige Vorgehensweise stattdessen zu verzerrten und nicht mehr vergleichbaren Daten führen wird.

## 1.2 Zielsetzung

Bisherige Studien, die sich mit einem Vergleich des Antwortverhaltens bei unterschiedlichen Befragungsmethoden beschäftigt haben, untersuchen zumeist einen Methodenvergleich zwischen einer schriftlichen bzw. postalischen Befragung sowie einer Online-Befragung. Zu dieser Vergleichskombination liegen entsprechend bereits Ergebnisse vor, ob bei einem Wechsel von einer schriftlichen Befragung auf eine online durchgeführte Befragung mit Effekten auf das Antwortverhalten und entsprechend mit Verzerrungen durch die Befragungsmethodik zu rechnen ist.

Ein Vergleich von Online-Befragungen und mündlichen Befragungen zur Ermittlung möglicher Unterschiede im Antwortverhalten stellt hingegen in dem Großteil der Studien zur Thematik die Ausnahme dar. Da aufgrund dieses Sachverhaltes wesentliche Erkenntnisse bezüglich möglicher Unterschiede im Antwortverhalten zwischen mündlichen Befragungen und Online-Befragungen fehlen, sollen mit der vorliegenden Arbeit diese Forschungslücke geschlossen und relevante Ergebnisse zur Vergleichbarkeit von mündlichen Befragungen mit online durchgeführten Befragungen geliefert werden. Nur so kann, wie oben bereits dargelegt wurde, zukünftig ausreichend bewertet werden, ob beispielsweise bei Längsschnittuntersuchungen ein Wechsel der Befragungsmethodik in Form einer Substitution von mündlichen Befragungen durch online durchgeführte Befragungen geeignet ist, oder ob diese beiden Befragungsmethoden zu völlig unterschiedlichen Ergebnissen führen, die unter der Begrifflichkeit „Methodeneffekt“ zusammengefasst werden können. In diesem Fall müsste von einem Wechsel der Befragungsmethode in laufenden Untersuchungen trotz verschiedener Vorteile von Online-Befragungen dringend abgeraten werden.

Zielsetzung dieser Arbeit ist es daher, einen Methodenvergleich zwischen einer mündlichen Befragung und einer Online-Befragung durchzuführen, um mögliche Unterschiede im Antwortverhalten zwischen den beiden Befragungsmethoden aufdecken zu können. Hierdurch kann bewertet werden, inwiefern es bei einer Substitution von mündlichen Befragungen durch Online-Befragungen zu Verzerrungen im Antwortverhalten der Befragten kommt, welche auf die Befragungsmethodik zurückgeführt werden können.

### 1.3 Vorgehensweise

In den Kapiteln zwei und drei wird in den thematischen Hintergrund der vorliegenden Arbeit eingeleitet. Da der zum Vergleich der beiden Befragungsmethoden herangezogene Fragebogen sich mit der momentan aktuell diskutierten Thematik „Klimaschutz und Lebensmittel“ beschäftigt, wird in Kapitel zwei der vorliegenden Arbeit einerseits auf den Klimawandel sowie auf mögliche Labelling-Strategien bei Lebensmitteln, welche in verschiedenen Ländern bereits angewendet werden, eingegangen. Hierdurch soll das für die später folgende Auseinandersetzung mit den Fragestellungen notwendige Hintergrundwissen geschaffen und eine kompakte Einleitung in den inhaltlichen Kontext des Fragebogens gegeben werden.

Kapitel drei hingegen geht dann detailliert auf die beiden Befragungsmethoden der mündlichen Befragung, nachfolgend als Face-to-Face-Interview bezeichnet, und der Online-Befragung ein. Hierbei werden sowohl theoretische Aspekte als auch mögliche Vor- und Nachteile der beiden Befragungsmethoden dargestellt. Weiterhin können bei mündlichen sowie bei online durchgeführten Befragungen verschiedene Formen unterschieden werden, die ebenfalls in diesem Kapitel voneinander abgegrenzt werden.

Für einen Vergleich der beiden Befragungsmethoden hinsichtlich möglicher Unterschiede im Antwortverhalten der Befragungspersonen sind verschiedene Fehlerarten relevant, die ebenfalls in diesem dritten Kapitel thematisiert werden. Neben einer grundsätzlichen Abgrenzung der verschiedenen Fehlerarten, die im Rahmen der Befragung als übergeordnete Erhebungsmethode der empirischen Sozialforschung auftreten können, wird hierbei ebenfalls aufgezeigt, welche Fehlerarten potentiell für mögliche Unterschiede im Antwortverhalten in Frage kommen.

Das vierte Kapitel gibt anschließend einen Literaturüberblick über Untersuchungen, die bereits einen Vergleich zwischen zwei Befragungsmethoden durchgeführt haben. Die Ergebnisse der Studien werden im weiteren Verlauf der Arbeit für die Hypothesenbildung herangezogen.

Das fünfte Kapitel umfasst den empirischen Teil der Arbeit. Hier werden in Kapitel 5.1 zuerst die aufgestellten Arbeitshypothesen aufgezeigt. Auf das Untersuchungsdesign wird anschließend in Kapitel 5.2 detailliert eingegangen. Im Anschluss daran werden die Ergebnisse der Untersuchung dargestellt. Begonnen wird dabei mit den Ergebnissen zu den soziodemographischen Variablen (Kapitel 5.3), anschließend werden die Ergebnisse zu den inhaltlichen Fragestellungen des Fragebogens aufgezeigt (Kapitel 5.4 und folgende). Der empirische Teil

der Arbeit wird mit einer ökonometrischen Analyse abgeschlossen, deren Ergebnisse im Kapitel 5.7 besprochen werden.

Eine Diskussion der Ergebnisse findet anschließend im sechsten Kapitel statt. Neben einer Einordnung der Resultate im Bezug auf die Ergebnisse der im vierten Kapitel dargestellten Studien findet weiterhin eine Diskussion und Reflektion der angewendeten Methodik statt.

Die Arbeit schließt mit einer Zusammenfassung der wesentlichen Aussagen der vorliegenden Arbeit ab. In diesem Kontext werden ebenfalls Sachverhalte und Fragestellungen aufgezeigt, die in zukünftigen wissenschaftlichen Untersuchungen aufgegriffen und tiefergehend analysiert werden sollten.

## 2 Klimawandel und Carbon Labelling-Strategien

Dieses Kapitel beschäftigt sich nachfolgend mit verschiedenen relevanten Begrifflichkeiten zur Thematik des Klimawandels sowie damit verbundener Carbon Labelling-Strategien und führt so grundlegend in den thematischen Hintergrund des in der Untersuchung verwendeten Fragebogens ein. Dieser Fragebogen wiederum ist die Grundlage für den Methodenvergleich zwischen einer mündlichen und einer online durchgeführten Befragung. Die Erläuterungen dieses Kapitels sollen im folgenden Verlauf der Arbeit dabei helfen, die verschiedenen Fragestellungen des Fragebogens und die latenten Hintergründe besser verstehen zu können.

### 2.1 Klimawandel

Grundsätzlich kann beim Stichwort des Klimawandels zwischen dem natürlichen Klimawandel sowie dem durch die Menschen verursachten, anthropogenen Klimawandel unterschieden werden. Auf beide wird nachfolgend eingegangen.

#### 2.1.1 Natürlicher Klimawandel

Eine wichtige Rolle im Zusammenhang mit dem Begriff des Klimawandels spielt der sogenannte Treibhauseffekt, auf den nachfolgend zuerst eingegangen wird. Bei dem Treibhauseffekt handelt es sich um einen natürlichen Prozess, der grundsätzlich das Leben auf der Erde überhaupt erst ermöglicht. Exemplarisch kann der Treibhauseffekt am Beispiel eines Glasgewächshauses bzw. Treibhauses erläutert werden. Hierbei werden kurzweilige elektromagnetische Strahlen in Form des Sonnenlichts durch die Glasscheiben des Gewächshauses hindurchgelassen, vom Innenraum absorbiert und in langweilige Strahlen umgewandelt, die zu einer Aufheizung des Raumes führen (PISTORIUS et al. 2006: 3f.). Beim Treibhauseffekt der Erde übernehmen Treibhausgase (THG) wie Kohlendioxid  $\text{CO}_2$ , Methan  $\text{CH}_4$ , Distickstoffoxid<sup>1</sup>  $\text{N}_2\text{O}$ , Ozon  $\text{O}_3$ , aber auch Wasserdampf  $\text{H}_2\text{O}$  die Funktion der Fensterscheibe. Den Hauptanteil am Treibhauseffekt hat mit etwa 66 % der Wasserdampf, gefolgt vom  $\text{CO}_2$  mit 29 % (SCHÖNWIESE 1996: 15). Eine schematische Darstellung für den Treibhauseffekt der Erde liefert Abbildung 1.

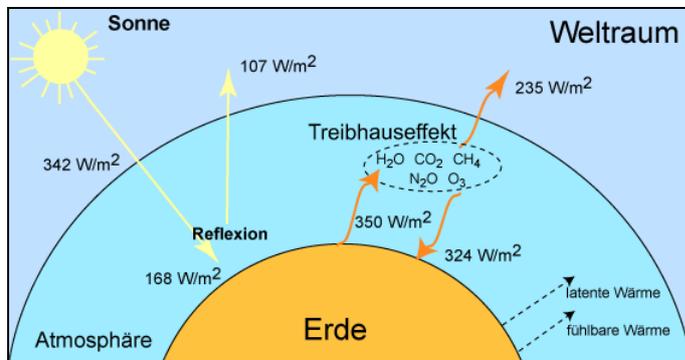
Die THG lassen die kurzweiligen Strahlen der Sonne passieren und absorbieren gleichzeitig die von der Erdoberfläche reflektierten langweiligen Strahlen. Etwa 70 % der Sonnenstrahlen werden dabei von der Erdoberfläche absorbiert, während die restlichen 30 % wieder in den Weltraum reflektiert werden. Die in der Atmosphäre verbleibende Wärme sorgt dabei für eine konstante Temperatur in Erdnähe (PISTORIUS et al. 2006: 5). Dieser natürliche Treibhauseff-

---

<sup>1</sup> Distickstoffoxid ist auch unter dem Trivialnamen Lachgas bekannt.

fekt ermöglicht es erst, dass die Durchschnittstemperatur auf der Erde um etwa 30°C höher liegt, als es ohne den Treibhauseffekt der Fall wäre. Daher kann ausgesagt werden, dass Leben auf der Erde nur aufgrund des natürlichen Treibhauseffektes möglich ist (STERN 2008: 9).

**Abbildung 1: Schematische Darstellung des Treibhauseffektes**



Quelle: HAMBURGER BILDUNGSSERVER 2007.

Das heutige Wissen über den Treibhauseffekt wurde im 19. Jahrhundert von den Wissenschaftlern Fourier, Tyndall und Arrhenius begründet. Fourier konnte in den 1820er Jahren nachweisen, dass die Atmosphäre für Strahlen, die von der Sonne kommen, permeabler ist als für von der Erdoberfläche reflektierte Infrarotstrahlen, wodurch in der Erdatmosphäre Wärme akkumuliert wird. Tyndall konnte in den 1850ern in einem nächsten Schritt die Molekültypen identifizieren, die heute als THG bekannt sind und eine Akkumulierung der Wärme und somit den Treibhauseffekt bewirken. Arrhenius konnte letztlich in einem weiteren Schritt zeigen, dass eine Verdopplung der CO<sub>2</sub>-Konzentration in der Erdatmosphäre zu einem signifikanten Anstieg der Durchschnittstemperatur auf der Erde führen würde. Heute ist bekannt, dass der Effekt des CO<sub>2</sub> auf das Weltklima logarithmisch mit seiner Konzentration in der Atmosphäre steigt (STERN 2008: 9).

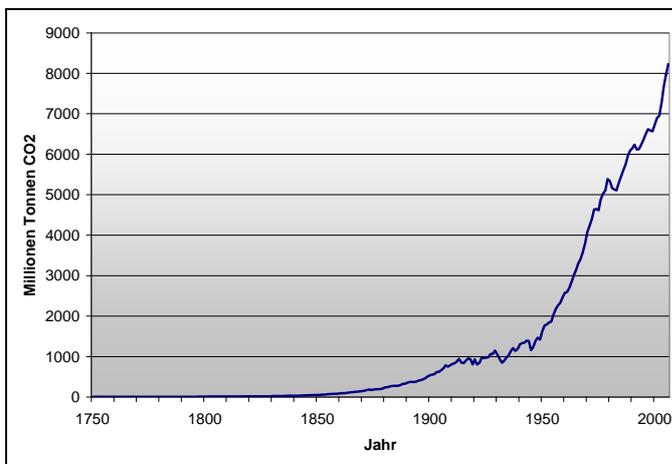
### 2.1.2 Anthropogener Klimawandel

Neben dem natürlichen Klimawandel zeigen sich seit mehreren Jahren immer mehr Belege für einen anthropogenen, das heißt durch Menschen verursachten Klimawandel und einer damit verbundenen Erwärmung der Erde. So hat sich seit der Industrialisierung der Gesellschaft, welche ab etwa 1750 begann, und einem damit verbundenen, veränderten menschlichen Handeln, die natürliche Zusammensetzung der Atmosphäre stark verändert (STERN 2008: 4). So stieg beispielsweise die Konzentration von CO<sub>2</sub> von einem vorindustriellen Wert von etwa 280 ppm<sup>2</sup> auf einen Wert von 379 ppm im Jahre 2005 an (IPCC 2007a: 2). Entsprechend kann

<sup>2</sup> ppm steht für „parts per million“ und kennzeichnet das Verhältnis von Treibhausgasmolekülen zu der Anzahl an Molekülen trockener Luft. So kennzeichnet der Wert von 280 ppm ein Verhältnis von 280 Treibhausgasmolekülen zu einer Million Molekülen trockener Luft (IPCC 2007a: 2).

ausgesagt werden, dass in den letzten 250 Jahren die atmosphärische Konzentration an CO<sub>2</sub> weltweit um etwa 36 % angestiegen ist. Der größte Anstieg der CO<sub>2</sub>-Konzentration in der Atmosphäre ist dabei in den letzten 50 Jahren zu verzeichnen (WTO und UNEP 2009: 3). Hauptursachen für diesen Konzentrationsanstieg sind neben der Verbrennung fossiler Brennstoffe die weltweite Abholzung der Wälder sowie Änderungen in der Landnutzung (STERN 2008: 4). Vor allem bei den CO<sub>2</sub>-Emissionen, die aus der Verbrennung fossiler Brennstoffe resultieren, lassen sich seit den 1950er Jahren deutliche Steigerungen verzeichnen, wie Abbildung 2 illustriert.

**Abbildung 2: Durch Verbrennung fossiler Brennstoffe verursachte CO<sub>2</sub>-Emissionen, Zeitraum 1751-2006**



Quelle: Eigene Darstellung basierend auf Daten von BODEN et al. 2009.

Neben CO<sub>2</sub> sind für den anthropogenen Klimawandel und dem damit verbundenen Temperaturanstieg bzw. der globalen Erwärmung hauptsächlich Treibhausgase wie Ozon, Methan, Distickstoffoxid, Halogenkohlenwasserstoffe sowie weitere Industriegase verantwortlich (IPCC 2007b: 135). Zwar kommen diese THG mit Ausnahme von Halogenkohlenwasserstoffen und anderen Industriegasen, wie oben bereits beschrieben wurde, grundsätzlich ebenfalls in der Erdatmosphäre vor, allerdings hat auch ihre Konzentration stark zugenommen. Dabei hat CO<sub>2</sub>, welches neben der Verbrennung von fossilen Brennstoffen durch die Abholzung der Wälder freigesetzt wird, einen Anteil von 77 % am anthropogenen Klimawandel. Veränderungen in der Landwirtschaft und der Landnutzung im Allgemeinen hingegen führen zu einer vermehrten Freisetzung von Methan und Distickstoffoxid, wobei die Methanemissionen zu 14 % für den anthropogenen Klimawandel verantwortlich gemacht werden. Einen deutlich geringeren Anteil am anthropogenen Klimawandel haben mit 9 % Stickstoffoxidemissionen, Ozonemissionen durch Autoabgase und andere Quellen sowie industrielle Emissionen von Halogenkohlenwasserstoffen sowie anderen Industriegasen (WTO und UNEP 2009: 3). CO<sub>2</sub> stellt hinsichtlich des anthropogenen Klimawandels somit das bedeutendste THG dar (IPCC

2007a: 2). Die insgesamt größten Emittenten von THG stellen die Industriestaaten dar, ihre absoluten Emissionen von CO<sub>2</sub> liegen etwa dreimal höher als die der Entwicklungsländer. Ausnahme hierbei stellen die THG dar, die durch Aktivitäten in den Bereichen Land- und Forstwirtschaft freigesetzt werden. In diesen Bereichen zeichnen sich die Entwicklungsländer aufgrund ihrer Wirtschaftsstruktur durch doppelt so hohe THG aus (WTO und UNEP 2009: 4). Werden die Emissionen an THG allerdings auf das Bruttoinlandsprodukt (BIP) bezogen, zeigt sich, dass mit zunehmendem BIP die relativen Emissionen an Treibhausgasen allgemein sowie CO<sub>2</sub> im Speziellen sinken (WORLD BANK 2008: 68).

Auch scheint insgesamt das Maximum der CO<sub>2</sub>-Emissionen noch nicht erreicht zu sein, wie die U.S. ENERGY INFORMATION ADMINISTRATION (EIA) beschreibt. Sie kommt bei Simulationen, die auf den momentanen, weltweiten Klimapolitiken beruhen, zu dem Ergebnis, dass bis zum Jahre 2030 die weltweiten CO<sub>2</sub>-Emissionen um etwa 40 % ansteigen werden (EIA 2009: 109). Bei einer Simulation der weltweiten Treibhausgasemissionen, ebenfalls auf Basis der momentanen Klimapolitiken bis zum Jahre 2050 wird von der WTO und dem UNEP davon ausgegangen, dass die Treibhausgasemissionen der OECD<sup>3</sup>-Staaten tendenziell konstant bleiben, während mit starken Steigerungen bei den Schwellenländern Brasilien, Russland, Indien und China, sowie beim Rest der Welt gerechnet wird (WTO und UNEP 2009: 5).

Nicht nur dieser prognostizierte Anstieg der THG-Emissionen ist dabei entscheidend, sondern ebenfalls der Sachverhalt, dass ein Großteil der Treibhausgase für einen langen Zeitraum in der Erdatmosphäre verbleibt, bevor sie abgebaut werden. Somit liegen zwischen der Emission der THG und ihrer letztlichen, finalen Wirkung auf den Treibhauseffekt teilweise mehrere hundert Jahre. Dies bedeutet entsprechend auch, dass selbst bei einer Reduzierung oder gar einer Eliminierung der THG-Emissionen in der nächsten Zeit mit einer weiteren Steigerung des anthropogenen Treibhauseffektes und somit des Klimawandels zu rechnen ist. Somit handelt es sich nach der World Trade Organization (WTO) und dem Umweltprogramm der Vereinten Nationen, dem United Nations Environment Programme (UNEP), bei der globalen Erwärmung neben einem Emissions-Problem ebenfalls um ein Problem der THG-Konzentration in der Erdatmosphäre (WTO und UNEP 2009: 4).

### **2.1.3 Folgen des Klimawandels**

Es ist heute allgemein anerkannt, dass die gestiegenen Konzentrationen an THG in der Erdatmosphäre im Zusammenhang mit dem anthropogenen Treibhauseffekt und einem Anstieg der

---

<sup>3</sup> OECD steht für „Organisation for Economic Co-Operation and Development“ (OECD 2010).

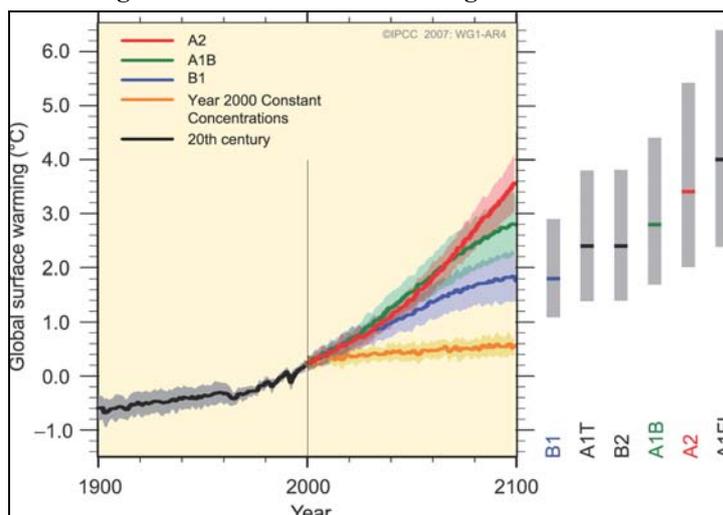
Durchschnittstemperatur auf der Erde stehen (IPCC 2007a: 2ff.; IPCC 2007c: 2ff.; SINN 2008: 1f.; STERN 2008: 4; GRABL 2009: 9). Die globale Durchschnittstemperatur kann dabei als ein grundsätzlicher Indikator für das Ausmaß des Klimawandels angesehen werden. So sind in den vergangenen 30 Jahren die globalen Temperaturen kontinuierlich um etwa  $0,2^{\circ}\text{C}$  per Dekade angestiegen (STERN 2008: 6f.). Seit der Industrialisierung hat sich die durchschnittliche Temperatur bereits um  $0,8^{\circ}\text{C}$  erhöht (SINN 2008: 1) und bringt das durchschnittliche Temperaturniveau an einen Wert, der nahe oder sogar über dem Höchstwert der letzten 12.000 Jahre liegt (STERN 2008: 7). Der lineare Erwärmungstrend der letzten 50 Jahre liegt mit einem Bereich von  $0,10^{\circ}\text{C}$  bis  $0,16^{\circ}\text{C}$  doppelt so hoch wie der Erwärmungstrend der letzten 100 Jahre. Die Folgen des anthropogenen Klimawandels sind dabei für uns bereits direkt spürbar: Elf von zwölf Jahren im Zeitraum 1995 bis 2006 waren die heißesten Jahre seit Beginn der Wetteraufzeichnungen im Jahre 1850.

Observationen seit 1961 zeigen ebenfalls, dass die Durchschnittstemperatur der Ozeane bis zu einer Tiefe von 3000 m angestiegen ist. Bisher wurden 80 % der zusätzlichen Wärme durch den anthropogenen Klimawandel von den Ozeanen aufgenommen, was zu einer Ausdehnung des Wassers und einem Anstieg der Pegel führt. Verstärkt wird der Anstieg der Pegel auch durch ein Abschmelzen der Gletscher in beiden Hemisphären. Diese Faktoren zeigen sich auch am jährlichen Anstieg der Meeresspiegel um circa 1,8 mm in den Jahren 1961 bis 2003. Deutlich höher ist der Anstieg des Meeresspiegels, wenn nur die Jahre 1993 bis 2003 betrachtet werden: Hier waren es im jährlichen Durchschnitt 3,1 mm. Allerdings ist momentan noch unklar, ob dies eine zufällige Abweichung oder eine Zunahme im Langzeittrend darstellt. Hinzu kommt, dass durch das Abschmelzen der Gletscher und Eisflächen eine Art Teufelskreis geöffnet wird. So sorgt das Abschmelzen beispielsweise dafür, dass natürliche Methanquellen geöffnet werden, die vorher durch das Eis verschlossen waren. Diese können nun ihre vorher eingeschlossenen, natürlichen Methanvorräte an die Atmosphäre abgeben, wodurch es letztlich ebenfalls zu einer Verstärkung des Treibhauseffektes kommt. Die damit verbundenen Temperatursteigerungen können dann zu einer weiteren Eisschmelze beitragen.

Auch hinsichtlich des Wetters zeigen sich zunehmend extremere Wetterbedingungen. So lassen sich in Nordeuropa, Nord- und Zentralasien sowie den östlichen Gebieten Nord- und Südamerikas signifikante Niederschlagssteigerungen feststellen. In anderen Gebieten der Erde wie der Sahelzone, dem Mittelmeerraum, Südafrika oder auch Südasien hingegen nehmen die Dürreperioden zu. Auch lassen sich zunehmend Phänomene wie Hitzewellen, Tsunamis und Wirbelstürme verzeichnen (IPCC 2007a: 5ff.; WTO und UNEP 2009: 9ff.; HEIMANN 2010: 1211).

Wie oben bereits dargelegt wurde, verbleiben die Treibhausgase für einen langen Zeitraum in der Erdatmosphäre. Daher ist aufgrund der für die Zukunft erwarteten Zunahme der THG-Emissionen ebenfalls mit einer weiteren Zunahme bzw. Verstärkung des Klimawandels bzw. der globalen Erwärmung auszugehen. Umfangreiche Prognosen, wie sich die Durchschnittstemperaturen bis zum Jahre 2100 entwickeln könnten, wurden vom INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE<sup>4</sup> (IPCC) vorgenommen. Das IPCC unterscheidet bei seinen Projektionen vier mögliche Szenarien. Das Szenario A1 geht dabei von einem starken ökonomischen Wachstum, einer Populationszunahme bis zum Jahre 2050 und einer anschließenden Verringerung der Weltbevölkerung sowie einer intensiven Nutzung fossiler Brennstoffe (Szenario A1FI) bzw. einem vollständigen Wechsel auf nicht-fossile Energieträger (A1T) aus. Das Szenario B1 übernimmt die Annahmen von A1 hinsichtlich der Bevölkerungsentwicklung, geht aber von einer Transformation der Gesellschaft zu einer sauberen und weniger CO<sub>2</sub>-intensiven Dienstleistungsgesellschaft mit einem geringeren Wirtschaftswachstum aus. Das Szenario A2 geht von einer Welt mit einem regional stark unterschiedlichen Wirtschaftswachstum bei einem insgesamt geringen Pro-Kopf-Wirtschaftswachstum aus, während das vierte Szenario B2 von regionalen und lokalen Lösungen ausgeht, die dem Gedanken der Nachhaltigkeit genügen. In diesem Szenario wird weiterhin von einem geringen, aber anhaltenden Wirtschaftswachstum ausgegangen (IPCC 2007c: 22; IPCC 2007d: 32; WTO und UNEP 2009: 7). Die für die verschiedenen Szenarien erwarteten Veränderungen der Durchschnittstemperatur in Abhängigkeit von der Zeit werden in Abbildung 3 dargestellt.

**Abbildung 3: Erwartete Veränderungen der Durchschnittstemperatur im Zeitverlauf nach Szenarien**



Quelle: IPCC 2010b.

<sup>4</sup> Beim IPCC handelt es sich um eine Institution, die vom UNEP und der WORLD METEOROLOGICAL ORGANIZATION mit dem Ziel gegründet wurde, wissenschaftlich fundierte Aussagen zum momentanen Stand des Klimawandels sowie daraus resultierende, potentielle ökologische und sozio-kulturelle Konsequenzen geben zu können (IPCC 2010a).

Je nach gewähltem Szenario wird dabei von einem Anstieg der Durchschnittstemperatur bis zum Jahre 2100 von circa 2°C (im Szenario B1) bis zu 4°C im Szenario A1FI ausgegangen, wie einmal anhand der Funktionen im linken Teil des Schaubildes sowie anhand der je nach Szenario erwarteten Durchschnittstemperaturen in 2100 samt Unsicherheitsbereich (graue Bereiche) im rechten Teil des Schaubildes dargestellt wird. Aus der Abbildung geht dabei ebenfalls hervor, dass selbst bei einer Konstanthaltung der THG-Konzentrationen in der Erdatmosphäre auf dem Stand des Jahres 2000, nach den Berechnungen des IPCC noch mit einem Anstieg der Durchschnittstemperatur von etwa 0,6°C gerechnet werden müsste (orange Linie).

Neben den oben dargestellten, bereits heute wahrnehmbaren Auswirkungen des Klimawandels werden vom IPCC mit steigenden Durchschnittstemperaturen für die Zukunft weitere, für das Leben auf der Erde bedeutsame Auswirkungen auf verschiedenste Bereiche und Ökosysteme erwartet.

So wird hinsichtlich der **Trinkwasserreserven** eine starke Divergenz erwartet. Während in den feucht-tropischen Gebieten und gemäßigten bzw. höheren Breitengraden der Erde mit einem Anstieg der Trinkwasserverfügbarkeit um etwa 10 bis 40 % gerechnet wird, geht das IPCC auf der Gegenseite davon aus, dass in den bereits heute trockenen Regionen der Erde nahe des Äquators die Trinkwasserreserven um möglicherweise 10 bis 30 % zurückgehen werden.

Auch bei den **Ökosystemen** werden deutliche Auswirkungen des Klimawandels durch das IPCC erwartet. Einflussfaktoren sind hier, neben den durch den anthropogenen Klimawandel verursachten Temperatursteigerungen, auch damit verbundene Aspekte wie Wetterextreme und -anomalien, Insektenplagen sowie weitere Determinanten wie Landnutzungsänderungen oder Umweltverschmutzungen. Insgesamt wird hierdurch erwartet, dass bereits bei einem Anstieg der Durchschnittstemperatur um 1,5-2,5°C etwa 20 bis 30 % der Tier- und Pflanzenarten der Welt ernsthaft gefährdet seien und möglicherweise unwiederbringlich verschwinden.

Durch den Klimawandel wird ebenfalls mit Auswirkungen auf die **Produktion der Land- und Forstwirtschaft** durch das IPCC gerechnet. So gehen die Experten davon aus, dass in den mittleren und hohen Breitengraden beider Hemisphären ein Temperaturanstieg von 1-3°C zu einer gesteigerten Getreideproduktion, abhängig von der Getreideart, führen kann. Bei einem stärkeren Temperaturanstieg wird allerdings von einer Abnahme des Produktionspotentials in diesen Gebieten ausgegangen. Weltweit betrachtet wird bei einem Temperaturanstieg von 1-3°C ebenfalls von einem höheren Produktionspotential bei Nahrungsmitteln im Allgemeinen ausgegangen. Allerdings sehen auch hier die Forscher des IPCC bei einem noch stärkeren Anstieg der Durchschnittstemperatur die Gefahr einer Abnahme des Produktionspoten-

tials. In den niedrigen Breitengraden nahe dem Äquator wird hingegen bereits bei einem Anstieg der Durchschnittstemperatur um 1-2°C von einer Abnahme des Produktionspotentials und einer damit verbundenen, möglichen Zunahme des Hungers ausgegangen.

Weiterhin wird mit Auswirkungen auf **Küstensysteme** und **Gebiete mit einem niedrigen Meeresspiegel** gerechnet. So rechnet das IPCC mit einer zunehmenden Gefahr für die Küsten durch Erosion sowie einen Anstieg des Meeresspiegels. Auch sehen die Experten eine steigende Gefahr für Millionen von Menschen durch zunehmende Flutkatastrophen. Gefährdet sind hierbei vor allem die Menschen in den großen Flussdeltas in Asien und Afrika sowie Bewohner kleiner Inseln.

Bezüglich der Folgen für **Industrien** und **Gesellschaften** ergibt sich nach dem IPCC ein sehr differenziertes Bild, das sehr von deren Lage und Größe determiniert wird. In der Summe betrachtet, gehen die Experten des IPCC allerdings von tendenziell negativen Effekten durch den Klimawandel und seinen Folgen aus.

Starke Auswirkungen des Klimawandels werden auch hinsichtlich des **Gesundheitsstatus** von Millionen Menschen erwartet. Hiervon sind nach dem IPCC vor allem diejenigen betroffen, die sich nur durch eine geringe Anpassungskapazität an die sich durch den Klimawandel verändernden Rahmenbedingungen auszeichnen. So wird erwartet, dass die Zahl der Menschen mit Mangelernährung und daraus resultierende Folgen stark zunehmen werden. Auch wird vom IPCC mit einer steigenden Zahl an Toten, Verletzten und Kranken durch Hitzewellen, Flutkatastrophen, Stürme, Feuer und Dürren gerechnet. Zunehmende Zahlen erwarten die Wissenschaftler auch bei Herz-Kreislauferkrankungen, verursacht durch hohe bodennahe Ozonkonzentrationen in Folge des Klimawandels. Mögliche positive Effekte des Klimawandels, wie einer reduzierten Zahl an Erfrierungen und Erfrierungstoten, können nach dem IPCC dadurch allerdings nicht ausgeglichen werden (IPCC 2007c: 11ff.; WTO und UNEP 2009: 17).

## **2.2 Carbon-Labeling-Strategien**

### **2.2.1 Carbon Labelling aus Konsumenten- und Unternehmenssicht**

Das zunehmende Wissen der Konsumenten über den anthropogenen Klimawandel und den zu erwartenden Folgen führt bei einem immer größer werdenden Teil der Konsumenten zu einem Interesse dahingehend, welche Treibhausgasemissionen von verschiedenen Produkten des alltäglichen Konsums während des Produktlebenszyklusses, also bei Herstellung, Verarbeitung, Transport, Handel, Benutzung und Entsorgung, freigesetzt werden.

Derartige Informationen werden zunehmend in Form sogenannter „Product Carbon Footprints“ (PCF) dem Konsumenten bereitgestellt. Nach BOLWIG und GIBBON meint „Carbon Footprint“ dabei die totale Summe an THG, die für eine gegebene Aktivität ausgestoßen werden, die Begrifflichkeit „Product“ kennzeichnet darüber hinaus den Bezug zu dem entsprechenden Produkt oder der Dienstleistung, für die die THG im Produktlebenszyklus anfallen (BOLWIG und GIBBON 2009a: 1).

Im April 2009 führte das Meinungsforschungsunternehmen Gallup im Auftrag der Europäischen Kommission eine Befragung zum Thema „Europäische Einstellungen zu den Themen nachhaltiger Konsum und Produktion“ durch. Im Rahmen der Befragung wurden insgesamt 26.500 Personen über 15 Jahren aus allen 27 EU-Staaten sowie Kroatien mündlich und per Telefon befragt. Als ein wesentliches Ergebnis kann festgehalten werden, dass 83 % der befragten Personen die Auswirkungen des Produktes auf die Umwelt bei der Kaufentscheidung als „sehr wichtig“ oder „ziemlich wichtig“ erachten. Damit sind die Umweltauswirkungen eines Produktes der wichtigste Kaufaspekt nach der wahrgenommenen Qualität (für 97 % der Befragten entscheidend) sowie dem Preis (für 89 % entscheidend) und liegen mit Abstand vor dem Markennamen als entscheidendes Kaufkriterium (dieser ist nur für 39 % der Befragten relevant). Bezüglich Produkten, die Elektroenergie oder fossile Brennstoffe benötigen, gaben weiterhin 77 % der Befragten an, beim Kauf derartiger Produkte immer oder zumindest häufig die Energieeffizienz dieser Produkte in die Kaufentscheidung einzubeziehen. 47 % der Befragten achten dabei auf Öko-Labels wie dem EU-Energieeffizienzlabel, welches beispielsweise bei Kühlschränken oder Leuchtmitteln zu finden ist. Etwa ein Viertel der Befragten beachtet beim Kauf derartige Labels hingegen überhaupt nicht.

Hinsichtlich der bereitgestellten Informationen eines Öko-Labels scheint es für den überwiegenden Teil der Befragten (38 %) auszureichen, dass ein derartiges Zeichen aussagt, dass ein Recycling des Produktes prinzipiell möglich ist. 32 % wünschen sich, dass ihnen ein solches Zeichen zeigt, dass die Rohstoffe eines entsprechend gekennzeichneten Produktes umweltfreundlich erzeugt wurden. Weitere 16 % wünschen Informationen darüber, ob die Umverpackung umweltfreundlich ist. Nur 10 % der Befragten gaben hingegen an, dass ein Öko-Label konkrete Informationen über den PCF eines Produktes enthalten sollte.

Dennoch wird von den Befragten die Einführung eines PCF als sehr wichtig erachtet. So gaben 72 % der Befragten an, dass ein PCF zukünftig rechtsverbindlich sein sollte. Besondere Zustimmung erhält diese Fragestellung dabei in Griechenland und Ungarn mit jeweils 90 % Zustimmung, in Deutschland liegt die Zustimmung mit einem Wert von 61 % unter dem EU-Schnitt. Die geringste Zustimmung erhält die Frage in Tschechien (47 %). Insgesamt zeichnet

sich über die EU dabei das Bild ab, dass von Befragten, die Öko-Labels aktiv in ihre Kaufentscheidung einbeziehen, eine obligatorische Einführung eines PCF präferiert wird. Befragten, die hingegen derartige Öko-Label nicht bei ihrer Kaufentscheidung berücksichtigen, reicht eine freiwillige Angabe eines PCFs auf der Produktverpackung. Eine Rolle spielt bei diesen Antworten sicherlich auch, dass 48 % der befragten Personen Angaben der Produzenten hinsichtlich der Umweltfreundlichkeit ihrer Produkte nicht vertrauen (EUROPEAN COMMISSION 2009: 10ff.).

Weiterhin ist nicht nur ein Interesse an PCFs auf der Konsumentenseite, sondern auch auf der Anbieterseite feststellbar. So sehen sich die Unternehmen zunehmend mit einer nachhaltigkeitsorientierten Konsumentenschicht unter dem Stichwort „LOHAS“<sup>5</sup> konfrontiert, für die Mittel und Wege gefunden werden müssen, die gesellschaftliche Verantwortung auf der Unternehmensseite verdeutlichen zu können (FREY und MÜHLBACH: 2009: 184f.). Einerseits entdecken diese Konsumenten nachhaltig produzierte Güter zunehmend für sich, andererseits geht von diesen Konsumenten auch ein starker Nachfrage-Pull aus. So werden von diesen zunehmend Produkte gefordert, die ihren sozio-ökologischen Anforderungen genügen (BILHARZ und BELZ 2008: 7). Hier geht es nach FREY und MÜHLBACH vor allem darum, mit möglichst geringen THG-Emissionen zu produzieren, regionale Produkte anzubieten sowie die Logistikkette zu optimieren. Hierbei kann aus Sicht der Autoren der PCF helfen, entsprechende Maßnahmen zu entwickeln, zu bewerten und abschließend die Ergebnisse präsentieren zu können. Zielsetzung ist dabei, die unternehmensbedingten THG-Emissionen zu erfassen und darauf aufbauend, reduzieren zu können (FREY und MÜHLBACH 2009: 184f.). Ähnlich sehen dies auch BOLWIG und GIBBON. Für sie bieten PCFs für Unternehmen die Chance, CO<sub>2</sub>- bzw. THG-Hotspots während der Produktion identifizieren zu können, um darauf aufbauend, kostengünstig Reduzierungsstrategien für CO<sub>2</sub> bzw. THG entwickeln zu können. Hieraus ergibt sich aus Autorensicht zusätzlich ein potentieller Nutzen in der Form, dass daraus signifikante Reduzierungen bei den Energiekosten möglich werden sowie auf mögliche, zukünftige Klimaschutzregelungen, durch geringere Anpassungsnotwendigkeiten schneller und kostengünstiger reagiert werden kann (BOLWIG und GIBBON 2009a: 2; BOLWIG und GIBBON 2009b: 8).

### **2.2.2 Definition des Product Carbon Footprints**

Eine einheitliche Definition des Product Carbon Footprints lässt sich in der Literatur nicht finden. Ursächlich hierfür ist auch, dass sich hinsichtlich einer möglichen Definition eines PCF verschiedene Fragen stellen, wie auch WIEDMANN und MINX aufzeigen. So muss einmal ent-

---

<sup>5</sup> LOHA steht für „Lifestyle of Health und Sustainability“ (FREY und MÜHLBACH: 184f.).

schieden werden, ob ein PCF ausschließlich CO<sub>2</sub>-Emissionen oder auch andere THG-Emissionen, wie beispielsweise Methan, umfassen sollte. Diesbezüglich stellt sich auch einmal die Frage, ob ein PCF ausschließlich Gase auf Kohlenstoffbasis berücksichtigen sollte, wie die Bezeichnung „Carbon“ vermuten lässt, oder stattdessen auch Gase umfassen sollte, die keinen Kohlenstoff enthalten, wie etwa Distickstoffoxid oder auch weitere Treibhausgase. Ebenfalls Klärungsbedarf besteht dahingehend, ob nur solche Gase berücksichtigt werden sollten, die überhaupt über ein Erderwärmungspotential verfügen. Eine weitere Frage, die bei einer Definition nach WIEDMANN und MINX unbedingt berücksichtigt werden sollte, ist die Frage nach den Systemgrenzen: Hier muss im Rahmen einer Definition des PCF festgelegt werden, ob der PCF nur direkt mit dem Produkt oder der Dienstleistung verbundene Emissionen berücksichtigen oder darüber hinaus zusätzlich auch indirekte Emissionen, die beispielsweise bei der Produktion von Hilfsstoffen anfallen, umfassen sollte. Die letzte Frage umfasst den Sachverhalt, anhand welcher Einheiten ein PCF definiert werden sollte. Nach den Autoren legt der Begriff „Footprint“ ein Flächenmaß nahe, so wie es bei dem sogenannten „Ecological Footprint“<sup>6</sup> getan wird. Dieser Frage wird von den Autoren dabei mehr Bedeutung beigemessen als der Entscheidung, ob nur die Emissionen quantifiziert werden sollen (beispielsweise in t CO<sub>2</sub>) oder ob auch die Auswirkungen dargestellt werden sollen (beispielsweise durch die Angabe von CO<sub>2</sub>-Äquivalenten<sup>7</sup> t CO<sub>2-E</sub>). Auf Basis dieser Fragestellungen und verschiedener Definitionsversuche in der Literatur geben sie abschließend sinngemäß folgende Definition eines PCF: *„Beim Product Carbon Footprint handelt es sich um ein Messinstrument zur Erfassung der Gesamtheit an CO<sub>2</sub>-Emissionen, die direkt oder indirekt durch eine Aktivität verursacht werden oder während des Lebenszyklusses eines Produktes anfallen“* (WIEDMANN und MINX 2007: 4). Aus Verfassersicht bietet es sich hierbei jedoch zusätzlich an, den Product Carbon Footprint aufgrund einer besseren Handhabbarkeit für den Verbraucher sowie eines einfacheren Verständnisses auf eine bestimmte Produktmenge, vor allem bei einem Bezug auf Lebensmittel, zu beziehen. Während bei langlebigen Konsumgütern wie Haushaltsgeräten oder Autos die betrachtete Produktmenge einfach nachvollziehbar ist, ist dies bei Lebensmitteln nicht unbedingt der Fall. So ist beispielsweise bei Angabe eines PCFs für Orangensaft nicht sofort für den Konsumenten ersichtlich, ob sich der PCF auf den gesamten Packungsin-

---

<sup>6</sup> Der „Ecological Footprint“ (EF) wird herangezogen, um den anthropogenen Naturverbrauch darzustellen. Die bei der Herstellung von Produkten oder Dienstleistungen entstehenden Energie- und Stoffströme werden dabei in Wasser- und Landflächen (in ha) umgerechnet. Diese sind nötig, um diese Ströme aufrecht erhalten zu können. Diese Wasser- und Landflächen ergeben in der Summe den EF von Produkten oder Dienstleistungen (WACKERNAGEL und REES 1997: 22f.).

<sup>7</sup> Mit Hilfe des sogenannten CO<sub>2</sub>-Äquivalents kann die Wirkung von THG zum anthropogenen Treibhauseffekt angegeben werden. Hierbei wird die Wirkung dieser THG in Relation zur Wirkung des CO<sub>2</sub> gesetzt und durch das entsprechende CO<sub>2</sub>-Äquivalent ausgedrückt (DIENER 1999: 3).

halt, eine Basiseinheit wie etwa einen Liter oder stattdessen nur auf ein Glas Orangensaft bezieht. Bei einer Kennzeichnung der Produkte sollte der PCF für eine optimale Nachvollziehbarkeit entsprechend in Relation zur betroffenen Produktmenge gesetzt werden.

### **2.2.3 Methodischer Hintergrund**

Wichtig sind bei der Ermittlung des PCFs einer Aktivität auch die Verfahren und Methoden im Hintergrund, die angewendet wurden, um den PCF bestimmen zu können. Hierbei spielen nach BOLWIG und GIBBON vor allem die sogenannte Life Cycle Analysis sowie die ISO-Normen 14040, 14044 und 14064 eine wichtige Rolle.

Bei der Life Cycle Analysis (LCA) handelt es sich um die Basismethode des Product Carbon Footprints, die dazu dient, die indirekten und direkten Treibhausgasemissionen einer Aktivität erfassen zu können. Da in der Summe mit einer LCA ebenfalls eine Ökobilanz eines Produktes oder einer Dienstleistung erstellt wird, werden durch die LCA zusätzlich sämtliche Stoff- und Energieströme erfasst, die durch die Aktivität verursacht werden. Bei der LCA handelt es sich dabei allerdings nicht um ein einheitliches Verfahren, sondern dient als Sammelbegriff für eine Reihe von Methoden zur Erstellung einer Ökobilanz einer Aktivität, um daraus Informationen über die emittierten Treibhausgase gewinnen zu können. Eine Übereinkunft, welche LCA-Methode angewendet werden sollte, existiert nicht. Daher gibt es auch keine allgemeingültige Vereinbarung, welche Methoden zur Kalkulation eines PCFs herangezogen werden sollten. Hieraus ergibt sich entsprechend, dass je nach verwendeter LCA der PCF eines Produktes oder einer Dienstleistung starke Varianzen aufweisen kann (BOLWIG und GIBBON 2009b: 9ff.).

Hinsichtlich der Durchführung von Life Cycle Analysen existieren zwei Standards der International Organization for Standardization: Die oben bereits benannten Normen ISO 14040 sowie ISO 14044 zum LCA bzw. zur Ökobilanzierung. Während die ISO 14040 sich mit Prinzipien und den Rahmenbedingungen für eine LCA-Implementierung in einem Unternehmen beschäftigt, regelt die ISO 14044 sowohl Anforderungen für eine LCA-Implementierung sowie gibt verschiedene Leitsätze zur Implementierung vor (BOLWIG und GIBBON 2009b: 9ff.; FREY und MÜHLBACH 2009: 191ff.). Ein Product Carbon Labelling-Ansatz, bei dem die beiden ISO-Normen angewendet wurden, ist das deutsche PCF Pilot-Projekt, an dem große Unternehmen wie etwa die Deutsche Telekom, BASF, Tchibo oder auch Tengelmann beteiligt waren (PCF PILOTPROJEKT DEUTSCHLAND 2009: 10ff.).

Die ISO-Normen sind dabei auch Grundlage für den britischen PAS<sup>8</sup> 2050-Standard, der im dritten Quartal 2008 vom British Standards Institute, den britischen Pendant des Deutschen Instituts für Normung, herausgegeben wurde (BOLWIG und GIBBON 2009b: 13) und beispielsweise von britischen Handelskonzern Tesco angewendet wird (CARBON TRUST 2008: 2f.). Etwa 100 Eigenmarken-Produkte aus den Bereichen Waschmittel, Orangensaft, Kartoffeln und Leuchtmittel sind dabei mit dem Label gekennzeichnet (TESCO 2009).

Die Norm ISO 14064 als dritte relevante ISO-Norm hat hingegen einen anderen Fokus. Ihr Inhalt bezieht sich nicht auf die Messung der gesamten Umweltwirkungen der Produktion, des Konsums und der Entsorgung über einen unspezifizierten Zeitraum, sondern fokussiert inhaltlich die gesamten THG-Emissionen der Unternehmung für eine bestimmte Aktivität innerhalb eines Jahres (BOLWIG und GIBBON 2009b: 11).

Für die Zukunft werden weitere PCF-Standards von Regierungen und internationalen Organisationen erwartet. So wird Ende 2011 ein PCF-Standard der ISO in Form der Norm ISO 14067 erwartet. Diese Norm soll Unternehmen als allgemein anerkannter Standard bei der Messung, Berichterstattung und Verifikation ihrer THG-Emissionen behilflich sein. Die Grundlagen für diese neue Norm werden ebenfalls durch die oben genannten Normen ISO 14040 und ISO 14044 bereitgestellt (BOLWIG und GIBBON 2009b: 14; ISO 2010).

Starke Ambitionen hinsichtlich der Implementierung eines PCF-Standards zeigt Frankreich. So wird geplant, bereits zu Beginn des Jahres 2011 ein verbindliches Zeichen für in Frankreich verkaufte Produkte und Dienstleistungen einzuführen, welches über die Umweltauswirkungen des entsprechenden Produktes bzw. der Dienstleistung Auskunft geben soll. Das Zeichen ist Ergebnis des sogenannten „Grenelle de l’Environment II“ (kurz: Grenelle II), einer Tagung im französischen Ort Grenelle, die im Sommer 2007 stattfand. Als Grundlage für das staatliche Zeichen wurden hier ebenfalls wie beim deutschen Pilotprojekt die ISO-Normen 14040 und 14044 beschlossen (BOLWIG und GIBBON 2009b: 16). Neben diesem von Staats wegen forcierten PCF gibt es in Frankreich bereits seit 2008 den privatwirtschaftlichen PCF mit der Bezeichnung „Indice Carbone Casino“ des Lebensmittelhändlers Casino. Mit dem Zeichen sind etwa 160 Produkte aus dem Eigenmarkensortiment der Bereiche Nahrungsmittel und Getränke gekennzeichnet. Angegeben wird dabei für den PCF eine konkrete Zahl, bezogen auf 100 g des fertigen Produktes. Zielsetzung des Zeichens soll aus Anbietersicht dabei sein, dass die Konsumenten verschiedene Produkte hinsichtlich des THG-Ausstoßes miteinander vergleichen können (GROUPE CASINO 2010).

---

<sup>8</sup> PAS steht für „Public Available Standard“ (BOLWIG und GIBBON 2009b: 13).

Zwei weitere privatwirtschaftliche PCFs, die sich in Europa finden lassen, sind einmal das „CO<sub>2</sub> approved by climatop“-Label, welches vom schweizerischen Handelskonzern Migros vergeben wird, sowie das seit 2009 vergebene schwedische Label „Climate Labelling for Food“, welches das Ergebnis einer Kooperation des schwedischen Bio-Händlers KRAV mit der Organisation Sigill Kvalitetssystem ist (KLIMATMÄRKNING 2010), die wiederum das Zeichen „Svenskt Sigill“ für geprüfte schwedische Lebensmittel vergibt (SVENSKT SIGILL 2010). Sowohl das schwedische Zeichen als auch das Zeichen von Migros verzichten dabei auf eine numerische Angabe des PCFs. Stattdessen werden jeweils die Produkte gekennzeichnet, die hinsichtlich der THG-Emissionen einen Vorteil gegenüber anderen Produkten des gleichen Produktranges aufweisen (MIGROS 2010; KLIMATMÄRKNING 2010).

In Tabelle 1 findet nachfolgend ein Vergleich der PCFs der Lebensmittelhändler Tesco und Casino mit dem schwedischen Zeichen, dem PCF des deutschen Pilotprojektes sowie dem geplanten staatlichen französischen PCF statt. Hinsichtlich des Zeichens von Migros werden von Seiten des Anbieters zu wenige Informationen zur genauen Methodik gegeben, weshalb eine Aufnahme des Zeichens in den Vergleich nicht sinnvoll erscheint.

Die an eine Darstellung von SCHLICH et al. (2009a) angelehnte Tabelle zeigt unter anderem auf, welche Produktarten von den verschiedenen PCF-Ansätzen betrachtet werden und ob über die Label weiterhin konkrete, numerische Angaben zum PCF gegeben werden. Ebenfalls findet eine Analyse der PCF-Ansätze dahingehend statt, ob es bei der Berechnung der PCFs Gutschriften für die Verwendung erneuerbarer Energien oder für ein einbezogenes Recycling gibt. Zusätzlich wird in der Tabelle dargestellt, welche Phasen der Wertschöpfungskette bei der Berechnung des PCFs überhaupt einbezogen werden. Hier zeigen sich zwischen den verschiedenen PCF-Ansätzen deutliche Unterschiede. So werden im Rahmen des deutschen PCF-Pilotprojektes Emissionen des Zwischenhandels sowie Emissionen, die durch die Nutzung und den Verbrauch der Ware beim Konsumenten entstehen, berücksichtigt und eingeschlossen. Anders sieht dies bei den PCFs der Lebensmittelhändler Tesco und Casino aus. Hier werden derartige Emissionen jeweils nicht in die Berechnungen des PCFs mit aufgenommen. Hierbei zeigt sich die Problematik, dass die verschiedenen PCFs aufgrund unterschiedlicher Systemgrenzen und unterschiedlicher angewandter Methodiken nur schwer zu vergleichen sind. So ist zu erwarten, dass aufgrund der ausgeschlossenen Prozessstufen des Zwischenhandels und der Endnutzung bzw. des Verbrauchs die angegebenen PCFs von Casino und Tesco niedriger sein werden als beispielsweise ein PCF, welcher nach der deutschen Methodik berechnet würde.

Tabelle 1: Vergleich verschiedener europäischer PCF-Ansätze

	Deutschland: PCF Pilotprojekt	Frankreich: Casino	Frankreich: Grenelle II	GB: Tesco	Schweden: Climate Labelling for Food
Label			?		
Methode	ISO 14040 und ISO 14044	ISO 14064	ISO 14040 und ISO 14044	PAS 2050	ISO 14040, ISO 14044 und ISO 14064
Angabe numeri- scher Wert für PCF	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Produktart	Lebensmittel und Non-Food-Artikel				Lebensmittel
Einbezug Primär- produktion	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	?	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Einbezug Trans- port und Distribu- tion	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	?	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Einbezug Zwi- schenhandel	<input checked="" type="checkbox"/>	<input type="checkbox"/>	?	<input type="checkbox"/>	<input type="checkbox"/>
Einbezug Nutzung und Verbrauch	<input checked="" type="checkbox"/>	<input type="checkbox"/>	?	<input type="checkbox"/>	<input type="checkbox"/>
Gutschrift durch erneuerbare Ener- gien	<input type="checkbox"/>	<input type="checkbox"/>	?	<input type="checkbox"/>	Ab 2012
Gutschrift durch Recycling	<input checked="" type="checkbox"/>	Bisher nicht	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	?
Implementation	Nicht geplant	2008	2011	2008	2009

Anmerkungen:  bedeutet „ja“,  bedeutet „nein“. Ein „?“ bedeutet, dass dieser Sachverhalt noch nicht abschließend geklärt ist.

Quelle: Eigene und erweiterte Darstellung auf der Grundlage von SCHLICH et al. 2009a.

Gerade hinsichtlich der Endnutzung sieht der HAUPTVERBAND DES DEUTSCHEN EINZELHANDELS (HDE) eine Kennzeichnung des individuellen PCFs von Produkten kritisch. So wird aus Sicht des HDEs in den verschiedenen PCF-Ansätzen das individuelle Verbraucherverhalten, beispielsweise beim individuellen Transport nach Hause, bei der Lagerung und auch bei der Zubereitung, nicht ausreichend berücksichtigt. Aus Sicht des HDEs kommt diesem individuellen Verbraucherverhalten jedoch ein signifikanter Einfluss auf den PCF eines Produktes zu. Daneben spielt es nach dem HDE für den PCF eines Produktes ebenfalls eine entscheidende Rolle, wie das Produkt transportiert wurde (per Schiff, Flugzeug, LKW oder Bahn) (HDE 2008: 3). Diesbezüglich merken MICHAELOWA und KRAUSE an, dass innerhalb dieser Transporttypen die verwendete Technologie einen starken Einfluss auf die THG-Emissionen ausübt (MICHAELOWA und KRAUSE 2000: 131f.). Weiterhin sieht der HDE grundsätzliche Probleme, bei hoch verarbeiteten Produkten, wie etwa Tiefkühl-Pizzen, überhaupt die korrekten THG-Emissionen bestimmen zu können. So besteht eine Tiefkühl-Pizza aus vielen verschiedenen Zutaten wie etwa Salami oder Paprika. Zur korrekten Bestimmung des PCF müssten entspre-

chend auch bereits die Wertschöpfungsketten der verwendeten Zutaten einbezogen werden, was aus Sicht des HDEs zu starken Komplexitätsproblematiken und hohen Kosten, die in keinem Verhältnis zum Nutzen stehen würden, führt (HDE 2008: 2ff.).

Probleme hinsichtlich der Kosten, die bei Ermittlung des PCFs eines Produktes anfallen, sehen auch BOLWIG und GIBBON. Sie erwarten, dass bei einer freiwilligen Einführung eines PCF-Labels dieses vor allem von großen und erfolgreichen Anbietern bzw. Produzenten genutzt wird. Für kleinere Produzenten erwarten sie stattdessen, dass diese aufgrund der hohen Kosten und der notwendigen personellen Ressourcen auf die Implementation eines solchen Labels verzichten werden. Hierdurch sehen sie vor allem auch Produzenten aus Entwicklungsländern benachteiligt (BOLLWIG und GIBBON 2009a: 3). BRENTON et al. beschreiben weiterhin, dass auch bei verbindlichen PCF-Ansätzen Produzenten in Entwicklungsländern von diesen Märkten mit derartigen PCF-Ansätzen, wie beispielsweise „Grenelle II“ in Frankreich, ausgeschlossen werden könnten. Aus ihrer Sicht sind Produzenten in Entwicklungsländern aufgrund der gewöhnlich kleinen Betriebsgrößen nicht in der Lage, die notwendigen Aufwendungen für die Ermittlung des PCFs eines Produktes stemmen zu können. Ebenfalls Nachteile für die Produzenten in Entwicklungsländern sehen die Autoren durch die längeren Transportwege zu potentiellen Abnehmermärkten (BRENTON et al. 2009: 244ff.).

Dass längere Transportwege sich nicht zwangsläufig negativ auf die THG-Emissionen auswirken müssen, zeigen SCHLICH et al.. Allerdings kommen sie in ihren Ausführungen zu dem Ergebnis, dass die freigesetzten THG-Emissionen mit einer zunehmenden Betriebsgröße sinken (SCHLICH et al. 2009b: 116ff.). Hierdurch wären die Entwicklungsländer mit ihren kleinstrukturierten Industrien dennoch benachteiligt. BRENTON et al. beschreiben hingegen, dass Entwicklungsländern aufgrund ihrer extensiven (Lebensmittel-)Produktion möglicherweise zukünftig eine stärkere weltweite Bedeutung zukommen kann (BRENTON et al. 2009: 243).

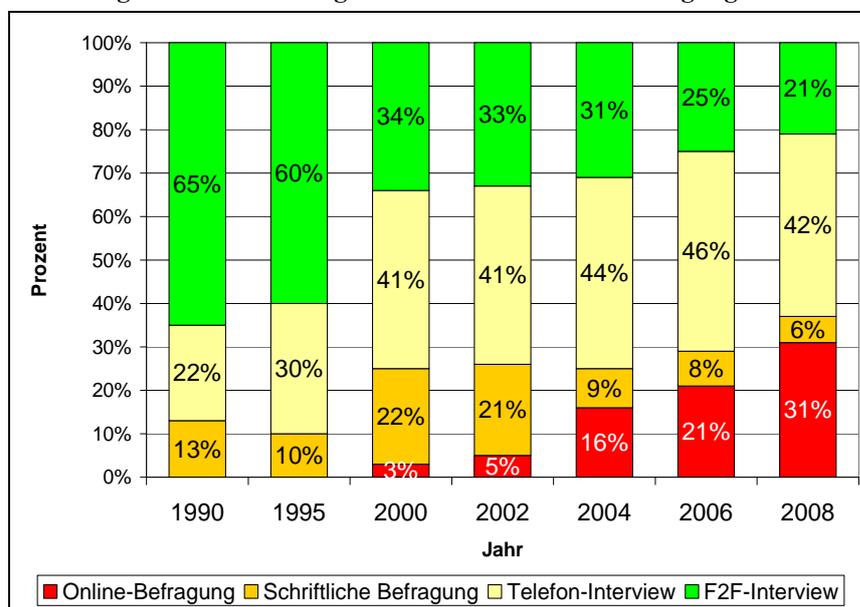
### 3 Relevante Methoden der empirischen Sozialforschung

Da entsprechend den in Kapitel 1.2 dargestellten Zielsetzungen dieser Arbeit die Wirkungen zweier Befragungsmethoden auf das Antwortverhalten analysiert werden sollen, wird in diesem Kapitel auf die beiden zu vergleichenden Befragungsmethoden der mündlichen Befragung bzw. des Face-to-Face-(F2F)-Interviews sowie der Online-Befragung eingegangen. Nach der Besprechung der F2F-Interviews in Kapitel 3.1 sowie der Online-Befragungen in Kapitel 3.2 wird in Kapitel 3.3 auf verschiedene, mögliche Fehlerarten eingegangen, die ursächlich für ein unterschiedliches Antwortverhalten von Befragungsteilnehmern in Abhängigkeit von der Befragungsmethode sein können.

#### 3.1 Face-to-Face-Interviews

Um Datenerhebungen im Rahmen der empirischen Sozialforschung durchführen zu können, wird allgemein ein Erhebungsinstrument benötigt, mit welchem die Werte der Merkmalsträger erfasst werden können (RAITHEL 2008: 65). Hierbei bieten sich je nach Untersuchungszweck und inhaltlicher Ausgestaltung der Untersuchung für den Forscher im Rahmen der Primärforschung verschiedene Erhebungsmethoden wie etwa die Inhaltsanalyse, die Beobachtung, die Befragung oder das Experiment an (SCHNELL et al. 2005: 319ff.).

**Abbildung 4: Entwicklung der Prozentanteile der Befragungsmethoden im Zeitraum 1990 bis 2008**



Quelle: Eigene Darstellung nach ADM 2008: 12.

Die lange Zeit am Häufigsten verwendete Erhebungsmethode ist dabei die Befragung in Form der mündlichen Befragung, die ebenfalls als F2F-Interview bezeichnet wird und bei der sich mindestens zwei Personen zu einem bestimmten Thema miteinander austauschen. Das F2F-Interview gilt dabei vielfach noch als Standardinstrument und „Königsweg“ der empirischen

Sozialforschung zur Ermittlung von Wissen, Fakten, Meinungen sowie Bewertungen von Sachverhalten im Rahmen sozialwissenschaftlicher Untersuchungen (ALEMANN 1977: 207ff.; HABERMEHL 1992: 157; SCHNELL et al. 2005: 321ff.; KROMREY 2006: 335; ATTESLANDER 2008: 101f.; DIEKMANN 2008: 434; DYKEMA et al. 2008: 240). Gerade im Bereich der Marktforschung zeigen sich in den letzten Jahren allerdings zunehmend Tendenzen dahingehend, dass etwa seit der Jahrtausendwende mit einem verstärkten Aufkommen von Internet-basierenden Befragungsmethoden der Anteil an F2F-Befragungen zugunsten des Anteils an Online-Befragungen zurückgeht (ADM 2008: 12; WIEGAND 2000: 12ff.), wie auch Abbildung 4 verdeutlicht.

### **Unterscheidung alltägliche und sozialwissenschaftliche Befragung**

Abzugrenzen von der sozialwissenschaftlichen Befragung, die in der vorliegenden Arbeit thematisiert wird, ist die sogenannte alltägliche Befragung. SCHEUCH liefert mit seiner Definition eines sozialwissenschaftlichen F2F-Interviews verschiedene Aspekte, die für dieses als relevant angesehen werden: „*Unter Interview als Forschungsinstrument sei hier verstanden ein planmäßiges Vorgehen mit wissenschaftlicher Zielsetzung, bei dem die Versuchsperson durch eine Reihe gezielter Fragen oder mitgeteilter Stimuli zu verbalen Informationen veranlasst werden soll*“ (SCHEUCH 1973: 70f.).

Nach den Ausführungen von ATTESLANDER bzw. ATTESLANDER und KOPP eignen sich die in der Definition genannten Kriterien jedoch nicht, um eine sozialwissenschaftliche Befragung von einer alltäglichen Befragung abgrenzen zu können. So sind sowohl die alltägliche als auch die sozialwissenschaftliche Befragung etwa dadurch geprägt, dass in irgendeiner Art und Weise im Rahmen eines Kommunikationsprozesses Informationen zwischen den beteiligten Akteuren ausgetauscht werden. So wird es nach ATTESLANDER im Bezug auf die alltägliche Befragung im Alltag kaum Gespräche geben, die nicht grundsätzlich durch Fragen, Antworten oder Gegenfragen geprägt sind, weshalb der Aspekt des Informationsaustauschs kaum zur Abgrenzung herangezogen werden kann. Auch andere Aspekte wie etwa eine systematische Vorbereitung und Durchführung oder eine Zielgerichtetheit reichen nach ATTESLANDER nicht aus, um die beiden Arten der mündlichen Befragung eindeutig voneinander abzugrenzen. So zeichnen sich viele alltägliche Befragungen ebenfalls durch eine systematische Durchführung bzw. eine systematische Vorbereitung aus. Als Beispiel kann hier ein Gespräch zwischen einem Arzt und seinem Patienten herangezogen werden, in dem der Arzt systematisch den Patienten nach Befindlichkeitsstörungen befragt.

Auch der Aspekt der Zielgerichtetheit eignet sich nicht, um eine wissenschaftliche Befragung von einer alltäglichen abzugrenzen. Dies kann ebenfalls wieder am Arzt-Patienten-Beispiel

aufgezeigt werden. So versucht der Arzt zielgerichtet durch seine Befragung, eine Diagnose zu stellen und daraus als Zielsetzung eine geeignete Therapie ableiten zu können.

Der relevante Unterschied zwischen beiden Befragungen ist stattdessen in der theoriegeleiteten Kontrolle der gesamten Befragung (und damit jedes Teilschritts) zu sehen. Einerseits soll dadurch gewährleistet werden, dass die Befragung als wissenschaftliche Methode eingesetzt werden kann. Andererseits soll dadurch feststellbar gemacht werden, inwieweit die erhobenen Daten durch verschiedene, während der Befragung eingewirkte Einflüsse verzerrt sind. Nur durch diese systematische Kontrolle jeder Phase der Befragung kann die Gültigkeit und die Verlässlichkeit von Ergebnissen gewährleistet werden (ATTESLANDER und KOPP 1999: 146ff.; ATTESLANDER 2008: 101ff.).

Als weiteres Unterscheidungsmerkmal kann nach ATTESLANDER und KOPP auch das System genannt werden, über welches Aussagen getroffen werden sollen: So ist es nicht Zielsetzung der sozialwissenschaftlichen Befragung, konkrete Aussagen über eine einzelne, bestimmte Person zu treffen. Hier wird stattdessen angestrebt, allgemeine theoretische Zusammenhänge zu überprüfen und entsprechend zu bestätigen bzw. zu falsifizieren (ATTESLANDER und KOPP 1999: 146ff.; ATTESLANDER 2008: 101ff.).

### **Theorie des Interviews**

Wesentliches Merkmal für eine Befragung in Form eines F2F-Interviews bzw. einer mündlichen Befragung ist die verbale Kommunikation zwischen den beteiligten Akteuren, wie auch ATTESLANDER in einer Definition beschreibt: „*Befragung bedeutet Kommunikation zwischen zwei oder mehreren Personen*“ (ATTESLANDER 2008: 101).

Im Unterschied zur oben angesprochenen alltäglichen Befragung spielt es für die Beantwortung der Fragen im Interview in diesem Zusammenhang eine bedeutsame Rolle, dass allgemein von einer hohen Künstlichkeit und Formalität der Befragungssituation ausgegangen wird (SCHOLL 1993: 127ff.; KROMREY 2006: 337ff.; ATTESLANDER 2008: 113f.; DIEKMANN 2008: 439).

So handelt es sich etwa nach KROMREY beim F2F-Interview um ein formalisiertes Verfahren der empirischen Sozialforschung, mit dessen Hilfe sozialwissenschaftliche Sachverhalte gemessen werden sollen. Die gesamte Kommunikation hat dabei einen rein instrumentellen Charakter und dient lediglich als Mittel zum gewünschten Zweck in Form der Antworten (KROMREY 2006: 338). Ähnlich sieht dies ATTESLANDER: *Durch verbale Stimuli (Fragen) werden verbale Reaktionen (Antworten) hervorgerufen. [...] Die Antworten beziehen sich auf erlebte und erinnerte soziale Ereignisse, stellen Meinungen und Bewertungen dar.*“ (ATTESLANDER 2008: 101).

Weiter wird von verschiedenen Autoren der Künstlichkeit einer Befragungs- bzw. Interview-situation eine besondere Bedeutung beigemessen. So werden von ALEMANN diesbezüglich drei von ihm als essentiell betrachtete Aspekte genannt: (i) Die miteinander interagierenden Personen sind im Normalfall **Fremde**, die noch nie einander wissentlich begegnet sind. Für eine der beiden Personen - den Interviewer - gibt es darüber hinaus festgelegte Rollenvorschriften, nach denen sie sich verhalten soll. Diese werden gewöhnlich in einer Interviewerschulung vermittelt (ALEMANN 1977: 208f.; KROMREY 2006: 338f.). (ii) Bei einer mündlichen Befragung bzw. einem F2F-Interview handelt es sich um eine **asymmetrische soziale Beziehung**. So geht von der interviewenden Person gewöhnlich der überwiegende Anteil der Aktivitäten aus, etwa indem sie Fragen an die andere Person richtet. Die andere Person in Form des Befragten hingegen verhält sich gewöhnlich weitestgehend passiv und gibt ausschließlich Auskunft hinsichtlich der gestellten Fragen (ALEMANN 1977: 208f.). Daher handelt es sich nach ATTESLANDER je nach Befragungsmethode beim Interviewten weniger um einen wirklichen Gesprächspartner als um einen Datenträger, von dem durch den Interviewer Daten abgerufen werden (ATTESLANDER 2008: 339). (iii) Ebenfalls unterscheidet sich die Situation des Interviews von natürlichen Interaktionen bzw. alltäglichen Befragungen dadurch, dass sie für den Befragten **sozial folgenlos** ist. Auf diesen Sachverhalt wird vom Interviewer häufig ausdrücklich hingewiesen (ALEMANN 1977: 208f.). SCHEUCH führt hinsichtlich einer solchen Folgenlosigkeitsbekundung allerdings an, dass diese auch zum Gegenteil in Form von Misstrauen bei dem Interviewten führen kann (SCHEUCH 1973: 68ff.). Weiterhin kann ausgesagt werden, dass eine soziale Folgenlosigkeit für den Befragten nur gewährleistet werden kann, so lange keine weiteren, ihm vertrauten Personen beim Interview mitanwesend sind<sup>9</sup> (DIEKMANN 2008: 439).

DIEKMANN greift die genannten Aspekte von ALEMANN auf und gibt eine entsprechende Definition für ein F2F-Interview. Danach handelt es sich bei einem Interview „*um eine künstliche, asymmetrische Interaktion unter Fremden mit der stillschweigenden Vereinbarung, dass keine dauerhafte Beziehung eingegangen wird*“ (DIEKMANN 2008: 439).

Einen weiteren relevanten Aspekt zur Theorie des Interviews beschreibt ATTESLANDER. So sollte bei einer Auseinandersetzung mit dem F2F-Interview als Befragungsmethode berücksichtigt werden, dass eine Befragung grundsätzlich eine soziale Situation darstellt. Zu dieser gehören nicht nur die beteiligten Personen am Interview, sondern auch die jeweilige Umgebung, in der das F2F-Interview bzw. eine andere Befragungsart stattfindet. Dadurch ergibt

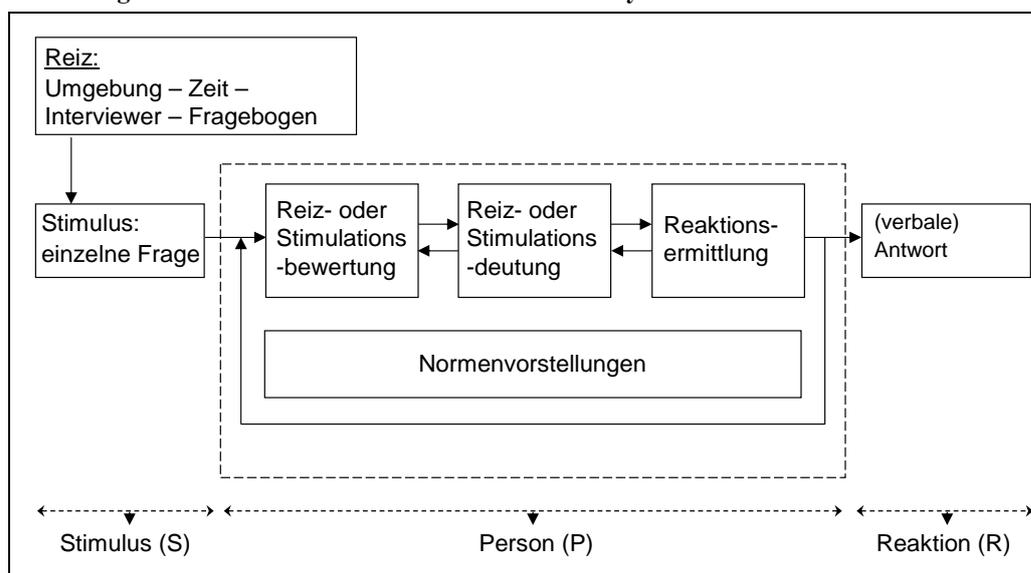
---

<sup>9</sup> Auf mögliche, aus der Anwesenheit Dritter resultierende Interviewfehler wird an späterer Stelle detaillierter eingegangen.

sich auch ein Problem des Interviews: So ist die Umgebung total und nie vollständig beobachtbar. Hieraus folgt, dass eine Totalkontrolle der sozialen Situation in Form des Interviews zwar anzustreben ist (siehe oben), der Beobachtungsfokus je nach herangezogenem Grundmodell allerdings an unterschiedlicher Stelle liegt. So beschreibt ATTESLANDER, dass in der Fachliteratur zwei Grundhaltungen in Form unterschiedlicher Reaktionsmodelle erkennbar werden. Ein Teil geht dabei von einem  $S \rightarrow R$ -Modell aus. Diesem liegt als Annahme zugrunde, dass ein zwingender und direkter Zusammenhang zwischen einem Stimulus  $S$  und einer bestimmten Reaktion  $R$  besteht. Anhänger dieser Grundhaltung proklamieren daher, höchstmögliche Kontrolle auf den Stimulus in Form des Fragebogens zu legen, um die Verlässlichkeit der Reaktion in Form der Antwort gewährleisten zu können. Daher wird unter Anhängern dieser Grundhaltung die Relevanz des Fragebogensaufbaus sowie der richtigen Fragenformulierung betont.

Neben den Anhängern dieser Grundhaltung beziehen sich nach ATTESLANDER zunehmend, wenn auch implizit, Autoren auf ein  $S \rightarrow P \rightarrow R$ -Modell, wobei das  $P$  für Person steht. Im Unterschied zum oberen Modell liegt diesem Modell die Annahme zugrunde, dass zwischen Stimulus  $S$  und Reaktion  $R$  keine zwingenden und unmittelbaren Beziehungen bestehen. Damit einher geht das Verständnis, dass die Antwort keine kausale Folge einer Frage ist, sondern die befragte Person mit ihren Vorstellungen, Erwartungen und Normen dazwischen geschaltet ist, wie Abbildung 5 präzisiert.

**Abbildung 5: Die Interviewsituation als Reaktionssystem**



Quelle: Eigene Darstellung nach ATTESLANDER 2008: 106.

Damit verbunden ist die Sichtweise, dass nicht mehr nur der Stimulus, sondern die gesamte Situation, in der die Untersuchung durchgeführt wird, möglichst einer systematischen Kontrolle zu unterziehen ist. Daher sollte vorab überlegt werden, welche Bedingungen möglicher-

weise systematisch erfasst werden sollten (ATTESLANDER 2008: 104ff.). Aus den vorangegangenen Ausführungen geht hervor, dass sowohl der Fragebogen im Falle des  $S \rightarrow R$ -Modells als auch weitere Reize wie etwa die Umgebung oder die Anwesenheit des Interviewers im  $S \rightarrow P \rightarrow R$ -Modell zu einer Beeinflussung des Antwortverhaltens führen können. So werden Erhebungstechniken, bei denen sich die Teilnehmer bewusst darüber sind, dass sie Gegenstand einer Untersuchung sind und daher die Möglichkeit haben, auf den Datenerhebungsvorgang aktiv zu reagieren, als reaktive Messverfahren bezeichnet. Entsprechend dazu wird der Sachverhalt, dass Personen in Erhebungssituationen auf einen der oben genannten Reize reagieren, als Reaktivität bezeichnet. Diese Reaktivität kann sich als Problem für die Validität der erhobenen Daten erweisen, da Reaktionen auf äußere Reize häufig nicht von inhaltlichen Reaktionen wie etwa einer sinnvollen Antwort auf eine Frage getrennt werden können (ALEMANN 1977: 221f.; SCHNELL et al. 2005: 353; BROSIUS et al. 2008: 129f.).

### Formen der mündlichen Befragung

Das hier besprochene F2F-Interview gehört, entsprechend seiner alternativen Bezeichnung als mündliche Befragung, zur Erhebungsform der Befragung. Hierbei handelt es sich nicht um ein uniformes Verfahren, sondern es lassen sich stattdessen innerhalb des Oberbegriffs „Befragung“ verschiedene Formen unterscheiden. Eine gute Möglichkeit zur systematischen Unterscheidung und Klassifikation liefert ALEMANN. In Anlehnung daran können vier Dimensionen der Befragung unterschieden werden:

1. Grad der Standardisierung,
2. Zahl der befragten Personen zur gleichen Zeit,
3. Art der Befragungssituation,
4. Häufigkeit der Befragung von Einzelpersonen (ALEMANN 1977: 216ff.).

So können die verschiedenen Befragungsmethoden einmal anhand ihres **Standardisierungs- bzw. Strukturierungsgrades** unterschieden werden. Nach DIEKMANN handelt es sich hierbei um ein Kontinuum mit den beiden Polen „vollständig strukturiert“ auf der einen sowie „unstrukturiert und offen“ auf der anderen Seite. Je nach Standardisierungs- bzw. Strukturierungsgrad unterscheiden sich die Befragungen dabei sehr deutlich voneinander. So werden bei einer vollständig strukturierten Befragung alle Fragen mit vorgegebenen Antwortmöglichkeiten in einer vorher festgelegten Reihenfolge gestellt. Standardisierte Befragungen finden dabei typischerweise in deskriptiven Studien Anwendung. Bei unstrukturierten und offenen Interviews werden dem Interviewer hingegen nur minimale Vorgaben gegeben, beispielsweise in Form eines Leitfadens. Fragenwortlaut und Ablauf der Befragung werden in diesem Fall durch das Ermessen des Interviewers in Abhängigkeit von der konkreten Befragungssituation

gesteuert (ALTOBELLI 2007: 36). Als Extremwert ist bei unstrukturierten Interviews nach DIEKMANN auch eine reine Nennung des Themas, ohne Leitfaden, denkbar (DIEKMANN 2008: 437). Zielsetzung solcher unstrukturierter Befragungen ist es, Sinnzusammenhänge, also die Meinungsstrukturen des Befragten, zu erfassen. Daher wird auch keine isolierte Reaktion auf die einzelnen Stimuli angestrebt, sondern offene Reaktionsmöglichkeiten des zu Befragenden. Diese daraus resultierenden Freiheitsspielräume auf Seiten des Interviewers sowie des Befragten sind bei stark strukturierten Befragungen hingegen durch den Fragebogen und die vorgeplante Vorgehensweise stark beschränkt. Dies kann sich dahingehend als problematisch erweisen, dass Fragen, mit denen Verständnisprobleme aufgedeckt werden könnten, im Unterschied zu offenen Befragungen in vollständig strukturierten Befragungen nicht zulässig sind. Nur bei Zurückfragen des Befragten kann daher überhaupt eine Klärung erzielt werden. ATTESLANDER führt bezüglich der häufig verwendeten Bezeichnung „unstrukturiert“ an, dass diese aus seiner Sicht nicht als zutreffend angesehen werden kann, da seines Erachtens keine soziale Situation ohne Struktur auskommt und so immer eine gewisse Strukturierung zwangsläufig vorliegen muss. Auch zweifelt er die, in der Literatur zur empirischen Sozialforschung allgemein gebräuchliche und daher in dieser Arbeit ebenfalls herangezogene, simultane Verwendung der Begriffe strukturiert und standardisiert an. Danach bezieht sich das Begriffspaar „strukturiert - nicht strukturiert“ seiner Ansicht nach ausschließlich auf die konkrete Interviewsituation und den Sachverhalt, ob Strukturierungshilfen wie ein Fragebogen oder ein Leitfaden herangezogen wurden. Wird demnach ein Fragebogen oder ein Interviewerleitfaden verwendet, kann von einem strukturierten Interview gesprochen werden, während bei einem unstrukturierten Interview auf diese Hilfsmittel verzichtet wird. Das Begriffspaar „standardisiert - nicht-standardisiert“ hingegen fokussiert nach ATTESLANDER nicht die Interviewsituation, sondern den Fragebogen als herangezogenes Instrument und unterscheidet, ob Antwortkategorien im Fragebogen verwendet bzw. vorgegeben werden oder nicht. Bei einem nicht-standardisierten Interview wird entsprechend auf eine vorgegebene Kategorisierung der Antworten verzichtet. Ein nicht-standardisierter Fragebogen kann daher nur Verwendung finden, wenn eine Vergleichbarkeit der Daten sowie Häufigkeitsverteilungen nicht Zielsetzung der Untersuchung oder nicht möglich sind (ATTESLANDER 2008: 124ff.).

Ein weiteres Unterscheidungsmerkmal für verschiedene Befragungsmethoden ist die **Zahl der befragten Personen zur gleichen Zeit** (ALEMANN 1977: 216). Hierbei kann zwischen den beiden möglichen Ausprägungen Einzelbefragung sowie Gruppenbefragung unterschieden werden. Während bei einer Einzelbefragung jeweils nur eine Untersuchungseinheit wie etwa eine Einzelperson oder auch ein einzelner Haushalt befragt wird, werden im Rahmen von

Gruppenbefragungen hingegen mehrere Untersuchungseinheiten zur gleichen Zeit interviewt. Das Einzelinterview stellt den Standardfall bei deskriptiven Studien dar, während Gruppenbefragungen hauptsächlich im Rahmen explorativer Untersuchungen Anwendung finden. Durch den Einsatz von Gruppenbefragungen erhofft man sich dabei, spontane Reaktionen und Assoziationen erfassen zu können sowie mögliche Antworthemmungen bei den Befragten abbauen zu können (ALTOBELLI 2007: 36). Ebenfalls werden Gruppenbefragungen herangezogen, um den Prozess der Meinungsbildung in Gruppen untersuchen zu können (ALEMANN 1977: 219). Aus der Kombination der verschiedenen Ausprägungen der beiden erläuterten Dimensionen „Strukturierungsgrad“ sowie „Zahl der befragten Personen“ können nach SCHNELL et al. sowie ATTESLANDER zusätzlich verschiedene Formen der mündlichen Befragung unterschieden werden. Entsprechend der Darstellung in Tabelle 2 lässt sich erkennen, dass es sich bei dem F2F-Interview, welches in der vorliegenden Arbeit thematisiert wird, um eine Einzelbefragung mit hohem Strukturierungsgrad handelt. Bei der an SCHNELL et al. sowie ATTESLANDER orientierten Einteilung handelt es sich dabei nur um eine Möglichkeit der Einteilung, die sich je nach Autor unterscheidet. So unterscheidet ALEMANN beispielsweise ausschließlich die beiden Formen des Gruppeninterviews sowie der Gruppendiskussion voneinander (ALEMANN 1977: 218f.). ATTESLANDER hingegen unterscheidet, simultan zu SCHNELL et al., ebenfalls anhand des Strukturierungsgrades der Interviewsituation drei unterschiedliche Formen der Gruppenbefragung. Die Bezeichnungen „Gruppeninterview“ sowie „Gruppenbefragung“ werden hier jedoch umgekehrt verwendet.

**Tabelle 2: Verschiedene Formen der mündlichen Befragung**

Strukturierungsgrad der Interviewsituation	Einzelbefragung	Gruppenbefragung
wenig	Experteninterview; exploratives Interview	Gruppendiskussion
teilweise	Leitfadengespräch	Gruppenbefragung
stark	Einzelinterview	Gruppeninterview

Quelle: SCHNELL et al. 2005: 323; ATTESLANDER 2008: 123.

Als dritte Dimension zur Unterscheidung von verschiedenen Befragungsmethoden kann die **Art der Befragungssituation** genannt werden. ALEMANN unterscheidet hierbei die mündliche von der schriftlichen Befragung (ALEMANN 1977: 216). Einen ähnlichen Ansatz verfolgt ATTESLANDER, indem er die Kommunikationsart anstatt der Befragungssituation als Unterscheidungsmerkmal heranzieht. Hierbei werden von ihm hinsichtlich der Kommunikationsart mündliche und schriftliche Befragungen unterschieden (ATTESLANDER 2008: 123).

Zuletzt lassen sich Befragungen anhand der Einteilung ALEMANNs hinsichtlich der **Häufigkeit der Befragung von Einzelpersonen** unterscheiden. Bei dieser vierten Dimension kann einmal zwischen einer einmaligen Befragung einer Einzelperson sowie einer Panel-Befragung unterschieden werden (ALEMANN 1977: 216). Im Rahmen eines Panels wird eine möglichst gleichbleibende Stichprobe verschiedener Merkmalsträger wie Einzelpersonen, aber auch Haushalte, Unternehmen oder Geschäfte im Zeitablauf mehrfach zu einem gleichbleibenden Sachverhalt befragt. Zielsetzung dieser Vorgehensweise ist, dadurch Veränderungen in bestimmten Variablen ermitteln zu können (GÜNTHER et al. 2006: 3ff.).

### **Aufgaben des Interviews**

Bei einem Interview lassen sich nach ATTESLANDER und KOPP drei grundlegende Aufgabenbereiche voneinander abgrenzen. Bei diesen handelt es sich (i) um die Entdeckung, (ii) um die Messung sowie (iii) um die Interpretation und Verfeinerung statistischer Beziehungen (ATTESLANDER und KOPP 1999: 157ff.).

Der Aufgabenbereich der **Entdeckung** zielt darauf ab, für die Untersuchung relevante Variablen aufzufinden, die entweder eine Abgrenzung der als relevant betrachteten Zielgruppe erlauben oder für die Herstellung theoretischer Zusammenhänge benötigt werden. Für diese Aufgabe eignet sich besonders eine unstrukturierte Interviewform wie etwa ein exploratives Interview, da hier verschiedene qualitative Aspekte eines Problems im Hinblick auf die genannten Zielsetzungen erleuchtet werden sollen (ATTESLANDER und KOPP 1999: 157ff.; KEPPER 2008: 184ff.).

Grundlegend befasst sich wissenschaftliche Tätigkeit damit, allgemeine theoretische Zusammenhänge zu überprüfen. Diesbezüglich ist die zweite Aufgabe des Interviews, das **Messen**, relevant. So ist in den meisten Fällen eine deskriptive Analyse sozialer Beziehungen, mit Ausnahme des ersten Aufgabenbereiches, nicht ausreichend. Stattdessen kann das Niveau der Erklärung nur durch eine Quantifizierung und entsprechend durch eine Zuordnung von Häufigkeitsverteilungen erreicht werden. Hier bietet sich ein strukturiertes Interview an, bei dem mit Hilfe eines standardisierten Fragebogens den Befragten durch den Interviewer alleamt die gleichen Fragen samt Antwortmöglichkeiten gegeben werden. Im Rahmen der oben dargestellten Reaktionsmodelle kann hierbei eine Antwort des Befragten als individuelle Reaktion auf den gleichen Reiz angesehen werden, wodurch eine Zusammenfassung zu Häufigkeiten und eine Zuordnung zu den entsprechenden Variablen ermöglicht wird (ATTESLANDER und KOPP 1999: 159f.).

Der dritte Aufgabenbereich von Interviews umfasst abschließend die **Interpretation und die Verfeinerung statistischer Beziehungen**. So beschreiben ATTESLANDER und KOPP, dass es

bei der Auswertung von Daten teilweise vorkommen kann, dass unerwartete statistische Beziehungen auftreten, die mit den zur Verfügung stehenden Hypothesen nicht erklärt werden können. Ebenfalls ist denkbar, dass einzelne Fälle auftreten, die statistisch sehr deutlich aus der Reihe fallen. Zur Klärung solcher Abweichungen oder Abweichler kann auch hier wieder ein unstrukturiertes Interview herangezogen werden. Somit kann zusammengefasst werden, dass bei qualitativen Aufgabenstellungen wie der Entdeckung sowie der Interpretation und der Verfeinerung statistischer Beziehungen unstrukturierte Interviews und bei quantitativen Aufgabenstellungen wie dem Messen strukturierte Interviews das jeweils geeignetere Verfahren darstellen (ATTESLANDER und KOPP 1999: 160).

### 3.2 Online-Befragungen

Online-Befragungen<sup>10</sup> erfreuen sich, auch durch eine zunehmende Verbreitung des Internets, einer steigenden Beliebtheit bei den Anwendern von Befragungen, etwa in Marktforschungsunternehmen oder auch in der Wissenschaft (WIEGAND 2000: 12) und haben zwischenzeitlich die Popularität erreicht, die vor wenigen Jahren CATI-Befragungen<sup>11</sup> hatten (WELKER und MATZAT 2009: 45). So zeigt sich etwa bei erneuter Betrachtung der Abbildung 4 auf Seite 21, dass der Anteil an online durchgeführten Befragungen seit dem Jahre 2000 massiv zugenommen hat und im Jahre 2008 bereits bei 31 % lag. Weiter wird etwa vom ARBEITSKREIS DEUTSCHER MARKT- UND SOZIALFORSCHUNGSINSTITUTE ADM davon ausgegangen, dass dies noch nicht der Höhepunkt der Entwicklungen ist und der Anteil online durchgeführter Befragungen zu Lasten der klassischen Erhebungstechniken wie F2F-, telefonischen oder schriftlichen Interviews zukünftig noch weiter ansteigen wird (ADM 2008: 12).

Online durchgeführte Befragungen bzw. Online-Befragungen, die auch als Web-Befragungen, Web Surveys, Online Questionnaires oder auch als Computerized Self-Administered Questionnaires (CSAQ) bezeichnet werden (TADDICKEN 2008: 37), bieten dabei nach einer Vielzahl von Autoren im Vergleich zum oben dargestellten F2F-Interview als klassisches Instrument der empirischen Sozialforschung entscheidende Vorteile (SCHOLL et al. 2002: 219; FAAS 2003: 120; DUFFY et al. 2005: 617; EVANS und MATHUR 2005: 196f.; WELKER et al. 2005: 8ff.; COUPER und COUTTS 2006: 218; SUE und RITTER 2007: 150f.; BANDILLA et al. 2009: 129; PÖTSCHKE 2009: 78).

---

<sup>10</sup> In der Literatur wird unter dem Begriff „online“ dabei gewöhnlich die direkte Verbindung eines Computers mit einer Datenverarbeitungsanlage bzw. dem Internet verstanden (TADDICKEN 2008: 39).

<sup>11</sup> Die Abkürzung CATI steht für „Computer Assisted Telephone Interview“ und meint eine Computerunterstützung bei der Durchführung von Telefoninterviews. Dadurch kann nach SCHNELL et al. der Erhebungs- und Auswertungsaufwand im Vergleich zu einer Durchführung ohne Computerunterstützung erheblich verringert werden (SCHNELL et al. 2005: 376f.).

### **Vor- und Nachteile von Online-Befragungen im Vergleich zu F2F-Interviews**

So wird im Zusammenhang mit Online-Befragungen als Vorteil häufig eine höhere Durchführungsgeschwindigkeit im Vergleich zu den klassischen Erhebungsmethoden genannt, da hier in kürzerer Zeit über das Internet mehr Teilnehmer befragt werden können als mit Hilfe des konventionellen F2F-Interviews (EVANS und MATHUR 2005: 196f.; WELKER et al. 2005: 8ff.; SUE und RITTER 2007: 150f.; BANDILLA et al. 2009: 129; PÖTSCHKE 2009: 78). Vor allem im Bereich der Marktforschung wird dieser Geschwindigkeitsvorteil geschätzt. WELKER und MATZAT konnten etwa in einer Studie ermitteln, dass von 80 % der Befragten aus dem Bereich der Marktforschung kürzere Projektlaufzeiten sowie eine schnellere Feldarbeit bei Online-Befragungen als große Vorteile angesehen werden (WELKER und MATZAT 2009: 39). Weiterhin können im Rahmen von Online-Befragungen trotz einer geringeren Antwortquote im Vergleich zu F2F-Interviews aufgrund der globalen Reichweite, die das Internet bietet, größere Stichproben generiert werden, während F2F-Interviews aufgrund des erforderlichen Erhebungsaufwandes nur eine begrenzte Stichprobengröße zulassen (EVANS und MATHUR 2005: 196ff.).

Auch bieten online durchgeführte Befragungen aus Sicht der Befragten den Vorteil, dass Auswertungen häufig automatisiert ablaufen, wodurch teilweise auf eine Vercodung, wie sie bei F2F-Befragungen notwendig ist, verzichtet werden kann. Hierdurch kann die Geschwindigkeit der Datenverarbeitung und -auswertung noch einmal gesteigert werden (WELKER und MATZAT 2009: 39). Um ähnliche Leistungen mit einer Erhebung auf der Basis von F2F-Interviews erreichen zu können, müsste eine hohe Zahl von Interviewern akquiriert werden, was sich im Vergleich negativ auf die zu veranschlagenden Kosten auswirken würde.

Dieser Kostenaspekt wird von verschiedenen Autoren ebenfalls als Vorteil von Online-Befragungen im Vergleich zu Untersuchungen auf der Basis von F2F-Interviews genannt. Neben dem oben genannten Aspekt einer, zumindest (halb-)automatisierten Auswertung kommt hier natürlich auch der Verzicht auf kostenintensive Interviewer zum Tragen, wodurch die variablen Kosten pro Interview deutlich gesenkt werden können (DUFFY et al. 2005: 617). Allerdings führen bezüglich der entstehenden Kosten VEHOVAR et al. an, dass eventuell zu beschaffende, zusätzlich notwendige Hard- und Software hierbei ebenfalls einkalkuliert werden muss (VEHOVAR et al. 2008: 280). Auch kommt es gerade im Vergleich zwischen telefonischen Befragungen und online durchgeführten Befragungen zu einer Umverteilung der anfallenden Verbindungsgebühren: Während bei telefonischen Befragungen die Verbindungskosten vom durchführenden Institut getragen werden, fallen bei Online-Befragungen die Verbindungskosten beim Teilnehmer der Befragung an (PÖTSCHKE 2009: 77). Allerdings kann

mit einer weiterhin zu erwartenden, zunehmenden Verbreitung von Flatrate-Tarifen, bei denen ausschließlich ein monatlicher Pauschalbetrag für die Internetnutzung zu entrichten ist, von solchen Kosten der Teilnahme abstrahiert werden. In diesem Fall spielen dann die Opportunitätskosten der Teilnahme eine entscheidendere Rolle.

COUPER sowie COUPER und COUTTS sehen die niedrigeren Kosten von Online-Erhebungen hinsichtlich der Qualität der online durchgeführten Befragungen nicht nur positiv. Sie beschreiben, dass mit Online-Befragungen und den damit verbundenen, möglicherweise niedrigeren Kosten insgesamt auch eine Absenkung der Erhebungsqualität verbunden sein kann. So ermöglichen die niedrigen Kosten es beinahe jeder Person mit Internetzugang, eigene Befragungen über das Internet durchzuführen, weshalb nach COUPER auch von einer Demokratisierung der empirischen Sozialforschung gesprochen werden kann (COUPER 2000: 465; COUPER und COUTTS 2006: 217f.). Allerdings wird diesbezüglich auch die Gefahr gesehen, dass es zukünftig schwieriger werden wird, gute von schlecht erhobenen Online-Befragungen unterscheiden zu können (COUPER 2000: 464f.; FAAS 2003: 120; COUPER und COUTTS 2006: 218). Durch die Möglichkeit, dass nun jeder online eine Befragung durchführen kann, sehen COUPER und COUTTS die Gefahr, dass dadurch insgesamt die Qualität von online durchgeführten Befragungen leiden könnte, da nicht von jedem die Regeln der empirischen Sozialforschung berücksichtigt bzw. befolgt werden. Simultan zu den Ausführungen AKERLOFS zum „Market for Lemons“ (siehe hierzu AKERLOF 1970: 488ff.) ist daher hier theoretisch denkbar, dass möglicherweise langfristig valide und zuverlässig gestaltete Online-Befragungen zu Lasten schlechter verschwinden, da der Nutzer bzw. Betrachter der Befragung berücksichtigen muss, eine schlecht durchgeführte Befragung zu erwischen.

Auch bieten Online-Befragungen durch das Medium Internet hinsichtlich des Fragebogens deutlich mehr Gestaltungsmöglichkeiten als die konventionellen Befragungsmethoden. So ist es möglich, Multimediainhalte wie etwa Bilder, Videos oder Audio-Dateien in den Online-Fragebogen zu integrieren, wodurch sich ganz neue Möglichkeiten im Rahmen der Befragung von Personen ergeben (WELKER et al. 2005: 8f.; BANDILLA et al. 2009: 129). Darüber hinaus können diese Multimediainhalte auch dazu beitragen, die Verständlichkeit bestimmter Fragestellungen zu erhöhen (PÖTSCHKE 2009:78). Allerdings beschreiben ZERBACK et al., dass bisher solche Designelemente kaum eingesetzt werden und daher bisher eher eine theoretische Möglichkeit der Fragebogengestaltung darstellen (ZERBACK et al. 2009: 28).

Im Hinblick auf die formale Gestaltung des Fragebogens ergeben sich bei der Anwendung von Online-Fragebögen, gerade im Vergleich zur schriftlichen Befragung, ebenfalls neue Möglichkeiten. Neben einer Festlegung der Fragenreihenfolge, von welcher der Befragungs-

teilnehmer nicht eigenmächtig abweichen kann, können Online-Fragebögen ebenfalls so programmiert werden, dass bestimmte Fragen vom Teilnehmer pflichtmäßig ausgefüllt werden müssen, was je nach Fragestellung durchaus gewollt sein kann (EVANS und MATHUR 2005: 196f.). Auch bieten nach BANDILLA und BOSNJAK Online-Fragebögen durch das Ausblenden irrelevanter Fragestellungen sowie einer ausgeklügelten Filterführung die Möglichkeit einer einfacheren Handhabung. Dies gilt vor allem im Vergleich zu schriftlichen Befragungen, bei denen durchaus die Möglichkeit besteht, dass vom Befragten Filterfragen nicht ausreichend beachtet werden, wodurch es zu Fehlbeantwortungen des Fragebogens kommen kann (BANDILLA und BOSNJAK 2000: 168).

Ein weiterer Vorteil, vor allem im Vergleich zum F2F-Interview, wird im möglichen Verzicht eines Interviewers bei Durchführung einer Online-Befragung gesehen. So kann ähnlich zu einer schriftlichen Befragung der Fragebogen hier von den Befragten selbst am Computer ausgefüllt werden. Daher wird in der Literatur davon ausgegangen, dass hier Beeinflussungen durch den Interviewer und somit Verzerrungen des Interviews, beispielsweise durch soziale Erwünschtheit, vermieden werden können (DUFFY et al. 2005: 617; EVANS und MATHUR 2005: 206; SUE und RITTER 2007: 150f.; PÖTSCHKE 2009: 78). Den Ausführungen zur Anwesenheit eines Interviewers folgend, beschreiben EVANS und MATHUR allerdings mehrere Aspekte, die für die Durchführung von Interviewer-gestützten F2F-Interviews sprechen. Sie sehen Vorteile dahingehend, dass eine persönliche Interaktion zwischen Befragtem und Interviewer möglich wird, wodurch beispielsweise eventuelle Unklarheiten durch Erläuterungen und Rückfragen des Interviewers beseitigt werden können. Auch können im Rahmen eines F2F-Interviews konkrete physikalische Stimuli, wie etwa Beispielprodukte oder Geschmacksproben, durch den Interviewer präsentiert und vorgeführt werden. Zuletzt sehen EVANS und MATHUR einen Vorteil in der Anwesenheit eines Interviewers dahingehend, dass hierdurch eine Kontrolle über die Umwelt der Befragung vorgenommen werden kann. So können durch den Interviewer die Umweltbedingungen während eines Interviews relativ gut gesteuert werden (EVANS und MATHUR 2005: 196ff.). Denkbar ist hier, dass vom Interviewer auf der Straße Befragungsteilnehmer akquiriert werden, für die Durchführung des Interviews vom Interviewer jedoch ein ruhigerer Ort vorgeschlagen wird, um Ablenkungen, etwa durch Straßenlärm, während des Interviews zu vermeiden. Eine solche Kontrolle der Umweltbedingungen ist bei einer Online-Befragung hingegen kaum möglich. So kann hier im Nachhinein kaum nachvollzogen werden, welche Umweltbedingungen während der Befragungsteilnahme herrschten: Beispielsweise kann durch den Forscher im Nachhinein nicht nachvollzogen werden, ob einer Online-Befragung durch den Teilnehmer eine hohe kognitive Aufmerksamkeit ge-

schenkt wurde oder ob diese möglicherweise nur nebenbei, beispielsweise während der Arbeit, beantwortet wurde. Je nachdem, welche Umweltbedingungen vorliegen, sind daher bei einem Befragungsteilnehmer bereits aufgrund dieser unterschiedlichen Bedingungen voneinander abweichende Antworten denkbar.

### **Formen von Online-Befragungen**

Simultan zu den Ausführungen zur mündlichen Befragung können auch bei der Online-Befragung verschiedene Formen voneinander abgegrenzt werden. Aufbauend auf den Ausführungen von SCHNELL et al., TADDICKEN sowie WELKER et al. können danach anhand der Art der Durchführung grundsätzlich folgende fünf Formen von Internet-Befragungen unterschieden werden:

1. Der Befragte bekommt einen Fragebogen per E-Mail zugesendet, druckt diesen aus, und sendet diesen anschließend per Briefcouvert zurück.
2. Der Befragte bekommt einen Fragebogen per E-Mail zugesendet und sendet diesen beantwortet ebenfalls per E-Mail zurück.
3. Der Befragte lädt einen Fragebogen von einem Server herunter und sendet diesen nach Beantwortung per E-Mail zurück.
4. Der Befragte nimmt an Befragungen in Chats oder Newsgroups teil.
5. Der Befragte beantwortet einen auf einem Server abgelegten Fragebogen online (SCHNELL et al. 2005: 381; WELKER et al. 2005: 73; TADDICKEN 2008: 39).

Diese fünf Formen können weiter unterteilt werden in sogenannte E-Mail-Befragungen (Formen 1, 2 und 3) sowie Online- bzw. Web-Befragungen (Formen 4 und 5). Bei E-Mail-Befragungen handelt es sich es sich um eine einfach anzuwendende Form der Befragung, die ähnlich der schriftlichen Befragung ist und grundsätzlich als deren Weiterentwicklung angesehen werden kann (TUTEN 1997: 5; ROSENTHAL 1998: 133f.; SCHNELL et al. 2005: 381). Aufgrund verschiedener Schwächen wie möglicher Darstellungsprobleme bei den Befragten wurde sie jedoch von SCHAEFER und DILLMAN nur als Übergangslösung auf dem Weg zu den Online-Befragungen vom Typ fünf gesehen und hat in der empirischen Sozialforschung mittlerweile nur noch eine sehr geringe Bedeutung (SCHAEFER und DILLMAN 1998: 392).

Die Durchführung von Befragungen in Chatrooms oder in Newsgroups kann nach TADDICKEN als eine Sonderform von Online-Befragungen angesehen werden. Im Unterschied zu der in der vorliegenden Arbeit hauptsächlich thematisierten Online-Befragung vom Typ fünf wird diese nicht auf einem Server abgelegt. Stattdessen ist hier ein Interviewer in die Kommunikationssituation integriert und stellt die interessierenden Fragen im Rahmen des Chats oder der Newsgroup. Hieran lässt sich erkennen, welche vielfältigen Möglichkeiten

Online-Befragungen, neben der Integration multimedial aufbereiteter Inhalte, bieten. Möglicherweise stellen sie daher eine zukunftsweisende Form von Online-Befragungen dar (TADDICKEN 2008: 41).

Online- bzw. Web-Befragungen vom fünften Typus, bei denen die Befragten einen, auf einem Server abgelegten, Fragebogen online beantworten, stellen momentan im Rahmen der empirischen Sozialforschung die Standardform von Online-Befragungen dar. Der Fokus liegt daher in der vorliegenden Arbeit auch auf diesem Typus von Online-Befragungen. Gewöhnlich handelt es sich dabei um HTML-basierte Befragungen, die teilweise mit Hilfe von Softwareapplikationen wie Java oder Flash dynamische Elemente wie Animationen in die HTML-Befragungen integrieren. Neben solchen HTML-Befragungen können nach WELKER et al. zusätzlich sogenannte ASCII-Befragungen unterschieden werden, die in ihrer Programmierung Ähnlichkeiten zu E-Mail-basierten Befragungsformen aufweisen und daher ebenfalls Darstellungsprobleme aufweisen können. Möglicherweise werden sie auch deshalb heutzutage kaum mehr angewendet (WELKER et al. 2005: 74f.; TADDICKEN 2008: 40f.).

Weiterhin können innerhalb dieses Typus von Online-Befragungen nach COUPER und COUTTS verschiedene Arten anhand ihres beabsichtigten Zwecks sowie des Auswahlmodus der Befragungsteilnehmer voneinander abgegrenzt werden, wie auch Tabelle 3 aufzeigt.

**Tabelle 3: Arten von Online-Befragungen**

Nicht-zufallsgesteuerte Ansätze	Zufallsgesteuerte Ansätze
1. Online-Befragungen zu Unterhaltungszwecken	4. Intercept-Befragungen
2. Befragungen mit uneingeschränkt selbstrekrutierter Teilnehmerschaft	5. Befragungen mit listen-basierten Stichproben aus Populationen mit einem hohen Abdeckungsgrad
3. Befragungen mit selbstrekrutierten Freiwilligen-Panels	6. Web-Option bei Mixed-Mode-Befragungen
	7. Befragungen mit einem vorrekrutierten Panel von Internetnutzern
	8. Befragungen mit einem vorrekrutierten Panel aus der Gesamtbevölkerung

Quelle: Eigene Darstellung nach COUPER und COUTTS 2006: 228.

Unterschieden werden danach nicht-zufallsgesteuerte Ansätze von zufallsgesteuerten Ansätzen mit jeweils mehreren Unterkategorien. Während bei den zufallsgesteuerten Ansätzen für die Mitglieder der Zielpopulation eine bekannte, von Null verschiedene Auswahlwahrscheinlichkeit berechnet werden kann, ist dies bei den nicht-zufallsgesteuerten Ansätzen hingegen nicht möglich. Daher sind Schlussfolgerungen oder Verallgemeinerungen im Bezug auf die Population auf Grundlage von inferenzstatistischen Verfahren nicht möglich, da hier das wesentliche Kriterium der Repräsentativität nicht gegeben ist.

Bei der ersten Art „Online-Befragungen zu Unterhaltungszwecken“ handelt es sich, wie der Name bereits vermuten lässt, nicht um Befragungen im wissenschaftlichen Sinne. Bei Online-Befragungen dieses Typus wird gewöhnlich nicht der Anschein erweckt, dass es sich um wissenschaftliche Untersuchungen mit repräsentativen Stichproben handelt. Stattdessen wird der Unterhaltungscharakter solcher Online-Befragungen deutlich herausgestellt und betont.

Der zweite Typus „Befragungen mit uneingeschränkt selbstrekrutierter Teilnehmerschaft“ verwendet zur Rekrutierung der Befragungsteilnehmer häufig offene Einladungen auf häufig besuchten Internetseiten oder Internetportalen. Bei solchen Befragungen gibt es gewöhnlich keine Beschränkungen hinsichtlich der Teilnahme: Jeder Interessierte kann beliebig oft teilnehmen, woraus ersichtlich wird, dass solche Befragungen im Rahmen wissenschaftlicher Untersuchungen eher ungeeignet erscheinen. Ursächlich hierfür ist, dass sich nach COUPER und COUTTS die Teilnehmer solcher selbstrekrutierter Befragungen auf relevante Weise von der Allgemeinbevölkerung unterscheiden. So weichen die soziodemographischen Variablen der Befragten häufig von denen der Allgemeinbevölkerung ab, wobei die Richtung der Unterschiede vom jeweiligen Befragungsthema abhängig ist.

Beim dritten Typus „Befragungen mit selbstrekrutierten Freiwilligen-Panels“ wird im Unterschied zum zweiten Typus ein Panel aus freiwilligen Teilnehmern zusammengestellt, die ebenfalls über gut besuchte Webportale oder Internetseiten rekrutiert werden. Da grundlegende soziodemographische Variablen gespeichert werden, können die potentiellen Panel-Teilnehmer später gezielt ausgewählt werden. Da die Teilnahme an späteren Befragungen ausschließlich auf der Basis von Einladungen erfolgt und durch personalisierte Zugangsdaten kontrolliert wird, kann hierbei eine mehrfache Teilnahme ausgeschlossen werden. Ein bekanntes Beispiel für diesen Typus stellt in Deutschland die vor einigen Jahren durchgeführte „Perspektive Deutschland“ dar.

„Intercept-Befragungen“ richten sich grundsätzlich nur an Besucher einer bestimmten Internetseite. Durch ein systematisches Auswahlverfahren wird hierbei jeder n-te Besucher der Seite eingeladen, an der Befragung, beispielsweise zur Gestaltung der Internetseite, teilzunehmen, wodurch sich per Definition die Auswahlgesamtheit auf die Besucher der Internetseite beschränkt. Auch hier stellt sich das Problem der Mehrfachteilnahme: Je häufiger die Internetseite von einer Person besucht wird, desto größer ist ihre Wahrscheinlichkeit, zur Teilnahme eingeladen zu werden. Um das Problem der Mehrfachteilnahme zu umgehen, wer-

den häufig Cookies<sup>12</sup> eingesetzt, die eine solche Mehrfachteilnahme nicht erlauben. Allerdings kann bei diesem Typ nicht ausgeschlossen werden, dass ausschließlich solche Personen teilnehmen, die grundlegend andere Interessen haben als Personen, die sich gegen eine Teilnahme an der Befragung entschieden haben (COUPER und COUTTS 2006: 228ff.).

„Befragungen mit listen-basierten Stichproben aus Populationen mit einem hohen Abdeckungsgrad“ als fünfter Typus richten sich ausschließlich an Teilpopulationen in der Gesamtbevölkerung, welche im Unterschied zur Gesamtbevölkerung, bei der noch nicht von einer flächendeckenden Verbreitung des Internets ausgegangen werden kann, einen sehr hohen oder gar vollständigen Internet-Abdeckungsgrad aufweisen. Wichtig ist bei diesem Typus das Vorliegen eines vollständigen Verzeichnisses der E-Mail-Adressen der Gruppenangehörigen, aus dem die Befragungsteilnehmer ausgesucht werden können. Solche Verzeichnisse sind beispielsweise an Hochschulen mit eigenen E-Mail-Adressen für alle Studierenden sowie Unternehmen mit eigenen Unternehmens-E-Mail-Adressen für alle Mitarbeiter denkbar (EVANS und MATHUR 2005: 208; SUE und RITTER 2007: 149). Über diese Verzeichnisse werden die potentiellen Teilnehmer ausgewählt und mit einer E-Mail zur Teilnahme eingeladen. Auch hier wird der Zugang zur Befragung üblicherweise kontrolliert, um zu verhindern, dass der Fragebogen mehrfach ausgefüllt oder an dritte Personen weitergegeben wird. Zwar führen COUPER und COUTTS hinsichtlich der Ergebnisse solcher Befragungen an, dass aufgrund der Ausgestaltung der Teilnehmerauswahl deren Ergebnisse nur sehr begrenzt auf die Gesamtbevölkerung verallgemeinerbar sind, für die eingeschlossenen Gruppen allerdings interessante Ergebnisse liefern können (COUPER und COUTTS 2006: 228ff.).

Auf die sog. „Mixed-Mode-Befragungen“ wird an dieser Stelle nicht weiter eingegangen werden. Wichtig ist hierbei nur, dass bei derartigen Befragungen dem Befragten das Internet und somit die Online-Befragung als ein Antwortmodus unter vielen angeboten wird.

Die beiden letzten Typen von Online-Befragungen, die auf vorrekrutierten Panels von Internetnutzern bzw. Personen aus der Gesamtbevölkerung basieren, ähneln dem oben dargestellten nicht-zufallsgesteuerten dritten Typus. Der Hauptunterschied besteht allerdings darin, dass sich bei der nicht-zufallsgesteuerten Methode das Panel aus selbstrekrutierten Freiwilligen zusammensetzt, während im Rahmen der zufallsgesteuerten Ansätze die anvisierten Panelmitglieder mit Hilfe von zufallsgesteuerten Auswahlverfahren bestimmt werden. Beispiel-

---

<sup>12</sup> Bei Cookies handelt es sich um kleine Textbausteine, die auf Aufforderung des Web-Servers auf der Festplatte des Nutzers gespeichert werden. Durch Cookies ist es dem Web-Server möglich, Internet-User wiederzuerkennen (CHRISTL 2005: 7ff.).

haft kann hier genannt werden, dass im Rahmen von Telefonbefragungen die per Zufallsstichprobe ausgewählten Teilnehmer dahingehend befragt werden, ob sie zukünftig bereit seien, an Online-Befragungen teilzunehmen. Während bei dem siebten Typus „Befragungen mit einem vorrekrutierten Panel von Internetnutzern“ ausschließlich Personen mit Internetzugang identifiziert und zu den Online-Befragungen eingeladen werden, bezieht der achte Typus „Befragungen mit einem vorrekrutierten Panel aus der Gesamtbevölkerung“ auch solche Personen ein, die bisher nicht über einen Zugang zum Internet verfügten. Im Rahmen einer Teilnahme bekommen die Teilnehmer dafür die notwendige technische Ausstattung gestellt. Nur bei diesem Typus ist es daher möglich, im Rahmen von inferenzstatistischen Verfahren Aussagen über eine Population zu treffen, die über die Gruppe der Internetnutzer hinaus geht.

Nachdem die beiden Befragungsformen F2F-Interview sowie Online-Befragung nun ausführlich diskutiert wurden, werden nachfolgend in Tabelle 4 noch einmal die beiden Formen hinsichtlich der Kriterien finanzieller Aufwand, zeitlicher Aufwand sowie verfolgte Untersuchungsziele gegenübergestellt, bevor im nachfolgenden Gliederungspunkt auf die verschiedenen Fehlerarten der Befragung eingegangen werden wird.

**Tabelle 4: Zusammenfassender Vergleich der beiden Befragungsformen F2F-Interview und Online-Befragung hinsichtlich ausgewählter Kriterien**

	Online-Befragung	F2F-Interview
Finanzieller Aufwand	Sehr hohe Fixkosten, geringe variable Kosten	Hohe Fixkosten, sehr hohe variable Kosten
Zeitlicher Aufwand	Gering	Hoch
Untersuchungsziele	Quantitatives Vorgehen, multimediale Stimuli, keine Kontrolle der Umwelt	Quantitatives und qualitatives Vorgehen, reale Stimuli, sehr gute Kontrolle der Umwelt möglich

Quelle: Eigene Darstellung in Anlehnung an TADDICKEN 2008: 51.

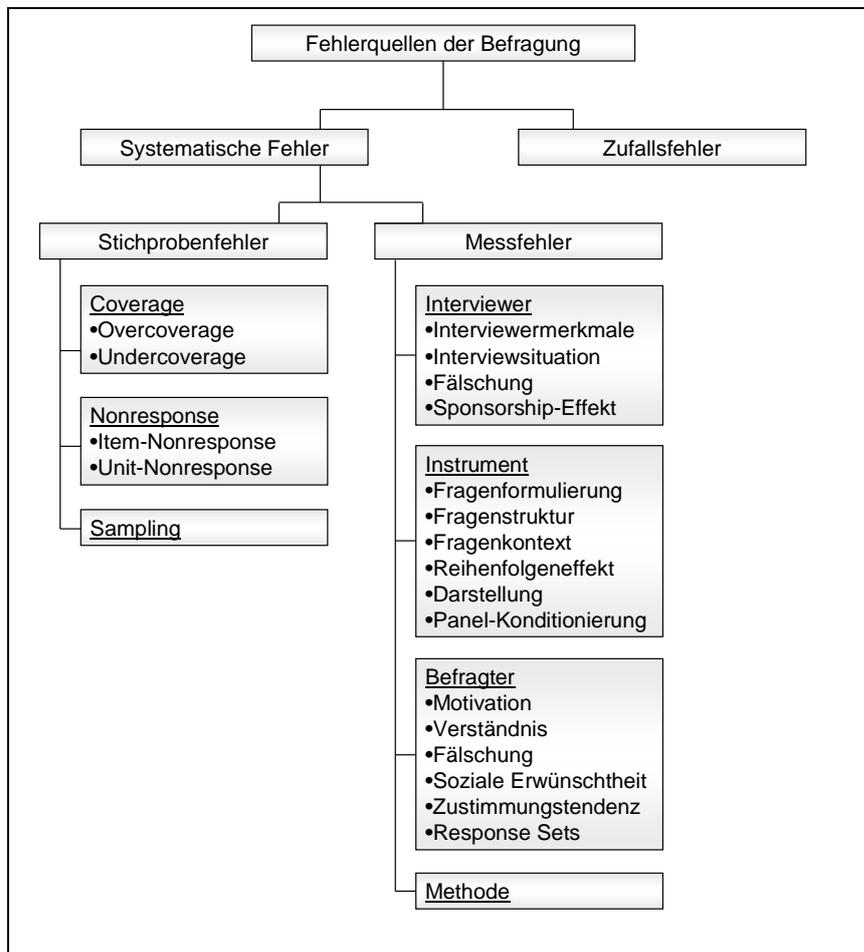
### 3.3 Fehlerarten in der Befragung

Die Güte von Daten, die im Rahmen von Online- bzw. mündlichen Befragungen erhoben werden, richtet sich grundsätzlich danach, wie nah die empirisch erhobenen Daten jeweils dem wahren Wert der zu messenden Variablen kommen. Hierbei kann es durch verschiedene Fehlerquellen zu deutlichen Beeinträchtigungen der Datengüte kommen.

Nach TADDICKEN ist dabei unter einem Fehler in der Umfrageforschung allgemein ein Störfaktor zu verstehen, der im Prozess der Datenerhebung auftritt und entsprechend für eine Verminderung der Datengüte sorgen kann (TADDICKEN 2008: 57). Ein Fehler im Rahmen der Befragung zeigt sich dabei nicht uniform, weshalb nachfolgend eine Morphologisierung des Fehlerbegriffs vorgenommen wird. Wie Abbildung 6 aufzeigt, kann eine Systematisierung des

Fehlerbegriffs damit begonnen werden, grundsätzlich zwischen einem Zufallsfehler und einem systematischen Fehler zu unterscheiden.

**Abbildung 6: Morphologie der Fehlerarten im Rahmen der Befragung**



Quelle: Eigene Darstellung auf der Grundlage von ATTESLANDER und KNEUBÜHLER 1975: 41; GROVES et al. 2004: 10; DIEKMANN 2008: 446ff.; TADDICKEN 2008: 58ff..

### 3.3.1 Zufällige und systematische Fehler

Eine derartige Unterteilung in zufällige und systematische Fehler wird von ATTESLANDER und KNEUBÜHLER in Anlehnung an COX (1961) vorgeschlagen (ATTESLANDER und KNEUBÜHLER 1975: 41f.). Auch BEREKOVEN et al. führen aus, dass bei der Messung statistischer Maße zwischen einem Zufallsfehler und einem systematischen Fehler unterschieden werden kann. Bei Zufallsfehlern handelt es sich danach um solche Fehler, die gleichmäßig um den wahren Wert einer Variablen streuen, so dass es per Saldo zu einem Ausgleich der Fehler kommt (BEREKOVEN et al. 2009: 57ff.). Zufallsfehler führen dennoch zu einer Beeinflussung der Reliabilität einer Messung (BAGOZZI 1984: 24). Die sogenannte Reliabilität bzw. die Zuverlässigkeit gibt dabei die Messgenauigkeit bzw. Präzision eines Instruments an. Die Reliabilität ist umso höher, je geringer der zu einem Messwert  $X$  gehörende Fehleranteil  $E$  ist. Eine perfekte Reliabilität würde demnach bedeuten, dass das Messinstrument in der Lage ist, den wahren

ren Wert  $T$  ohne jeglichen Messfehler  $E$  zu erfassen, so dass gilt:  $X=T$ . Somit kennzeichnet die Reliabilität den Grad der Genauigkeit, mit dem ein zu prüfendes Merkmal gemessen wird (BORTZ und DÖRING 2006: 196). Entsprechend ist die Reliabilität auch ein Maß für die Reproduzierbarkeit von Messergebnissen (DIEKMANN 2008: 250). Ein perfekt reliabler Test müsste nach wiederholter Anwendung bei denselben Personen zu den exakt gleichen Ergebnissen führen, gleichbleibende wahre Werte der zu messenden Variablen vorausgesetzt. Dem folgend könnte alternativ ausgesagt werden, dass eine perfekte Korrelation beider Messreihen vorliegt. Weichen die Ergebnisse wiederholter Tests hingegen voneinander ab bzw. sind sie unkorreliert, können neben zufälligen Einflüssen und somit zufälligen Fehlern auch systematische Verzerrungen und damit systematische Fehler dafür verantwortlich gemacht werden (BORTZ und DÖRING 2006: 196). BROSIUS et al. führen hierzu aus, dass durch zufällige Fehler die Genauigkeit der Ergebnisse beeinflusst wird, während es durch systematische Fehler zu einer Verzerrung der Ergebnisse insgesamt kommt (BROSIUS et al. 2008: 223).

Systematische Fehler verteilen sich im Unterschied zu den zufälligen Fehlern nicht gleichmäßig um den wahren Wert, sondern konzentrieren sich in eine Richtung. Zwar ist es nach BEREKOVEN et al. prinzipiell möglich, dass sich einzelne systematische Fehler in ihrer Wirkung gegenseitig aufheben, allerdings ist genauso gut eine Verstärkung denkbar und möglich (BEREKOVEN et al. 2009: 61ff.). Bei systematischen Fehlern kann weiterhin zwischen Stichproben- und Messfehlern unterschieden werden, wie TADDICKEN darstellt (TADDICKEN 2008: 58). Auf diese beiden voneinander abzugrenzenden Fehlerarten wird nachfolgend eingegangen.

### **3.3.2 Stichprobenfehler**

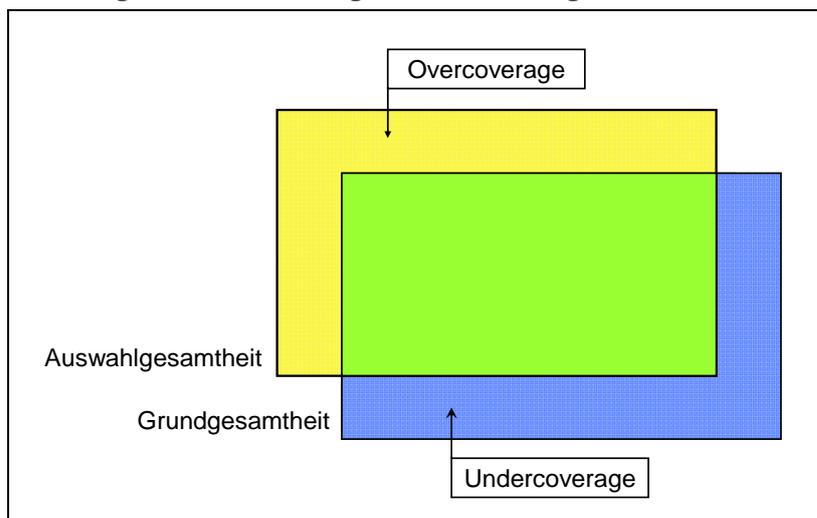
Entsprechend der Abbildung 6 können Stichprobenfehler grundsätzlich in Coverage-, Sampling- und Nonresponse-Fehler unterteilt werden. Diese stellen nicht nur bei mündlichen Befragungen wie F2F-Interviews, sondern auch bei Online-Befragungen eine bedeutsame Fehlerquelle dar.

#### **Coverage-Fehler**

Der sogenannte Coverage-Fehler tritt auf, wenn die zugrunde liegende Auswahlgesamtheit, aus der die Stichprobe gezogen wird, nicht mit der Grundgesamtheit identisch ist und somit Unterschiede zwischen der Auswahlgesamtheit und der Grundgesamtheit vorliegen (COUPER 2000: 467; DIEKMANN 2008: 417; TADDICKEN 2008: 59). Beispielhaft für das Auftreten von Coverage-Fehlern kann nach DIEKMANN und TADDICKEN eine bevölkerungsrepräsentative, telefonische Befragung herangezogen werden, die auf der Grundlage von Telefonbüchern bzw.

nationalen Telefonverzeichnissen durchgeführt wird. Zwar kann nach BEREKOVEN et al. inzwischen von einer flächendeckenden Verbreitung von Telefonanschlüssen in Deutschland ausgegangen werden (BEREKOVEN et al. 2009: 102), jedoch können durch die dargelegte Vorgehensweise nur Personen bzw. Haushalte erreicht werden, deren Telefonanschluss in einem Telefonbuch verzeichnet ist bzw. die trotz des hohen Verbreitungsgrades überhaupt über einen Telefonanschluss verfügen. Dem folgend, werden alle Personen bzw. Haushalte ohne Eintrag ins Telefonbuch bzw. ohne Telefonanschluss von einer Teilnahme an der Befragung ausgeschlossen. Hieran zeigt sich, dass die Auswahlgesamtheit, in Form der im Telefonbuch verzeichneten Personen bzw. Haushalte, sowie die Grundgesamtheit (alle Bürger in Deutschland) nicht deckungsgleich sind (DIEKMANN 2008: 417; TADDICKEN 2008: 59). Um im Rahmen inferenzstatistischer Verfahren jedoch Aussagen aus der Stichprobe über die Grundgesamtheit ableiten zu können, hätte im Rahmen einer Zufallsauswahl jedes Element der Grundgesamtheit eine berechenbare, von Null verschiedene Wahrscheinlichkeit haben müssen, in die Stichprobe gelangen zu können (LÖFFLER 1999: 17; SCHNELL et al. 2005: 272ff.). Dies ist hier jedoch durch die Differenz von Auswahl- und Grundgesamtheit nicht gegeben.

**Abbildung 7: Undercoverage und Overcoverage**



Quelle: Eigene Darstellung in Anlehnung an SCHNELL et al. 2005: 272.

Beim Coverage-Fehler können weiterhin zwei verschiedene Arten unterschieden werden: Das sogenannte Undercoverage sowie das Overcoverage. Beide werden in Abbildung 7 zur Vereinfachung graphisch dargestellt. Entsprechend zu den oberen Ausführungen bezüglich der Telefonbefragung besteht Undercoverage, wenn von der Auswahlgesamtheit (gelbes Rechteck in der Abbildung) nicht alle Elemente der Grundgesamtheit (blaues Rechteck) erfasst werden. Undercoverage wird dementsprechend in der Abbildung durch den blauen, überstehenden Bereich dargestellt. Overcoverage tritt hingegen auf, wenn Elemente zur Auswahlgesamtheit, al-

lerdings nicht zur Grundgesamtheit gehören. Overcoverage wird in der Abbildung entsprechend durch den gelben überstehenden Bereich repräsentiert.

Neben anderen, beispielsweise telefonischen oder mündlichen Befragungsmethoden sind nach BAUR und FLORIAN, COUPER und COUTTS sowie DIEKMANN online durchgeführte Befragungen besonders von solchen Coverage-Fehlern betroffen (BAUR und FLORIAN 2009: 111ff.; COUPER und COUTTS 2006: 219; DIEKMANN 2008: 417). Beispielsweise beschreiben COUPER und COUTTS, dass die meisten Fragestellungen, die im Rahmen von Online-Erhebungen beantwortet werden sollen, sich auf eine Grundgesamtheit beziehen, die über die Population der Internetnutzer hinausgeht (COUPER und COUTTS 2006: 219). Dabei ist es aufgrund verschiedener Sachverhalte häufig jedoch nicht möglich, von der Auswahlgesamtheit in Form der Internetnutzer auf die Grundgesamtheit in Form der Gesamtbevölkerung zu schließen: So wird in der Literatur zur Evaluierung von Online-Erhebungsverfahren mehrfach beschrieben, dass davon ausgegangen werden muss, dass die Verteilung soziodemographischer Merkmale „typischer“ Internetnutzer nicht bevölkerungsrepräsentativen Verteilungen folgt. So handelt es sich bei den „typischen“ Internetnutzern tendenziell eher um Männer, jüngere Personen mit einem höheren Bildungsgrad sowie einem höheren Haushaltseinkommen. Diese soziodemographischen Merkmale sind somit im Vergleich zur Gesamtbevölkerung in der Internetnutzer-Population und entsprechend der Auswahlgesamtheit überrepräsentiert (BANDILLA et al. 2001: 11; BERRENS et al. 2003: 4; FAAS 2003: 125ff.; COUPER und COUTTS 2006: 221f.; SCHOEN 2004: 31; DUFFY et al. 2005: 619; FALLOWS 2005: 2ff.; WELKER et al. 2005: 70; AGOF 2009: 7ff.; BAUR und FLORIAN 2009: 112f.; INITIATIVE D21 2009a: 10ff.). Werden daher Online-Befragungen eingesetzt, um mit Hilfe inferenzstatistischer Verfahren auf Variablenverteilungen in der Grundgesamtheit schließen zu können, werden in der Auswahlgesamtheit tendenziell Frauen, ältere Menschen, Personen mit einem niedrigen Bildungsgrad und Haushaltseinkommen nicht genügend berücksichtigt. Es kann somit zum Undercoverage kommen. Ebenfalls relevant ist hierbei der Sachverhalt, dass auch im Jahre 2009 in Deutschland noch nicht alle Bundesbürger über einen Zugang zum Internet verfügen bzw. das Internet nutzen, wodurch ebenfalls Auswahl- und Grundgesamtheit nicht deckungsgleich sind. So kann laut der Arbeitsgemeinschaft Online Forschung e.V. (AGOF) davon ausgegangen werden, dass im Jahre 2009 lediglich 65,1 % der Bundesbürger über 14 Jahren das Internet nutzen (AGOF 2009: 7ff.). Auf etwas höhere Nutzungszahlen kommt die Studie der INITIATIVE D21, einem gemeinnützigen Verein zur Förderung von praxisnahen non-profit-Projekten aus dem Bereich der Informations- und Kommunikationstechnologien (INITIATIVE D21 2009b). Hier konnte ermittelt werden, dass knapp 70 % der deutschen Bevölkerung über 14 Jahre das Internet nut-

zen. Allerdings zeigen sich zwischen den Bundesländern starke Unterschiede in der Internetnutzung. So wird das Internet am häufigsten von Bewohnern der Stadtstaaten Bremen und Berlin genutzt. 74,2 % bzw. 73,3 % der Bevölkerung nutzen hier jeweils das Internet. Am geringsten ist diese Zahl in den Bundesländern Mecklenburg-Vorpommern und Sachsen-Anhalt mit 61,9 % bzw. 60,7 % (INITIATIVE D21 2009a: 10ff.).

Auch existieren im Unterschied beispielsweise zu den Registern der Einwohnermeldeämter oder zu Verzeichnissen aller Telefonanschlüsse in Deutschland keine Verzeichnisse über die Nutzer des Internets (COUPER und COUTTS 2006: 219f.), was TADDICKEN unter dem „Problem der symbolischen Verfügbarkeit“ zusammenfasst (TADDICKEN 2008: 61). So besteht neben einem mangelhaften Coverage der Zielpopulation ein weiteres bedeutsames Problem darin, die Auswahlgesamtheit überhaupt erst identifizieren zu können, um anschließend daraus eine Zufallsstichprobe ziehen zu können (COUPER und COUTTS 2006: 219). Ausnahmen bilden hierbei die in Kapitel 3.2 aufgezeigten Arten von Online-Befragungen, die auf eingrenzbaaren Grundgesamtheiten wie Studierende einer Universität oder auf einem vorrekrutierten Teilnehmerpool basieren.

Eine Berechnung bzw. Quantifizierung des Fehlers durch Undercoverage ist nach folgender, von GROVES et al. gegebener Formel möglich:

$$\mu_C - \mu = \frac{U}{N}(\mu_C - \mu_U) \quad (1)$$

$\mu$  steht dabei für den Mittelwert einer unbestimmten Variablen in der Grundgesamtheit mit  $N$  Merkmalsträgern,  $\mu_C$  und  $\mu_U$  sind die Mittelwerte der Variablen in der Auswahlgesamtheit (covered) sowie in der Undercoverage-Menge, also außerhalb der Auswahlgesamtheit.  $U$  steht für die Anzahl der Merkmalsträger in der Undercoverage-Menge (GROVES et al. 2004: 83f.). Der Coveragefehler, der sich als Differenz zwischen dem Mittelwert der Auswahlgesamtheit und dem der Grundgesamtheit darstellt, berechnet sich entsprechend aus dem Undercoverage-Anteil in der Befragung, multipliziert mit der Mittelwertdifferenz zwischen Auswahlgesamtheit und Elementen der Undercoverage-Gruppe (DIEKMANN 2008: 417).

Neben dem Problem des Undercoverage kann bei Online-Befragungen auch Overcoverage auftreten. BAUR und FLORIAN geben verschiedene Ursachen an, wie es dazu kommen kann, dass Merkmalsträger zur Auswahlgesamtheit, allerdings nicht zur Grundgesamtheit gehören. So ist es beispielsweise denkbar, dass Befragte an einer Online-Befragung mehrfach teilnehmen, um ihrer Meinung ein stärkeres Gewicht zu geben, obwohl dies nicht vorgesehen ist. Eine weitere mögliche Ursache für Overcoverage kann auch sein, dass Personen an einer Online-Befragung teilnehmen, die nicht zur definierten Grundgesamtheit gehören. So ist etwa

denkbar, dass ausgewählte Befragte eine Einladung zur Teilnahme an einer Online-Befragung erhalten, diese aber an Dritte weitergeben. Begegnet werden kann dieser Problematik aber über den Einbau spezieller Filterfragen in den Fragebogen, mit dem Teilnehmer außerhalb der Grundgesamtheit identifiziert und entsprechend herausgefiltert werden können (BAUR und FLORIAN 2009: 117f.).

### **Sampling-Fehler**

Die Stichproben-Ziehung stellt nach GROVES et al. einen entscheidenden Punkt im Forschungsprozess von Befragungen dar, da hier wesentliche Fehler begangen werden können. So kann es durch die Auswahl der Stichprobenelemente zu Verzerrungen in den Variablen kommen (GROVES et al. 2004: 93f.). Wird eine Stichprobe gezogen, ist nach TADDICKEN davon auszugehen, dass die Werte der Stichprobe nicht exakt mit den Werten der Grundgesamtheit übereinstimmen, sondern von denen abweichen. Diese Abweichungen bzw. Ungenauigkeiten werden als Sampling-Fehler bezeichnet (TADDICKEN 2008: 62) und sind unter anderem abhängig von der Stichprobengröße, der Größe der Grundgesamtheit sowie von der Art der Stichprobenziehung (BORTZ 2005: 86).

### **Nonresponse-Fehler**

Unter dem Begriff „Nonresponse“ wird allgemein der Ausfall von Befragungspersonen zusammengefasst. Unterschieden werden kann dabei zwischen dem sogenannten Item-Nonresponse, bei dem es zu einem Ausfall der Befragten bei einzelnen Fragen kommt, sowie dem Unit-Nonresponse, bei dem es zu einem Totalausfall des Befragten, beispielsweise durch Verweigerung, kommt (PÖTSCHKE 2009: 80). TADDICKEN führt diesbezüglich an, dass prinzipiell bei jeder Befragung davon ausgegangen werden muss, dass nicht alle in der Stichprobe enthaltenen Probanden den Fragebogen vollständig ausfüllen bzw. überhaupt an der Befragung teilnehmen. Mögliche Gründe hierfür sind beispielsweise die Abwesenheit der Befragten aufgrund von Krankheit oder Urlaub, Zeitdruck, fehlende Motivation oder die Antwortverweigerung bei heiklen Fragen, beispielsweise zum Einkommen, zur politischen Einstellung oder zur Sexualität (TADDICKEN 2008: 63; BAUR und FLORIAN 2009: 123).

Mit zunehmendem Unit-Nonresponse, also dem Totalausfall von Befragungspersonen, einher geht das Absinken der Ausschöpfungsquote als Quotient aus der Anzahl der durchgeführten Interviews und der geplanten Stichprobengröße (DIEKMANN 2008: 418). Geringe Ausschöpfungsquoten stellen nach SCHNEEKLOTH und LEVEN sowie PÖTSCHKE dabei an sich kein Problem dar. Häufig gehen mit dem Ausfall von Personen jedoch systematische Verzerrungen der statistischen Parameter wie Mittelwerten, Anteilen oder Intervallen einher (SCHNEEKLOTH und LEVEN 2003: 50f.; PÖTSCHKE 2009: 80f.). Entsprechend muss die Stichprobe ausreichend

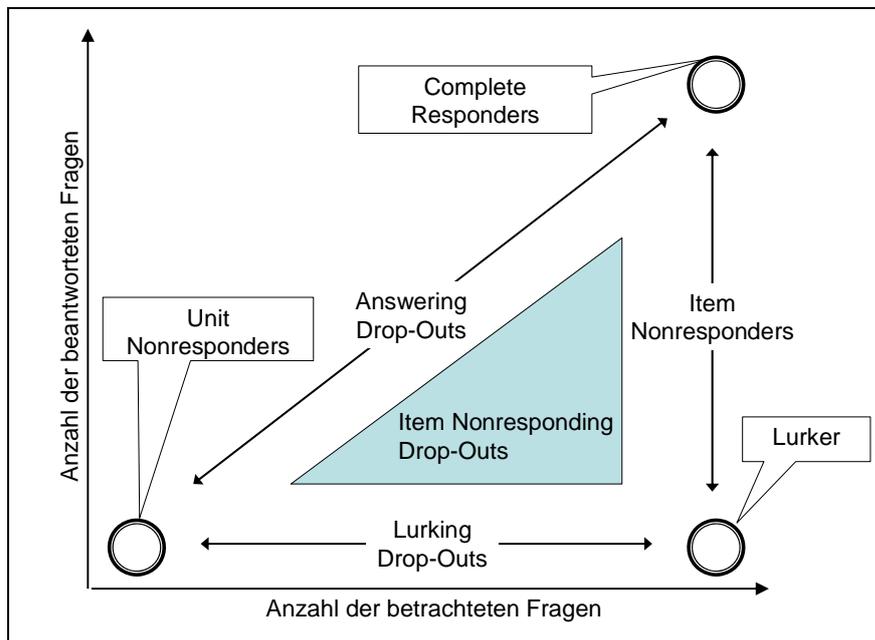
groß sein, um solche Verzerrungen durch Nonresponse ausgleichen zu können (TADDICKEN 2008: 63). Der Nonresponse-Fehler ist dabei jedoch nicht nur von der Teilnehmerquote, sondern auch von der Größe des Unterschieds zwischen Teilnehmern und Nichtteilnehmern abhängig (COUPER und COUTTS 2006: 223). Zusätzlich muss nach BAUR und FLORIAN bei fehlenden Werten zwischen realen fehlenden Werten, die beispielsweise durch Nichtzutreffen, Nicht-Wissen oder Meinungslosigkeit der Befragten zustande kommen, sowie der Weigerung, auf heikle, unsinnige oder uninteressante Fragen zu antworten, unterschieden werden. Während die realen, fehlenden Werte aus der Perspektive der Stichproben- und Datenqualität durchaus wünschenswert sind, um daraus resultierende Variablenverzerrungen zu vermeiden, können Antwortverweigerungen hingegen stark verzerrend sein (BAUR und FLORIAN 2009: 122f.). Um allerdings Aussagen hinsichtlich der Teilnehmerquote treffen zu können, müsste auch hier die Größe der Auswahlgesamtheit bekannt sein, aus der die Stichprobe gezogen wird. Während dies bei telefonischen oder auch schriftlichen Befragungen durchaus möglich ist, können bei Online-Befragungen hierzu keine bzw. kaum Aussagen getroffen werden. Ausnahme sind hierbei wiederum die Arten von Online-Befragungen, die auf Listen oder vorrekrutierten Panels basieren. Denn nur wenn die Auswahlgesamtheit sowie die Auswahlwahrscheinlichkeit bekannt sind, kann der Nonresponse-Fehler nachvollziehbar gemessen und bewertet werden (COUPER und COUTTS 2006: 223).

Im Rahmen von Online-Befragungen können neben dem Unit- und Item-Nonresponse durch zusätzliche, im Hintergrund erhobene Daten detailliertere Formen des Nonresponse unterschieden werden. Möglich wird dies beispielsweise durch sogenannte Logfiles, anhand denen BOSNJAK et al. sieben Response- bzw. Nonresponse-Typen identifizieren. Unterschieden werden die verschiedenen Typen, wie auch Abbildung 8 darstellt, einmal anhand der Anzahl der betrachteten Fragen des Online-Fragebogens sowie der Anzahl der beantworteten Fragestellungen. BOSNJAK führt hinsichtlich dieser detaillierteren Unterteilung an, dass diese dem tatsächlichen Geschehen bei Online-Befragungen eher gerecht wird als die einfache Unterteilung in Item- und Unit-Nonresponders (BOSNJAK 2003: 59).

Complete Responders sind entsprechend der Abbildung Befragte, die alle Fragen des Fragebogens ansehen sowie beantworten. Unit-Nonresponders als Gegenteil zu den Complete Responders sind simultan zu den oberen Ausführungen Personen, die nicht an der Befragung teilnehmen und entsprechend einen Totalausfall darstellen. Ursächlich für Unit-Nonresponse kann, wie oben bereits dargestellt wurde, einmal sein, dass die Personen eine Teilnahme an der Befragung verweigern. Zusätzlich ist nach BOSNJAK et al. denkbar, dass die potentiellen Befragten aus verschiedenen Gründen von einer Teilnahme abgehalten werden. Answering

Drop-Outs bezeichnen Personen, die alle angesehenen Fragen beantwortet haben, die Befragung jedoch vorzeitig abbrechen.

Abbildung 8: (Non-)Response-Typen in Online-Befragungen



Quelle: Eigene Darstellung nach BOSNIAK et al. 2001: 11.

Die sogenannten Lurker hingegen haben sich alle Fragen der Befragung betrachtet, allerdings keine davon beantwortet. Lurking Drop-Outs weisen Ähnlichkeiten zu den Answering Drop-Outs sowie den Lurkern auf. Solche Teilnehmer schauen sich einen Teil der Fragen an, ohne diese zu beantworten und brechen die Teilnahme an der Befragung vorzeitig ab. Nach BOSNIAK et al. weisen Lurker und Lurking Drop-Outs zumindest ein gewisses Interesse an der Thematik oder der Gestaltung des Fragebogens auf, da sonst von einem sofortigen Abbruch auszugehen sei. Die Item-Nonresponders betrachten alle Fragen des Fragebogens, beantworten selektiv allerdings nur einen Teil der Fragen.

Als letzten Response- bzw. Nonresponse-Typ können nach BOSNIAK et al. die Item-Nonresponding Drop-Outs als einen Mix aus Answering Drop-Outs und Item-Nonresponders abgegrenzt werden. Individuen dieses Typs betrachten nur einen Teil der Fragen des Fragebogens und beantworten von diesen Fragen wiederum nur einen Teil. Zusätzlich bricht dieser Typus ebenfalls vor Erreichen des Fragebogenendes die Befragung ab. Die Autoren gehen bei diesem Typus nicht von mangelnder Motivation aus, sondern dass sie sich durch eine oder mehrere Fragen empfindlich gestört gefühlt haben, und daher vorzeitig die Teilnahme an der Online-Befragung beendet haben (BOSNIAK et al. 2001: 11ff.).

Neben den oben genannten, von Seiten des Befragten stammenden möglichen Ursachen für Nonresponse wie die Abwesenheit zum Befragungstermin, Zeitdruck, fehlende Motivation

oder Antwortverweigerungen, können für Online-Befragungen zusätzlich spezielle Ursachen für Nonresponse abgegrenzt werden. So kann die Nutzung von Multimediaelementen, die prinzipiell als ein Vorteil von Online-Befragung angesehen wird, für höhere Nonresponse-Raten sorgen. Ursächlich hierfür ist, dass für eine korrekte Präsentation solcher Fragebögen häufig aktuelle Versionen von Internetbrowsern oder Erweiterungen wie Java-Applets notwendig sind, die einen Teil der Zielpopulation davon abhalten, an der Befragung teilzunehmen (SCHWARZ und REIPS 2001: 75ff.; COUPER und COUTTS 2006: 224). Auch rechenintensive Graphiken (LOZAR MANFREDA et al. 2002) sowie unzuverlässige, langsame Internetverbindungen können zu einer Erhöhung des Nonresponse beitragen (COUPER und COUTTS 2006: 224).

Verschiedene Studien haben sich mit der Größe des Nonresponse-Problems bei Online-Befragungen auseinandergesetzt. In diesen Studien wird dabei gewöhnlich die Rücklauf- bzw. Teilnahmequote bei Online-Befragungen mit der Rücklaufquote einer parallel durchgeführten alternativen Befragungsmethode verglichen. Da bei F2F-Interviews keine Aussagen über die Rücklaufquote gegeben werden können, werden in diesen Studien hauptsächlich auf dem Postweg versendete schriftliche Befragungen als Vergleichsgröße herangezogen. In einem Großteil der Studien kann dabei für Online-Befragungen nur eine niedrigere Rücklauf- bzw. Teilnahmequote im Vergleich zu den alternativen Befragungsformen ermittelt werden (SCHULDT und TOTTEN 1994: 37f.; WEIBLE und WALLACE 1998: 23; COUPER et al. 1999: 46f.; KWAK und RADLER 2002: 262f.; REIPS und FRANEK 2004: 74f.; BECH and KRISTENSEN 2009: 3). HAYSLETT und WILDEMUTH haben in ihrer Untersuchung ebenfalls eine geringere Teilnahmequote bei Online-Befragungen im Vergleich zur parallel durchgeführten postalischen Befragung ermittelt. Zusätzlich konnten sie nachweisen, dass die Versendungsart des Einladungsschreibens auch Einfluss auf die Teilnahmequote ausübt. So war die Teilnahmequote bei der Online-Befragung nochmals niedriger, wenn statt mit einer E-Mail mit einem Brief zur Teilnahme an der Befragung eingeladen wurde (HAYSLETT und WILDEMUTH 2004: 85). In anderen Studien hingegen konnten bei Online-Befragungen eine höhere Teilnahmequote (COBANOGLU et al. 2001: 449; KIERNAN et al. 2005: 249) bzw. eine etwa gleich hohe Rücklauf- bzw. Teilnahmequote (KAPLOWITZ et al. 2004: 97ff.; SHIH und FAN 2007: 64) ermittelt werden.

### **3.3.3 Messfehler**

Während in früheren Ausführungen zu der Qualität von Befragungen hauptsächlich die oben dargestellten Stichprobenfehler thematisiert wurden, rücken zunehmend auch die sogenannten Messfehler in den Fokus der Betrachtungen zur Qualität von Befragungen (COUPER und

COUTTS 2006: 227). Die Messung spielt dabei in der empirischen Sozialforschung eine bedeutsame Rolle. Sie dient nach SCHNELL et al. zur Beschreibung sozialer Sachverhalte sowie zur Überprüfung sozialwissenschaftlicher Theorien, um daraus wiederum soziale Sachverhalte erklären zu können. Eine Erfüllung beider Aufgaben ist nur mit Hilfe von Messungen möglich (SCHNELL et al. 2005: 138). Die Messung in der empirischen Sozialforschung entspricht dabei grundsätzlich der naturwissenschaftlichen Messung (TADDICKEN 2008: 66). Unter dem Begriff der Messung kann dabei die Zuordnung von Zahlen zu bestimmten Objekten verstanden werden (STEVENS 1959: 19). Die Zielsetzung ist letztlich die Erhebung exakter und möglichst fehlerfreier Daten (TADDICKEN 2008: 66).

Nach LORD und NOVICK setzt sich eine, mit einer Methode 1 gemessene Variable  $X_1$  aus dem wahren Wert  $T_1$  („true score“) sowie dem Messfehler  $E_1$  („Error-Variable“) zusammen (LORD und NOVICK 1968: 34; DIEKMANN 2008: 261f.), wie Gleichung (2) aufzeigt:

$$X_1 = T_1 + E_1 \quad (2)$$

Theoretisch ist es daher möglich, den Messfehler zu bestimmen, wenn der wahre Wert vom entsprechenden Messwert abgezogen wird. Da der wahre Wert in aller Regel jedoch unbekannt ist, kann eine genaue Ermittlung des Messfehlers nur über die Durchführung von Antwortvergleichen erzielt werden (KÖLTRINGER 1997: 23f.). Die Ursachen für Messfehler können dabei einmal bei dem Interviewer, dem Instrument, dem Befragten oder auch der angewendeten Befragungsmethode liegen (KOOLWIJK 1969: 870; SINGER et al. 1983: 68ff.; ESSER 1986: 316ff.; JOINSON 2001: 587ff.; COUPER und COUTTS 2006: 227ff.; DIEKMANN 2008: 446ff.; TADDICKEN 2008: 67ff.; BEREKOVEN et al. 2009: 61f.; TADDICKEN 2009: 87ff.).

### **Messfehler durch den Interviewer**

Messfehler, die durch die Anwesenheit eines Interviewers bzw. dessen Interviewereinfluss entstehen, können nach TADDICKEN auch als Interviewereffekte bezeichnet werden. Unterschieden werden können dabei grundsätzlich zwei verschiedene Typen dieser Interviewereffekte: Einmal interaktionale Interviewereffekte sowie nicht-interaktionale Interviewereffekte (TADDICKEN 2008: 70). Bei den interaktionalen Interviewereffekten entstehen Messfehler bzw. Verzerrungen durch die Interaktion zwischen Interviewer und Befragtem. Mögliche Ursachen hierfür sind eine Beeinflussung des Befragten durch äußere Merkmale des Interviewers wie Geschlecht, Alter oder Kleidung sowie durch sein Auftreten gegenüber dem Befragten. Hierzu gehört auch das suggestive Vorbringen der Fragen durch den Interviewer, wodurch eine bestimmte Antwort provoziert werden kann. Auch die Kenntnis des Auftraggebers einer Studie kann zu systematischen Antwortverzerrungen führen, was unter dem Begriff des Sponsorship-Effekts in der Literatur beschrieben wird. So ist es denkbar, dass Befragte bei

Kenntnis des Auftraggebers anders antworten, als sie es ansonsten getan hätten (SINGER et al. 1983: 68ff.; DIEKMANN 2008: 466ff.; TADDICKEN 2008: 70; BEREKOVEN et al. 2009: 62).

Messfehler bzw. systematische Verzerrungen der Antworten können auch durch den Interviewer selbst entstehen. In diesem Fall wird von nicht-interaktionalen Interviewereffekten gesprochen (TADDICKEN 2008: 70). Mögliche Ursachen sind hierbei, dass beispielsweise durch den Interviewer Auswahlpläne nicht eingehalten werden oder auch Interviews bewusst gefälscht werden. Auch können eigenmächtige Änderungen der Fragenformulierungen durch den Interviewer zu Verzerrungen der Ergebnisse führen (DIEKMANN 2008: 466f.; TADDICKEN 2008: 70; BEREKOVEN et al. 2009: 62).

Auch durch die Interviewsituation als solches kann es zu Messfehlern kommen. So können nach DIEKMANN die Situation, in der das Interview zeitlich und räumlich stattfindet sowie die vorherrschende emotionale Stimmung einen bedeutsamen Einfluss auf Antworten der Befragten ausüben. Ein weiterer bedeutsamer Faktor der Interviewsituation kann bei F2F-Interviews die Anwesenheit Dritter sein. Hierdurch kann es ebenfalls zu systematischen Antwortverzerrungen kommen (DIEKMANN 2008: 468).

Entsprechend den Ausführungen können sowohl in F2F-Interviews als auch in telefonischen Befragungen Messfehler durch die Anwesenheit eines Interviewer auftreten. Nur bei postalischen bzw. schriftlichen Befragungen sowie bei Online-Befragungen können durch ein Fehlen des Interviewers solche Interviewereinflüsse ausgeblendet werden (DIEKMANN 2008: 466; TADDICKEN 2008: 70). Allerdings führt DIEKMANN diesbezüglich an, dass die Anwesenheit von Interviewern auch zu einer Vermeidung verzerrter Antwortreaktionen beitragen kann: So können bei bestimmten Fragen Verzerrungen durch Verständnisprobleme seitens des Befragten vermieden werden (DIEKMANN 2008: 466).

### **Messfehler durch das Instrument**

Auch durch das verwendete Instrument kann es zu systematischen Verzerrungen bzw. Messfehlern kommen, im Falle von Befragungen also durch den verwendeten Fragebogen (TADDICKEN 2008: 71). Ausgelöst werden können derartige Messfehler beispielsweise durch die Formulierung der Fragen, die Auswahl der vorgegebenen Antwortkategorien, die Fragenreihenfolge sowie den Kontext, in den die Fragestellungen eingebettet sind (SUDMAN und BRADBURN 1982: 264ff.; SCHNELL et al. 2005: 353f.; DIEKMANN 2008: 458ff.). Ebenfalls spielen nach BEREKOVEN et al. Effekte durch die Reihenfolge der Antwortmöglichkeiten bzw. der Items einer Fragenbatterie eine wichtige Rolle (BEREKOVEN et al. 2009: 93). So stehen in-

nerhalb eines Fragebogens Fragen immer im Kontext mit den anderen Fragestellungen. Hierdurch kann es zu einer Beeinflussung der Beantwortung von Fragen kommen. Im Rahmen des sogenannten „Halo-Effektes“ bzw. „Ausstrahlungseffektes“ kann sich so die Beantwortung der Folgefragen entweder an den vorhergehenden Fragestellungen oder auch an bereits gegebenen Antworten orientieren. Das Auftreten dieses Effektes kann dabei auf zwei Ursachen zurückgeführt werden: Einmal werden Fragen bzw. Items durch andere Fragen bzw. andere Items einer Itematterie in einen Sinnzusammenhang gestellt, auf den der Befragte reagiert. (ATTESLANDER und KNEUBÜHLER 1975: 54f.; SCHNELL et al. 2005: 342; DIEKMANN 2008: 464f.). Weiterhin konnten SCHWARZ et al. sowie FUCHS in experimentellen Untersuchungen jeweils herausfinden, dass Items in Fragebatterien bzw. Antwortmöglichkeiten signifikant häufiger ausgewählt werden, wenn sie zu Beginn einer Batterie oder Liste stehen. Sie sprechen sich daher jeweils für eine Randomisierung aus, um dadurch hervorgerufene Verzerrungen in den Antworten zu vermeiden (SCHWARZ et al. 1991: 203ff.; FUCHS 2003: 29ff.). Die zweite Ursache für das Auftreten des „Halo-Effektes“ kann das mit der ersten Ursache verknüpfte Bemühen des Befragten sein, seine Antworten möglichst konsequent zu halten (SCHNELL et al. 2005: 342).

Neben den bereits in Kapitel 3.2 diskutierten Vorteilen einer einfachen und effektiven Filterführung sowie der Vorgabe eines pflichtmäßigen Beantwortens von Fragestellungen bieten Online-Befragungen häufig weiterhin den Vorteil, dass prinzipiell eine automatische Randomisierung innerhalb des Fragebogens bzw. innerhalb der Fragestellungen vorgenommen werden kann, wodurch mögliche Reihenfolgeeffekte kontrolliert werden können (PEPELS 2008: 194).

Bei der Darstellung der verschiedenen Arten von Online-Befragungen in Kapitel 3.2 wurde aufgezeigt, dass von einem Teil der verschiedenen Online-Befragungen auf ein vorrekrutiertes Panel zurückgegriffen wird. Hierbei kann es zu einem weiteren Messfehler kommen. So sind Panel- oder Längsschnitt-Befragungen, egal ob offline oder online durchgeführt, anfällig für einen weiteren Typus instrumentenbezogener Messfehler: Die sogenannte Panel-Konditionierung. Diese tritt, entsprechend der Vorgehensweise in Panelbefragungen, durch eine regelmäßige Befragung der selben Panelmitglieder auf (KALTON und CITRO 1993, zitiert in: COUPER und COUTTS 2006: 227). Durch die mehrfache Teilnahme an zurückliegenden Befragungen des Panels kann sich das Wissen oder das Verhalten von Teilnehmern systematisch von den Teilnehmern unterscheiden, die erstmals an einer Befragung des Panels teilnehmen. Auch hinsichtlich der Einstellungen sind Unterschiede zwischen erstmaligen und regelmäßigen Teilnehmern möglich und denkbar. Ursache für die Panel-Konditionierung kann einmal sein, dass

die Teilnehmer zusätzliches Wissen durch die Informationen erhalten, die im Interview vermittelt werden. Weitere Ursachen sind eine aus der Teilnahme entstehende Vorbereitung für zukünftige Befragungen, eine stärkere Auseinandersetzung mit der untersuchten Thematik sowie Lernprozesse hinsichtlich den Anforderungen des Panels (TOEPOEL et al. 2009: 73).

### **Messfehler durch Befragte**

Messfehler durch Befragte sind Messfehler, die durch die Befragten selbst und nicht durch das Instrument oder den Interviewer verursacht werden. TADDICKEN spricht hierbei simultan zu ihren Ausführungen zum Interview von den sogenannten Befragteneffekten. Auch hier kann eine Unterteilung in nicht-interaktionale sowie interaktionale Befragteneffekte vorgenommen werden (TADDICKEN 2008: 69).

Nicht-interaktionale Effekte können durch eine niedrige Motivation zur Teilnahme oder eine geringe Auskunftsfähigkeit verursacht werden. Auch ein fehlendes Erinnerungsvermögen in Verbund mit verzerrten oder falschen Antworten seitens des Probanden kann zu der Entstehung dieser Messfehler beitragen (COUPER und COUTTS 2006: 227; TADDICKEN 2008: 69; BEREKOVEN et al. 2009: 62).

Die interaktionalen Befragteneffekte ergeben sich auch hier aus der Interaktion zwischen Interviewer und Befragtem. Hier sind als Ursachen bewusste Falschbeantwortungen der Befragten mit der Zielsetzung der Ergebnisverfälschung, eine Zustimmungstendenz seitens des Befragten gegenüber dem Interviewer oder auch die sogenannte soziale Erwünschtheit zu nennen (ESSER 1986: 316f.; TADDICKEN 2008: 69; BEREKOVEN et al. 2009: 62).

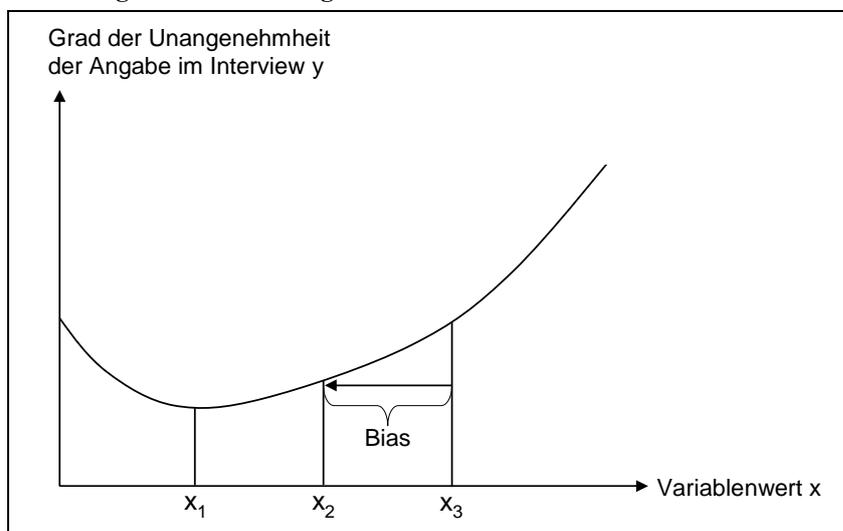
In diesem Zusammenhang sind weiterhin die sogenannten Response-Sets zu nennen, die ebenfalls zu einer Verzerrung der Antworten führen können und Ausdruck sowohl von interaktionalen sowie nicht-interaktionalen Befragteneffekten sein können. Hierunter sind systematische Antwortmuster von Befragten, wie etwa Mittelwert- oder auch Extremwerttendenzen, zu verstehen (CRONBACH 1946: 476; HUI und TRINANDIS 1985: 253; DIEKMANN 2008: 451).

Unter der sogenannten Zustimmungstendenz wird in der Literatur eine systematische Zustimmung-Reaktion des Befragten ohne Bezug auf den Frageninhalt verstanden (ESSER 1986: 316). Ursächlich hierfür kann einmal eine in der Persönlichkeit des Befragten fest verankerte grundlegende Neigung sein, die letztlich beim Antwortverhalten durchschlägt (BASS 1956: 299). Aber auch die Mehrdeutigkeit des Frageninhaltes kann zu Zustimmungstendenzen führen (BASS 1955: 622; JACKMAN 1973: 330).

Auch bei der sozialen Erwünschtheit sind zwei mögliche Ursachen denkbar: Einmal kann soziale Erwünschtheit als ein in der Persönlichkeit des Befragten verankertes Bedürfnis nach sozialer Anerkennung, im Falle der Befragung durch den Interviewer oder weitere anwesende Personen, beschrieben werden. Daneben kann soziale Erwünschtheit auch als eine durch die Datenerhebungssituation beeinflusste Abweichung von dem wahren Wert der Variablen verstanden werden, welcher aus Konsequenzbefürchtungen vom Befragten nicht offen gelegt wird (ESSER 1986: 317). DIEKMANN beschreibt hierzu, dass fast alle sozialen Aktivitäten und Eigenschaften Bewertungen unterzogen werden. Diese Bewertungen sind individuell und abhängig von sozialer Klassenzugehörigkeit, Bildungsgrad sowie weiteren Merkmalen. Das Maximum der Bewertungen dieser sozialen Aktivitäten oder Eigenschaften kann als Ort sozialer Erwünschtheit bezeichnet werden. Je weiter nun der tatsächliche Variablenwert von diesem Ort der sozialen Erwünschtheit abweicht, desto unangenehmer wird die Angabe der entsprechenden Antwort (bzw. des wahren Wertes) empfunden (DIEKMANN 2008: 447).

DIEKMANN stellt in Anlehnung an die Ausführungen von KOOLWIJK das Ausmaß der Unangenehmheit der Angabe im Interview ( $Y$ ) in Abhängigkeit von dem im Interview offenbarten Variablenwert  $X$  graphisch dar (siehe hierzu Abbildung 9). Anhand dieser Darstellung kann auch die Verzerrung durch soziale Erwünschtheit (der sogenannte „social desirability-Effekt“) einfach aufgezeigt werden. Hinsichtlich des Variablenwertes  $X$  wird vom Befragten möglicherweise der Wert  $X_1$  als sozial (und somit gesellschaftlich) gewünscht angenommen (DIEKMANN 2008: 448). Positive sowie negative Abweichungen von diesem Wert werden vom Befragten als abweichend von der Norm betrachtet und gelten daher für sein Empfinden als weniger sozial wünschenswert (KOOLWIJK 1969: 870).

**Abbildung 9: Verzerrung durch soziale Erwünschtheit**



Quelle: Eigene Darstellung in Anlehnung an DIEKMANN 2008: 450.

Wichtig ist hierbei der tatsächliche Variablenwert des Befragten  $X_3$ : Dieser entspricht aus Befragtersicht nicht dem als sozial wünschenswert betrachteten Variablenwert  $X_1$ , sondern weicht von diesem ab. Denkbar ist nun, dass vom Befragten ein Kompromiss zwischen dem wahren Wert und dem sozial erwünschten Wert hergestellt und der Wert  $X_2$  im Rahmen der Befragung genannt wird. Diese Angabe ist mithin durch den Effekt der sozialen Erwünschtheit verzerrt: Die Differenz zwischen  $X_3$  und  $X_2$  entspricht dabei der Verzerrung bzw. dem Bias durch die soziale Erwünschtheit (DIEKMANN 2008: 448).

Hinsichtlich des Effektes der sozialen Erwünschtheit führt DIEKMANN an, dass dieser bei vielen Fragen nicht oder nur in geringem Maße auftreten wird. Je heikler eine Fragestellung jedoch für den Befragten ist (die Funktion in Abbildung 9 wird dabei nach oben verschoben), desto stärker muss mit einem Einfluss des Effektes der sozialen Erwünschtheit gerechnet werden.

Zusätzlich kommt hier auch noch der Aspekt zum Tragen, dass häufig der wahre Variablenwert den befragten Personen selbst nicht genau bekannt ist. Zusätzlich können bei retrospektiven Verhaltensangaben noch Erinnerungsfehler hinzukommen. So kann nach DIEKMANN angenommen werden, dass bei Fragestellungen nach der Häufigkeit von Aktivitäten in der Vergangenheit häufig Schätzstrategien angewendet werden. Aus diesem Grund kann der Bereich der vom Befragten als subjektiv wahr empfundenen Angaben relativ groß ausfallen. Vermutet werden kann in diesen Fällen, dass vom Befragten ein Variablenwert aus diesem Bereich genannt wird, welcher dem Ort der sozialen Erwünschtheit am nächsten kommt (DIEKMANN 2008: 448f.).

### **Messfehler durch die angewendete Methode**

Im Unterschied zu Messfehlern, die durch das Instrument, also beispielsweise den Fragebogen, verursacht werden, entstehen die sogenannten Methodeneffekte durch die Verwendung einer bestimmten Befragungsmethode, also beispielsweise der telefonischen Befragung oder der in dieser Arbeit fokussierten Online-Befragung (TADDICKEN 2008: 72). Nach SCHNELL et al. entstehen Methodeneffekte daher auch „*durch die Auswirkungen der Datenerhebungsmethode auf die Ergebnisse der Messungen*“ (SCHNELL et al. 2005: 160). Eine exakte Definition des Begriffs wird allerdings in der Literatur nach TADDICKEN nicht vorgenommen. Auch werden Methodeneffekte häufig nur unsauber von den Messfehlern, die durch das Instrument entstehen, abgegrenzt (TADDICKEN 2008: 73; TADDICKEN 2009: 87f.).

Wie oben bereits dargestellt wurde, kann der Messwert einer Variablen  $X_l$ , gemessen mit der Methode 1, in den wahren Wert  $T_l$  sowie den Fehlerterm  $E_l$  zerlegt werden:

$$X_1 = T_1 + E_1 . \quad (2)$$

$X_1$  hat dabei einen endlichen Wert, der in einem bestimmten erwarteten Bereich liegen wird. Es wird weiterhin angenommen, dass die Variablen  $X_1$  sowie  $X_j$  exakt das selbe Konstrukt mit den unterschiedlichen Methoden 1 und  $j$  messen. Daher ist davon auszugehen, dass die wahren Werte  $T_1$  sowie  $T_j$  identisch sind:  $T_1=T_j$ . Daher gilt auch:

$$X_j = T_1 + E_j . \quad (3)$$

In diesem Fall liegen keine systematischen Methodeneffekte vor. Hinsichtlich der klassischen Testtheorie kann daher ausgesagt werden, dass die beiden Variablen  $X_1$  und  $X_j$  zu dem wahren Wert  $T$  äquivalent sind. Liegen jedoch systematische Methodeneffekte vor, kann die Gleichung (3) nicht gehalten werden und muss durch nachfolgende Gleichung (4) ersetzt werden, welche immer gültig ist:

$$X_j = T_1 + (T_j - T_1) + E_j . \quad (4)$$

Durch den Klammerausdruck  $(T_j - T_1)$  wird in dieser Gleichung dabei die Differenz zwischen den beiden unterschiedlich erhobenen, wahren Werten ausgedrückt. Eine vorliegende Differenz repräsentiert dabei einen systematischen Effekt, welcher sich durch die Anwendung der Methode  $j$  anstatt der Methode 1 ergibt. Daher gilt auch:

$$M_j \equiv T_j - T_1 . \quad (5)$$

Gleichung (5) kann daher auch als Methodeneffekt von  $X_j$  bezeichnet werden.  $M_j$  steht dabei für den Methodeneffekt (POHL et al. 2008: 48).

Nach KEUSCH lassen sich grundsätzlich vier verschiedene Ursachen für Methodeneffekte voneinander abgrenzen. Die von KEUSCH gewählten Bezeichnungen zeichnen sich dabei teilweise durch Redundanzen zu Bezeichnungen anderer Autoren aus:

- Medienbezogene Faktoren,
- Faktoren der Informationsübermittlung,
- Interviewereffekte,
- Befragungssituation.

Medienbezogene Faktoren beziehen sich auf das verwendete Medium sowie damit verbundene soziale Konventionen und Traditionen. Hier spielen beispielsweise die Vertrautheit und die Erfahrungen des Befragten mit dem Medium eine wichtige Rolle oder auch das Bewusstsein des Befragten, wer eine Kontrollposition im Rahmen der Befragung besitzt. So hat die Vertrautheit des Befragten mit dem Medium Einfluss auf den Befragungsverlauf: Je geringer die Erfahrung des Befragten mit dem Medium ist, desto zurückhaltender wird er wahrscheinlich Informationen von sich preisgeben. Auch macht sich gerade bei Online-Befragungen der Sachverhalt bemerkbar, dass Online-Befragungen im Unterschied zu F2F-Interviews nebenbei

beantwortet werden können. Auch muss möglicherweise erwartet werden, dass Online-Teilnehmer im Schnitt länger überlegen, was die gegebenen Antworten weniger spontan macht (JOINSON 2001: 587; KEUSCH 2007; ANONYMUS 2008)

Faktoren der Informationsübermittlung beziehen sich hingegen auf kognitive Stimuli, welche die Auskunftspersonen durch die Befragungsmethode erhalten. Diese kognitive Stimuli haben dabei maßgebliche Auswirkungen auf die Beantwortung der gestellten Fragen: So spielt es eine wichtige Rolle, ob eine Frage selbst gelesen oder vom Interviewer vorgelesen wird. Weiterhin spielt es eine Rolle, ob verbal oder nonverbal kommuniziert wird. Daneben sollten nach KEUSCH gerade bei Online-Befragungen auch noch Effekte durch die Reihenfolge der Fragen sowie durch die Gestaltung des Fragebogens berücksichtigt werden. Entsprechend muss die Gestaltung des Online-Fragebogens äußerst durchdacht erfolgen (KEUSCH 2007; ANONYMUS 2008).

Die sogenannten Interviewereffekte beziehen sich, wie der Begriff bereits vermuten lässt, auf die Anwesenheit eines Interviewers im Rahmen der Befragungssituation. Sie sind von den von TADDICKEN ebenfalls als Interviewereffekte bezeichneten Messfehlern durch den Interviewer abzugrenzen. Die Anwesenheit des Interviewers hat dabei Auswirkungen auf die wahrgenommene Anonymität einer Befragung: Diese wird aus Befragtersicht im Rahmen von Online-Befragungen aufgrund des Fehlens eines Interviewers größer sein als bei F2F-Interviews (GROVES et al. 2004: 142f.; KEUSCH 2007; ANONYMUS 2008).

Auch kommt es im Rahmen von Online-Interviews und dem damit verbundenen Wegfall des Interviewers zu einer Verschiebung der Kontrolle über den Befragungsprozess: Während in F2F-Interviews die Kontrolle der Befragungssituation in Händen des Interviewers liegt, wird im Rahmen von Online-Befragungen die Kontrolle dem Befragten selbst übertragen. Daher werden gewöhnlich in Online-Befragungen weniger sozial erwünschte Antworten gegeben. Auch die Zustimmungstendenz ist daher bei Online-Befragungen gewöhnlich geringer ausgeprägt (KEUSCH 2007; ANONYMUS 2008).

Auch der Befragungssituation als letztem Ursachenkomplex kommt nach KEUSCH eine wichtige Bedeutung bei dem Auftreten von Methodeneffekten zu. So spielen grundsätzlich die Rahmenbedingungen, unter denen eine Befragung stattfindet, eine entscheidende Rolle für das Auftreten von Methodeneffekten. Einmal ist hier der Ort, an dem die Befragung durchgeführt wird, zu nennen. So sind je nach Befragungsort verschiedene Störeinflüsse denkbar, die sich auf das Befragungsergebnis auswirken können. Weiterhin relevant für das Entstehen von Methodeneffekten ist auch die vom Befragten empfundene Legitimität der Befragung: On-

line-Befragungen bzw. F2F-Befragungen bekannter Organisationen wird wahrscheinlich eine höhere Aufmerksamkeit durch die Befragten geschenkt, als Befragungen, deren Auftraggeber vollkommen unbekannt ist (KEUSCH 2007; ANONYMUS 2008).

Anhand der vorhergehenden Ausführungen wird dabei ersichtlich, dass, wie oben bereits beschrieben wurde, eine genaue Abgrenzung zwischen Methodeneffekten und Messfehlern, die aus anderen Quellen her rühren, häufig nur sehr schwer möglich ist. Daher sind diesbezüglich auch gewisse Redundanzen zu verzeichnen. Oftmals besteht nach TADDICKEN auch ein enger Zusammenhang zwischen beiden Fehlerarten. Dennoch besteht der entscheidende Unterschied darin, dass Messfehlern, die auf dem verwendeten Instrument beruhen, im Unterschied zu Methodeneffekten, entgegengewirkt werden kann: So können beispielsweise durch eine Randomisierung der Antwortvorgaben entsprechende, aus einer nicht vorgenommenen Randomisierung entstehende, Instrumenten-Messfehler eliminiert werden. Bestehen bleiben im Anschluss entsprechend nur die Methodeneffekte (TADDICKEN 2008: 74).

Die Relevanz von Methodeneffekten ist bisher relativ wenig erforscht, was nach TADDICKEN am Sachverhalt liegt, dass Methodeneffekte nur schwierig nachweisbar sind (TADDICKEN 2008: 72). Dies zeigt sich möglicherweise auch anhand der Forschungslage: So gibt es bisher nur wenige Untersuchungen, die einen Methodenvergleich zwischen einer Online-Befragung und F2F-Interviews vorgenommen haben (ANONYMUS 2008). Andererseits führen nach TADDICKEN Methodeneffekte möglicherweise auch nur zu einer relativ geringen Verzerrung der Antworten, weshalb der Aufwand für den Nachweis von Methodeneffekten häufig nicht als lohnenswert bewertet wird.

Allerdings muss hierzu angemerkt werden, dass Methodeneffekte untrennbar mit der jeweiligen Befragungsmethode verbunden sind und somit eine heikle Fehlerquelle von Befragungen bilden. Denn im Unterschied zu Messfehlern, die aus dem verwendeten Instrument resultieren kann ihnen nicht durch eine sorgfältige Heran- und Vorgehensweise vorgebeugt werden (TADDICKEN 2008: 73).

#### **4 Literaturüberblick: Empirische Ergebnisse bisheriger Studien zu Methodeneffekten**

Die Fragestellung nach Methodeneffekten ist im Bereich der Sozialwissenschaften prinzipiell keine neue Fragestellung. So haben sich bereits verschiedene Wissenschaftler mit diesem Themenkomplex beschäftigt und es lässt sich entsprechend eine gewisse Anzahl an Studien finden, die der Fragestellung nachgegangen sind, ob sich zwischen den jeweils betrachteten Befragungsmethoden Unterschiede in verschiedener Hinsicht, wie beispielsweise der Rücklaufgeschwindigkeit oder auch der Ausschöpfungsquote, ergeben bzw. ermitteln lassen. Der Großteil dieser Studien vergleicht hierbei jedoch traditionelle Befragungsmethoden wie telefonische Befragungen und schriftlich-postalische Befragungen miteinander. Aufgrund des Sachverhaltes, dass es sich bei Internet-basierten Befragungen um eine relativ neue Befragungsmethodik handelt, die erst seit etwa der Jahrtausendwende Verbreitung findet, werden Online-Befragungen überhaupt erst in neueren Arbeiten zu möglichen Methodeneffekten einbezogen.

Häufig findet hierbei ein Methodenvergleich zwischen Online-Befragungen und postalischen bzw. schriftlichen Befragungen oder auch telefonischen Befragungen statt. Als Hintergrund dieser Vergleichskombination kann angesehen werden, dass, wie bereits in Kapitel 3.2 dargestellt werden konnte, Online-Befragungen im Vergleich zu traditionellen Befragungsmethoden wie der schriftlichen Befragung wesentliche Vorteile wie Zeit- und Kostenvorteile zugesprochen werden. Vor allem seit Beginn des neuen Jahrtausends wird aufgrund dieser Vorteile darüber nachgedacht, schriftliche, aber auch telefonische Befragungen durch Online-Befragungen zu substituieren. Eine theoretisch in Frage kommende Substitution traditioneller Befragungsmethoden durch die neue Form der Online-Befragung lässt dabei schnell die Frage aufkommen, ob Online-Befragungen überhaupt als ernsthaftes Befragungsinstrument zur Durchführung empirischer Untersuchungen geeignet sind. Die Studien dieser Zeit fokussieren entsprechend mögliche Vor- und Nachteile von Online-Befragungen sowie mögliche Auswirkungen auf das Antwortverhalten im Vergleich zu schriftlichen oder telefonischen Befragungen. Insgesamt zeichnet sich dabei ein sehr kritisches Bild hinsichtlich der Anwendbarkeit von Online-Befragungen ab. Eine derart kritische und ablehnende Haltung im Bereich der Sozialwissenschaften stellt dabei nichts Neues dar, wie DIEKMANN bemerkt. So wurde bereits bei Einführung der telefonischen Befragung kontrovers darüber diskutiert, ob telefonische Befragungen überhaupt im Rahmen wissenschaftlicher Untersuchungen geeignet sind und valide Ergebnisse liefern (DIEKMANN 2008: 502).

Ein Teil der Studien, die einen Methodenvergleich zwischen einer Online-Befragung und einer weiteren traditionellen Befragungsform wie der telefonischen oder der postalischen bzw. schriftlichen Befragung vornehmen, betrachtet dabei zusätzlich zu den oben genannten Forschungsfragen weitere Aspekte wie etwa die gegebene Zahl an „weiß nicht“-Antworten oder auch die Zahl der Item- oder der Unit-Nonresponses, anhand derer die Güte von Online-Befragungen bemessen werden soll. Beispiele sind hier etwa die Studien von HUDSON et al., die eine Online-Befragung mit einer telefonischen Befragung hinsichtlich des Item-Nonresponse verglichen haben. Als Ergebnis halten die Autoren fest, dass zwischen den Befragungsmethoden keine Verzerrungen durch den Item-Nonresponse nachweisbar waren (HUDSON et al. 2004: 237ff.). BECH und KRISTENSEN konnten bei einer Befragung von Menschen mit einem Lebensalter über 50 Jahren nachweisen, dass im Rahmen der beiden verglichenen Befragungsmethoden Online und Postalisch die Online-Befragung eine signifikant niedrigere Anzahl an Item-Nonresponse sowie „weiß nicht“-Antworten lieferte. Während häufig mit Online-Befragungen Kostenvorteile verbunden werden, ergab sich bei den Autoren hingegen das Bild, dass aufgrund der geringen Teilnahmequote der älteren Menschen an der Online-Befragung sich letztlich höhere Kosten für die Online-Befragung ergaben, um eine geeignete Stichprobengröße zu erhalten (BECH und KRISTENSEN 2009: 1ff.). Ein niedrigerer Item-Nonresponse im Rahmen einer Online-Befragung konnte weiterhin in der Studie von KWAK und RADLER im Vergleich zu einer parallel durchgeführten schriftlichen Befragung festgehalten werden (KWAK und RADLER 2002: 257ff.).

Verschiedene Studien gehen jedoch explizit auf die Fragestellung ein, ob sich zwischen den zu vergleichenden Befragungsmethoden Unterschiede im Antwortverhalten ermitteln lassen. Eine frühe Studie hierzu stammt von COUPER et al., die einen Methodenvergleich zwischen einer postalischen Befragung mit einem Vorläufer der heute bekannten Online-Befragung, nämlich der E-Mail-Befragung, durchführen. Die E-Mail-Befragung zeichnet sich dabei prinzipiell durch starke Ähnlichkeiten zur schriftlichen Befragung aus, wird allerdings am Computer, beispielsweise mit einer Textverarbeitung, ausgefüllt und anschließend an den Absender per E-Mail zurückgesendet. COUPER et al. kommen in ihrem Vergleich dabei zu dem Ergebnis, dass im Vergleich zur E-Mail-Befragung bei der postalischen Befragung eine signifikant höhere Responserate zu verzeichnen war. Während die Datenqualität, gemessen beispielsweise am Nonresponse, ähnlich war, konnten sie ermitteln, dass in der Tendenz die Teilnehmer der E-Mail-Stichprobe positivere Antworten gaben (COUPER et al. 1999: 39ff.).

Zu einem ähnlichen Ergebnissen kommen auch ROWE et al., die in ihrer Untersuchung eine postalische Befragung mit einer Online-Befragung vergleichen. Sie konnten ebenfalls ermit-

teln, dass über alle Fragen hinweg die Antworten der Online-Befragung positiver ausfielen als die Antworten des Papierfragebogens, wodurch sich signifikante Unterschiede im Antwortverhalten zwischen den beiden betrachteten Stichproben ergeben. Daneben konnten sie weiterhin einen signifikanten Unterschied der beiden Befragungsmethoden hinsichtlich der demographischen Variablen nachweisen. So zeigen sie, dass an der Online-Befragung vor allem jüngere Menschen, Männer sowie Personen aus reicheren Regionen teilgenommen haben. Sie bestätigen somit das häufig mit Online-Befragungen verbundene Coverage-Problem (ROWE et al. 2006: 352ff.).

Starke Wirkungen der Befragungsmethode auf das Antwortverhalten der Befragungsteilnehmer konnten auch BANDILLA et al. ermitteln. In ihrer Befragung, in der sie insgesamt 1.485 Personen mit Hilfe eines Fragebogens zu Umweltthemen befragten, konnten sie zwischen Teilnehmern, die den Fragebogen online ausgefüllt hatten und Teilnehmern, die an der schriftlichen Befragung teilgenommen hatten, bei einem großen Teil der Ergebnisse starke Unterschiede ermitteln. Keine Unterschiede konnten sie hingegen ermitteln, wenn homogene Gruppen, wie beispielsweise Gruppen von Rentnern oder Gruppen von Studierenden untereinander verglichen wurden. Hier waren die Antworten über die beiden untersuchten Befragungsmethoden hinweg gleich (BANDILLA et al. 2001: 7ff.).

Kleinere Unterschiede im Antwortverhalten konnten weiterhin von CARINI et al., HAYSLETT und WILDEMUTH sowie GAUPP et al. in ihren Methodenvergleichen ermittelt werden. CARINI et al. haben in ihrer Untersuchung einen Vergleich zwischen einer Online-Befragung und einer schriftlichen Befragung anhand des Datensatzes des National Surveys of Student Engagement (NSSE) vorgenommen, in dem 58.288 Studierende aus den USA zu ihrem ehrenamtlichen Engagement befragt wurden. Hinsichtlich möglicher Antwortunterschiede kommen sie dabei zum Schluss, dass nicht durchgängig Effekte auf das Antwortverhalten ermittelbar sind. Konnten hingegen Unterschiede ermittelt werden, waren diese sehr gering. Eine Ausnahme bildeten hierbei Fragestellungen, die sich mit dem Internet bzw. mit Computern beschäftigten. Hier waren im Unterschied zu den anderen Fragestellungen deutliche Effekte der Befragungsmethode auf das Antwortverhalten dahingehend erkennbar, dass Teilnehmer der Online-Befragung die Fragestellungen zu diesem Themenfeld signifikant besser und positiver beantworteten (CARINI et al. 2003: 1ff.). HAYSLETT und WILDEMUTH konnten in ihrer Studie ebenfalls nur bei einem Item einen signifikanten Unterschied im Antwortverhalten zwischen den Teilnehmern der Online-Befragung und den Teilnehmern der schriftlichen Befragung ermitteln. Bei allen anderen Items ihres Fragebogens hingegen zeigten sich keine Unterschiede. Allerdings zeigte sich auch hier wie bei COUPER et al., dass im Vergleich zu der Online-Befragung

die Response-Rate der postalischen Befragung signifikant höher lag. Wurde hingegen an der Online-Befragung teilgenommen, wurde diese signifikant schneller beantwortet, wie anhand einer durchgeführten ANOVA von den Autoren ermittelt werden konnte (HAYSLETT und WILDEMUTH 2004: 73ff.).

Grundsätzlich kein signifikanter Einfluss der Methode auf das Antwortverhalten, mit einer Ausnahme, konnte von GAUPP et al. nachgewiesen werden, die 84 bildungsbenachteiligte Schüler schriftlich, telefonisch oder per Internet befragten. Der signifikante Unterschied ergab sich dabei jedoch nicht im Zusammenhang mit der Online-Befragung, sondern im Vergleich der schriftlichen mit der telefonischen Befragung. Bei der Datenqualität, gemessen über die Anzahl der nichtbeantworteten Fragestellungen, zeigte sich in der Untersuchung weiterhin, dass im Vergleich die Zahlen der nichtbeantworteten Fragen bei der telefonischen und der Online-Befragung signifikant geringer waren und diese somit mehr auswertbare Fragen bereitstellten. GAUPP et al. führen diesbezüglich jedoch an, dass dieser ermittelte Unterschied möglicherweise auch auf das niedrige Bildungsniveau der Schüler bzw. deren fehlenden Schreibkenntnisse zurückzuführen sind (GAUPP et al. 2006: 12ff.).

Diesen Studien, die ein unterschiedliches Antwortverhalten der Befragungsteilnehmer je nach Befragungsmethode ermitteln konnten, können ebenfalls Studien gegenübergestellt werden, in denen von den Autoren keine Unterschiede im Antwortverhalten nachgewiesen werden konnten.

So verglichen COBANOGLU et al. in ihrer Untersuchung eine schriftliche Befragung mit einer Fax-Befragung und einer Online-Befragung. Die Stichprobe setzte sich dabei aus 300 Professoren zusammen. Die Ergebnisse können so zusammengefasst werden, dass hinsichtlich der Rücklaufgeschwindigkeit die Fax-Befragung die schnellste Befragungsmethode vor der Online-Befragung und der schriftlichen Befragung war. Die höchste Rücklaufquote konnte im Unterschied zu den oben dargestellten Studien von COUPER et al. und HAYSLETT und WILDEMUTH hier bei der Online-Befragung festgestellt werden, gefolgt von der schriftlichen Befragung. Bei der Überprüfung der Antworten auf mögliche Unterschiede zwischen den drei Befragungsmethoden mit Hilfe von t-Tests kommen sie dabei zu dem Resultat, dass keine signifikanten Antwortunterschiede zwischen den drei Stichproben nachweisbar sind (COBANOGLU et al. 2001: 441ff.). Zu diesem Ergebnis kommen ebenfalls BERRENS et al., die einer telefonische Befragung eine Online-Befragung gegenüber gestellt haben. Dazu befragten sie 1.699 Personen in den USA per Telefon sowie weitere 13.034 Personen über einen Online-Fragebogen (BERRENS et al. 2003: 1ff.).

Inhaltlich keine Unterschiede ergaben sich auch bei KIERNAN et al.. Insgesamt wurden hier 274 Personen entweder online oder postalisch befragt. Allerdings konnten die Autoren nachweisen, dass von den Teilnehmern der Online-Befragung bei quantitativen Fragen eine größere Anzahl an Fragen beantwortet wurde sowie bei qualitativen Fragen die Antworten länger und ausführlicher waren (KIERNAN et al. 2005: 245ff.).

Neben diesen Studien, die im Allgemeinen überprüfen, ob zwischen verschiedenen Befragungsmethoden Unterschiede im Antwortverhalten ersichtlich werden, lassen sich noch weitere Studien aufzeigen, die bewusst überprüft haben, inwiefern sich Methodeneffekte aufgrund von sozialer Erwünschtheit bzw. einer geringeren Anonymität ergeben. JOINSON konnte hierzu bei einer mit 82 Studierenden durchgeführten Untersuchung nachweisen, dass sowohl in einem anonymen als auch in einem nicht-anonymen Untersuchungsaufbau die Teilnehmer der schriftlichen Befragung signifikant sozial erwünschter antworteten als die Studierenden, die an der Online-Befragung teilgenommen hatten. Hinsichtlich des unterschiedlichen Grades an Anonymität kommt der Autor jedoch zu dem Schluss, dass diesbezüglich keine Effekte nachweisbar sind (JOINSON 1999: 433ff.). KREUTER et al. kommen in ihrer Studie ebenfalls zu dem Ergebnis, dass die Teilnehmer in der Online-Befragung durchaus bereit waren, auch sozial weniger erwünschte Dinge über sich preis zu geben als die Teilnehmer in der CATI-Befragung und einer automatisierten, telefonischen Befragung ohne Interviewer (KREUTER et al. 2008: 847ff.). TADDICKEN, die jeweils eine Internetbefragung mit einer schriftlichen und einer telefonischen Befragung verglichen hat, konnte ebenfalls ermitteln, dass die Befragten in der Online-Befragung offenere und ehrlichere Antworten gaben sowie weniger sozial erwünscht antworteten (TADDICKEN 2009: 91ff.).

Während beispielsweise JOINSON, wie oben dargestellt, keine Effekte hinsichtlich des Grades an Anonymität nachweisen konnte, kommen diesbezüglich REIPS und FRANEK zu einem anders lautenden Ergebnis. Aus ihrem Methodenvergleich zwischen einer Online-Befragung und einer schriftlichen Befragung mit insgesamt 655 Teilnehmern folgern sie, dass umso sozial erwünschter geantwortet wurde, je geringer die Anonymität aus Befragtersicht war. Hieraus schließen die Autoren, dass bei einer Kontrolle der Anonymität in der schriftlichen Befragung gleichlautende Ergebnisse wie in der Online-Befragung zumindest theoretisch möglich seien (REIPS und FRANEK 2004: 67ff.). Dies kann beispielsweise durch die Ergebnisse von FUCHS bestätigt werden, der sich mit den ablaufenden kognitiven Prozessen bei schriftlichen Befragungen sowie Online-Befragungen beschäftigt hat. FUCHS konnte im Antwortverhalten hierbei keine Unterschiede zwischen den beiden Befragungsmethoden feststellen. Hieraus schließt der Autor, dass die ablaufenden kognitiven Prozesse, welche zur Entstehung der

Antworten bei den Befragten führen, zwischen den beiden Befragungsmethoden grundsätzlich gleich sind. Daher sind höchstwahrscheinlich bei einem Wechsel der Befragungsmethode keine Antwortverzerrungen zu erwarten (FUCHS 2003: 19ff.).

Während sich nach den bisherigen Ausführungen ergibt, dass mögliche Unterschiede zwischen Online-Befragungen und schriftlichen oder telefonischen Befragungen stark erforscht sind und in der Wissenschaft bereits ausführlich diskutiert wurden, zeigt sich hinsichtlich möglicher Unterschiede zwischen Online-Befragungen und F2F-Befragungen ein anderes Bild. Hier lassen sich nur wenige Studien finden, die bereits einen solchen Methodenvergleich durchgeführt haben.

Ein früher Spezialfall stellt hierbei die Studie von SCHOLL et al. aus dem Jahre 2002 dar. Die Autoren führen hierbei zwar einen Vergleich zwischen einer Online-Befragung sowie einer F2F-Befragung durch. Für den Vergleich ziehen sie jedoch keine quantitative Erhebung heran, was der Usus bei derartigen Vergleichen ist, sondern greifen auf eine qualitative Erhebung zurück. Eingebettet wird die Gesamtuntersuchung in den Kontext der Marktforschung. Die für die Befragten vermeintliche Zielstellung ist es, bei der Entwicklung eines neuen Rasierapparates mitzuhelfen. Im Rahmen der F2F-Befragungen wurden die Personen dazu mit qualitativen Interviews befragt, während in der Online-Befragung Online-Interviews über einen Instant-Messenger<sup>13</sup> mit den zu befragenden Personen geführt wurden. Nach der Auswertung der Interviews kommen die Autoren dabei zu dem Ergebnis, dass die gefundenen Ergebnisse vergleichbar sind und die unterschiedlichen Befragungsmethoden wohl keinen Einfluss auf das Antwortverhalten der Befragten hatten (SCHOLL et al. 2002: 210ff.).

In einer anderen Studie verglich SCHOEN zwei Online-Befragungen mit einer F2F-Befragung. Themenschwerpunkt war hierbei unter anderem die Bundestagswahl 2002. Insgesamt wurden dabei von SCHOEN 38.426 Personen befragt. 3.263 Personen wurden dabei mündlich befragt, weitere 1.065 gehörten einem geschlossenen Online-Panel an, während weitere 34.098 Personen an einer offenen Online-Befragung zum Wahlverhalten teilgenommen hatten. Hinsichtlich der Ergebnisse lässt sich festhalten, dass die Teilnehmer der beiden Online-Befragungen deutlich stärker politisch involviert waren und sich zudem subjektiv merklich kompetenter fühlten als die Teilnehmer der mündlich befragten Zufallsstichprobe. Hinsichtlich des politischen Hintergrundwissens konnte SCHOEN herausarbeiten, dass die Teilnehmer der F2F-Be-

---

<sup>13</sup> Bei Instant-Messengern handelt es sich um eine Software, die das sogenannte Instant Messaging erlaubt. Das Instant Messaging ermöglicht es, in Echtzeit mit anderen Teilnehmern kommunizieren zu können. Die Kommunikation erfolgt dabei per Tastatur und wird als „Chatten“ (englisch für plaudern) bezeichnet (BÄCHLE 2006: 122).

fragungen verschiedene Fragestellungen zum politischen System in Deutschland, wie etwa dem Staatsaufbau oder dem Wahlsystem, sechs bzw. elf Prozentpunkte seltener korrekt beantworteten als ihre per Online-Erhebung befragten Pendanten. SCHOEN führt diesen Wissensvorsprung dabei auf die Möglichkeit der Online-Befragungs-Teilnehmer zurück, unbeobachtet von einem Interviewer mit einem Griff zum Lexikon oder durch Nutzung einer Internetsuchmaschine aufklaffende Wissenslücken schließen zu können (SCHOEN 2004: 27ff.).

Einen ähnlichen Effekt bezüglich der Beantwortung von Wissensfragen konnten auch DUFFY et al. nachweisen. Sie befragten im Jahre 2005 in Großbritannien 2.057 Personen per F2F-Befragung sowie weitere 4.131 Personen über eine Online-Befragung. Die zum Vergleich herangezogenen Fragestellungen wurden dabei in bereits bestehende Panel-Untersuchungen integriert und später davon isoliert einer vergleichenden Betrachtung unterworfen. Bei einer, in die anderen Befragungen integrierten, wissensorientierten Fragestellung zum Thema Cholesterin zeigte sich bei den Teilnehmern der Online-Befragung hier ein höheres Wissen in der Form, dass die Fragestellung signifikant häufiger korrekt beantwortet wurde. Auch hier führen DUFFY et al. als Erklärungsmöglichkeit an, dass die online befragten Teilnehmer möglicherweise mangels einer Kontrolle durch den Interviewer die Möglichkeit genutzt haben, in Lexika oder mittels Internetsuchmaschinen die korrekte Antwort zu recherchieren. Unterschiede ergaben sich auch bei einer weiteren Fragestellung zum Wahlverhalten zwischen den beiden Befragungssamples. So gab in der Online-Befragung ein größerer Teil der Befragten an, liberal orientierte Parteien zu wählen als in der F2F-Stichprobe (DUFFY et al. 2005: 615ff.).

Unterschiedliche Angaben hinsichtlich des Wahlverhaltens konnte auch FAAS mit dem gleichen Datensatz aufdecken, den auch SCHOEN (siehe oben) für seine Studie herangezogen hat. Beide Studien kommen dabei zu dem Schluss, dass die jeweils nachgewiesenen Unterschiede im Wahlverhalten dabei durch die, hinsichtlich der demographischen Variablen, unterschiedlich zusammengesetzten Stichproben zustande kommen. Beide wenden daraufhin eine Gewichtung des Datensatzes an, um die Verteilungen der zum Methodenvergleich herangezogenen Stichproben jeweils anzugleichen. Bezogen ist die Gewichtung dabei jeweils auf die demographischen Variablen Alter sowie Geschlecht und folgt dabei bevölkerungsrepräsentativen Verteilungen. Während entsprechend FAAS die Stichproben nach der deutschen Gesamtbevölkerung gewichtet, ziehen DUFFY et al. die demographischen Verteilungen Großbritanniens bei den relevanten Variablen zur Gewichtung heran. Diese Vorgehensweise ist in den Studien dabei nicht gleichermaßen zielführend: Während sich bei FAAS die Befragungssamples auch noch nach der Gewichtung hinsichtlich des Wahlverhaltens signifikant voneinander

unterschieden, führte bei DUFFY et al. die Gewichtung dazu, dass nach der Gewichtung zwischen den beiden verglichenen Befragungssamples F2F und Online keine Unterschiede im angegebenen Wahlverhalten mehr nachgewiesen werden konnten. Über die Gewichtung konnten somit Verzerrungen, die auf unterschiedlich zusammengesetzten Stichproben beruhten, beseitigt werden (FAAS 2003: 120ff.; DUFFY et al. 2005: 615ff.).

Eine relativ neue Studie, die einen Vergleich zwischen einer F2F-Befragung sowie einer Online-Befragung durchführt, ist die Untersuchung von HEERSWEGH und LOOSVELDT aus dem Jahre 2008. Sie untersucht unter anderem, inwiefern sich zeitliche Unterschiede bei der Beantwortung der Fragebögen in den beiden Teilstichproben ergeben. Dabei konnten sie ermitteln, dass von den insgesamt 3.255 Teilnehmern dieser Studie diejenigen, die der Online-Stichprobe angehörten, signifikant weniger Zeit benötigten, um die gegebenen Fragen zu beantworten als die Versuchsteilnehmer, die im Rahmen der F2F-Interviews befragt wurden. Sie konnten somit den in der Literatur mehrfach genannten Vorteil von Online-Befragungen hinsichtlich einer schnelleren Datenverfügbarkeit und Durchführbarkeit bestätigen (HEERSWEGH und LOOSVELDT 2008: 836ff.).

## 5 Empirische Analyse

In diesem Kapitel wird nachfolgend die empirische Datenanalyse gemäß den Zielsetzungen der vorliegenden Arbeit durchgeführt. Begonnen wird dabei mit der Darstellung der Arbeitshypothesen in Kapitel 5.1. Im Anschluss daran wird in Kapitel 5.2 detailliert auf die angewendete Methodik sowie das Untersuchungsdesign eingegangen, welches in der vorliegenden Arbeit aufgrund der Zielsetzung eine essentielle Rolle spielt. Daran anknüpfend werden die Ergebnisse der Untersuchung in den Kapiteln 5.3 bis 5.7 dargestellt sowie eine Verifizierung oder Falsifizierung der Arbeitshypothesen vorgenommen.

### 5.1 Arbeitshypothesen

Die in diesem Kapitel vorgenommene empirische Analyse liegt den nachfolgend dargestellten Arbeitshypothesen zugrunde. Diese wurden, aufbauend auf der in Kapitel vier durchgeführten Literaturanalyse und den dort dargestellten, relevanten Ergebnissen aufgestellt. Die Arbeitshypothesen sollen dazu beitragen, die mit dieser Arbeit verknüpften Forschungsfragen ziel führend zu beantworten. So gibt es bisher nur wenige Studien, die überhaupt einen Methodenvergleich zwischen einer Online-Befragung sowie einer F2F-Befragung durchgeführt haben. Hier werden bezüglich Online-Befragungen häufig Vorteile wie ehrlichere Antworten aufgrund einer höheren Anonymität genannt. Diese ehrlicheren Antworten, vor allem bei heiklen Fragen, konnten, wie im Literaturüberblick gezeigt wurde, bei Methodenvergleichen zwischen Online-Befragungen sowie telefonischen oder schriftlichen Befragungen bestätigt werden. Bei den wenigen bisherigen Methodenvergleichen zwischen Online- und F2F-Befragungen steht hingegen eine derartige Überprüfung noch aus. Geeignet scheint hierbei vor allem eine Frage nach dem Einkommen, bei der es nach SCHNELL et al. und BEREKOVEN et al. immer noch um eine heikle Frage handelt (SCHNELL et al. 2005: 339; BEREKOVEN et al. 2009: 97): Daher wird hier zur Überprüfung folgende Forschungshypothese F1 aufgestellt:

**Hypothese F1:** Bei Fragen zum Einkommen als Repräsentant für heikle Fragestellungen sind aufgrund der unterschiedlichen Anonymität in beiden Befragungsmethoden Unterschiede im Antwortverhalten dahingehend erkennbar, dass in der Online-Befragung niedrigere Angaben gemacht werden.

?

Weiterhin werden als Vorteile von Online-Befragungen eine schnellere Durchführbarkeit der Untersuchung sowie eine schnellere Datenverfügbarkeit genannt. Allerdings lässt sich hierzu nur eine Studie finden, die sich mit der Fragestellung zu Zeitvorteilen von Online-Befragungen beschäftigt hat. Daher wird als zweite Forschungshypothese folgende Hypothese F2 aufgestellt:

**Hypothese F2:** Die Befragungsmethode Online zeichnet sich im Vergleich zu F2F durch eine signifikant kürzere Interviewdauer aus.

?

Des Weiteren wurde im Literaturüberblick herausgearbeitet, dass sich häufig bei Wissensfragen signifikante Unterschiede zwischen Online- und F2F-Befragungen dahingehend ergeben, dass von den Teilnehmern der Online-Erhebung signifikant häufiger korrekte Antworten auf die Fragen gegeben werden. Hierbei stellt sich jedoch die Frage, ob dieser Sachverhalt auch in relativ homogenen und hoch gebildeten Befragungssamples überhaupt eine Rolle spielt und ob nicht stattdessen erwartet werden kann, dass aufgrund eines hohen Bildungsniveaus derartige Antwortunterschiede keine Rolle spielen. Entsprechend dieser Überlegungen wird folgende dritte Forschungshypothese F3 aufgestellt:

**Hypothese F3:** Hinsichtlich der Beantwortung von Wissensfragen sind keine signifikanten Unterschiede zwischen den beiden Befragungssamples zu erwarten.

?

Zusätzlich fehlen momentan auch noch belastbare Ergebnisse dahingehend, ob neben diesem Spezialfall der Wissensfragen auch bei anderen Fragetypen bzw. bei Fragestellungen im Allgemeinen mit Antwortverzerrungen zu rechnen ist, wenn die Befragungsmethode von F2F auf Online umgestellt wird. Ein derartiger Methodenwechsel stellt dabei nicht unbedingt nur eine theoretisch anzunehmende Möglichkeit dar. So versprechen Online-Befragungen gegenüber ihren mündlich durchgeführten Pendanten verschiedene Vorteile. Auch die zunehmende Substitution von F2F-Befragungen durch Online-Befragungen zeigt die Relevanz einer derartigen Forschungsfrage auf. Diesen Ausführungen folgend soll weiterhin folgende Forschungshypothese F4 als Hauptforschungsziel dieser Arbeit verifiziert oder falsifiziert werden:

**Hypothese F4:** Auch bei anderen Fragetypen als Wissensfragen lassen sich eindeutig Methodeneffekte in Form von Antwortunterschieden zwischen Online- und F2F-Befragungen nachweisen.

?

## 5.2 Darstellung des Untersuchungsdesigns

In diesem Gliederungspunkt wird nachfolgend auf das in der Untersuchung angewendete Forschungsdesign zum Einfluss der Befragungsmethode auf das Antwortverhalten eingegangen.

### 5.2.1 Split-Ballot-Technik

Für die Durchführung eines Methodenvergleichs zwischen F2F-Interviews und Online-Befragungen sowie die Entscheidung über die Annahme beziehungsweise Ablehnung der oben unter 5.1 dargestellten Arbeitshypothesen wird in der vorliegenden Arbeit die sogenannte Split-

Ballot-Technik herangezogen. Bei diesem Verfahren handelt es sich um eine in der empirischen Literatur häufig vorgeschlagene und angewendete Vorgehensweise in Form eines Methodenexperiments (BANDILLA et al. 2001; SCHOLL et al. 2002; CARINI et al. 2003; FAAS 2003; FUCHS 2003; SCHOEN 2004; DUFFY et al. 2005;): Bei der Split-Ballot-Technik handelt es sich um ein Verfahren, bei dem ursprünglich zwei Substichproben einer Grundgesamtheit mit unterschiedlichen Frageformulierungen befragt wurden, um dadurch hervorgerufene Unterschiede in dem Antwortverhalten der Befragten identifizieren zu können. So untersuchte CANTRIL mit Hilfe der Split-Ballot-Technik die Wirkung verschiedener Worte in Fragestellungen auf das Antwortverhalten (CANTRIL 1947: 23ff.). Neuere Anwendungen dieser Technik schließen inzwischen ebenfalls weiterführende methodische Fragestellungen wie etwa die Auswirkungen der Befragungsmethode auf das Antwortverhalten mit ein. Dabei ist die allgemeine und grundlegende Vorgehensweise so, dass zwei möglichst homogene Samples aus einer Grundgesamtheit unter Zuhilfenahme eines geeigneten Fragebogens zu einer bestimmten Thematik befragt werden. Die eine Teilstichprobe wird dabei mit der ersten der beiden zu vergleichenden Befragungsmethoden, hier den F2F-Interviews, befragt, während die zweite Teilstichprobe mit der zweiten, zu vergleichenden Befragungsmethode, hier den Online-Befragungen, befragt wird. Im Anschluss daran werden die beiden Teilstichproben mit Hilfe sowohl deduktiver als auch induktiver statistischer Verfahren untersucht und auf mögliche Unterschiede im Antwortverhalten hin analysiert (CANTRIL 1947: 23ff.; JAROS und MASON 1969: 106; SCHNELL et al. 2005: 351f.; PETERSEN 2008: 325ff.; TADDICKEN 2008: 159f.).

### 5.2.2 Auswahl der Zielpopulation

Um entsprechend der Zielsetzung dieser Arbeit mit Hilfe der oben dargestellten Split-Ballot-Technik auf die Erhebungsmethode zurückzuführende Unterschiede im Antwortverhalten verschiedener Teilstichproben analysieren zu können, ist es notwendig, dass diese aus der gleichen Grundgesamtheit stammen. TADDICKEN führt diesbezüglich an, dass nur so eine Gleichverteilung der „wahren Werte“ in den Teilstichproben sichergestellt werden kann. Ansonsten können zwischen den Teilstichproben unterschiedliche Messwerte nicht ausschließlich auf die Methodeneffekte zurückgeführt werden, die jeweils aus der eingesetzten Befragungsmethode resultieren, sondern können möglicherweise auch auf unterschiedliche „wahre Werte“ in den Teilstichproben zurückgeführt werden. Diesbezüglich sieht TADDICKEN auch methodische Probleme bei den meisten bisherigen Untersuchungen, die einen Vergleich zwischen einer Online-Befragung und einer anderen Erhebungsmethode vorgenommen haben: So stellt sich für einen Methodenvergleich das Problem dar, dass, entsprechend den Ausführungen zum Undercoverage in Kapitel 3.3.2, angenommen werden muss, dass sich die Verteilungen sozio-

demographischer Merkmale von Internetnutzern von denen der Gesamtbevölkerung, unabhängig vom Land, in dem die Erhebung durchgeführt wurden, unterscheiden (TADDICKEN 2005: 161ff.). Häufig werden jedoch diese unterschiedlichen Verteilungen der soziodemographischen Variablen in den Methodenvergleichen nicht ausreichend berücksichtigt. Auch ist nach TADDICKEN davon auszugehen, dass sich Nutzer und Nichtnutzer des Internets signifikant in relevanten psychographischen Variablen voneinander unterscheiden. Beide Aspekte können entsprechend den oberen Ausführungen zu einer Ungleichverteilung der „wahren Werte“ der Teilstichproben führen. Problematisch erweist sich dieser Sachverhalt dahingehend, dass Unterschiede im Antwortverhalten damit nicht zwangsläufig auf die unterschiedlichen Erhebungsmodi, sondern eventuell auch auf die Ungleichverteilung der „wahren Werte“ der Teilstichproben zurückgeführt werden können (TADDICKEN 2008: 161ff.).

Aufgrund des erheblichen Erhebungsaufwandes, um oben dargestellte Probleme umgehen bzw. korrigieren zu können, wurde die dieser Arbeit zu Grunde liegende empirische Untersuchung zum Vergleich zweier Erhebungsverfahren von Beginn an als nicht bevölkerungsrepräsentative Untersuchung geplant. Um dennoch möglichst aussagefähige Ergebnisse zu erhalten, wurde stattdessen auf eine gut erreichbare Spezialpopulation ausgewichen. So wurden als Grundgesamtheit, aus der die beiden Teilstichproben gezogen werden sollten, ausschließlich Studierende der Justus-Liebig-Universität Gießen (JLU) definiert. Mit dieser Definition sind mehrere, wesentliche Vorteile verbunden: Zum einen kann davon ausgegangen werden, dass es sich bei Studierenden im Vergleich zur Gesamtbevölkerung um eine relativ homogene Population handelt. Beispielsweise kann angenommen werden, dass sich Studierende hinsichtlich ihres Bildungsniveaus, ihres Familienstatus und ihrer Lebensweise bzw. ihres Lebensabschnitts im Vergleich zu einer Grundgesamtheit, welche die gesamte deutsche Bevölkerung umfasst, deutlich weniger unterscheiden (KHLAVNA 2005: 15; AUSPURG und ABRAHAM 2007: 283; SAUER et al. 2009: 2). Mit der Definition der Studierendenschaft der JLU als Grundgesamtheit wird weiter angenommen, dass sich die Studierenden daher bezüglich ihrer psychographischen Variablen wie den Einstellungen annäherungsweise ähneln. Dies hat den Vorteil, dass mögliche, daraus resultierende Verzerrungen, wie oben beschrieben, vermieden werden können.

Für die Durchführung eines Methodenvergleichs bietet eine Fokussierung auf Studierende der JLU ebenfalls den wesentlichen Vorteil, dass hier davon ausgegangen werden kann, dass der weit überwiegende Teil der Studierenden über einen Zugang zum Internet verfügt bzw. diesen

ebenfalls regelmäßig nutzt. So nutzten im Jahre 2008 bundesweit bereits 87 % der Schüler<sup>14</sup> und Studierenden täglich das Internet. Weitere 12 % nutzten das Internet mindestens einmal wöchentlich und lediglich ein Prozent der Schüler und Studierenden nutzte das Internet seltener als einmal wöchentlich (STATISTISCHES BUNDESAMT 2009: 116). Simultan zu dem Anstieg der Nutzerzahlen (siehe oben) kann hier eine Verschiebung der Anteile zugunsten der täglichen Internetnutzung vermutet werden.

Für eine regelmäßige Internetnutzung und eine gewisse Internetnutzungskompetenz spricht auch der Sachverhalt, dass im Bereich der Justus-Liebig-Universität verschiedene, fachbereichsübergreifende oder -spezifische Internetportale wie etwa Stud.IP und Flexnow!<sup>15</sup>, SPIC<sup>16</sup>, k-MED<sup>17</sup> oder auch das Bibliothekssystem samt OPAC Verwendung finden.

Weiterhin lässt eine Fokussierung auf die Studierendenschaft einen einfachen Zugang zur Grundgesamtheit und somit zur Stichprobenziehung zu, was ebenfalls, zumindest im Kontext dieser Arbeit, als wesentlicher Vorteil gegenüber einer bevölkerungsrepräsentativen Untersuchung angesehen werden kann. Neben der einfachen Durchführbarkeit von F2F-Interviews spielen hierbei auch gute Zugangsmöglichkeiten zur Zielgruppe über das Internet sowie eine gute Durchführbarkeit der Online-Befragung eine bedeutsame Rolle. Wie bereits in Kapitel 3.3.2 ausführlich dargestellt wurde, stellt der Coverage-Fehler ein wesentliches Problem von Internetbefragungen dar. So existieren im Unterschied zur Gesamtbevölkerung (hier können grundsätzlich Angaben vom Einwohnermeldeamt eingeholt werden) keine Listen darüber, wer Zugang zum Internet hat bzw. das Internet nutzt. Durch die Beschränkung auf Studierende der JLU kann dieses Problem allerdings über die Nutzung des Rundmail-Services „Hinweise der JLU“ vernachlässigt werden. Mit Immatrikulation an der JLU erhält jeder Studierende vom Hochschulrechenzentrum der JLU eine E-Mail-Adresse im Format „Vorname.Nachname@Fachgebietskennung.uni-giessen.de“ (JLU 2007). An diese E-Mailadressen können über den Rundmail-Dienst „Hinweise der JLU“ Mitteilungen von Angehörigen der JLU an andere Organisationsangehörige in großer Zahl versendet werden. Ebenfalls als Vorteil einer Nutzung des Services für die vorliegende Arbeit kann betrachtet werden, dass vor Versand einer Mitteilung die entsprechende Zielgruppe, wie etwa Studierende, Mitarbeiter oder Professoren, ausgewählt werden kann. Die gute Erreichbarkeit über die „Hinweise der JLU“ sprach ebenfalls für eine Beschränkung der Grundgesamtheit auf Studierende aller Fachberei-

---

<sup>14</sup> Schüler ab 16 Jahre (STATISTISCHES BUNDESAMT 2009: 116).

<sup>15</sup> Stud.IP und Flexnow! werden an verschiedenen Fachbereichen der Justus-Liebig-Universität verwendet (JLU 2009a; JLU 2009b).

<sup>16</sup> SPIC steht für „Students' Personal Information Center“ und wird am Fachbereich 02 Wirtschaftswissenschaften eingesetzt (JLU 2009c).

<sup>17</sup> k-MED wird am Fachbereich 11 Medizin eingesetzt (JLU 2009d).

che und Semester der JLU. Ausgeschlossen wurden lediglich Studierende in Graduiertenprogrammen, Promotionsstudierende bzw. wissenschaftliche Mitarbeiter, die eine Promotion anstreben. Ursächlich hierfür ist, dass diese nicht unbedingt als Student an der JLU immatrikuliert sein müssen und somit auch nicht über die „Hinweise der JLU“, sofern Studierende als Empfänger ausgewählt werden, erreicht werden können. Um etwaige, bereits dadurch entstehende Verzerrungen durch Undercoverage ausschließen zu können, wurden diese daher von der Zielpopulation ausgeschlossen.

### 5.2.3 Fragebogen

An den in der Untersuchung zu verwendenden Fragebogen wurden bereits zu Beginn der Untersuchungskonzeption verschiedene, als relevant betrachtete Anforderungen gestellt. So wurde aufgrund der verfolgten Zielsetzung eine möglichst einfache Umsetzbarkeit des Fragebogens sowohl für die F2F-Interviews als auch die Online-Erhebung als essentiell betrachtet. Auch sollte der Fragebogen eine für Studierende interessante Thematik behandeln, um einerseits das Interesse der Studierenden gewinnen und andererseits eine hohe Abbruchquote vermeiden zu können. Diese beiden Aspekte scheinen dabei sowohl für die F2F-Interviews, als auch für die Online-Untersuchung relevant. So muss in der Online-Befragung auf die motivierende Funktion eines Interviewers, die sich positiv auf das Teilnahmeverhalten und die Abbruchquote auswirken kann, verzichtet werden (ATTESLANDER 2008: 147ff.; SCHNELL et al. 2005: 358ff.; DIEKMANN 2008: 514).

Im Hinblick auf die inhaltliche Gestaltung eines Fragebogens sollten verschiedene grundlegende Aspekte berücksichtigt werden. So ist nach DIEKMANN bei der Fragenformulierung darauf zu achten, dass diese kurz, verständlich und hinreichend präzise formuliert sind (DIEKMANN 2008: 479). Fachausdrücke, sofern sie bei der angepeilten Zielgruppe der Untersuchung nicht gebräuchlich sind, sollten nicht verwendet werden. Weiterhin sollten sogenannte „Suggestivfragen“ vermieden werden, die durch ihre Formulierung bereits eine bestimmte Beantwortung provozieren. Ähnlich ist dabei die Faustregel von DILLMAN zu verstehen, dass Fragen grundsätzlich neutral formuliert sein sollten und daher keine besonders belasteten Worte wie etwa „Kommunist“ oder „Bürokratie“ enthalten sollten. Neben einer einfachen Formulierung der Fragen sollte allerdings auch auf eine möglichst einfache Handhabbarkeit der Fragen für den Interviewer geachtet werden (DILLMAN 1978: 95ff.; CONVERSE und PRESSER 1986: 11ff.; SCHNELL et al. 2005: 334ff.).

Bezüglich der Gestaltung von Online-Befragungen gelten nach WELKER et al. zwar grundsätzlich die gleichen Anforderungen an die Fragebogengestaltung, die Fragenabfolge sowie die Formulierung der Fragen wie bei anderen, bewährten Befragungsmethoden. Ähnlich wie

bei schriftlichen Befragungen müssen Online-Befragungen allerdings so konzipiert und gestaltet werden, dass die Befragten sich auch ohne das Vorhandensein eines Interviewers im Fragebogen zurecht finden sowie die Fragen ohne diesen verstehen und beantworten können. Grundsätzlich gilt weiterhin, dass Online-Fragebögen immer auch für weniger erfahrene Internetnutzer gut handhabbar sein sollten, weshalb nach WELKER et al. neben etwa optisch hervorgehobenen Buttons und vordefinierten Eingabefeldern ebenfalls eine klare Filterführung sowie eindeutige und verständliche Instruktionen genutzt werden sollten (WELKER et al. 2005: 77). Auch wenn, wie oben bereits dargestellt wurde, bei der Zielpopulation der Untersuchung von einer gewissen Erfahrung mit dem Internet ausgegangen werden kann, wurden selbstverständlich diese grundsätzlichen Regeln dennoch eingehalten. Auch SCHNELL et al. führen aus, dass die Konzeption eines Online-Fragebogens ähnlich zu der Fragebogenkonzeption einer schriftlichen Befragung erfolgen sollte. Sie sowie THEOBALD geben zusätzlich zu bedenken, dass auf besonders aufwendige graphische Darstellungen und Animationen häufig verzichtet werden sollte. Ursächlich hierfür ist, dass es einerseits je nach verwendetem Internetbrowser oder Betriebssystem zu Darstellungsproblemen kommen kann. Andererseits können durch solche Gestaltungsmöglichkeiten die Ladezeiten einer Online-Befragung empfindlich verlängert werden, was sich negativ auf die Teilnahmebereitschaft bzw. die Nonresponse-Quote auswirkt (SCHNELL et al. 2005: 382f.; THEOBALD 2007: 106ff.).

Da die Zielsetzung dieser Arbeit hauptsächlich ein Vergleich zwischen einer Online-Befragung und einer mündlichen Befragung hinsichtlich des Antwortverhaltens sein sollte, wurde aufgrund dieser eher methodisch als inhaltlich ausgerichteten Fragestellung auf die Entwicklung eines eigenen Fragebogens verzichtet und stattdessen auf einen bereits bestehenden Fragebogen mit der Thematik „Carbon Labelling bei Lebensmitteln“ zurückgegriffen. Dieser Fragebogen wurde im Rahmen des am Fachbereich 09 der JLU angebotenen Master-Profilmoduls MP 74 „Demoskopische Methoden der Marktforschung“ von den Studierenden entwickelt und dort in einer mündlichen Befragung bereits angewendet. Der Fragebogen wurde ebenfalls ausgewählt, da er bereits die oben besprochenen Hinweise zur Formulierung von Fragen berücksichtigt.

Während im Modul dabei inhaltliche Fragestellungen zur „Carbon Labelling“-Thematik beantwortet werden sollten, dient dieser Fragebogen hier vor allem zur Beantwortung der unter 5.1 dargestellten Arbeitshypothesen und somit eher methodisch ausgerichteten Fragestellungen. Da es sich bei der Thematik des Fragebogens jedoch um ein aktuelles und zukunftsorientiertes Thema handelt, werden relevante und interessante, inhaltliche Ergebnisse des Fragebogens in der nachfolgenden Ergebnispräsentation dargelegt und besprochen.

Da die, im Rahmen des Moduls durchgeführte, Befragung nicht auf Studierende der JLU beschränkt worden war und zusätzlich bereits einige Zeit zurücklag, konnten die Ergebnisse zu einer Vereinfachung der Vorgehensweise im Rahmen der hier verfolgten, übergeordneten Fragestellung nicht herangezogen werden. Da sich der Fragebogen bereits in der Befragung, die im Rahmen des Moduls durchgeführt wurde, bewähren konnte, wurde auf einen umfangreichen Pre-Test verzichtet. Gänzlich konnte auf einen Pre-Test dennoch nicht verzichtet werden, da für die neuere Untersuchung verschiedene kleinere Änderungen am Fragebogen vorgenommen wurden.

So wurde bei verschiedenen Fragestellungen die Anzahl der vorgegebenen Antwortmöglichkeiten auf fünf bzw. sechs Vorgaben bzw. Items reduziert. Eliminiert wurden dabei die Antwortvorgaben, die in der studentischen Untersuchung bereits keine eindeutigen Ergebnisse gezeigt hatten bzw. deren Aussagekraft in einem Methodenvergleich als begrenzt betrachtet wurde. Ebenfalls sollte durch die Reduzierung der Antwortvorgaben die notwendige Bearbeitungsdauer des Fragebogens reduziert und dadurch vor allem im online durchgeführten Teil eine geringe Unit-Nonresponse-Quote erzielt werden. Ebenfalls wurden Änderungen im Abschnitt der soziodemographischen Variablen vorgenommen. Während diese in der ursprünglichen Untersuchung auf eine sehr breite gesellschaftliche Stichprobe ausgerichtet waren, sollten mit der neueren Untersuchung entsprechend der ausgewählten Zielpopulation nur Studierende der JLU Gießen befragt werden. Hierbei bot es sich an, die soziodemographischen Variablen an die veränderte Zielpopulation anzupassen. So wurden die Einkommensklassen, in denen sich die Befragten einordnen sollten, erwartbaren studentischen Einkommensverhältnissen angepasst. Neu aufgenommen wurden weiter soziodemographische Variablen zum Fachbereich, zum momentan angestrebten Abschluss sowie zum ehrenamtlichen Engagement. Um ebenfalls die Durchführung der Interviews für die studentischen Interviewer vereinfachen zu können, wurden im Vergleich zum Basis-Fragebogen noch Intervieweranweisungen eingebaut. Diese geben, soweit nötig, den Interviewern jeweils relevante Hinweise zu den verschiedenen Fragestellungen. Dabei wurde für die Intervieweranweisungen in Abgrenzung zu den Fragestellungen und Antwortvorgaben eine kursive Schrift gewählt, um diese eindeutig von den Fragestellungen des Fragebogens abgrenzen zu können.

Entsprechend den Hinweisen von SCHNELL et al. besteht der im Anhang abgedruckte Fragebogen für die F2F-Befragung aus mehreren Themenblöcken mit insgesamt 21 Fragen (inklusive Einstiegs-Frage sowie Gewinnspiel) (SCHNELL et al. 2005: 343f.). Während im ersten Themenmodul grundlegend verschiedene politische Themen sowie Klimaschutz im Allgemeinen thematisiert werden, wird im zweiten Themenmodul auf den Product Carbon Footprint

anhand verschiedener Fragestellungen im Detail eingegangen. In den beiden letzten Themenblöcken werden abschließend soziodemographische Angaben von den Befragten gefordert. So sollten die Befragten im dritten Themenblock etwa Fragen zu ihrem Einkaufsverhalten und abschließend in Themenblock vier Fragestellungen zu allgemeinen soziodemographischen Variablen beantworten.

Im Fragebogen kamen dabei mit Ausnahme einer Fragestellung, die als offene Frage formuliert wurde, ausschließlich geschlossene Fragestellungen zu Einsatz. Während bei offenen Fragen eine Antwort in den eigenen Worten des Befragten erwartet wird und entsprechend keine Antwortmöglichkeiten vorgegeben werden, muss bei geschlossenen Fragen (meist mindestens) eine der vorgegebenen Antworten ausgewählt werden. Weiter lassen sich je nachdem, welche Informationen durch eine Frage von Befragten gewonnen werden sollen, die vier grundsätzlichen Fragetypen „Einstellungs- bzw. Meinungsfragen“, „Überzeugungsfragen“, „Verhaltensfragen“ und „Fragen nach Befragteneigenschaften“ unterscheiden.

„Einstellungs- bzw. Meinungsfragen“ beziehen sich dabei auf eine positive oder negative Beurteilung bestimmter Statements und Items durch den Befragten, häufig unter Anwendung einer Ratingskala. „Überzeugungsfragen“ hingegen fragen danach, was Befragte für wahr oder falsch halten. Beziehen können sich Überzeugungsfragen nach SCHNELL et al. dabei einmal auf das Wissen der Befragten zu bestimmten Sachverhalten oder auch auf Problembereiche, zu denen niemand eine „richtige“ Antwort weiß. Während „Überzeugungsfragen“ ausschließlich auf Ansichten über kognitiv erfahrene Sachverhalte basieren, beziehen sich „Verhaltensfragen“ konkret auf eigene Erfahrungen oder eigenes Verhalten. Der vierte Fragentyp „Fragen zu Befragteneigenschaften“ entspricht den oben angesprochenen Fragestellungen zu soziodemographischen Variablen und umfasst im Allgemeinen Fragen nach demographischen und personalen Eigenschaften der Befragten wie etwa Geschlecht, Alter, Ausbildung, Beruf, Einkommen, Familienstand und ähnlichem. Über diese Variablen wird grundsätzlich versucht, statistische Zusammenhänge zwischen demographischen Eigenschaften von Personen und ihren Einstellungen, Überzeugungen und Verhaltensweisen zu ermitteln (DILLMAN 1978: 80ff.; SCHNELL et al. 2005: 325ff.; ATTESLANDER 2008: 140ff.; DIEKMANN 2008: 471ff.)

Um einen Methodenvergleich durchführen zu können, musste der Fragebogen neben der Papier-Version ebenfalls in einer Online-Version programmiert werden. Hierfür kann auf spezielle Software-Pakete wie etwa GlobalPark oder LimeSurvey zurückgegriffen werden, die für die Durchführung von Online-Befragungen entwickelt wurden. Im Rahmen der vorliegenden Untersuchung wurde dabei LimeSurvey ausgewählt, da dieses beispielsweise im Ver-

gleich zu GlobalPark über das Hochschulrechenzentrum der JLU kostenlos genutzt werden kann (KOMM 2009).

Bei LimeSurvey handelt es sich um eine Softwarelösung, welche die Darstellung verschiedenster Fragetypen erlaubt. Aufgrund dieser vielfältigen Darstellungsmöglichkeiten und geringen Restriktionen im Hinblick auf die Anzahl der Fragen und Antwortmöglichkeiten können daher mit LimeSurvey auch aufwendige Online-Befragungen realisiert werden (LIMESURVEY 2009). Ebenfalls als Vorteil von LimeSurvey kann die intuitive Handhabbarkeit bei der Fragebogenprogrammierung, die gute Dokumentation sowie der einfache Datenexport in andere Programme wie Microsoft Excel oder SPSS<sup>18</sup> betrachtet werden. Auch bei der Darstellung des fertigen Fragebogens zeichnet sich LimeSurvey einerseits durch eine hohe Kompatibilität an verschiedene Browser<sup>19</sup> sowie Bildschirmauflösungen aus, andererseits auch durch eine einfache und intuitive Bedienbarkeit, wie Abbildung 10 darstellt.

**Abbildung 10: Darstellung einer Frage in LimeSurvey**

Carbon Labelling und dessen Bedeutung beim Lebensmittelkauf

0%  100%

Teil 1

\*Nachfolgend werden Dir nun verschiedene politische Aufgabenbereiche in Deutschland vorgestellt. Bitte sage mir anhand der Skala, wie wichtig Dir persönlich eine zusätzliche Stärkung dieser Aufgabenbereiche jeweils ist. Eine 1 bedeutet "sehr wichtig", eine 5 hingegen bedeutet "unwichtig", dazwischen kannst Du Deine Meinung abstimmen.

	sehr wichtig 1	wichtig 2	etwas wichtig 3	weniger wichtig 4	unwichtig 5
Bekämpfung von Kriminalität	<input type="radio"/>				
Schutz von Tier- und Pflanzenarten	<input type="radio"/>				
Maßnahmen zum Klimaschutz	<input type="radio"/>				
Verbesserung des Zusammenlebens mit Ausländern	<input type="radio"/>				
Förderung der Wirtschaft	<input type="radio"/>				
Bekämpfung der Arbeitslosigkeit	<input type="radio"/>				

Ein 1 bedeutet "sehr wichtig", eine 5 hingegen bedeutet "unwichtig". Dazwischen kannst Du Deine Meinung abstimmen.

Weiter >>

[Umfrage verlassen und löschen]

Quelle: Eigene Aufnahme.

Um mögliche Verzerrungen im Antwortverhalten durch unterschiedliche Frageformulierungen bzw. eine stark abweichende Darstellungsweise des Online-Fragebogens im Vergleich

<sup>18</sup> Mit Erscheinen der Programmversion 17 wurde SPSS von der Herstellerfirma SPSS Inc. in PASW bzw. PASW Statistics umbenannt. PASW steht dabei für Predictive Analytics Software (SPSS 2009). Wegen der hohen Bekanntheit der Bezeichnung SPSS wird diese jedoch in der vorliegenden Arbeit weiterverwendet.

<sup>19</sup> Die Darstellung des Online-Fragebogens wurde mit Hilfe der Internetbrowser Mozilla Firefox 3.5, Microsoft Internet Explorer 8 sowie Apple Safari 4 überprüft.

zum Fragebogen für die F2F-Befragung vermeiden zu können, wurde bei der Umsetzung des Fragebogens in LimeSurvey stark auf eine gleich lautende Fragenformulierung sowie auf eine Darstellungsweise geachtet, die sich im Vergleich zum F2F-Fragebogen in möglichst wenigen Details unterscheidet. Als weiteren Vorteil bietet LimeSurvey die Möglichkeit, vorgegebene Antwortmöglichkeiten bei geschlossenen Fragestellungen zu randomisieren, also in unterschiedlicher Reihenfolge darstellen zu können. Hierdurch kann dem in Kapitel 3.3.3 dargestellten Halo-Effekt sowie dem daraus ebenfalls resultierenden Methodeneffekt ebenfalls entgegengewirkt werden.

Da die Problematik des Halo-Effektes nicht nur die Online-Befragung, sondern auch die F2F-Interviews betrifft, musste hierfür eine Lösung gefunden werden, um daraus resultierende Verzerrungen vermeiden zu können. Daher wurden entsprechende Hinweise für die Interviewer bei den Intervieweranweisungen gegeben: Neben dem Hinweis für die Interviewer, dass bei den Fragenbatterien die Reihenfolge der verschiedenen Items mit jedem Interview zu variieren sei, wurde noch eine zusätzliche Spalte bei den Fragenbatterien eingebaut, in welcher vom Interviewer das jeweilige Start-Item markiert werden sollte. Diese zusätzlich eingefügte Spalte verfügt dabei über zwei wichtige Funktionen: Zum einen fungiert sie als Erinnerungstütze für die Interviewer, mit jedem Interview den Startpunkt innerhalb der Fragenbatterien zu wechseln, zum anderen wird durch die Notwendigkeit, den jeweiligen Startpunkt zu markieren, eine gewisse Kontrollfunktion ausgeübt.

#### **5.2.4 Durchführung der Befragungen**

Bereits während der ersten, frühzeitigen Planungen zur Durchführung eines Methodenvergleichs entsprechend der Zielsetzung dieser Arbeit konnte der damit verbundene, erhebliche Aufwand abgeschätzt werden. Daher wurde relativ frühzeitig entschieden, die F2F-Interviews mit Hilfe von studentischen Interviewern durchzuführen. Gewonnen werden konnten die studentischen Interviewer über eine Ausschreibung per E-Mail, die über den E-Mail-Verteiler des Dekanats des Fachbereichs 09 „Agrarwissenschaften, Ökotrophologie und Umweltmanagement“ versendet wurde. Eine Begrenzung auf den Fachbereich 09 bei der Interviewerauswahl wurde vorgenommen, da hier gewisse notwendige Fachkenntnisse und thematische Zusammenhänge entweder als bereits vorhanden vorausgesetzt werden konnten oder aber relativ schnell aufgrund des Vorwissens hätten erarbeitet werden können. Da erwartet wurde, dass vor allem intrinsisch motivierte Studierende mit einem Interesse an der Thematik des Carbon Labellings zu einem Erfolg der F2F-Befragungen hinsichtlich der Auswahl der Befragten, der Sorgfalt der Durchführung und der Dokumentation beitragen würden, wurde explizit darauf hingewiesen, dass eine monetäre Entlohnung nicht möglich sei. Stattdessen wurde jedem In-

interviewer ein qualifiziertes Arbeitszeugnis bei erfolgreicher Durchführung der Interviews sowie ein SPSS-Einführungsseminar in Aussicht gestellt. Mit der Anwerbung studentischer Interviewer ist auch der Vorteil verbunden, dass dadurch in einem fest definierten Zeitrahmen eine größere Stichprobe an verschiedenen Orten gleichzeitig gezogen werden konnte.

Mit der Entscheidung für eine interviewergestützte Befragung verbunden war allerdings die Notwendigkeit von zusätzlichen Vorarbeiten. So ist nach SCHNELL et al. eine Interviewerschulung unerlässlich, wenn, wie in der vorliegenden Arbeit, freiwillige Studierende als Laien-Interviewer zum Einsatz kommen. Die Verwendung von freiwilligen Laien-Interviewern anstatt eines professionellen Interviewerstabes stellt dabei in der akademischen Forschung eine typische Vorgehensweise dar (SCHNELL et al. 2005: 351f.). Entsprechend den Hinweisen von FRIEDRICHS und FOWLER umfasste die an zwei Nachmittagen durchgeführte Interviewerschulung folgende Bereiche:

- Erläuterung der geplanten Untersuchung,
- Erläuterung des Fragebogens (einschließlich aller Fragen und Besonderheiten einzelner Fragen),
- Erläuterungen zur Dokumentation der gegebenen Antworten,
- Einübung der Kontaktaufnahme mit Befragten, sowie
- Verhaltensregeln für die Erhebungssituation (wie beispielsweise Neutralität im verbalen als auch nonverbalen Verhalten) (FRIEDRICHS 1973: 214; FOWLER 1984: 115).

Neben der theoretischen Abhandlung des Fragebogens enthielt die Interviewerschulung auch praktische Elemente in Form von mehreren Fragenrunden sowie Übungsinterviews. Während in der Interviewerausschreibung noch das Carbon Labelling und dessen Bedeutung beim Lebensmittelkauf als relevantes Untersuchungsthema angegeben wurde, ist den studentischen Interviewern erst während der Schulungsaktion der Vergleich der zwei Befragungsmethoden als wirkliches Untersuchungsziel präsentiert und erläutert worden. Auch wurden sie darauf hingewiesen, den wahren Untersuchungsgegenstand während der gesamten Untersuchung für sich zu behalten. Dieses Vorgehen wurde gewählt, um den Methodenvergleich als Zielsetzung der Untersuchung nicht publik zu machen, um eventuell aus einer Offenlegung der Zielsetzung resultierende Verzerrungen vermeiden zu können. Dass dieser Sachverhalt nicht vollkommen irrelevant ist, zeigt folgende Überlegung: Die Interviewerausschreibung wurde an alle Studierende des Fachbereichs 09 versendet, die nach der offiziellen Studierendenstatistik einen Anteil von etwa 10 Prozent an den gesamten Studierenden der JLU haben (JLU 2009e: 2). Da soziale Kontakte nicht nur auf den eigenen Fachbereich begrenzt sind, ist eine Weitergabe der Zielsetzung des Methodenvergleichs durchaus denkbar und eine bewusste Teilhabe

an beiden Teilstichproben und somit eine bewusste Beeinflussung des Ergebnisses nicht auszuschließen.

Entsprechend der Theorie zur oben dargestellten Split-Ballot-Technik müssen zur Durchführung eines Methodenvergleichs zwei Teilstichproben aus einer definierten Grundgesamtheit gezogen werden. Um die Aussagekraft der Ergebnisse gewährleisten zu können, muss dabei, den Überlegungen zur Split-Ballot-Technik folgend, sichergestellt werden, dass jedes Element der Grundgesamtheit nur in eine der beiden Teilstichproben jeweils gelangen kann. Auf die dieser Arbeit zu Grunde liegende Untersuchung bezogen bedeutet das, dass sichergestellt werden musste, dass jeder Studierende der JLU nur einmal an einer der beiden Befragungen teilnehmen konnte. Eine dahingehende Gestaltung der Untersuchung, die dies sicherstellen konnte, offenbarte sich als äußerst schwierig. Da Befragungen zum Schutze der Befragten häufig anonymisiert durchgeführt werden (siehe hierzu DIEKMANN 2008: 85ff.), fiel die Elimination von Studierenden, die an beiden Befragungen teilgenommen haben, beispielsweise anhand von Vor- und Nachnamen und/oder E-Mail-Adressen, als Möglichkeit heraus. Daher musste auf eine alternative Vorgehensweise zurückgegriffen werden. Während zu Untersuchungsbeginn noch eine simultane Durchführung der beiden Befragungen angedacht worden war, wurde aufgrund der beschriebenen Stichprobenproblematik diese geplante Vorgehensweise zugunsten zweier aufeinander folgender Untersuchungswellen aufgegeben. Zwar erlaubt diese sequentielle Vorgehensweise ebenfalls keine eindeutige Identifizierung redundanter Elemente in den beiden Teilstichproben und somit von Studierenden, die an beiden Befragungen teilgenommen haben, allerdings wurde durch entsprechende Interviewerhinweise sowie eine zusätzliche Filterfrage im Online-Fragebogen versucht, der Problematik dadurch begegnen zu können. So wurde am Ende des F2F-Fragebogens in den Intervieweranweisungen der Hinweis für die Interviewer eingebaut, die Befragten darauf hinzuweisen, aufgrund spezifischer Gründe bitte nicht mehr an einer, in den nächsten Wochen folgenden, Online-Befragung zur gleichen Thematik teilzunehmen. Auf die Relevanz dieses Punktes wurde explizit in der Interviewerschulung noch einmal eingegangen sowie in einem Rollenspiel geübt.

In der Online-Befragung wurde den eigentlichen Fragestellungen des Fragebogens eine Filterfrage vorgeschaltet, um der oben angesprochenen Problematik der doppelten Teilnahme an beiden Befragungsaktionen besser begegnen zu können. In der Filterfrage wurde entsprechend abgefragt, ob der Befragte bereits an einem mündlichen Interview zur Thematik teilgenommen habe. Falls nicht, wurde die Online-Befragung mit den relevanten Fragestellungen fortgesetzt, ansonsten wurde die Befragung beendet.

Neben der Möglichkeit der Teilnahme an beiden Teilbefragungen muss ebenfalls die Möglichkeit betrachtet werden, mehrfach an einer der beiden Teilbefragungen teilzunehmen. Während eine Mehrfachteilnahme bei den F2F-Interviews grundsätzlich nicht auszuschließen ist, allerdings aufgrund der pro Standort gleichbleibenden Interviewer als eher unwahrscheinlich betrachtet werden kann, musste vor allem bei der Online-Befragung aufgrund des einfachen Zugangs versucht werden, eine Mehrfachteilnahme auszuschließen. Hierfür bietet LimeSurvey verschiedene Möglichkeiten, wie etwa das Setzen von Cookies, an, um mehrfache Teilnahmen an einer Online-Befragung zu verhindern. Diese Möglichkeit wurde bei der Programmierung des Fragebogens berücksichtigt, wodurch entsprechend eine Mehrfachteilnahme bei der Online-Befragung ausgeschlossen werden konnte.

Wie oben bereits angesprochen wurde, sind die beiden zu vergleichenden Teilbefragungen nicht parallel, sondern sequentiell durchgeführt worden. Begonnen wurde dabei mit der F2F-Befragung, die aufgrund des Beginns der vorlesungsfreien Zeit im geplanten Befragungszeitraum zuerst forciert wurde. Der Zeitraum für die Durchführung der F2F-Interviews erstreckte sich vom 01.07.2009 bis zum 31.07.2009. Als Standorte für die Interviews wurden dabei verschiedene Einrichtungen der JLU wie etwa Bibliotheken oder auch des Studentenwerks ausgewählt:

- Große Mensa, Otto-Behagel-Straße;
- Hochschulrechenzentrum/ Zweigbibliothek im Chemikum, Heinrich-Buff-Ring;
- Universitätsbibliothek (UB), Otto-Behagel-Straße;
- Universitätshauptgebäude, Ludwigstraße;
- Zweigbibliothek im Philosophikum II, Karl-Glöckner-Straße;
- Zweigbibliothek Recht und Wirtschaft, Licher Straße;
- Zeughaus, Senckenbergstraße.

Über die Durchführung der Befragungen an diesen verschiedenen Standorten in Gießen sollte sichergestellt werden, dass alle Fachbereiche mit der Befragung relativ gut erreicht werden konnten. Um dabei eine möglichst heterogene Verteilung der Interviewer zwischen den verschiedenen Befragungsstandorten und an den verschiedenen Befragungstagen sicherzustellen, wurden diese vom Autor entsprechend disponiert. Zusätzlich zu Hilfsmitteln, wie Klemmbrettern oder auch Stiften bekamen die Interviewer ebenfalls Namensschildern ausgehändigt, die sie während ihrer Interviewertätigkeit am entsprechenden Befragungsort tragen sollten. Hierdurch sollte einerseits das Interesse der Studierenden an der Befragung gestärkt werden, andererseits der seriöse Charakter der Untersuchung unterstrichen werden.

Für die Durchführung der Interviews sollten von den studentischen Interviewern „typische“ Studierende ausgewählt werden. Diese wurden von den Interviewern gezielt angesprochen, auf ihre Zugehörigkeit zu der in Kapitel 5.2.2 dargestellten Zielgruppe überprüft sowie ihre Bereitschaft zur Teilnahme an der Befragung zur Thematik „Carbon Labelling und dessen Bedeutung beim Lebensmittelkauf“ eingeholt. Weiter wurde den Befragten an dieser Stelle für die vollständige Beantwortung des Fragebogens ein Schokoriegel sowie eine Gewinnmöglichkeit in Form eines Amazon.de-Gutscheins in Aussicht gestellt. Neben der Zielsetzung, dadurch die Affinität der angesprochenen Studierenden zur Teilnahme an der Befragung zu erhöhen und möglichst vollständig ausgefüllte Fragebögen zu erhalten, wurde bewusst auch nicht ausgeschlossen, dadurch weitere interessierte Studierende zu einer Teilnahme an den Interviews bewegen zu können, die nicht direkt angesprochen wurden. In der Literatur werden für Incentives zwar unterschiedliche Wirkungen diskutiert (COOK et al. 2000: 826ff.; DILLMAN 2007: 151), allerdings wurde wie in der Umfragepraxis üblich, auch in dieser Untersuchung von einer Erhöhung der Antwortrate bei Anreizvergabe ausgegangen und daher Incentives eingesetzt (PÖTSCHKE 2009: 84).

Aufgrund der relativ großen Anzahl von Itembatterien mit unterschiedlichen Skalen wurden den Befragten zusätzlich laminierte Ausdrücke der unterschiedlichen Skalen in die Hand gegeben. Diese sollten zusammen mit der vollständigen verbalen Erklärung der jeweiligen Skalen durch den Interviewer zu einer Vereinfachung für den Befragten führen. Die vier verwendeten Skalen sowie eine zusätzlich verwendete Darstellung mit Labels sind beispielhaft im Anhang abgebildet.

Insgesamt wurden von den studentischen Interviewern im Befragungszeitraum 189 Studierende der JLU Gießen befragt. Die angestrebte Mindestzahl an Interviews in Höhe von 120 Interviews konnte somit deutlich überboten werden.

Nach der Beendigung der F2F-Interviews wurde die Online-Befragung für die Teilnahme geöffnet. Im Unterschied zur F2F-Befragung, bei der von Anfang an aufgrund des enormen Erhebungsaufwands ausschließlich eine begrenzte Stichprobengröße eingeplant wurde, sollte die online durchgeführte Vergleichsbefragung mit einer deutlich umfangreicheren Stichprobe durchgeführt werden. Die Einladung zur Online-Befragung wurde am 06.08.2009 über den E-Mail-Verteiler „Hinweise der JLU“ an alle Studierenden der JLU Gießen versendet<sup>20</sup>.

Bei Teilnahme aller Studierenden, abzüglich derer, die bereits an den F2F-Interviews teilgenommen hatten, wäre somit theoretisch eine Vollerhebung möglich gewesen. Die E-Mail ent-

---

<sup>20</sup> Nach Angaben des Hochschulrechenzentrums HRZ wurde die Einladung zur Online-Befragung an 22.281 Studierende, inklusive denen in Graduierten- und Promotionsprogrammen, versendet (KÜTH 2009).

hielt dabei neben einem Text, in dem der Verfasser, die Zielsetzung sowie die voraussichtliche Teilnahmedauer genannt wurden, einen Link, der direkt zur Online-Befragung führte. Um auch bei der Online-Befragung die Teilnahmebereitschaft zu steigern, wurde ebenfalls wieder die Verlosung eines Amazon.de-Einkaufsgutscheins unter denjenigen Teilnehmern in Aussicht gestellt, die den Online-Fragebogen vollständig ausgefüllt haben. Auf die Gewinnmöglichkeit wurde dabei sowohl in der Einladungs-E-Mail als auch auf dem Startbildschirm der Befragung noch einmal hingewiesen.

Möglich war eine Teilnahme an der Online-Befragung im Zeitraum vom 06.08.2009 bis zum 31.08.2009. In den letzten Tagen dieses Zeitraumes konnten jedoch keine weiteren Teilnahmen an der Online-Befragung mehr registriert werden, weshalb die Umfrage, die ursprünglich wie die F2F-Interviews ebenfalls 31 Tage zur Teilnahme geöffnet sein sollte, beendet wurde. Dieser Zeitraum erscheint dabei ausreichend, um mögliche Verzerrungen durch eine zu kurze Feldzeit der Online-Befragung vermeiden zu können. So gibt GADEIB an, dass bei Feldzeiten von sieben Tagen und mehr bei Online-Befragungen nicht mehr mit einer derartigen Verzerrung zu rechnen ist (GADEIB 2005). Für einen geringen Einfluss des vorzeitigen Abbruchs spricht auch das Ergebnis von BAUR und FLORIAN. Diese konnten anhand einer Online-Befragung, die an drei deutschen Hochschulen durchgeführt wurde, zeigen, dass nur innerhalb der ersten Tage Studierende an der Online-Befragung teilnahmen. Je weiter die Einladung zur Teilnahme zurücklag, desto geringer war die Wahrscheinlichkeit, dass überhaupt noch teilgenommen wurde (BAUR und FLORIAN 2009: 120f.).

An der Online-Befragung haben im oben genannten Zeitraum insgesamt 652 Studierende teilgenommen. Auf die 22.281 Empfänger der Rundmail „Hinweise der JLU“ bezogen, entspricht dies einer Rücklaufquote von etwa drei Prozent. In der Literatur werden zur Rücklaufquote sehr unterschiedliche Leitzahlen genannt. So werden bei schriftlichen Befragungen unter Verwendung von Erinnerungsschreiben teilweise Rücklaufquoten von etwa 20 bis 30 Prozent genannt. Aufgrund der anonymen Ausgestaltung des verwendeten Online-Fragebogens wären solche häufig herangezogenen Erinnerungsschreiben zur Erhöhung der Rücklaufquote jedoch nicht einsetzbar gewesen. Ein möglicher Erklärungsansatz für die relativ geringe Rücklaufquote könnte dabei im Rahmen dieser Untersuchung die starke Nutzung des „Hinweise der JLU“-Verteilers für verschiedenste Ankündigungen und Hinweise sein. Daraus resultierend wird unter Umständen der einzelnen E-Mail mit einem Hinweis (wie in diesem Fall der Einladung zur Online-Befragung) nur noch ein geringes Interesse geschenkt. Entsprechend zu den Ausführungen oben sinkt damit verbunden auch die Wahrscheinlichkeit der Reaktion auf eine E-Mail aus dem Verteiler im Zeitablauf stark.

Insgesamt konnten über beide Untersuchungsmethoden hinweg letztlich 841 Studierende befragt werden.

### **5.2.5 Aufbereitung und Auswertung der Daten**

Nach Beendigung der beiden Befragungen mussten vor der eigentlichen Datenanalyse die Daten zuerst aufbereitet werden. Hierfür wurden entsprechend eines, für den Fragebogen erstellten Codeplans die gegebenen Antworten der F2F-Interviews in die Statistiksoftware SPSS (Statistical Package for Social Sciences) eingegeben und auf Vollständigkeit, Konsistenz, Verständlichkeit sowie Vergleichbarkeit hin überprüft. Zielsetzung dieser Vorgehensweise ist es dabei, offenkundig fehlerhafte Daten, die beispielsweise aus einer bewussten Fehlbeantwortung der Fragen her rühren, aus dem Datensatz eliminieren zu können (KUB 2007: 167ff.). Von den 189 F2F-Interviews erwiesen sich dabei zwei aufgrund von unvollständigen Angaben bei den soziodemographischen Angaben als nicht verwendbar und wurden entsprechend aus dem Datensatz eliminiert. Ein weiterer Fall musste ebenfalls ausgeschlossen werden, da hier eine Promotion als momentan angestrebter Studienabschluss angegeben wurde. Da Studierende in einem Promotionsstudium nicht zur definierten Zielpopulation zählen (siehe hierzu Kapitel 5.2.2), blieben nach Eliminierung dieses Falls insgesamt somit noch 186 Datensätze der Teilstichprobe F2F übrig.

Für eine Überprüfung und Auswertung der online mit Hilfe der Befragungssoftware LimeSurvey erhobenen Datensätze mussten diese nach Beendigung der Online-Befragung zuerst in SPSS exportiert werden. Auch hier mussten bei der Überprüfung auf Vollständigkeit, Konsistenz, Verständlichkeit sowie Vergleichbarkeit mehrere Datensätze ausgeschlossen werden: So füllten von den insgesamt 652 Teilnehmern der Online-Befragung 517 den Fragebogen im Internet vollständig aus, was einer Quote von 79,3 % entspricht. Davon musste ein weiterer Datensatz eliminiert werden, da von der entsprechenden Person bei der vorgeschalteten Filterfrage angegeben wurde, bereits an einem der vorausgegangenen F2F-Interviews teilgenommen zu haben. Diese Eliminierung wurde dabei als unbedingt notwendig erachtet, um eine aus der Zugehörigkeit zu beiden Teilstichproben resultierende Ergebnisverzerrung vermeiden zu können. Zwei weitere Teilnehmer der Online-Befragung mussten bei der Überprüfung der Daten ebenfalls aus dem endgültigen Datensatz herausgenommen werden, da sie, simultan zu den oberen Ausführungen, als Studierende in Graduierten- bzw. Promotionsprogrammen nicht zur definierten Grundgesamtheit gehörten. Abzüglich dieser drei Fälle umfasste die zum Methodenvergleich herangezogene Online-Stichprobe letztlich 514 Fälle. Für den Methodenvergleich konnten somit insgesamt 700 Fälle herangezogen werden.

Die weitere Auswertung der Daten zur Annahme oder Ablehnung der aufgestellten Hypothesen erfolgte mit SPSS 15 und 18 sowie dem Programm G\*Power<sup>21</sup> 3.1. Zur Darstellung der Ergebnisse in Graphiken und Abbildungen wurde weitestgehend Microsoft Excel, vorwiegend in der Version 2003, verwendet.

### **5.3 Ergebnisse zu den soziodemographischen Variablen**

Nachfolgend werden in diesem Kapitel die verschiedenen Ergebnisse der Untersuchung dargestellt, begonnen wird dabei mit den soziodemographischen Daten der Untersuchung. Im Anschluss daran wird auf die Dauer der Fragenbeantwortung eingegangen. Kapitel 5.3.3. beschäftigt sich anschließend mit der Notwendigkeit, eine Gewichtung des Datensatzes vorzunehmen. Abgeschlossen wird das Kapitel durch die Darstellung der Ergebnisse einer zurückgestellten Fragestellung zum Einkommen.

#### **5.3.1 Soziodemographische Daten**

Im vierten und somit letzten Teil des Fragebogens wurden verschiedene soziodemographische Variablen erfasst. Mit der Darstellung der Verteilungen dieser Variablen soll dabei begonnen werden. Einmal wurden die Teilnehmer dabei zu allgemeinen soziodemographischen Variablen wie Alter und Geschlecht sowie zu studiumsbezogenen soziodemographischen Variablen wie Fachbereichszugehörigkeit und angestrebtem Studienabschluss befragt. Ebenfalls wurden die Befragten in den beiden Befragungen jeweils gebeten, Angaben zu ihrem ehrenamtlichen Engagement sowie zum monatlichem Nettoeinkommen zu machen.

Die nachfolgend präsentierten Ergebnisse der soziodemographischen Variablen beziehen sich dabei nur auf die Teilnehmer, die entweder in der Online-Befragung den Fragebogen vollständig ausgefüllt haben oder bei den F2F-Interviews das Interview nicht vorzeitig abgebrochen haben. Bei der Ergebnispräsentation der jeweiligen soziodemographischen Variablen wird weiterhin so vorgegangen, dass zuerst auf die Häufigkeitsverteilung in der Gesamtstichprobe, also der zusammengesetzten Stichprobe aus F2F- und Online-Sample eingegangen wird. Da sich möglicherweise zwischen den beiden Teilstichproben bereits deutliche Unterschiede in den Häufigkeitsverteilungen zeigen, findet zusätzlich jeweils eine Gegenüberstellung der Teilergebnisse statt. Durch diese Vorgehensweise lassen sich möglicherweise bereits relevante Schlüsse für einen Methodenvergleich zwischen einer F2F- und einer Online-Erhebung ziehen.

---

<sup>21</sup> Bei G\*Power handelt es sich um ein frei verfügbares Programm zur Analyse von Teststärken (FAUL et al. 2007: 175ff.).

Von den insgesamt 700 befragten Studierenden in beiden Teilbefragungen mit einem vollständigen Fragebogen<sup>22</sup>, waren 503 Studierende weiblich und 197 Studierende männlich. Der Anteil an weiblichen Studierenden liegt entsprechend mit 71,9 % deutlich über dem Anteil der männlichen Studierenden mit 28,1 %. Deutliche Unterschiede zeigen sich dabei auch in der Häufigkeitsverteilung zwischen beiden Teilstichproben, wie auch Tabelle 5 entnommen werden kann.

**Tabelle 5: Häufigkeitsverteilung der Variablen "Geschlecht" in den Teilstichproben sowie in der Grundgesamtheit der JLU**

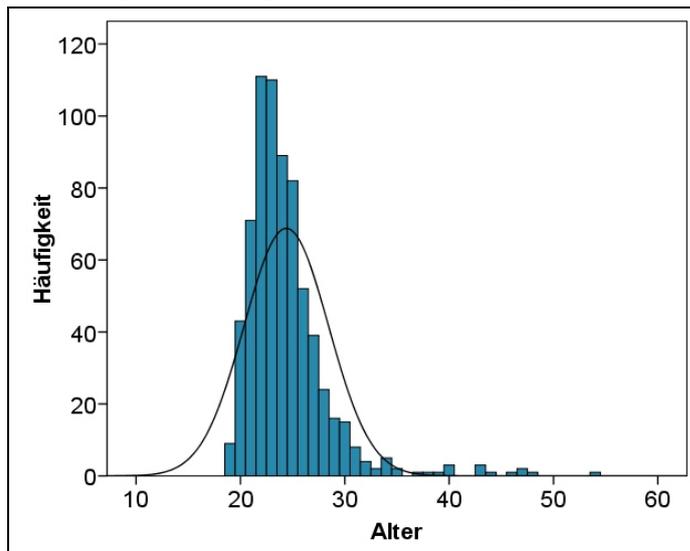
	F2F-Stichprobe		Online-Stichprobe		JLU-Grundgesamtheit
	Häufigkeit	Anteil	Häufigkeit	Anteil	Anteil
männlich	85	45,7 %	112	21,8 %	35,6 %
weiblich	101	54,3 %	402	78,2 %	64,4 %
<b>Gesamt</b>	<b>186</b>	<b>100,0 %</b>	<b>514</b>	<b>100,0 %</b>	<b>100,0 %</b>

Quelle: Auswertung der eigenen Erhebung, eigene Berechnungen auf Grundlage JLU 2009e: 2ff..

Während in den F2F-Interviews die Verteilung zwischen beiden **Geschlechtern** mit 45,7 % männlichen und 54,3 % weiblichen Studierenden noch relativ ähnlich ist, zeigen sich in der Online-Befragung sehr deutliche Abweichung davon: So waren hier 78,2 % derer, die den Online-Fragebogen vollständig ausgefüllt haben oder an einem Interview bis zum Schluss teilgenommen haben, weiblich und lediglich 21,8 % männlich. Zur Einordnung der dargestellten Ergebnisse werden in Tabelle 5 zusätzlich die Anteile der männlichen und weiblichen Studierenden an der JLU angegeben. Zwar ist auch hier der Anteil der weiblichen Studierenden deutlich höher als der Anteil der männlichen, allerdings ist die Ungleichverteilung zwischen beiden Geschlechtern weniger stark ausgeprägt als in der online durchgeführten Befragung.

Ebenfalls wurden die Studierenden im vierten Teil des Fragebogens zu ihrem **Alter** befragt. Daher enthält die Variable „Alter“ das von den Studierenden geäußerte Alter zum Zeitpunkt der Befragung. Zu dieser Fragestellung gaben insgesamt 697 Studierende Auskunft. In der aus beiden Teilstichproben zusammengeführten Gesamtstichprobe nimmt die Variable dabei Werte von 19 bis 54 Jahren an und weist ein arithmetisches Mittel von 24,39 Jahren auf. Der Median, der die Stichprobe in zwei gleich große Hälften aufteilt, liegt bei 24,00 Jahren. Für die Standardabweichung konnte weiter ein Wert von 4,04 und für die Varianz ein Wert von 16,34 berechnet werden. Die Verteilung der Variablen „Alter“ in der gezogenen Gesamtstichprobe wird nachfolgend durch das Histogramm in Abbildung 11 dargestellt.

<sup>22</sup> Hierzu werden selbstverständlich entsprechend den oberen Ausführungen auch Fragebögen von Studierenden gezählt, bei denen bestimmte Fragen bewusst ausgelassen wurden. Vielmehr bezieht sich diese Formulierung auf vorzeitig abgebrochene Interviews bzw. unvollständig ausgefüllte Online-Fragebögen.

**Abbildung 11: Histogramm zur Verteilung der Variablen "Alter" in der Gesamtstichprobe (n=697)**

Quelle: Auswertung der eigenen Erhebung.

Bei einem Vergleich der beiden Teilstichproben hinsichtlich der Variablen „Alter“ zeigen sich weiter geringfügige Unterschiede in der deskriptiven Statistik, wie Tabelle 6 darstellt. So nimmt die Variable in der offline erhobenen Stichprobe die Werte von 19 bis 47 Jahren an, in der Online-Stichprobe sogar Werte von 19 bis 54 Jahren. Auch beim Mittelwert sind mit 24,04 Jahren beim F2F-Sample und 24,52 Jahren beim Online-Sample geringfügige Änderungen erkennbar: So waren die Teilnehmer der Online-Befragung dem arithmetischen Mittel nach im Durchschnitt geringfügig älter als die Teilnehmer der F2F-Interviews.

**Tabelle 6: Deskriptive Statistik der Variablen "Alter", unterteilt nach Befragungsmodus**

	F2F-Stichprobe (n=186)	Online-Stichprobe (n=511)
Minimum	19	19
Maximum	47	54
Arithm. Mittel	24,04	24,52
Median	23	24
Getr. Mittel	23,68	24,04
SD	3,63	4,18
Varianz	13,18	17,46

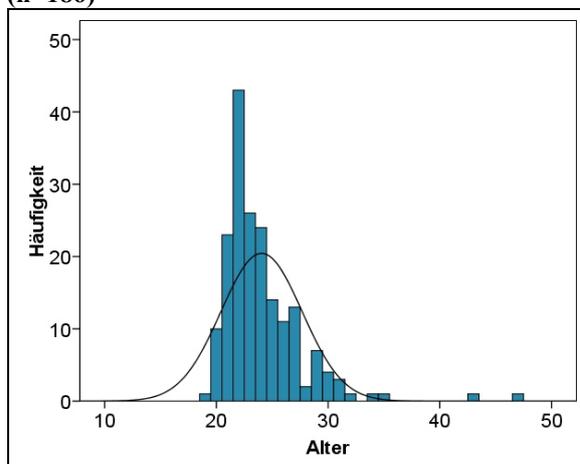
Anmerkungen: Arithm. Mittel = Arithmetisches Mittel; Getr. Mittel = 5% Getrimmtes Mittel; SD = Standardabweichung.

Quelle: Auswertung der eigenen Erhebung.

Als ein Problem des arithmetischen Mittels kann jedoch dessen Störanfälligkeit für Ausreißer angesehen werden (CLEFF 2008: 42; TOUTENBURG und HEUMANN 2008: 60). Daher werden zusätzlich zur Beschreibung der Verteilung der Variablen die beiden Parameter Median sowie

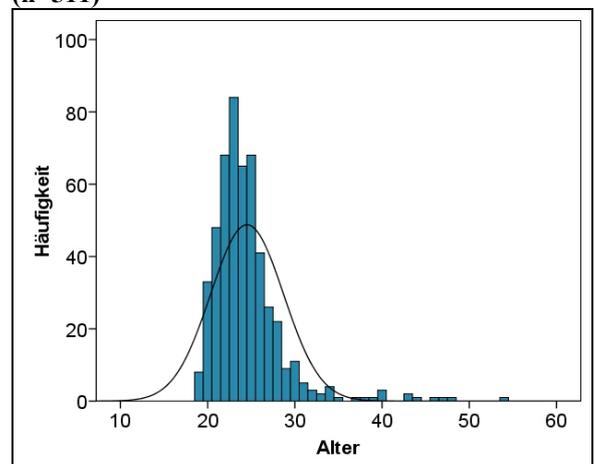
das 5% Getrimmte Mittel<sup>23</sup> herangezogen. Anhand dieser beiden statistischen Kennzahlen ist ebenfalls erkennbar, dass die Teilnehmer der Online-Befragung im Vergleich zu denen der F2F-Befragung eine etwas höhere Altersstruktur aufweisen. Die Häufigkeitsverteilungen der Variablen „Alter“ für die beiden Teilstichproben werden nachfolgend in Abbildung 12 sowie Abbildung 13 dargestellt. Da verschiedene statistische Testverfahren eine Normalverteilung der einbezogenen Variablen voraussetzen, wurde die Verteilung der Variablen „Alter“ in der Gesamt- sowie in den Teilstichproben jeweils mit Hilfe des Kolmogorov-Smirnov-Tests<sup>24</sup> auf Normalverteilung überprüft (siehe hierzu Tabelle A 1 und Tabelle A 2 im Anhang).

**Abbildung 12: Histogramm zur Verteilung der Variablen "Alter" in der F2F-Stichprobe (n=186)**



Quelle: Auswertung der eigenen Erhebung.

**Abbildung 13: Histogramm zur Verteilung der Variablen "Alter" in der Online-Stichprobe (n=511)**



Quelle: Auswertung der eigenen Erhebung.

Hierbei konnte jeweils eine Signifikanz von 0,000 ermittelt werden. Dies kann dahingehend interpretiert werden, dass eine Normalverteilung sowohl in den beiden Teilstichproben als auch der zusammengeführten Gesamtstichprobe nicht gegeben ist.

Ob weiterhin signifikante Unterschiede in den Häufigkeitsverteilungen zwischen den beiden Teilstichproben vorliegen, wurde mit Hilfe des Kolmogorov-Smirnov Z-Tests<sup>25</sup> überprüft. Hier konnte eine exakte Signifikanz von 0,028 ermittelt werden (siehe Tabelle A 3 im Anhang). Die im Test formulierte Nullhypothese, dass beide Stichproben aus Grundgesamtheiten mit gleicher Verteilung stammen, muss entsprechend abgelehnt werden: Die Häufigkeitsverteilungen der beiden Stichproben unterscheiden sich signifikant voneinander.

<sup>23</sup> Durch die Berechnung des Getrimmten Mittels kann die Wirkung von Ausreißern auf den Mittelwert verringert oder teilweise sogar eliminiert werden. Vor der Berechnung des Mittelwertes werden hierbei die jeweils fünf Prozent kleinsten und größten Werte aus der Analyse eliminiert (CLEFF 2008: 42).

<sup>24</sup> Mit statistischen Testverfahren wie dem **Kolmogorov-Smirnov-Test** für die Güte der Anpassung oder dem Shapiro-Wilk-Test kann überprüft werden, ob empirisch gewonnenes Datenmaterial der Forderung nach Normalverteilung genügt (KÖHLER et al. 2007: 139).

<sup>25</sup> Der **Kolmogorov-Smirnov Z-Test** überprüft, ob zwei unabhängige Zufallsstichproben aus Grundgesamtheiten mit gleicher Verteilung stammen. Überprüft werden vom Test jegliche Abweichungen der Verteilungen, weshalb häufig auch vom sogenannten Omnibus-Test gesprochen wird (JANSSEN und LAATZ 2007: 576f.).

Neben diesen allgemeinen soziodemographischen Variablen zum Geschlecht sowie zum Alter wurden die befragten Studierenden ebenfalls gebeten, Fragestellungen zu verschiedenen studiumsbezogenen soziodemographischen Variablen zu beantworten. Auf diese wird nun nachfolgend eingegangen werden.

Begonnen wird dabei mit der Darstellung der Ergebnisse zur Fachbereichsangehörigkeit, die in Fragestellung 4.3 des Fragebogens erhoben wurde. Während die meisten Studiengänge relativ einfach einem bestimmten **Fachbereich** zugeordnet werden können, ist dies bei bestimmten Studiengängen hingegen teilweise nicht problemlos möglich. So ist eine eindeutige Zuordnung von Lehramts-Studiengängen oder Magister-Studiengängen aufgrund der Vielzahl an möglichen Fächerkombinationen und der daraus resultierenden Zuordnung zu mehreren Fachbereichen häufig schwierig. Einer möglichen, daraus eventuell resultierenden Problematik wurde im Fragebogen dadurch entgegengewirkt, dass von den befragten Studierenden jeweils der Fachbereich angegeben werden sollte, in dem sie die meisten Veranstaltungen besuchen bzw. besucht haben. Die Ergebnisse zu dieser Fragestellung werden nachfolgend in Tabelle 7, nach den verschiedenen Stichproben aufgeschlüsselt, dargestellt.

**Tabelle 7: Häufigkeitsverteilungen der Variablen "Fachbereich" in den Teilstichproben sowie in der Grundgesamtheit der JLU**

	F2F-Stichprobe		Online-Stichprobe		JLU-Grundgesamtheit
	Häufigkeit	Anteil	Häufigkeit	Anteil	Anteil
FB 01	13	7,0%	4	0,8%	7,1%
FB 02	31	16,7%	13	2,5%	7,6%
FB 03	46	24,7%	69	13,4%	15,5%
FB 04	5	2,7%	18	3,5%	3,8%
FB 05	22	11,8%	45	8,8%	17,0%
FB 06	3	1,6%	31	6,0%	6,4%
FB 07	10	5,4%	45	8,8%	7,4%
FB 08	16	8,6%	38	7,4%	6,5%
FB 09	34	18,3%	185	36,0%	9,9%
FB 10	1	0,5%	32	6,2%	6,0%
FB 11	5	2,7%	34	6,6%	12,7%
<b>Gesamt</b>	<b>186</b>	<b>100,0%</b>	<b>514</b>	<b>100,0%</b>	<b>100,0%</b>

Anmerkungen: Eine Liste mit den vollständigen Fachbereichs-Bezeichnungen findet sich in Tabelle A 4 des Anhangs.

Quelle: Auswertung der eigenen Erhebung; JLU 2009e: 2ff..

Bei Vergleich der beiden Teilstichproben mit der JLU-Grundgesamtheit zeigen sich deutliche Unterschiede. Während die Häufigkeitsverteilungen in der Online-Stichprobe mit Ausnahme der Überrepräsentation des Fachbereichs 09 „Agrarwissenschaften, Ökotrophologie und Umweltmanagement“ zumindest tendenziell mit der JLU-Grundgesamtheit übereinstimmen bzw. in deren Richtung deuten, zeigen sich bei Betrachtung der F2F-Stichprobe hingegen deutliche Unterschiede zur JLU-Grundgesamtheit. Hier sind es vor allem die Fachbereiche 05, 06, 10

und 11, die im Vergleich zur Grundgesamtheit deutlich unterrepräsentiert sind sowie die Fachbereiche 02, 03 und 09, die deutlich überrepräsentiert sind. Ob diese Verzerrungen aus einer unterschiedlich ausgeprägten Teilnahmebereitschaft der Studierenden dieser Fachbereiche oder doch aus der Auswahl der Befragungsstandorte resultieren, kann im Nachhinein jedoch nicht mehr eindeutig geklärt werden.

Nach der Fragestellung zum Fachbereich wurden die Studierenden in Frage 4.4 zu ihrem momentan angestrebten **Studienabschluss** befragt. Die Ergebnisse der beiden Teilstichproben werden in Tabelle 8 dargestellt. Zum Vergleich werden zusätzlich die Anteile der verschiedenen Studiengänge an der JLU aufgezeigt. So sind in beiden Teilstichproben die Bachelorstudierenden deutlich überrepräsentiert, während Lehramts-Studierende sowie Studierende, die ein Staatsexamen anstreben, im Vergleich zur Uni-Grundgesamtheit deutlich unterrepräsentiert sind. Starke Unterschiede zwischen der F2F-Stichprobe und der Online-Befragung im Vergleich zur JLU-Grundgesamtheit lassen sich ebenfalls hinsichtlich der Merkmalsausprägung Master erkennen. So gaben in den F2F-Interviews lediglich 3,2 % der Studierenden an, momentan einen Master-Abschluss anzustreben. An der Online-Befragung war mit 21,0 % der Anteil der Studierenden mit angestrebtem Master-Abschluss deutlich höher und lag somit auch deutlich über dem Anteil der Masterstudierenden an der JLU, die hier bisher nur einen Anteil von 4,6 % ausmachen. Ein möglicher Erklärungsansatz hierfür kann der oben dargestellte, jeweils hohe Anteil an Studierenden des Fachbereichs 09 „Agrarwissenschaften, Ökotoxikologie und Umweltmanagement“ sein. So sind am Fachbereich 09 die nach der Bologna-Reform geforderten neuen Studiengänge bereits im Jahre 2002 implementiert worden (JLU 2002).

**Tabelle 8: Häufigkeitsverteilungen der Variablen "Momentan angestrebter Abschluss" in den Teilstichproben sowie in der Grundgesamtheit der JLU**

	F2F-Stichprobe		Online-Stichprobe		JLU-Grundgesamtheit
	Häufigkeit	Anteil	Häufigkeit	Anteil	Anteil
Bachelor	86	46,2%	172	33,5%	25,9%
Master	6	3,2%	108	21,0%	4,7%
Lehramt	27	14,5%	48	9,3%	21,9%
Staatsexamen	26	14,0%	72	14,0%	25,6%
Diplom	28	15,1%	89	17,3%	16,6%
Magister	13	7,0%	25	4,9%	5,5%
<b>Gesamt</b>	<b>186</b>	<b>100,0%</b>	<b>514</b>	<b>100,0%</b>	<b>100,0%</b>

Quelle: Auswertung der eigenen Erhebung.

Andere Fachbereiche der JLU mit einem Angebot an Diplom- und Magister-Studiengängen, die entsprechend der Reform abzuschaffen sind, haben teilweise erst vor kürzerer Zeit ihre Studiengänge umgestellt. So wurden an der JLU im Sommersemester 2009 insgesamt 26

Masterstudiengänge angeboten, von denen alleine 11 Studiengänge dem Fachbereich 09 zugeordnet werden können (JLU 2009f). Weitere Unterschiede zwischen den Teilstichproben und der JLU-Grundgesamtheit können in ähnlicher Weise erklärt werden: So sind die Studienabschlüsse in einem gewissen Umfang natürlich mit den Fachbereichen verknüpft. Dort konnten bereits deutliche Abweichungen der Teilstichproben von der JLU-Grundgesamtheit ermittelt werden, die sich möglicherweise weiter auf die Ergebnisse der Studienabschlüsse auswirken können.

Zusätzlich zu den oben dargestellten allgemeinen und studienbezogenen soziodemographischen Angaben wurden die Studierenden in Frage 4.6 des Fragebogens ebenfalls gebeten, Angaben zu ihren **monatlichen Nettoeinkünften** zu geben. Hierzu wurden in der Fragestellung in beiden Teilbefragungen die in Tabelle 9 dargestellten Antwortmöglichkeiten vorgegeben. Da es sich nach SCHNELL et al. sowie BEREKOVEN et al. (SCHNELL et al. 2005: 339; BEREKOVEN et al. 2009: 97) sowie der aufgestellten Forschungshypothese F1 bei Fragen zum Einkommen grundsätzlich um heikle Fragen handelt, wurden hier entsprechend der theoretischen Ausführungen in Kapitel 3.3.3 Unterschiede im Antwortverhalten nicht ausgeschlossen. Daher stellt Tabelle 9 nachfolgend die Häufigkeitsverteilungen der verschiedenen Kategorien in den beiden Teilstichproben sowie in der Gesamtstichprobe dar.

**Tabelle 9: Häufigkeitsverteilungen der Variablen "Verfügbares Nettoeinkommen" in den Teilstichproben sowie in der Gesamtstichprobe**

	F2F-Stichprobe		Online-Stichprobe		Gesamtstichprobe	
	Häufigkeit	Anteil	Häufigkeit	Anteil	Häufigkeit	Anteil
unter 400 €	41	22,0%	145	28,2%	186	26,6%
400 - 600 €	76	40,9%	181	35,2%	257	36,7%
601 - 800 €	34	18,3%	91	17,7%	125	17,9%
801 - 1200 €	16	8,6%	66	12,8%	82	11,7%
1201 - 2500 €	5	2,7%	13	2,5%	18	2,6%
über 2500 €	1	0,5%	1	0,2%	2	0,3%
Keine Angabe	13	7,0%	17	3,3%	30	4,2%
<b>Gesamt</b>	<b>186</b>	<b>100,0%</b>	<b>514</b>	<b>100,0%</b>	<b>700</b>	<b>100,0%</b>

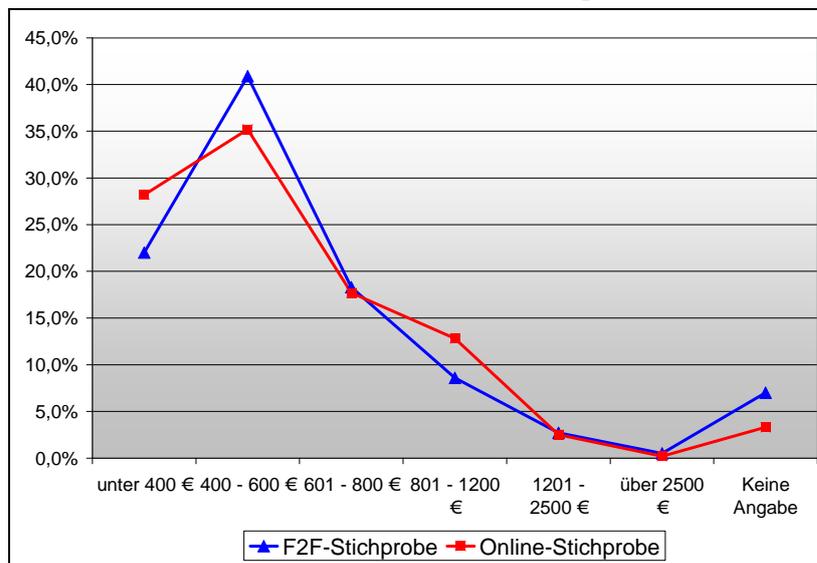
Quelle: Auswertung der eigenen Erhebung.

So lassen sich zwischen beiden Teilstichproben leichte Unterschiede im Antwortverhalten erkennen. In einer Graphik dargestellt, zeigen sich ebenfalls leicht unterschiedliche Kurvenverläufe für die beiden Teilstichproben (siehe Abbildung 14).

So zeigen sich vor allem in den unteren beiden Kategorien „unter 400 €“ und „400 - 600 €“ sowie in der Kategorie „801 - 1200 €“ Unterschiede. Ob sich beide Befragungssamples hinsichtlich des Antwortverhaltens signifikant voneinander unterscheiden, wurde weiterhin aufgrund der Kategorisierung der Variablen mit dem nichtparametrischen U-Test von Mann und

Whitney<sup>26</sup> überprüft. Hierbei konnte kein signifikanter Unterschied zwischen den beiden Teilstichproben nachgewiesen werden. Entsprechend der Nullhypothese des Tests stammen beide Stichproben somit aus der gleichen Grundgesamtheit. Die Teststatistik des Mann-Whitney U-Test kann der Tabelle A 5 im Anhang entnommen werden. Da allerdings nicht ausgeschlossen werden kann, dass sich dieses Ergebnis aufgrund der hinsichtlich der soziodemographischen Variablen unterschiedlich zusammengesetzten Teilstichproben ergeben hat, wird die Falsifizierung oder Verifizierung der Forschungshypothese F1 zurückgestellt. Diese wird an einem fortgeschritteneren Punkt der Arbeit in Kapitel 5.3.4 wieder aufgegriffen werden.

**Abbildung 14: Graphische Darstellung der Häufigkeitsverteilungen der Variablen „Verfügbares Nettoeinkommen“ der beiden Teilstichproben**



Quelle: Auswertung der eigenen Erhebung.

### 5.3.2 Dauer der Fragenbeantwortung

In der Literatur zu Internet-gestützten Befragungsmethoden wird vielfach als Vorteil eine schnellere Durchführbarkeit derartiger Erhebungsmethoden im Vergleich zu anderen Befragungsformen wie etwa den telefonischen, schriftlichen oder auch persönlichen Befragungen beschrieben (SCHOLL et al. 2002: 219; FAAS 2003: 121; DUFFY et al. 2005: 617; EVANS und MATHUR 2005: 197ff.; WELKER et al. 2005: 8; SUE und RITTER 2007: 150f.; DIEKMANN 2008: 552; TADDICKEN 2008: 49; BANDILLA et al. 2009: 129; WELKER und MATZAT 2009: 39). Neben kürzeren Feldzeiten im Allgemeinen spielen auch auf Ebene der einzelnen Interviews

<sup>26</sup> Bei dem **U-Test** von Mann und Whitney handelt es sich um einen nichtparametrischen Signifikanztest, der als verteilungsunabhängiges Gegenstück zum parametrischen t-Test angesehen werden kann. Der Test überprüft dabei, ob zwei Stichproben die gleiche Verteilung aufweisen und somit aus einer Grundgesamtheit stammen. Die Vorgehensweise des Tests ist dabei, dass die  $n=n_1+n_2$  Beobachtungen der Größe nach geordnet und jeweils in eine Rangfolge gebracht werden. Anschließend wird aus den Werten die sog. Rangsumme gebildet, die zur Berechnung der Teststatistiken  $U_1$  und  $U_2$  notwendig ist. Die kleinere der beiden Teststatistiken wird anschließend zur Entscheidungsfindung herangezogen (SACHS und HEDDERICH 2006: 391ff.; BORTZ et al. 2008: 200ff.).

kürzere Zeiten für die Beantwortung des Fragebogens eine Rolle. So beschreibt KAYA, dass auch bei Verwendung des gleichen Fragebogens bei persönlichen Befragungen mit einer längeren Beantwortungsdauer als bei der entsprechenden Online-Befragung zu rechnen ist (KAYA 2007: 52ff.). Verantwortlich für diesen Sachverhalt ist dabei nach BEREKOVEN et al. die Anwesenheit des Interviewers bei den F2F-Interviews. So muss dieser im Unterschied zu einer Online-Befragung nicht nur die Fragestellungen vorlesen und möglicherweise verbal notwendige Instruktionen geben, sondern ebenfalls die vom Befragten gegebenen Antworten notieren (BEREKOVEN et al. 2009: 98ff.).

Aufgrund dieser Ausführungen soll in der vorliegenden Arbeit ermittelt werden, wie viel Zeit von den Befragten durchschnittlich für die Beantwortung des Fragebogens aufgebracht wurde. Weiterhin soll darüber hinaus entsprechend der Forschungshypothese F2 ermittelt werden, ob zwischen den beiden unterschiedlich erhobenen Teilbefragungen ein signifikanter Unterschied in der durchschnittlichen Beantwortungszeit dahingehend nachgewiesen werden kann, dass die Interviewdauer in der Online-Befragung signifikant kürzer ist.

Zur Beantwortung dieser Fragestellungen wurden daher bei jedem Interview jeweils die Start- und die Endzeit festgehalten. So waren die Interviewer der F2F-Befragungen angewiesen, jeweils die Uhrzeit des Beginns sowie des Endes des Interviews auf dem Fragebogen in entsprechend vorgesehenen Feldern zu vermerken. Auch wurden die Interviewer während der Interviewerschulung auf die Relevanz dieser Angaben hingewiesen, so dass fast alle Interviews vollständig mit diesen Angaben ausgefüllt wurden. Auch wurde an dieser Stelle auf die Bedeutung genauer Angaben hierbei hingewiesen.

Im Rahmen der Online-Befragung mit LimeSurvey konnten diese jeweiligen Start- und Endzeiten jeweils automatisch erfasst werden. Die Berechnung der jeweiligen Interviewdauer erfolgte durch Differenzbildung aus Start- sowie Endzeit. In Tabelle 10 werden für beide Teilstichproben verschiedene Kennzahlen der deskriptiven Statistik präsentiert.

**Tabelle 10: Deskriptive Statistik der Interviewdauer, unterteilt nach Befragungsmodus**

	F2F-Stichprobe (n=184)	Online-Stichprobe (n=514)
Minimum	00:05 h	00:04 h
Maximum	01:13 h	00:46 h
Arith. Mittel	00:15 h	00:10 h
Median	00:14 h	00:09 h
Getr. Mittel	00:14 h	00:10 h
SD	00:07 h	00:05 h

Anmerkungen: Arith. Mittel = Arithmetisches Mittel; Getr. Mittel = 5% Getrimmtes Mittel; SD = Standardabweichung.

Quelle: Auswertung der eigenen Erhebung.

So konnten bei den Online-Interviews für die Interviewdauer Werte von 4 bis hin zu 46 Minuten ermittelt werden, womit sich eine Spannweite von 42 Minuten ergibt. Deutlich weiter auseinander geht der Maximal- und Minimalwert und somit die Spannweite hingegen bei der F2F-Stichprobe: Entgegen den theoretischen Ausführungen konnte hier mit 1:08 h eine deutlich größere Spannweite als in der Online-Stichprobe ermittelt werden. Vor allem im Vergleich erscheint der ermittelte Maximalwert der F2F-Stichprobe sehr hoch und wenig plausibel. Stattdessen würde ein höherer Maximalwert eher bei einem Teilnehmer der Online-Befragung vermutet werden, welcher möglicherweise den Fragebogen im Hintergrund neben einer anderen, als relevanter erachteten Tätigkeit ausfüllt. Da im Nachhinein jedoch nicht mehr ermittelt werden konnte, ob es sich bei dem Wert trotz Zweifel um einen korrekten Wert oder einen Schreibfehler von Seiten des Interviewers handelt (die Transkription in SPSS erfolgte korrekt), wurde hier neben dem arithmetischen Mittel ebenfalls das 5% Getrimmte Mittel berechnet, um die Aussagekraft nicht durch Ausreißer zu verzerren. Wie aus Tabelle 10 hervorgeht, zeigen sich bei einem Vergleich der beiden statistischen Größen zwischen der persönlich und der online erhobenen Stichprobe deutliche Unterschiede. So lagen sowohl das arithmetische Mittel als auch das 5% Getrimmte Mittel der F2F-Stichprobe jeweils über den entsprechenden Werten der Online-Stichprobe – die durchschnittliche Dauer der F2F-Interviews war somit höher als die durchschnittliche Bearbeitungszeit des Online-Fragebogens.

Um auf dieser Erkenntnis aufbauend eine Entscheidung zur Annahme oder Ablehnung der Hypothese *„Die Befragungsmethode Online zeichnet sich im Vergleich zu F2F durch eine signifikant kürzere Interviewdauer aus.“* vornehmen zu können, wurde die Anwendung eines statistischen Tests forciert. Für die Auswahl des passenden Testverfahrens gibt es verschiedene Anforderungen wie etwa Normalverteilung oder Skalenniveau der Daten. So konnte bei Überprüfung der relevanten Variablen „Interviewdauer“ durch den Kolmogorov-Smirnov-Test mit einer Irrtumswahrscheinlichkeit von 0,0001 keine Normalverteilung ermittelt werden (siehe hierzu Tabelle A 6 im Anhang). Diese gilt jedoch als Voraussetzung für die Anwendung von t-Tests (KÖHLER et al. 2007: 89ff.). Daher musste die Auswahl der anzuwendenden Testverfahren auf nichtparametrische Tests beschränkt werden. Angewendet wurde zur Entscheidungsfindung der sogenannte Mann-Whitney U-Test, der als Skalenniveau mindestens ordinal skalierte Daten voraussetzt. Der Tests formuliert die Nullhypothese  $H_0$ , dass die im Test zu vergleichenden Stichproben aus einer Grundgesamtheit stammen und sich daher hinsichtlich der durchschnittlichen Interviewdauer nicht unterscheiden. Nach Testdurchführung kommt dieser zum Resultat, dass mit einer Irrtumswahrscheinlichkeit von 0,001 die oben genannte Nullhypothese zugunsten der Alternativhypothese  $H_1$ , dass beide Stichproben aus unterschiedlichen Grundgesamtheiten stammen, abzulehnen ist (vgl. Tabelle A 7 im Anhang). Ent-

sprechend kann ausgesagt werden, dass sich die beiden Teilstichproben hinsichtlich der Interviewdauer signifikant voneinander dahingehend unterscheiden, dass die Befragungsdauer in der Online-Befragung signifikant kürzer ist. Daher ist die aufgestellte Forschungshypothese „Die Befragungsmethode Online zeichnet sich im Vergleich zu F2F durch eine signifikant kürzere Interviewdauer aus“ anzunehmen und zu bestätigen.

**Hypothese F2:** Die Befragungsmethode Online zeichnet sich im Vergleich zu F2F durch eine signifikant kürzere Interviewdauer aus.



### 5.3.3 Gewichtung der Fälle

Bereits in Kapitel 5.3.1 konnte dargestellt werden, dass sich die beiden, mit unterschiedlichen Methoden erhobenen Teilstichproben bezüglich der soziodemographischen Variablen wie Alter, Geschlecht, Fachbereich und angestrebtem Abschluss teilweise sehr deutlich voneinander unterscheiden. Weiterhin konnten nicht nur zwischen den Teilstichproben, sondern auch im Vergleich zur Grundgesamtheit der Justus-Liebig-Universität deutliche Differenzen aufgezeigt werden. Da jedoch die zum Methodenvergleich herangezogene Split-Ballot-Technik voraussetzt, dass die verwendeten Daten aus einer Grundgesamtheit stammen, wurde dies ebenfalls mit Hilfe statistischer Tests überprüft. Übereinstimmend zu vorherigen Ausführungen mussten auch hier entsprechend den Testvoraussetzungen die geeigneten Verfahren ausgewählt werden. Für die nominalskalierten Variablen Geschlecht, Fachbereich sowie angestrebter Bildungsabschluss wurde daher der  $\chi^2$ -Test nach Pearson<sup>27</sup> zur Überprüfung herangezogen. Als Ergebnis des  $\chi^2$ -Tests kann für die drei getesteten Variablen festgehalten werden, dass mit einer Irrtumswahrscheinlichkeit von 0,001 die aufgestellte Nullhypothese abzulehnen ist und somit die beobachteten Unterschiede auch in der Grundgesamtheit vorhanden sind (vgl. Tabelle A 8 im Anhang). Entsprechend den Ausführungen von KÖHLER et al. bedeutet dies, dass es zwischen den Verteilungen der beiden Stichproben jeweils signifikante Unterschiede zu beobachten gibt und die Daten entsprechend inhomogen sind (KÖHLER et al. 2007: 101ff.).

Auch bei der Überprüfung der Alters-Variablen musste auf nichtparametrische Tests zurückgegriffen werden. Zwar weist die Variable grundsätzlich ein intervallskaliertes und somit me-

<sup>27</sup> Der  $\chi^2$ -Test nach Pearson kann zur Überprüfung der Nullhypothese herangezogen werden, dass in einer Grundgesamtheit eine statistische Unabhängigkeit zwischen zwei nominalverteilten Merkmalen besteht. Im Rahmen des Tests werden dazu die beobachteten Werte  $B_i$  mit den erwarteten Werten  $E_i$  einer Kreuztabelle verglichen. Für die Entscheidungsfindung wird die Testgröße  $\chi^2$  durch die Formel

$$\chi^2 = \sum_{i=1}^n \frac{(B_i - E_i)^2}{E_i}$$

berechnet und anschließend mit dem theoretischen  $\chi^2$ -Wert verglichen. Dieser kann Tabellen entnommen werden (KÄHLER 2008: 205ff.; KÖHLER et al. 2007: 99ff.).

trisches Skalenniveau auf, allerdings konnte wie bereits in Kapitel 5.3.1 analysiert wurde, keine Normalverteilung festgestellt werden, wodurch eine wesentliche Voraussetzung für die Anwendung des t-Tests verletzt ist. Auch hier wurden daher der Mann-Whitney U-Test sowie der Kolmogorov-Smirnov-Test für zwei Stichproben zur Überprüfung auf signifikante Unterschiede zwischen den beiden Samples herangezogen. Auch hier muss als Ergebnis festgehalten werden, dass sich die Variablenverteilung in beiden Stichproben signifikant voneinander unterscheidet (vgl. Tabelle A 9 im Anhang).

Um dennoch die Split-Ballot-Technik anwenden zu können, wurde daher eine Gewichtung der Daten vorgenommen. Aus diesem Grund wird nachfolgend etwas genauer auf die Theorie zur Gewichtung eingegangen.

### **Grundlagen der Gewichtung**

Unter einem Gewicht kann allgemein ein multiplikativer Faktor verstanden werden, der auf einige oder auch alle Teilnehmer einer Umfrage angewendet wird. Die Gewichte bzw. Gewichtungsfaktoren können dabei grundsätzlich Werte von kleiner, gleich oder auch größer eins annehmen. Die Anwendung einer Gewichtung ändert dabei durch Multiplikation die relative Wichtigkeit der Befragten bzw. des Stichprobenelements. Daher kann auch nur von gewichteten Daten gesprochen werden, wenn alle Gewichte ungleich eins sind. Zielsetzung einer Gewichtung ist letztlich, dass ein Stichprobenprofil einer interessierenden Variablen nach Durchführung der Gewichtung dem tatsächlichen Profil in der Grundgesamtheit ähnlicher ist als vor der Gewichtung (GABLER 2006: 128ff.).

Bei Gewichten lassen sich weiterhin grundsätzlich zwei Arten unterscheiden: Einmal sogenannte Designgewichte sowie Anpassungsgewichte. Erstere dienen dazu, Verzerrungen zu korrigieren, welche aus dem Erhebungsdesign resultieren. Ein klassisches Anwendungsbeispiel für Designgewichte sind sogenannte Transformationsgewichte in zweistufigen Zufallsauswahlen. Hierbei werden auf der ersten Stufe per Zufallsauswahl Haushalte und anschließend auf Stufe zwei innerhalb dieser Haushalte die zu interviewenden Personen ausgewählt. Innerhalb eines Haushaltes nimmt so die Auswahlwahrscheinlichkeit einer Person mit zunehmender Haushaltsgröße ab, woraus resultiert, dass Personen aus Einpersonenhaushalten gegenüber solchen aus beispielsweise Vier- oder Fünfpersonenhaushalten systematisch überrepräsentiert sind. Um diese designbedingten unterschiedlichen Auswahlwahrscheinlichkeiten korrigieren zu können, werden entsprechend Designgewichte berechnet. Die Anwendung von Designgewichten lässt sich dabei mit Schätzfunktionen in Verbindung bringen, wie nachfolgende Ausführungen von GABLER erläutern:

So sei eine interessierende Gesamtheit  $U$  von  $N$  Einheiten, beispielsweise Personen, gegeben:

$$U = \{1, \dots, N\}. \quad (6)$$

Das interessierende Merkmal habe weiter die unbekanntenen Ausprägungen  $Y_1, \dots, Y_N$ , die im Parametervektor  $\theta$  zusammengefasst sind:

$$\theta = (Y_1, \dots, Y_N) \quad (7)$$

Der zu schätzende Parameter

$$T = T(\theta) \quad (8)$$

ist eine Funktion des Parametervektors, beispielsweise

$$T = \bar{Y}. \quad (9)$$

Weiter wird die Wahrscheinlichkeit, dass eine Stichprobe  $s$  und damit eine Teilmenge von  $U$ , ausgewählt wird, mit  $p(s)$  bezeichnet.

Um nun das interessierende Merkmal  $T(\theta)$  schätzen zu können, wird eine Schätzfunktion  $t(s, \theta)$  herangezogen. Diese darf nur von  $Y_i$ -Werten abhängen, für die  $i \in s$  ist.

Solch eine Funktion heißt linear, wenn

$$t(s, \theta) = \sum_{i \in s} a_{si} Y_i + b_s \quad (10)$$

ist. Häufig ist dabei  $b_s = 0$ .

Auch sollten die Gewichte  $a_{si}$  so gewählt werden, dass die Schätzfunktion  $t(s, \theta)$  möglichst nahe am zu schätzenden Wert  $T(\theta)$  liegt (GABLER 2006: 192ff.).

Anpassungsgewichte oder auch Redressementgewichte dienen im Unterschied zu den Designgewichten nicht dazu, aus dem Erhebungsdesign resultierende, bekannte Eigenschaften zu kompensieren. Stattdessen sollen über Anpassungsgewichte selektive Stichprobenausfälle ausgeglichen werden, die beispielsweise daraus resultieren, dass bestimmte Personen schlechter erreicht werden können oder es überproportional ablehnen, an einem Interview teilzunehmen. So wird versucht, mit Anpassungsgewichten die gemeinsame Verteilung ausgewählter Merkmale in der Stichprobe an die entsprechende Verteilung in der Grundgesamtheit anzunähern bzw. übereinstimmen zu lassen. Häufig wird diese Vorgehensweise gewählt, um trotz selektiver Ausfälle eine repräsentative Stichprobe zu gewinnen, welche es erlaubt, im Rahmen der schließenden Statistik Schlussfolgerungen über die angezielte Grundgesamtheit zu ziehen (RÖSCH 1994: 11; RUBIN und THOMAS 1996: 255; GABLER 2006: 129ff.; FAAS und SCHOEN 2009: 148).

Wenn die gemeinsame Verteilung der Anpassungsvariablen bekannt ist, kann durch eine einfache Soll/Ist-Gewichtung die Stichprobenverteilung an die Verteilung der Grundgesamtheit angepasst werden. Diese Art der Gewichtung wird auch als „nachträgliche Schichtung“ be-

zeichnet. Für den Fall, dass zwei Anpassungsvariablen verwendet werden, lautet die allgemein zu verwendende Formel hierbei:

$$w_{ijk} = \frac{1}{\pi_k} \frac{N_{ij}}{\hat{N}_{ij}} \quad \text{mit} \quad \hat{N} = \sum_{k \in s_{ij}} \frac{1}{\pi_k}, \quad (11)$$

wobei  $\pi_k$  die Inklusionswahrscheinlichkeiten erster Ordnung bezeichnen und  $s_{ij}$  alle Einheiten der Stichprobe aus  $ij$ -ter Zelle enthält (GABLER 2006: 139ff.). Zellen bezeichnen dabei die Kombinationen einzelner Merkmalsausprägungen der Anpassungsvariablen. Für die Gewichtung spielt es nun eine Rolle, welcher Zelle die Merkmalsträger der Stichprobe zugeordnet werden. Ebenfalls notwendig für das technische Gelingen einer Gewichtung ist, dass jede Zelle mit mindestens einem Fall der Stichprobe besetzt ist, da sonst eine Gewichtung mit dem entsprechenden Gewicht nicht möglich ist (RÖSCH 1994: 11ff.).

### Auswahl der Variablen zur Gewichtung

Während anfangs eine Gewichtung der Daten mittels Anpassungsgewichten dahingehend präferiert wurde, eine der beiden Stichproben an die entsprechend andere Stichprobe anzugleichen, wurde von dieser Alternative relativ schnell wieder Abstand genommen. Ursächlich hierfür sind mehrere Überlegungen, die gegen diese Vorgehensweise sprechen: Einmal stellt sich bei der angedachten Vorgehensweise grundsätzlich die Frage, welche der beiden Stichproben als Gewichtungsbasis herangezogen werden sollte – die durch Interviewer erhobene F2F-Stichprobe oder die online erhobene Stichprobe? Hierbei ist es nur sehr schwer möglich, plausible Kriterien festzulegen, anhand derer die Auswahl einer Stichprobe als Gewichtungsbasis erfolgen könnte. Weiterhin kann sich bei dieser Vorgehensweise das Problem ergeben, dass sich etwaige Verzerrungen in der, als Gewichtungsbasis herangezogenen, Stichprobe durch diese Vorgehensweise auch auf die andere, möglicherweise noch unverzerrte Teilstichprobe übertragen könnten.

Daher wurde von dieser Vorgehensweise abgesehen und stattdessen eine Gewichtung der Daten gewählt, welche in den beiden Teilstichproben die Häufigkeitsverteilungen zweier Variablen an die entsprechenden Häufigkeitsverteilungen in der JLU-Grundgesamtheit annähert. Hierbei finden die oben dargestellten Anpassungs- oder Redressmentgewichte Anwendung. Als Anpassungsvariablen wurden nach verschiedenen Überlegungen die soziodemographischen Variablen „Fachbereich“ sowie „Geschlecht“ herangezogen, da diese eindeutig der offiziellen Studierendenstatistik der JLU Gießen entnommen werden können. Da sich nach RÖSCH eine Gewichtung mit mehr als zwei Anpassungsvariablen außerordentlich schwierig

gestaltet (RÖSCH 1994: 16), wurde nach Abwägen die auf die Einbeziehung der Variablen „Angestrebter Studienabschluss“ verzichtet.

Um die Durchführung der Gewichtung durch eine Verringerung der Merkmalsausprägungen der Anpassungsvariablen weiter vereinfachen zu können, wurde die Variable „Fachbereich“ zusätzlich dahingehend aggregiert, dass anstatt der einzelnen Fachbereiche vier thematisch und inhaltlich ähnliche Fachbereichs-Gruppen gebildet wurden. Die Zusammensetzung dieser vier Gruppen in der neugebildeten Variablen „Klassifizierter Fachbereich“ wird nachfolgend in Tabelle 11 aufgezeigt.

Der Fachbereich 09 stellt dabei im Unterschied zu den anderen Fachbereichen, die zu Clustern zusammengefasst wurden, eine eigene Gruppe. Leitgedanke für diese Vorgehensweise war einerseits, im Rahmen induktiver statistischer Verfahren Besonderheiten des Fachbereichs 09, beispielsweise im Rahmen der Kontingenzanalyse, besser herausarbeiten zu können. Andererseits wurde ebenfalls ein stärkeres Interesse an der Thematik sowie eine höhere Kenntnis verschiedener Sachverhalte bei Angehörigen des Fachbereichs 09 vermutet, was neben dem Aspekt des überproportionalen Anteils Studierender des Fachbereichs 09 in den beiden Teilstichproben auch für die verwendete Gruppierung sprach.

**Tabelle 11: Zusammensetzung der Fachbereichs-Gruppen**

	Überbegriff klassifizierte Fachbereiche	Fachbereiche
FB_Zus 1	Wirtschafts- und Rechtswissenschaften	01 und 02
FB_Zus 2	Sozial-, Sprach- und Kulturwissenschaften	03, 04, 05 und 06
FB_Zus 3	Mathematik, Naturwissenschaften sowie Medizin	07, 08, 10 und 11
FB_Zus 4	Agrarwiss., Ökologie und Umweltm.	09

Quelle: Eigene Darstellung.

Über die Gewichtung der beiden unterschiedlich erhobenen Teilstichproben anhand der Anpassungsvariablen konnte letztlich in beiden eine Merkmalsverteilung entsprechend jener in der JLU Grundgesamtheit erzielt werden. Die Verteilung der beiden Variablen „Klassifizierter Fachbereich“ sowie „Geschlecht“ in den Teilstichproben vor und nach der Gewichtung unterscheiden sich entsprechend deutlich, wie jeweils Tabelle 12 und Tabelle 13 darstellen. Um zusätzlich die Wirksamkeit der Gewichtung aufzeigen zu können, werden in den Tabellen außerdem die jeweiligen Häufigkeitsverteilungen in der Grundgesamtheit der JLU dargestellt.

**Tabelle 12: Änderung der Häufigkeitsverteilung der Variablen "Geschlecht" nach Durchführung der Gewichtung**

	F2F-Stichprobe (n=186)		Online-Stichprobe (n=514)		JLU- Grundgesamtheit
	Vorher	Nachher	Vorher	Nachher	
Männlich	45,7%	35,7%	21,8%	35,7%	35,6%
Weiblich	54,3%	64,3%	78,2%	64,3%	64,4%
Gesamt	100,0%	100,0%	100,0%	100,0%	100,0%

Quelle: Auswertung der eigenen Erhebung, JLU 2009e: 2ff..

Durch die Gewichtung mittels Anpassungs- bzw. Redressementgewichten verändern sich die Häufigkeitsverteilungen der Variablen „Geschlecht“ sowie „Klassifizierter Fachbereich“ teilweise sehr deutlich und entsprechen nach der Gewichtung der Grundgesamtheit der JLU, wie aus beiden Tabellen jeweils entnommen werden kann. Dies bedeutet weiterhin, dass sich die beiden Teilstichproben hinsichtlich der beiden Merkmale „Klassifizierter Fachbereich“ sowie „Geschlecht“ nicht mehr voneinander unterscheiden.

**Tabelle 13: Tabelle 8: Änderung der Häufigkeitsverteilung der Variablen "Klassifizierter Fachbereich" nach Durchführung der Gewichtung**

	F2F-Stichprobe (n=186)		Online-Stichprobe (n=514)		JLU- Grundgesamtheit
	Vorher	Nachher	Vorher	Nachher	
FB_Zus 1	23,7%	14,6%	3,3%	14,6%	14,6%
FB_Zus 2	40,9%	42,7%	31,7%	42,7%	42,7%
FB_Zus 3	17,2%	32,7%	29,0%	32,7%	32,7%
FB_Zus 4	18,3%	9,9%	36,0%	9,9%	9,9%
<b>Gesamt</b>	<b>100,0%</b>	<b>100,0%</b>	<b>100,0%</b>	<b>100,0%</b>	<b>100,0%</b>

Quelle: Auswertung der eigenen Erhebung, JLU 2009e: 2ff..

Daher wird für die weitere Analyse zur Untersuchung der Wirkungen der Befragungsmethode auf das Antwortverhalten davon ausgegangen, dass die Voraussetzungen der Split-Ballot-Technik durch die Gewichtung der Daten erfüllt werden und sich die beiden gewichteten Teilstichproben nur noch hinsichtlich des verwendeten Befragungsmodus unterscheiden.

### 5.3.4 Beantwortung der zurückgestellten Frage zum Einkommen

In Kapitel 5.3.1 musste eine Verifizierung oder Falsifizierung der Forschungshypothese F1 noch zurückgestellt werden. Ursache hierfür war, dass eine korrekte Entscheidung hinsichtlich der Hypothese nicht sichergestellt werden konnte, da sich in den beiden Teilstichproben hinsichtlich der soziodemographischen Variablen gewisse Unterschiede gezeigt hatten. Nach der Gewichtung des Datensatzes, wie im vorherigen Unterkapitel beschrieben, kann nun allerdings die Fragestellung erneut aufgegriffen werden. Tabelle 14 zeigt nachfolgend die Prozentanteile der einzelnen Einkommenskategorien vor und nach der Gewichtung des Gesamtdatensatzes auf.

Auch bei der Einkommensvariablen zeigen sich durch die Gewichtung gewisse Veränderungen. Allerdings fallen diese tendenziell eher kleiner aus. Zur erneuten Überprüfung, ob sich die beiden Teilstichproben Online und F2F hinsichtlich der Angaben zu dieser Variablen signifikant unterscheiden, wurde wiederum der Mann-Whitney U-Test herangezogen. Dieser bestätigt auch nach der Gewichtung, dass zwischen den beiden Teilstichproben keine signifikan-

ten Unterschiede hinsichtlich der Verteilung der Einkommensklassen vorhanden sind (siehe hierzu die Teststatistik in Tabelle A 10 des Anhangs).

**Tabelle 14: Prozentanteile der Einkommensklassen in den beiden Befragungssamples vor und nach der Gewichtung**

	F2F-Stichprobe		Online-Stichprobe	
	Vorher	Nachher	Vorher	Nachher
unter 400 €	22,0%	24,4%	28,2%	26,1%
400 - 600 €	40,9%	40,8%	35,2%	34,6%
601 - 800 €	18,3%	16,8%	17,7%	17,1%
801 - 1200 €	8,6%	7,3%	12,8%	14,4%
1201 - 2500 €	2,7%	2,5%	2,5%	2,8%
über 2500 €	0,5%	0,3%	0,2%	0,2%
Keine Angabe	7,0%	7,9%	3,3%	4,8%
<b>Gesamt</b>	<b>100,0%</b>	<b>100,0%</b>	<b>100,0%</b>	<b>100,0%</b>

Quelle: Auswertung der eigenen Erhebung.

Da sich die Angaben der befragten Studierenden nicht unterscheiden, lassen sich keine Hinweise auf Antwortunterschiede zwischen den beiden Teilstichproben finden. Somit lassen sich keine Hinweise darauf finden, dass in der Online-Befragung ehrlicher geantwortet wurde. Aus diesem Grund muss die erste Forschungshypothese *„Bei Fragen zum Einkommen als Repräsentant für heikle Fragestellungen sind aufgrund der unterschiedlichen Anonymität in beiden Befragungsmethoden Unterschiede im Antwortverhalten dahingehend erkennbar, dass in der Online-Befragung niedrigere Angaben gemacht werden“* abgelehnt werden:

**Hypothese F1:** Bei Fragen zum Einkommen als Repräsentant für heikle Fragestellungen sind aufgrund der unterschiedlichen Anonymität in beiden Befragungsmethoden Unterschiede im Antwortverhalten dahingehend erkennbar, dass in der Online-Befragung niedrigere Angaben gemacht werden.



## 5.4 Ergebnisse der inhaltlichen Fragestellungen des Fragebogens

Nachfolgend wird nun auf ausgewählte Fragestellungen des Fragebogens eingegangen und analysiert, ob zwischen den beiden Teilstichproben signifikante Unterschiede erkennbar sind, die durch die versuchte Ausschaltung äußerer Störeinflüsse auf den Befragungsmodus zurückgeführt werden können.

### 5.4.1 Fragen zur Relevanz der Stärkung verschiedener politischer Aufgabebereiche

In der ersten Frage des Fragebogens wurden die Teilnehmer jeweils befragt, wie wichtig ihnen eine zusätzliche Stärkung verschiedener politischer Aufgabebereiche in Deutschland ist. Vorgegeben wurde den Teilnehmern bei dieser Einstellungsfrage eine fünfstufige Skala mit

den Polen eins („sehr wichtig“) bis fünf („unwichtig“), die als Skala 1 im Anhang abgedruckt ist. Tabelle 15 stellt für die Items der Fragestellung die jeweiligen Mittelwerte der Teilstichproben F2F und Online sowie die sich daraus ergebenden Mittelwertdifferenzen dar. Um entsprechend der Zielsetzungen dieser Arbeit weiter ermitteln zu können, ob sich die Teilstichproben hinsichtlich der jeweiligen Item-Bewertungen signifikant voneinander unterscheiden, sollte ursprünglich anschließend der t-Test für ungepaarte Stichproben durchgeführt werden. Der t-Test ermöglicht es dabei, signifikante Unterschiede zwischen den Mittelwerten zweier ungepaarter Stichproben, hier der beiden Teilstichproben, zu ermitteln. Aufgrund der Ausschaltung sonstiger Einflüsse hätten Unterschiede im Antwortverhalten entsprechend auf die Befragungsmethodik zurückgeführt werden können. Hierbei musste jedoch beachtet werden, dass bei dieser sowie weiteren Fragestellungen des Fragebogens mehrere Vorgaben des t-Tests verletzt worden wären. So handelt es sich bei den mit Hilfe von mehrstufigen Skalen erhobenen Daten ausschließlich um ordinalskalierte Daten. Zwar können diese nach BAUR aufgrund der von Befragten als gleich groß wahrgenommenen Skalenabstände häufig wie metrisch skalierte Daten verwendet werden (BAUR 2008: 281), dennoch wird eine weitere, wesentliche Vorgabe des t-Tests, die Normalverteilung, ebenfalls nicht erfüllt (siehe hierzu die Ergebnisse des Kolmogorov-Smirnov-Tests auf Normalverteilung in Tabelle A 11 des Anhangs). Daher wurde als Alternative zum t-Test der nichtparametrische U-Test von Mann und Whitney herangezogen, der ebenfalls zum Vergleich von Mittelwerten verwendet werden kann. Dieser Test benötigt im Unterschied zum t-Test ausschließlich ordinal skalierte Daten, weshalb eine Normalverteilung entsprechend nicht gegeben sein muss.

Neben den Mittelwerten der Item-Bewertungen in den beiden Teilstichproben sowie den sich daraus ergebenden Mittelwertdifferenzen stellt Tabelle 15 jeweils die Ergebnisse des Mann-Whitney U-Tests für die verschiedenen Items der Fragestellung dar.

**Tabelle 15: Mittelwerte, Mittelwertdifferenzen sowie Ergebnisse des Mann-Whitney U-Tests für die Items der Fragestellung zur Relevanz einer zusätzlichen Stärkung verschiedener politischer Aufgabenbereiche in Deutschland**

	Arithmetisches Mittel			Mann-Whitney U-Test <sup>a</sup>
	F2F	Online	Differenz	
Bekämpfung der Arbeitslosigkeit	1,69	1,75	0,06	n.s.
Schutz von Tier- und Pflanzenarten	2,03	2,01	-0,02	n.s.
Verbesserung des Zusammenlebens mit Ausländern	1,86	2,07	0,21	0,02
Maßnahmen zum Klimaschutz	1,65	1,69	0,04	n.s.
Bekämpfung von Kriminalität	2,12	2,10	-0,02	n.s.
Förderung der Wirtschaft	2,28	2,26	-0,02	n.s.

Anmerkungen: Weitere statistische Kennzahlen des Mann-Whitney U-Tests finden sich in Tabelle A 12 des Anhangs; „n.s.“ steht für „nicht signifikant“.

a. Gruppierungsvariable: Befragungssample

Quelle: Auswertung der eigenen Erhebung.

Mit Ausnahme des Items „*Verbesserung des Zusammenlebens mit Ausländern*“ konnten dabei keine signifikanten Unterschiede in der Bewertung der verschiedenen Items durch den Mann-Whitney U-Test ermittelt werden. Beim genannten Item lassen sich deutliche Mittelwertdifferenzen erkennen: So wurde dem Item im Rahmen der F2F-Interviews mit einem Durchschnittswert von 1,86 eine höhere Wichtigkeit bescheinigt als in der Online-Befragung. Hier wurde die Wichtigkeit einer zusätzlichen Stärkung dieses politischen Aufgabenbereiches mit einem Durchschnittswert von 2,07 etwas weniger stark bewertet. Da diesem Item in Deutschland immer noch eine gewisse Relevanz zuzurechnen ist, sind durchaus Methodeneffekte denkbar, die auf einer höheren sozialen Erwünschtheit im Rahmen der F2F-Interviews basieren können.

Insgesamt zeigt sich bei allen Itembewertungen der Fragestellung sowohl in der F2F- als auch in der Online-Befragung eine hohe Zustimmung dahingehend, dass eine zusätzliche Stärkung dieser politischen Aufgabenbereiche in Deutschland immer noch als wichtig erachtet wird.

#### **5.4.2 Fragestellungen zum Themenkomplex „Klimaschutz“**

##### **Einstellungsfrage zum Themenkomplex „Klimaschutz“**

In der nächsten Fragestellung, die ebenfalls als Einstellungsfrage konzipiert wurde, sollten die Befragten jeweils angeben, inwieweit sie verschiedenen Aussagen zum Thema Klimaschutz zustimmen. Die bei dieser Frage verwendete Skala (siehe Skala 2 im Anhang) umfasste wiederum fünf Stufen von eins („stimme vollkommen zu“) bis fünf („lehne vollkommen ab“) und ließ als zusätzliche Antwortmöglichkeit „nicht relevant“ für Items zu, die aus Befragtersicht als nicht zutreffend bzw. bedeutsam angesehen wurden. Diese Antwortmöglichkeit wurde entsprechend der Möglichkeiten in SPSS als fehlender Wert codiert, wodurch bei den Berechnungen des Mittelwertes sowie bei Durchführung der Signifikanztests eine Nicht-Berücksichtigung erzielt werden konnte.

Die Mittelwerte der beiden Teilstichproben sowie die sich ergebenden Mittelwertdifferenzen werden in Tabelle 16 aufgezeigt. Bereits die deutlicher ausgeprägten Mittelwertdifferenzen der beiden Items „*Für den Klimaschutz bin ich bereit, auf Flugreisen zu verzichten*“ und „*Klimaschutz ist alleinige Sache der Politik*“ lassen stärkere Unterschiede im Antwortverhalten der Befragten zwischen den beiden Teilstichproben erahnen. Diese Vermutung bestätigt sich bei Betrachtung der Ergebnisse des Mann-Whitney U-Tests. Im Unterschied zu den anderen Items dieser Fragestellung konnten für die beiden genannten Items signifikante bzw.

höchst signifikante Unterschiede in der zentralen Tendenz der Item-Bewertungen ermittelt werden.

So wurde im Rahmen der F2F-Interviews das Item „Für den Klimaschutz bin ich bereit, auf Flugreisen zu verzichten“ mit einem arithmetischen Mittel von 3,57 eher abgelehnt als in der Online-Befragung. Hier wurde dem Item mit einem Mittelwert von 3,01 stärker zugestimmt. Der Mann-Whitney U-Test bestätigt bei diesem Item weiter, dass sich die beiden Teilstichproben hinsichtlich der zentralen Tendenz der Bewertungen höchst signifikant voneinander unterscheiden.

**Tabelle 16: Mittelwerte, Mittelwertdifferenzen sowie Ergebnisse des Mann-Whitney U-Tests für die Items der Einstellungsfrage zum Themenkomplex „Klimaschutz“**

	Mittelwert			Mann-Whitney U-Test <sup>a</sup>
	F2F	Online	Differenz	
Auch mit kleinen Dingen wie sparsamer heizen und öfters mal das Auto stehen lassen kann schon viel erreicht werden.	1,56	1,64	0,08	n.s.
Wenn ich als Einzelner etwas für den Klimaschutz tue, ändert das auch nichts.	4,11	4,03	-0,08	n.s.
Für den Klimaschutz bin ich bereit, auf Flugreisen zu verzichten.	3,57	3,01	-0,56	0,001
Ich bin nur dann bereit, mein Konsumverhalten für den Klimaschutz zu ändern, wenn alle anderen dies auch tun.	4,03	3,93	-0,10	n.s.
Klimaschutz ist alleinige Sache der Politik.	4,59	4,4	-0,19	0,013
Um das Klima zu schützen, sollten höhere Steuern auf Benzin und Treibstoffe erhoben werden.	3,28	3,26	-0,02	n.s.

Anmerkungen: Weitere statistische Kennzahlen des Mann-Whitney U-Tests finden sich in Tabelle A 13 des Anhangs

a. Gruppierungsvariable: Befragungssample

Quelle: Auswertung der eigenen Erhebung.

Unter Zuhilfenahme des Konstrukts der sozialen Erwünschtheit hätte hier möglicherweise eher erwartet werden können, dass in den F2F-Interviews diesem Item stärker als in den Online-Interviews zugestimmt wird. Aber auch das vorgefundene Bewertungsmuster kann erklärt werden: So werden möglicherweise die mit Flugreisen verbundenen Attribute wie Entspannung, Spaß oder das Kennenlernen fremder Städte, Länder, Kulturen oder Sprachen unter dem Oberbegriff „Kosmopolitismus“ aus Studierendensicht als sozial wünschenswerter betrachtet als der Verzicht auf solche Flugreisen. Durch Interviewer, die ebenfalls dem studentischen Milieu angehören, kann diese soziale Erwünschtheit weiter verstärkt werden. Diesem Erklärungsmuster folgend, würde ein Verzicht auf Flugreisen möglicherweise eher einem fehlenden Interesse an fremden Sprachen, Ländern, Kulturen oder auch globalen Zusammenhängen gleichgesetzt werden, was gerade aus studentischer Sicht möglicherweise nicht unbedingt als wünschenswert angesehen wird.

Durch das Fehlen eines Interviewers im Rahmen der Online-Befragung kann hingegen von einer deutlichen Verringerung sozial erwünschter Antworten ausgegangen werden. Hierdurch könnte entsprechend die größere Zustimmung zum Verzicht auf Flugreisen erklärt werden: Es besteht hier keine Gefahr, sich als umweltbewusst zu Lasten kosmopolitischer Denkweisen offenbaren zu müssen.

Bei dem Item „*Klimaschutz ist alleinige Sache der Politik*“ konnte ebenfalls mit Hilfe des Mann-Whitney U-Tests ein signifikanter Unterschied hinsichtlich der zentralen Tendenz der Item-Bewertung zwischen den beiden Teilstichproben F2F und Online ermittelt werden. Während in den F2F-Interviews diesem Item mit einem Durchschnittswert von 4,59 eine sehr deutliche Ablehnung entgegengebracht wird, fällt die Ablehnung in der Online-Befragung mit einem arithmetischen Mittel von 4,4 immer noch deutlich, allerdings etwas abgeschwächt, aus. Der nachgewiesene Unterschied im Antwortverhalten kann hierbei möglicherweise ebenfalls der Erhebungsmethodik auf Grundlage der sozialen Erwünschtheit geschuldet sein. So ist bei diesem Item durchaus denkbar, dass im Rahmen der F2F-Interviews dahingehend sozial erwünscht geantwortet wird, dass einer Alleinverantwortung der Politik in Sachen Klimaschutz stärker als in der Online-Befragung widersprochen wird.

### **Verhaltensfrage zum Themenkomplex „Klimaschutz“**

In der Frage 1.3 des Fragebogens wurden die Befragten anschließend befragt, inwiefern und wie häufig sie bereits verschiedene, im Alltag einfach umzusetzende Maßnahmen zum Klimaschutz unternommen haben. Für diese Fragestellung, die sieben Items umfasste, wurde wie bei den vorhergehenden Fragen ebenfalls eine fünfstufige Skala vorgegeben, die in diesem Fall von „Tue ich oft“ (Skalenpunkt 1) bis „Tue ich nicht und habe es auch nicht vor“ (Skalenpunkt 5) reichte (siehe Skala 3 im Anhang). Entsprechend zu der Vorgehensweise bei den vorherigen Fragestellungen wurde auch hier mit Hilfe des Mann-Whitney U-Tests überprüft, ob sich zwischen der Online- und der F2F-Stichprobe signifikante Unterschiede im Antwortverhalten ermitteln lassen. Tabelle 17 zeigt zur Beantwortung dieser Frage daher einmal die Mittelwerte der beiden Teilstichproben, die sich daraus ergebenden Mittelwertdifferenzen sowie die Ergebnisse des Mann-Whitney U-Tests auf. Anhand der Ergebnisse des Mann-Whitney U-Tests zeigen sich bei drei der sieben Items signifikante Unterschiede.

So wurde von den Befragten im Rahmen der Online-Befragung bei dem Item „*Beim Kauf von Haushaltsgeräten auf den Energieverbrauch achten*“ entsprechend des Mittelwertes von 1,75 weitaus häufiger angegeben, darauf zu achten als dies von den Teilnehmern der F2F-Interviews angegeben wurde, wie auch der dort ermittelte Mittelwert von 2,21 zeigt. Auch beim Item „*Mit der Familie oder Freunden über das Thema Globale Erwärmung sprechen*“ konnte

mit Hilfe des Mann-Whitney U-Tests ein signifikanter Unterschied in den Verteilungen der beiden Stichproben ermittelt werden. Auch anhand des Mittelwertes zeigen sich diese Unterschiede entsprechend: Während mit einem Mittelwert von 2,56 die Teilnehmer der Online-Befragung im Durchschnitt angeben, „manchmal“ bis „eher selten“ mit Freunden oder der Familie über das Thema Globale Erwärmung zu sprechen, wird von den Teilnehmern der F2F-Interviews seltener dieses Thema diskutiert: Hier konnte nur ein Durchschnittswert von 2,80 registriert werden.

Beim Item „Aus Klimaschutzgründen auf Fernreisen verzichten“ konnten durch den Mann-Whitney U-Test ebenfalls höchst signifikante Unterschiede ermittelt werden. Die Nullhypothese, dass beide Verteilungen aus der gleichen Grundgesamtheit stammen, kann mit einer Irrtumswahrscheinlichkeit von 0,001 abgelehnt werden, was dahingehend interpretiert werden kann, dass sich die Verteilungen dieses Items in beiden Stichproben höchst signifikant voneinander unterscheiden.

**Tabelle 17: Mittelwerte, Mittelwertdifferenzen sowie Ergebnisse des Mann-Whitney U-Tests für die Items der Verhaltensfrage zum Themenkomplex „Klimaschutz“**

	Mittelwert			Mann-Whitney U-Test
	F2F	Online	Differenz	
Häufiger die Heizung runterdrehen, weniger warmes Wasser verbrauchen	1,84	1,81	-0,03	n.s.
Energiesparlampen kaufen	1,94	2,07	0,13	n.s.
Beim Kauf von Haushaltsgeräten auf Energieverbrauch achten	2,21	1,75	0,56	0,001
Produkte aus der näheren Umgebung kaufen	2,28	2,07	-0,21	n.s.
Das Auto stehen lassen und zu Fuß gehen bzw. Fahrrad oder öffentliche Verkehrsmittel nutzen	1,57	1,47	-0,10	n.s.
Mit der Familie oder Freunden über das Thema Globale Erwärmung sprechen	2,80	2,56	-0,24	0,003
Aus Klimaschutzgründen auf Fernreisen verzichten	4,06	3,45	-0,61	0,001

Anmerkungen: Weitere statistische Kennzahlen des Mann-Whitney U-Tests finden sich in Tabelle A 14 des Anhangs

a. Gruppierungsvariable: Befragungssample

Quelle: Auswertung der eigenen Erhebung.

Während in den F2F-Interviews angegeben wurde, bislang nicht auf Flugreisen zu verzichten, es aber für die Zukunft durchaus in Erwägung zu ziehen, was anhand eines Durchschnittswertes von 4,06 ersichtlich wird, wurde im Rahmen der Online-Interviews angegeben, in der Tendenz dies etwas häufiger zu tun. Hier lag das arithmetische Mittel bei 3,45. Auch hier ist durchaus denkbar, dass der Unterschied im Antwortverhalten bei diesem Item auf die Methode bzw. die Anwesenheit eines Interviewers und somit auf das Konstrukt der sozialen Erwünschtheit zurückgeführt werden können. Zum einen können hier die in der vorherigen Fragestellung aufgezeigten Erklärungsansätze zum unterschiedlichen Antwortverhalten hinsichtlich des Items zum Flugverzicht aufgegriffen werden. Ein alternativer Erklärungsan-

satz könnte hier entsprechend der theoretischen Erläuterungen in Kapitel 3.3.3 lauten, dass die befragten Studierenden es immer noch, trotz der zunehmenden Relevanz der Thematik, als sozial erwünscht erachten, ein möglichst normales, von Klimaschutzaspekten unbeeinflusstes Leben zu führen und sich entsprechend dieser Maxime zu verhalten. Dieses Erklärungsmuster kann dabei auch für die ermittelten signifikanten Unterschiede bei den beiden anderen Items dieser Fragestellung herangezogen werden.

#### 5.4.3 Verhaltensfrage zum Themenkomplex „Grüner“ Strom

Mit der Frage 1.4 des Fragebogens wurden die Befragten mit einer weiteren Verhaltensfrage konfrontiert: So sollten sie jeweils angeben, ob sie momentan oder in der Vergangenheit bereits „grünen“ Strom bezogen haben. Die Bezeichnung „grüner“ Strom steht dabei als Überbegriff für Elektrizität, die aus erneuerbaren Energiequellen wie Wind, Sonne, Biomasse oder Wasserkraft stammt (SINN 2008: 4). Möglich waren bei dieser Fragestellung die in Tabelle 18 dargestellten Antwortmöglichkeiten. Die Tabelle stellt weiterhin die Häufigkeiten sowie die Prozentanteile der Antwortmöglichkeiten, aufgeteilt nach dem Befragungssample, dar.

**Tabelle 18: Häufigkeiten sowie Prozentanteile der Fragestellung "Beziehst Du momentan 'Grünen' Strom oder hast Du schon mal 'Grünen' Strom bezogen?"**

	F2F-Stichprobe		Online-Stichprobe	
	Anzahl	Prozent	Anzahl	Prozent
Nein	114	60,96%	296	57,59%
Ja, momentan	12	6,42%	54	10,51%
Ja, aber nicht mehr	16	8,56%	22	4,28%
Weiß nicht	45	24,06%	142	27,63%
<i>Gesamt</i>	<i>187</i>	<i>100,00%</i>	<i>514</i>	<i>100,00%</i>

Quelle: Auswertung der eigenen Erhebung.

Bei der Betrachtung der Ergebnisse lässt sich erkennen, dass der überwiegende Teil der befragten Studierenden bisher noch keinen „Grünen“ Strom bezogen hat. Auch der Anteil der Studierenden, die momentan oder in der Vergangenheit bereits klimafreundlichen Strom nachgefragt haben, ist mit 14,98 % (F2F-Sample) bzw. 14,79 % (Online-Sample) ähnlich hoch – allerdings ist im direkten Vergleich der Anteil der Befragten, die momentan „Grünen“ Strom beziehen in der Online-Befragung höher. Umgekehrt verhält es sich in der F2F-Stichprobe: Hier ist der Anteil derer, die momentan „Grünen“ Strom beziehen, zu Lasten derer, die in der Vergangenheit „Grünen“ Strom bezogen haben, ausgeprägter. Ebenfalls ein leichter Unterschied ist hinsichtlich der „weiß nicht“-Antworten zwischen den beiden Befragungssamples erkennbar.

Um überprüfen zu können, ob das Antwortverhalten hinsichtlich des Bezugs von „Grünem“ Strom in einem Zusammenhang mit der angewendeten Befragungsmethode steht, wurde wei-

terhin der  $\chi^2$ -Test nach Pearson durchgeführt. Hierbei konnte ein signifikanter Zusammenhang zwischen dem Antwortverhalten sowie der Befragungsmethode ermittelt werden (vgl. Tabelle A 15 im Anhang). Um die Stärke des ermittelten Zusammenhangs bestimmen zu können, wurde zusätzlich Cramér's  $V^{28}$  als Maß für die Stärke des Zusammenhangs zwischen den beiden Variablen berechnet. Hierbei konnte für Cramér's  $V$  ein Wert von 0,107 ermittelt werden, was dahingehend interpretiert werden kann, dass nur ein schwacher Zusammenhang zwischen den beiden Variablen besteht.

#### 5.4.4 Resultate des Quiz zum Themenkomplex „Klimaschutz“

Zu Beginn des zweiten Fragebogenabschnitts wurden die Befragten in beiden Teilstichproben jeweils gebeten, an einem Quiz zum Thema Klimaschutz teilzunehmen. Neben der Feststellung des Wissens der Teilnehmer rund um diesen Themenkomplex zeichnet sich diese als Überzeugungsfrage formulierte Fragestellung ebenfalls durch ihre Eignung aus, mögliche Unterschiede in den Antworten, die auf Methodeneffekte zurückgeführt werden können, aufdecken zu können.

Hinsichtlich der Methodeneffekte spielt auch hier wieder die An- bzw. Abwesenheit des Interviewers eine gewichtige Rolle: So übernimmt dieser im Rahmen der F2F-Interviews die Kontrolle über die Interviewsituation sowie auch über die Beantwortung der Fragestellung. Dadurch können entsprechend äußere Einflüsse, die zu einer Verzerrung des Ergebnisses führen könnten, wie etwa die Anwesenheit weiterer Personen oder die Nutzung von Medien wie Büchern oder dem Internet, kontrolliert werden. In Online-Befragungen hingegen ist eine derartige Kontrolle nicht möglich (TADDICKEN 2008: 51). Dieser Sachverhalt macht den Vergleich der Quiz-Ergebnisse besonders interessant: Denn möglicherweise wurden im Rahmen der Online-Befragung von den Teilnehmern Internet-Suchmaschinen oder Online-Enzyklopädien verwendet, welche in kurzer Zeit dem Nutzer viele Informationen bereit stellen. Da es sich bei den Quizfragen nicht zwangsläufig um voraussetzbares Allgemeinwissen handelt, kann ein daher rührender Einfluss und somit ein Methodeneffekt (mangels einer Kontrolle der Interviewsituation durch den Interviewer) nicht ausgeschlossen werden.

---

<sup>28</sup> Zur Ermittlung und Beschreibung eines statistischen Zusammenhangs zwischen zwei Merkmalen X und Y kann der von Cramér vorgeschlagene Koeffizient V herangezogen werden, der als Weiterentwicklung des sogenannten Phi-Koeffizienten angesehen werden kann und wie folgt definiert ist:

$$V = \sqrt{\frac{\chi^2}{N \cdot (L - 1)}}, \text{ mit } L = \min(k, m),$$

wobei k die Anzahl der Zeilen und m die Anzahl der Spalten bezeichnet. Cramér's  $V$  kann Werte von 0 bis 1 annehmen, wobei 1 für einen perfekten Zusammenhang und 0 für keinen Zusammenhang zwischen X und Y steht (BENNINGHAUS 2001: 212f.; BORTZ et al. 2008: 355ff.).

Im Rahmen der Fragestellung wurden die Befragten in fünf Items mit verschiedenen Aussagen zum Klimaschutz konfrontiert, die sie als richtig oder falsch kennzeichnen sollten. Als alternative Auswahlmöglichkeit wurde zusätzlich in beiden Teilstichproben noch „weiß nicht“ angeboten. Die verschiedenen Items inklusive der jeweils korrekten Antwort werden in Tabelle 19 aufgezeigt. Weiterhin wird in der Tabelle dargestellt, wie viel Prozent der Befragten in den beiden Teilstichproben jeweils die korrekte Antwort gegeben haben. Bei Vergleich der Prozentanteile der korrekt gegebenen Antworten zwischen den beiden Teilstichproben zeigen sich dabei mit einer Ausnahme insgesamt sehr ähnliche Ergebnisse. Dies bestätigen auch die Ergebnisse des jeweils durchgeführten  $\chi^2$ -Tests, die ebenfalls in der Tabelle aufgeführt sind (für die Teststatistik siehe Tabelle A 16 im Anhang).

**Tabelle 19: Resultate des Quiz zum Klimaschutz, aufgeteilt nach Teilstichproben.**

	Korrekte Antwort	Korrekt gegeben Antworten in Prozent		Ergebnis $\chi^2$ -Test
		F2F	Online	
CO <sub>2</sub> , auch Kohlendioxid genannt, ist ein Gas, das bei der Verbrennung von Erdöl, Erdgas, Benzin, Kohle oder auch Holz entsteht.	Richtig	97,2%	96,0%	n.s.
Wenn CO <sub>2</sub> in die Atmosphäre gelangt, wirkt es dort wie eine Isolierung. Dadurch steigt die Durchschnittstemperatur an und das Klima verändert sich.	Richtig	79,3%	74,0%	n.s.
Es gibt keinen vom Menschen verursachten Klimawandel.	Falsch	90,7%	88,0%	n.s.
Der Ausstoß an Treibhausgasen ist bei der Erzeugung von tierischen Lebensmitteln wie Fleisch, Milch und Eiern wesentlich höher als bei pflanzlichen Erzeugnissen.	Richtig	62,2%	73,0%	0,009
Die menschliche Ernährung trägt nur minimal zum Klimawandel bei.	Falsch	78,2%	79,3%	n.s.

Quelle: Auswertung der eigenen Erhebung.

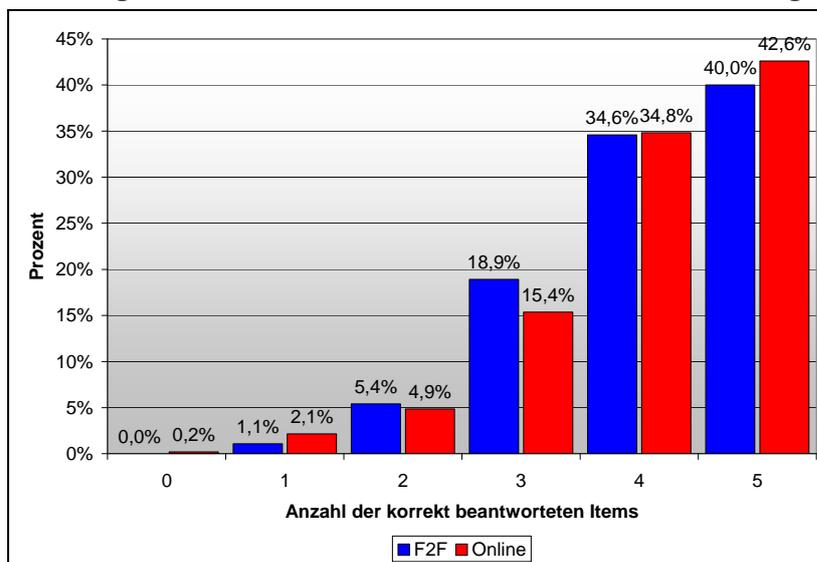
Nur bei dem Item „*Der Ausstoß an Treibhausgasen ist bei der Erzeugung von tierischen Lebensmitteln wie Fleisch, Milch und Eiern wesentlich höher als bei pflanzlichen Erzeugnissen*“ konnte demnach ein hochsignifikanter Zusammenhang zwischen der Befragungsmethode und dem Antwortverhalten ermittelt werden. Allerdings ergibt sich bei der Berechnung von Cramér's V nur ein Wert von 0,117, welcher dahingehend interpretiert werden kann, dass der mit Hilfe des  $\chi^2$ -Tests ermittelte Zusammenhang nur schwach ausgeprägt ist.

Aufgrund des Sachverhalts, dass bei diesem Item nur ein relativ schwach ausgeprägter sowie bei den anderen Items kein Zusammenhang zwischen dem Antwortverhalten und der Befragungsmethode ermittelt werden konnte, lässt den Schluss zu, dass möglicherweise nicht die

Befragungsmethode ursächlich dafür ist, sondern stattdessen andere Möglichkeiten in Betracht gezogen werden müssen. Hierauf wird in Kapitel 5.7 noch einmal eingegangen.

Bei einer Analyse der Fragestellung dahingehend, wie viele Items von den Befragten jeweils korrekt beantwortet wurden, zeigen sich ebenfalls keine starken Unterschiede zwischen den beiden Teilstichproben, wie aus Abbildung 15 ersichtlich wird. So konnten jeweils etwa 40,0 % alle fünf Items korrekt beantworten, vier Items stellten immerhin für knapp 35,0 % der Befragten keine Schwierigkeit dar. Drei Viertel der Befragten konnten somit mindestens vier der fünf Items korrekt beantworten.

**Abbildung 15: Anzahl der korrekt beantworteten Items dieser Fragestellung**



Quelle: Auswertung der eigenen Erhebung.

Auch hier sollte weiterhin mit einem geeigneten Testverfahren überprüft werden, ob die Anzahl der korrekt gegebenen Antworten von der Befragungsmethodik abhängt. Ausgewählt wurde hierfür wiederum der Mann-Whitney U-Test, da, wie Abbildung 15 bereits vermuten lässt, eine Normalverteilung der Daten nicht zu erwarten ist. Diese Annahme konnte durch den Kolmogorov-Smirnov-Test auf Normalverteilung entsprechend bestätigt werden (siehe Tabelle A 17 im Anhang). Als Ergebnis des Mann-Whitney U-Tests, dessen Ergebnisse in Tabelle A 18 des Anhangs tabellarisch dargestellt werden, kann festgehalten werden, dass kein signifikanter Unterschied hinsichtlich der zentralen Tendenz bei der Anzahl der korrekt gegebenen Antworten zwischen den beiden Teilstichproben ermittelt werden konnte. Somit kann ausgesagt werden, dass beide Teilstichproben aus der gleichen Grundgesamtheit stammen. Entsprechend konnte auch hier kein Hinweis für einen Methodeneffekt ermittelt werden, der auf dem Fehlen eines Interviewers in der Online-Befragung und der damit verbundenen geringeren Kontrolle der Interviewsituation beruht.

### 5.4.5 Überzeugungsfrage zu verschiedenen Gütezeichen

Neben dem oben besprochenen Quiz wurden die Befragten mit einer weiteren Überzeugungsfrage konfrontiert, deren Frageninhalt die Kenntnis verschiedener Gütezeichen<sup>29</sup> aus dem Bereich der Ernährungsindustrie bzw. des Lebensmittelhandels umfasste. Anhand einer vierstufigen Skala mit den Endpolen eins („Kenne ich gut“) bis vier („Noch nie gesehen“) sollten die Befragten dabei jeweils ihr Wissen bzw. ihre Kenntnis über die verschiedenen Gütezeichen einordnen. Die verwendete Skala wird im Anhang durch Skala 4 dargestellt.

**Tabelle 20: Kenntnis der Befragten über verschiedene Zeichen aus der Bereich der Ernährungsindustrie bzw. des Lebensmittelhandels**

	Mittelwert F2F	Mittelwert Online	Differenz	Ergebnis Mann-Whitney U-Test
Zeichen 1: 	1,23	1,21	0,02	n.s.
Zeichen 2: 	2,11	2,1	0,01	n.s.
Zeichen 3: 	1,75	1,64	0,11	n.s.
Zeichen 4: 	3,21	2,75	0,46	0,001
Zeichen 5: 	3,76	3,59	0,17	0,009
Anmerkungen: Weitere statistische Kennzahlen des Mann-Whitney U-Tests finden sich in Tabelle A 19 des Anhangs				

Quelle: Auswertung der eigenen Erhebung.

Tabelle 20 zeigt weiterhin die ausgewählten Zeichen sowie die berechneten Mittelwerte der beiden Teilstichproben. Während die Zeichen eins bis vier real existierende Zeichen darstellen, die jedem Befragten beim Lebensmittelkauf begegnen können, stellt Zeichen fünf ein selbstentwickeltes und in Deutschland nicht real existierendes Zeichen für einen PCF eines Produktes dar. Dieses Zeichen wurde 2008 im Rahmen des Moduls MP 74 des Masterprogramms am Fachbereich 09 der Justus-Liebig-Universität entwickelt und weist Ähnlichkeiten

<sup>29</sup> Eine Übersicht und Abgrenzung verschiedener Gütezeichen aus dem Lebensmittel-Bereich bietet HENKEL (2008: 12ff.).

zum ebenfalls 2008 präsentierten PCFs von Tesco in Großbritannien auf (TESCO 2009). Vor allem bei diesem unbekanntem Zeichen können Unterschiede im Antwortverhalten durch soziale Erwünschtheit nicht ausgeschlossen werden. So ist es durchaus denkbar, dass die Befragten in den F2F-Interviews vor dem Interviewer nicht zugeben möchten, dieses Zeichen nicht zu kennen. Als verstärkender Effekt kommt hinzu, dass die vorher gezeigten Zeichen real existent sind und sich dieser Eindruck möglicherweise auf das Zeichen übertragen könnte.

Bei Betrachtung der Mittelwerte zeigt sich, dass zumindest die Zeichen eins bis drei in den beiden Stichproben eine ähnliche Bekanntheit aufweisen, in der Online-Befragung sich die Befragten marginal jedoch eine bessere Kenntnis der Zeichen postulieren. Die höchste Bekanntheit weist dabei das deutsche Bio-Siegel mit einem Mittelwert von 1,23 bzw. 1,21 auf. Darauf folgen das Fairtrade-Zeichen mit einem Mittelwert von 1,75 bzw. 1,64 sowie das CMA-Gütezeichen mit einem Mittelwert von 2,11 bzw. 2,10. Mit Hilfe des hier weiterhin herangezogenen Mann-Whitney U-Tests konnten bei diesen drei Zeichen keine signifikanten Unterschiede hinsichtlich der Verteilung bzw. der zentralen Tendenz ermittelt werden. Ein Methodeneffekt ist, auch im Hinblick auf die sehr nah zusammen liegenden Mittelwerte, entsprechend nicht nachzuweisen.

Anders sieht dies bei den Zeichen vier („Marine Stewardship Council MSC<sup>30</sup>“) und fünf („Product Carbon Footprint“) aus. Hier zeigen sich bereits anhand der Mittelwerte starke Differenzen zwischen den beiden Teilstichproben. So ist die Bekanntheit des MSC-Zeichens in der F2F-Stichprobe deutlich schwächer ausgeprägt als in der Online-Stichprobe, wie auch die Mittelwertdifferenz von 0,46 erkennen lässt. Auch beim PCF ist ein gleichgerichteter, allerdings schwächer ausgeprägter Sachverhalt anhand der Mittelwertdifferenz von 0,17 erkennbar. Durch den angewendeten Mann-Whitney U-Test kann entsprechend die Nullhypothese, dass sich die Verteilungen der beiden Stichproben nicht voneinander unterscheiden, mit einer Irrtumswahrscheinlichkeit von 0,001 bzw. 0,009 abgelehnt. Entgegen der obigen Vermutung, dass aufgrund eines möglichen Interviewereinflusses die Bekanntheit der Zeichen in der F2F-Stichprobe höher ausfallen könnte, zeigt sich über alle fünf Zeichen hinweg das gegenteilige Bild: Die Bekanntheit ist in beiden Stichproben jeweils ähnlich groß (Zeichen eins bis drei) oder liegt bei dem MSC-Zeichen sowie dem PCF unter dem Wert der Online-Befragung. Zur Erklärung kann hier alternativ möglicherweise ebenfalls der in Kapitel 5.4.4 dargestellte Aspekt hinsichtlich der fehlenden Situationskontrolle durch den Interviewer in der Online-Befra-

---

<sup>30</sup> Bei dem „Marine Stewardship Council“, kurz MSC, handelt es sich um eine unabhängige Organisation, die nachhaltig arbeitende Fischerei zertifiziert. Die Zertifizierung ist dabei anhand des blauen MSC-Zeichens zu erkennen (MSC 2010a; MSC 2010b).

gung angeführt werden: So ist es durchaus denkbar, dass die durch den Mann-Whitney U-Test nachgewiesenen Unterschiede hinsichtlich der Bekanntheit der beiden Zeichen aus der Nutzung von Internet-Suchmaschinen oder Online-Enzyklopädien her rühren. Ob diese dabei auf der Methodik beruhen, muss im weiteren Verlauf der Arbeit geklärt werden.

#### 5.4.6 Fragestellungen zum Themenkomplex „Carbon Labelling“

##### Überzeugungsfrage zum Carbon Labelling

In der Frage 2.3 des Fragebogens wurden die Befragten nochmals mit dem Zeichen fünf, also dem Product Carbon Footprint-Zeichen, konfrontiert. In dieser offen formulierten Fragestellung, deren genaue Formulierung dem Fragebogen im Anhang entnommen werden kann, wurden die Befragten jeweils gebeten, anzugeben, was aus ihrer Sicht das Zeichen bedeuten könne bzw. für was es stehe. Da im vorherigen Abschnitt signifikante Unterschiede in der Bekanntheit des Zeichens zwischen der F2F- und der Online-Stichprobe nachgewiesen werden konnten, soll in diesem Abschnitt nun entsprechend überprüft werden, ob der ermittelte Unterschied in der Bekanntheit bzw. des Wissens über dieses imaginäre Zeichen anhand der offenen Fragestellung bestätigt werden kann.

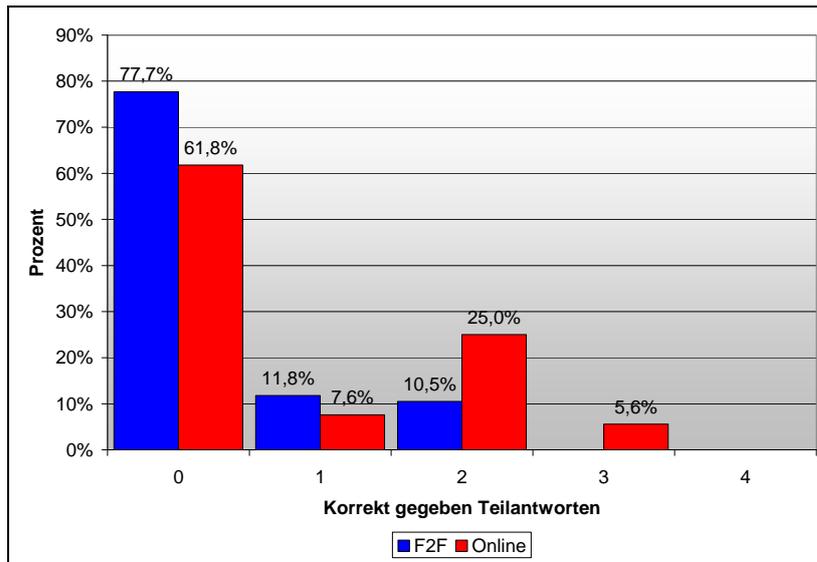
Die Vorgehensweise bei der Auswertung dieser Fragestellung ist wie folgt: Die gegebenen Antworten der Befragten in der offenen Frage wurden mit einer Musterlösung verglichen, die auf den Ausführungen des Kapitels 2.2 beruht und im grauen Kasten aufgezeigt wird.

• Das Zeichen gibt an, welche Menge an klimaschädigenden Emissionen (zum Beispiel CO <sub>2</sub> )	<i>[Menge CO<sub>2</sub>]</i>
• im gesamten Lebenszyklus des Produktes, also beispielsweise bei Herstellung, Verarbeitung, Transport, Verwendung und Entsorgung freigesetzt werden.	<i>[Produktionsstufen]</i>
• Man nennt dies den CO <sub>2</sub> -Fußabdruck oder „Carbon Footprint“ eines Produktes.	<i>[Bezeichnung Carbon Footprint bzw. CO<sub>2</sub>-Fußabdruck]</i>
• Die Mengenangabe bezieht sich auf 100g Produktmasse.	<i>[Bezug auf bestimmte Produktmasse].</i>

Die Musterlösung besteht dabei aus vier Teilen, die jeweils einen inhaltlichen Schwerpunkt fokussieren. Diese inhaltlichen Schwerpunkte werden jeweils rechts neben den einzelnen Teilen in den eckigen Klammern genannt und beziehen sich einmal auf die Menge an CO<sub>2</sub>, die ausgestoßen wird und im Zeichen durch die 125g dargestellt wird. Des Weiteren bezieht sich der nächste Schwerpunkt auf verschiedene Produktionsstufen, in denen CO<sub>2</sub> freigesetzt werden kann. Der dritte Schwerpunkt bezieht sich auf die Bezeichnung des dargestellten Zeichens

als „CO<sub>2</sub>-Fußabdruck“ bzw. „Product Carbon Footprint“, während der letzte Schwerpunkt die Relativierung der Mengenangabe an CO<sub>2</sub> auf eine bestimmte Produktmasse fokussiert. Wurde von einem Befragten der jeweilige Schwerpunkt in seiner Erklärung genannt, wurde dafür ein Punkt vergeben. Bei Nennung aller vier relevanter Aspekte konnte ein Befragter somit einen Score von vier Punkten erreichen. Die prozentualen Verteilungen der Scores beider Teilstichproben stellt Abbildung 16 dar.

**Abbildung 16: Prozentanteile der korrekt gegebenen Teilantworten nach Stichproben**



Quelle: Auswertung der eigenen Erhebung.

Danach konnte der überwiegende Teil der Befragten sowohl in der F2F- als auch in der Online-Stichprobe nichts mit dem Zeichen verbinden und entsprechend keine der vier möglichen Teillösungen korrekt benennen. In der F2F-Stichprobe ist der Anteil dieser Befragten mit 77,7 % dabei deutlich höher als in der Online-Stichprobe: Hier waren es immerhin nur 61,8 %, die mit dem Zeichen keine korrekte Assoziation verbinden konnten. Auch ist die Zahl der Befragten, die immerhin zwei Teillösungen korrekt geben konnten, mit 25,0 % deutlich höher als in der F2F-Stichprobe. Über beide Stichproben hinweg gesehen waren es dabei vor allem die beiden Teillösungen „Menge CO<sub>2</sub>“ und „Produktionsstufe“, die von den Befragten korrekt benannt wurden. Darüber hinaus waren weiterhin 5,6 % der Befragten der Online-Stichprobe in der Lage, drei Aspekte korrekt zu benennen, während dies in der F2F-Befragung niemanden gelang. Wie aus der Abbildung ebenfalls entnommen werden kann, gelang es über beide Teilstichproben hinweg keinem Befragten, alle vier Aspekte korrekt zu benennen.

Das unterschiedliche Wissen der Befragten hinsichtlich des Product Carbon Footprint-Zeichens bestätigt sich weiterhin bei Vergleich der Mittelwerte der beiden Teilstichproben: Wäh-

rend in der F2F-Stichprobe ein Mittelwert von durchschnittlich 0,33 korrekten Teilantworten ermittelt werden konnte, gaben in der Online-Stichprobe die Befragten durchschnittlich 0,75 korrekte Antworten. Die unterschiedliche Verteilung der erreichten Score-Werte bestätigt weiterhin auch der Mann-Whitney U-Test. Nach diesem ergibt sich mit einer Irrtumswahrscheinlichkeit von 0,001 ein höchstsignifikanter Unterschied in den Verteilungen der beiden Teilstichproben (vgl. Tabelle A 20 im Anhang). Als möglicher Erklärungsgrund kann hier wie bei den vorherigen Überzeugungsfragen wieder ein Methodeneffekt auf Basis der in der Online-Befragung fehlenden Kontrolle durch den Interviewer genannt werden (siehe hierzu Kapitel 3.3.3). Wie bei den vorangegangenen Überzeugungsfragen ist es auch hier denkbar, dass von einem Teil der Befragten das Internet zu einer kurzfristigen Recherche genutzt wurde. Naheliegend ist dies auch unter dem Gesichtspunkt, dass neben bei der Beantwortung eines Online-Fragebogens nur ein neues, zusätzliches Browserfenster geöffnet werden muss, um sich kurzfristig relevante Informationen beschaffen zu können. Diese Vermutung kann dabei zumindest punktuell durch die Antwort eines Befragten in der Online-Befragung bestätigt werden: „Nach kurzer Google-Info weiß ich nun auch, was es bedeutet, aber ich hätte ansonsten keine Ahnung gehabt“. Dass entsprechend des Zitats über Internetsuchmaschinen schnell eine Vielzahl an Informationen zum Befragungsthema „Carbon Labelling“ zu finden ist, kann durch den Verfasser der vorliegenden Arbeit bestätigt werden: So liefert die weltweit am häufigsten verwendete Suchmaschine Google (CHEN et al 2009: 10; WEBHITS 2010) innerhalb von 0,09 Sekunden Suchzeit über eine Million Ergebnisse zu diesem Thema (GOOGLE 2010). Nicht unwahrscheinlich ist dabei, dass diese Vorgehensweise von einem größeren Teil der Befragten angewendet worden ist. Hierdurch kann möglicherweise der ermittelte Unterschied hinsichtlich der unterschiedlichen Verteilung des Wissens zum Product Carbon Footprint-Zeichen in den beiden Teilstichproben erklärt werden. Da dies entsprechend auf die Nichtkontrollierbarkeit der Befragungssituation im Rahmen der Online-Befragung zurückzuführen wäre, kann bei dieser Fragestellung ebenfalls ein Einfluss der Befragungsmethodik auf die Ergebnisse vermutet werden.

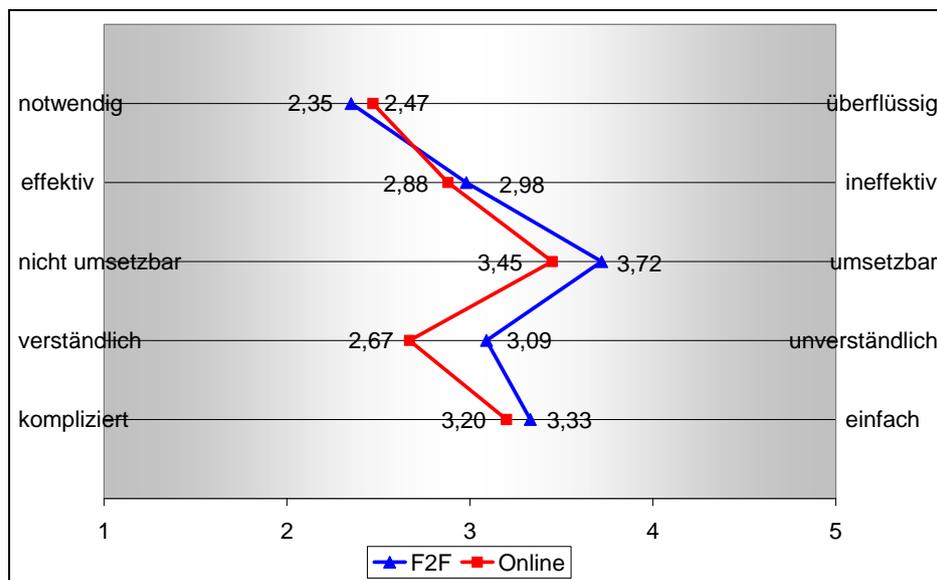
Ob sich allerdings auch die Einstellungen der Befragten in den beiden Teilstichproben signifikant voneinander unterscheiden, und somit einen Hinweis auf einen Methodeneffekt liefern, soll in der folgenden Fragestellung analysiert werden. Hierfür spielte es eine wichtige Rolle in beiden Teilbefragungen, dass ein grundsätzlich einheitliches Wissen bezüglich des Product Carbon Footprint-Zeichens vorausgesetzt werden konnte. Daher wurde den Befragten vor der Fortsetzung der Befragung anhand der auf Seite 110 dargestellten Definition das Carbon Footprint-Zeichen erläutert. In der Online-Befragung wurde dafür eine eigene Textseite vor-

gesehen, die von den Befragten bewusst weitergeklickt werden musste. In der F2F-Befragung wurde die Erklärung hingegen den Befragten vorgelesen.

### Einstellungen zum Carbon Footprint-Zeichen

Das in der vorangegangenen Fragestellung bereits schwerpunktmäßig thematisierte Product Carbon Footprint-Zeichen sollte von den Befragten im Rahmen der Fragestellung 2.4 in Form eines semantischen Differentials bewertet werden. Hierdurch sollten, simultan zu anderen Fragestellungen des Fragebogens, mögliche Unterschiede, die auf Methodeneffekten basieren können, herausgearbeitet werden. Die genaue Fragestellung lautete dabei „*Hältst Du ein solches ‚CO<sub>2</sub>-Zeichen‘ für...*“, gefolgt von den jeweiligen Eigenschaftspaaren, mit deren Hilfe das Zeichen aus Befragtensicht eingeordnet werden sollte. Als zusätzliche Antwortmöglichkeit wurde weiterhin ein „weiß nicht“ gegeben, um Verzerrungen durch Befragte zu vermeiden, die bei dieser Fragestellung keine Wertung abgeben wollten oder konnten. Die vorgegebenen Eigenschaftspaare werden in Abbildung 17 samt den für beide Stichproben ermittelten semantischen Differentialen dargestellt.

**Abbildung 17: Semantische Differentiale der beiden Teilstichproben zum Product Carbon Footprint-Zeichen**



Quelle: Auswertung der eigenen Erhebung.

Die semantischen Differentiale ergeben sich dabei aus den Stichprobenmittelwerten, welche ebenfalls in der Abbildung angegeben sind. Hierbei lassen sich vor allem bei den drei Eigenschaftspaaren „*nicht umsetzbar/umsetzbar*“, „*verständlich/unverständlich*“ und „*kompliziert/einfach*“ Unterschiede in den Bewertungen zwischen den Teilstichproben erkennen. Signifikante Unterschiede konnten durch den Mann-Whitney U-Test jedoch nur bei den beiden Ei-

genschaftspaaren „*nicht umsetzbar/umsetzbar*“ und „*verständlich/unverständlich*“ nachgewiesen werden (vgl. Tabelle A 21 im Anhang).

Insgesamt zeigt sich bei der Betrachtung der semantischen Differentiale, dass von den Befragten der Online-Stichprobe in der Tendenz eher die linken Eigenschaften präferiert werden bzw. das semantische Differential nach links verschoben ist, während von den Teilnehmern der F2F-Interviews eher die Eigenschaften auf der rechten Seite bevorzugt werden. Da die positiven und negativen Attribute nicht auf einer Seite des semantischen Differentials angeordnet sind, kann entsprechend aufgrund des Fehlens eines eindeutigen inhaltlichen Trends ein Methodeneffekt auf Basis sozialer Erwünschtheit bzw. von Interviewereffekten prinzipiell ausgeschlossen werden. Stattdessen könnten die ermittelten Unterschiede im Antwortverhalten bei den beiden Eigenschaftspaaren „*nicht umsetzbar/umsetzbar*“ und „*verständlich/unverständlich*“ durch verschiedene Faktoren der Informationsübermittlung wie etwa die graphische Darstellung des Fragebogens eine Rolle spielen. Denkbar ist etwa, dass trotz gleicher Darstellung der Fragestellung in beiden Fragebögen, im Rahmen der Online-Befragung aufgrund des Leseverhaltens von links nach rechts (MAYERHOFER 2006: 474) Itempunkte tendenziell im linken Bereich gewählt werden, während im Rahmen der F2F-Interviews dieser Effekt durch die unterstützenden Erklärungen des Interviewers unterdrückt bzw. abgeschwächt wird.

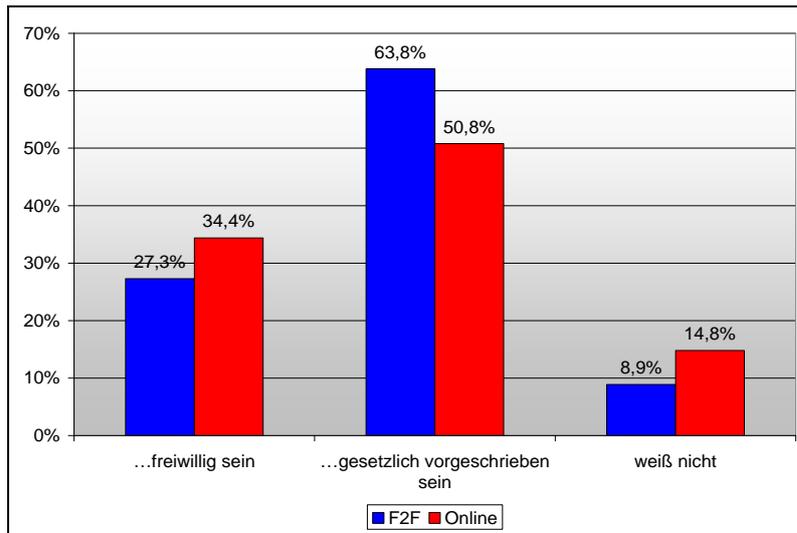
### **Fragestellungen zur rechtlichen Regelung des Product Carbon Footprint-Zeichens**

Welche gesetzlichen Regelungen von den Befragten bezüglich eines solchen Product Carbon Footprint-Zeichens bevorzugt würden, wurde in Fragestellung 2.5 des Fragebogens erhoben. Während in dem ersten Teil der Fragestellung erhoben wurde, ob ein solches Zeichen freiwillig oder stattdessen gesetzlich vorgeschrieben sein sollte, ging es im zweiten Teil der Fragestellung um die mögliche graphische Ausgestaltung des Zeichens.

So sollte von den Befragten hier entschieden werden, ob das Zeichen anhand einer konkreten Mengenangabe den durch das Lebensmittel verursachten CO<sub>2</sub>-Ausstoß angeben sollte oder stattdessen eine vereinfachte Form des Zeichens verwendet werden sollte, welche ohne eine konkrete Mengenangabe auskommt und stattdessen nach dem Prinzip der Ampel eine schlechte, durchschnittliche oder gute CO<sub>2</sub>-Bilanz durch die Farben Rot, Gelb und Grün symbolisiert. Wie aus der Abbildung 18 hervorgeht, zeigen sich hinsichtlich der Fragestellung, ob ein solches Zeichen freiwillig oder gesetzlich vorgeschrieben sein sollte, gewisse Unterschiede zwischen den beiden Teilstichproben. Diese Unterschiede können durch den herangezogenen  $\chi^2$ -Test nach Pearson bestätigt werden. Dieser kommt zu dem Ergebnis, dass die Be-

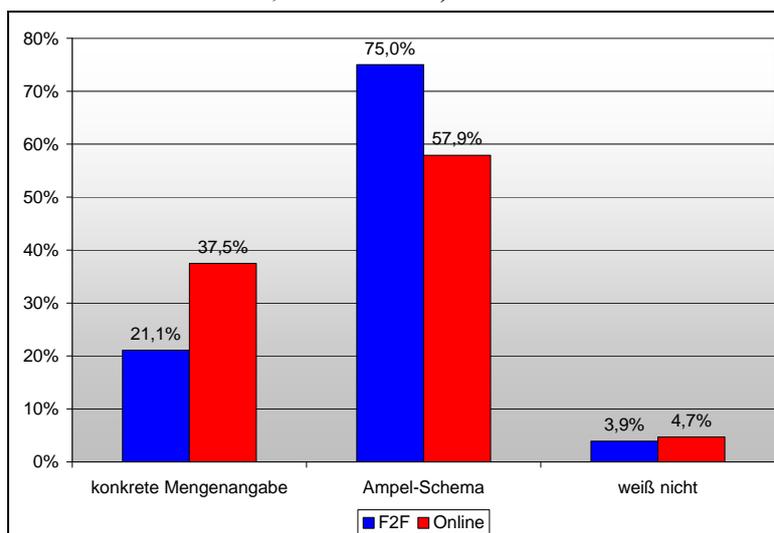
fragungsmethode mit einer Irrtumswahrscheinlichkeit von 0,005 einen höchst signifikanten Einfluss auf das Antwortverhalten hat, da durch den Signifikanztest ein höchst signifikanter Zusammenhang nachgewiesen werden konnte (vgl. Tabelle A 22 im Anhang). Bei der Ermittlung der Stärke des Zusammenhangs mit Hilfe des Zusammenhangsmaßes Cramér's V zeigt sich anhand des berechneten Wertes von 0,122 allerdings, dass dieser ermittelte Zusammenhang nur schwach ausgeprägt ist.

**Abbildung 18:** Häufigkeitsverteilungen zur Fragestellung, ob ein CO<sub>2</sub>-Zeichen freiwillig oder gesetzlich vorgeschrieben sein sollte (F2F n=186; Online n=514)



Quelle: Auswertung der eigenen Erhebung.

**Abbildung 19:** Häufigkeitsverteilungen der Fragestellung zur Gestaltung eines CO<sub>2</sub>-Zeichens (F2F n=186; Online n=514)



Quelle: Auswertung der eigenen Erhebung.

Ebenfalls Unterschiede zwischen den beiden Teilstichproben und somit zwischen den Befragungsmethoden zeigen sich bei der zweiten Teilfrage nach der Ausgestaltung des Zeichens. Zwar wird grundsätzlich in beiden Stichproben eine Ampel-ähnliche Ausgestaltung des Zei-

chens bevorzugt, allerdings ist der Anteil der Befragten, der diese Ausgestaltung präferiert, in den F2F-Interviews deutlich höher als in der Online-Befragung, wie Abbildung 19 ebenfalls darstellt.

Auch bei der zweiten Teilfrage konnte mit einer Irrtumswahrscheinlichkeit von 0,001 ein höchst signifikanter Zusammenhang zwischen der Frage nach der Ausgestaltung des Zeichens und dem Befragungssample durch den  $\chi^2$ -Test nach Pearson ermittelt werden (vgl. hierzu Tabelle A 23 im Anhang). Allerdings deutet auch hier wiederum der Wert des Zusammenhangsmaßes Cramérs V von 0,159 lediglich auf einen schwach ausgeprägten Zusammenhang hin.

#### **5.4.7 Zahlungsbereitschaft für Milch mit einer reduzierten CO<sub>2</sub>-Emission**

In der letzten inhaltlich-thematisch ausgerichteten Fragestellung des Fragebogens wurden die Befragten zu ihrer Zahlungsbereitschaft für Milch befragt. Hierbei sollten die Befragten einmal angeben, wie viel sie normalerweise für einen Liter Milch bezahlen. Anschließend wurden die Befragten zusätzlich gebeten, anzugeben, wie hoch ihre Zahlungsbereitschaft für einen Liter Milch sei, bei dessen Produktion 30 % weniger CO<sub>2</sub>-Emissionen freigesetzt wurden. Aus diesen beiden Werten wurde anschließend durch Differenzbildung die Mehr- oder Minderzahlungsbereitschaft für die Milch mit den niedrigeren CO<sub>2</sub>-Emissionen im Vergleich zu der konventionell erzeugten Milch berechnet. Die Differenzbildung erlaubt es dabei, von unterschiedlichen Preisniveaus, welche auf unterschiedlichen Fettgehalten, Marken oder Prozessqualitäten basieren, abstrahieren zu können. Für die im Rahmen eines Methodenvergleichs interessierende Größe der Zahlungsbereitschaftsdifferenz kann dabei möglicherweise erwartet werden, dass sie in den F2F-Interviews aufgrund von Methodeneffekten, die auf die Anwesenheit eines Interviewers zurückgeführt werden können, durchschnittlich einen höheren Wert als in Online-Befragungen annehmen wird. So ist durchaus denkbar, dass Befragte in den F2F-Interviews aufgrund von sozialer Erwünschtheit für die Milch mit den niedrigeren CO<sub>2</sub>-Emissionen eine höhere Mehrzahlungsbereitschaft äußern werden als die Befragten in den Online-Interviews. Die deskriptive Statistik zu der neu gebildeten Variablen „Zahlungsbereitschaftsdifferenz“ zeigt Tabelle 21 auf.

So zeigen sich hinsichtlich der Minimal- und der Maximalwerte zwischen beiden Teilstichproben sehr deutliche Unterschiede. Während in der F2F-Stichprobe im Minimum die Befragten gleich viel für die Milch mit den niedrigeren CO<sub>2</sub>-Emissionen zahlen würden, ist in der Online-Stichprobe die Differenz bei mindestens einer befragten Person sogar negativ: Diese Person würde entsprechend für die Milch mit den niedrigeren Emissionen an CO<sub>2</sub> sogar nur

weniger als für die konventionell produzierte Milch zahlen wollen. Hinsichtlich des Maximalwertes liegt in der F2F-Stichprobe die maximale Mehrzahlungsbereitschaft für die klimafreundlichere Milch bei 1,41 €, während sie in der Online-Stichprobe sogar 2,00 € beträgt. Aufgrund der unterschiedlichen Spannweiten ergeben sich auch die unterschiedlichen Werte beim arithmetischen Mittel, beim Median sowie beim 5 % getrimmten Mittel. Auch ist die Standardabweichung in der Online-Stichprobe größer, was dahingehend interpretiert werden kann, dass die verschiedenen Messwerte gestreuter um das arithmetische Mittel angeordnet sind.

**Tabelle 21: Deskriptive Statistik der Variablen "Zahlungsbereitschaftsdifferenz"**

	F2F-Stichprobe (n=168)	Online-Stichprobe (n=458)
Minimum	0,00	-0,85
Maximum	1,41	2,00
Arith. Mittel	0,262	0,246
Median	0,213	0,200
Getr. Mittel	0,243	0,236
SD	0,221	0,310

Anmerkungen: Arithm. Mittel = Arithmetisches Mittel;  
Getr. Mittel = 5% Getrimmtes Mittel; SD = Standardabweichung.

Quelle: Auswertung der eigenen Erhebung.

Um überprüfen zu können, ob sich beide Teilstichproben hinsichtlich ihrer Mehrzahlungsbereitschaft signifikant voneinander unterscheiden, wurde zuerst überprüft, ob die Verteilungen der Variablen der Normalverteilung folgen. Der hierfür herangezogene Kolmogorov-Smirnov-Test konnte dabei weder für die beiden Teilstichproben noch für die Gesamtstichprobe eine Normalverteilung ermitteln (vgl. Tabelle A 24 sowie Tabelle A 25 im Anhang), weshalb auch bei dieser Fragestellung wieder der Mann-Whitney U-Test zur Entscheidungsfindung herangezogen wird. Dieser kommt dabei zu dem Ergebnis, dass sich die Differenzen der Zahlungsbereitschaften zwischen den beiden Teilstichproben nicht signifikant voneinander unterscheiden (vgl. Tabelle A 26 im Anhang). In der zentralen Tendenz unterscheiden sich somit die Befragten der beiden Stichproben „F2F“ und „Online“ hinsichtlich der Mehrzahlungsbereitschaft für die Milch mit den reduzierten CO<sub>2</sub>-Emissionen nicht voneinander. Entsprechend kann die oben geäußerte Möglichkeit eines Methodeneffektes bei dieser Fragestellung nicht bestätigt werden.

## 5.5 Zwischenfazit

Im Rahmen der vorangegangenen Ausführungen des Kapitels 5.4 wurde mit Hilfe nichtparametrischer Tests überprüft, ob bei der Beantwortung der inhaltlichen Fragen signifikante Unterschiede im Antwortverhalten zwischen den beiden zu vergleichenden Befragungssamples

bzw. Teilstichproben erkennbar wurden. Diese ermittelten Unterschiede können dabei entsprechend der methodischen Vorgehensweise, sonstige verzerrende Einflüsse auszuschalten, grundsätzlich als Indiz für Methodeneffekte interpretiert werden.

**Tabelle 22: Zusammenstellung der Items bzw. Fragestellungen, bei denen signifikante Unterschiede zwischen den Teilstichproben ermittelt werden konnten**

Frage-Nr.	Fragestellung	Item	Test	Fragentyp
1.1	Stärkung verschiedener politischer Aufgabenbereiche in Deutschland	Verbesserung des Zusammenlebens mit Ausländern	MW	Einstellungsfrage
1.2	Einstellungsfrage zum Themenkomplex "Klimaschutz"	Für den Klimaschutz bin ich bereit, auf Flugreisen zu verzichten Klimaschutz ist alleinige Sache der Politik	MW	Einstellungsfrage
1.3	Verhaltensfrage zum Themenkomplex "Klimaschutz"	Beim Kauf von Haushaltsgeräten auf Energieverbrauch achten Mit der Familie oder Freunden über das Thema Globale Erwärmung sprechen Aus Klimaschutzgründen auf Fernreisen verzichten	MW	Verhaltensfrage
1.4	Bezug von "Grünem" Strom		CQ	Verhaltensfrage
2.1	Quiz zum Themenkomplex "Klimaschutz"	Der Ausstoß an Treibhausgasen ist bei der Erzeugung von tierischen Lebensmitteln wie Fleisch, Milch und Eiern wesentlich höher als bei pflanzlichen Erzeugnissen.	CQ	Überzeugungsfrage
2.2	Überzeugungsfrage zu verschiedenen Gütezeichen	Marine Stewardship Council (MSC)-Zeichen Eigenes Product Carbon Footprint-Zeichen	CQ	Überzeugungsfrage
2.3	Überzeugungsfrage zum Product Carbon Footprint-Zeichen		MW	Überzeugungsfrage
2.4	Einstellungen zum Product Carbon Footprint-Zeichen	Eigenschaftspaar "nicht umsetzbar/umsetzbar" Eigenschaftspaar "verständlich/unverständlich"	MW	Einstellungsfrage
2.5	Fragestellung zur rechtlichen Regelung des Product Carbon Footprint-Zeichens Fragestellung zur Ausgestaltung des Product Carbon Footprint-Zeichens	Freiwillig vs. gesetzlich vorgeschrieben vs. weiß nicht Konkrete Mengenangabe vs. Ampel-Schema vs. weiß nicht	CQ	Einstellungsfrage

Anmerkungen: MW = Mann-Whitney U-Test; CQ = Chi-Quadrat-Test nach Pearson.

Quelle: Eigene Zusammenstellung.

Tabelle 22 stellt zusammenfassend noch einmal alle Fragen und Items dar, bei denen signifikante Unterschiede im Antwortverhalten zwischen den beiden Teilstichproben ermittelt werden konnten. Wie aus der Tabelle hervorgeht, konnte bei den angegebenen Fragestellungen bzw. Items ein signifikanter Unterschied im Antwortverhalten zwischen beiden Teilstichpro-

ben (Fragestellungen, in denen der Mann-Whitney U-Test herangezogen wurde), bzw. ein signifikanter Zusammenhang zwischen dem Antwortverhalten und der Befragungsmethode (Fragestellungen, bei denen der  $\chi^2$ -Test verwendet wurde) ermittelt werden.

Allerdings muss an dieser Stelle kritisch hinterfragt werden, ob die ermittelten Unterschiede bei den in der Tabelle dargestellten Fragestellungen bzw. Items ausreichen, um pauschal Methodeneffekte zwischen den beiden Befragungsmethoden Face-to-Face und Online bestätigen zu können. Zwar lassen sich die Antworten und Häufigkeitsverteilungen dieser Items und Fragestellungen durchaus dahingehend interpretieren, dass Methodeneffekte vorliegen, die zu einem unterschiedlichen Antwortverhalten in den beiden Teilstichproben geführt haben.

Weiterhin stellt sich hierbei die Frage, ob bei der Annahme, dass die ermittelten Antwortunterschiede wirklich auf Methodeneffekten basieren, nicht bei einer größeren Zahl an Fragestellungen bzw. Items ein unterschiedliches Antwortverhalten hätte nachgewiesen werden müssen. Vor allem bei den als Einstellungs- bzw. Verhaltensfragen konzipierten Itembatterien 1.1, 1.2 und 1.3 konnten nur bei einigen wenigen Items Unterschiede nachgewiesen werden. Bei der Einstellungsfrage zum PCF (Frage 2.4 des Fragebogens) konnten ebenfalls keine konsistenten Unterschiede zwischen den beiden Teilstichproben nachgewiesen werden. Auch bei den Überzeugungs- und Verhaltensfragen des Fragebogens konnten weiterhin keine eindeutigen Ergebnisse ermittelt werden. Bei der Annahme des pauschalen Vorliegens von Methodeneffekten wäre stattdessen davon auszugehen gewesen, dass auch bei der in Kapitel 5.4.7 besprochenen Fragestellung zur Zahlungsbereitschaft für klimafreundliche Milch sich signifikante Unterschiede zwischen den beiden Teilstichproben hätten ergeben müssen.

Bisher wurde davon ausgegangen, dass durch die Wahl der Studierendenschaft der JLU Gießen hinsichtlich der Einstellungen eine in sich homogene Grundgesamtheit ausgewählt wurde. Möglicherweise muss jedoch aufgrund der uneindeutigen und unsicheren Ergebnisse bezüglich des Vorliegens von Methodeneffekten von dieser Annahme jedoch Abstand genommen werden. So ist es durchaus denkbar, dass die bei den oben dargestellten Fragestellungen und Items ermittelten Unterschiede im Antwortverhalten doch auf unterschiedliche Einstellungen der Befragten zurückgeführt werden können. Da für die Beantwortung der verschiedenen Fragestellungen möglicherweise nicht alle Einstellungsdimensionen relevant sind, ist nicht auszuschließen, dass Unterschiede in den Teilstichproben hinsichtlich relevanter Einstellungsdimensionen nicht alle Fragestellungen und Items gleichermaßen determinieren. Beispielsweise ist denkbar, dass sich Unterschiede in den Einstellungsdimensionen zwischen beiden Teilstichproben bei manchen Fragestellungen stärker auswirken als bei anderen. Dieser Sachverhalt könnte dabei das inkonsistente Bild bezüglich des Vorliegens von Methodeneffekten

fekten beeinflussen. Daher wird nachfolgend analysiert, ob sich die beiden Teilstichproben hinsichtlich relevanter Einstellungsdimensionen voneinander unterscheiden. Darauf aufbauend wird analysiert, inwieweit diese möglicherweise für die ermittelten Unterschiede im Antwortverhalten der beiden Teilstichproben verantwortlich gemacht werden können.

## 5.6 Einstellungsdimensionen der Befragten

Um mögliche Einstellungsdimensionen der Befragten ermitteln zu können, sind vor allem die Einstellungsfragen 1.1 und 1.2 relevant, die als Item-Batterien jeweils sechs verschiedene Items enthalten. Während die Einstellungsfrage 1.1 sich dabei allgemein auf verschiedene politische Aufgabenbereiche und somit allgemeine Politikeinstellungen bezieht, geht, wie oben bereits dargelegt wurde, die Frage 1.2 explizit auf die Einstellungen zum Themenkomplex „Klimawandel“ ein. Eine Extraktion der möglichen Einstellungsdimensionen kann durch die Faktorenanalyse erfolgen. Deren Zielsetzung ist, die Zahl der Ausgangsvariablen (in diesem Fall die einzelnen Items der Fragebatterien) auf eine geringere Zahl an Items zu reduzieren und dadurch latente Zusammenhänge zwischen den einzelnen Variablen aufdecken zu können.

Im Gegensatz zu den vorherigen Analysen wird bei der faktorenanalytischen Auswertung nachfolgend auf eine Gewichtung des Datensatzes verzichtet, da diese für die Ermittlung der Einstellungsdimensionen der Befragten hier nicht relevant ist und aus Sicht des Verfassers für die Zielsetzung sogar hinderlich sein könnte. Weiterhin musste vor der Durchführung einer solchen faktorenanalytischen Auswertung zuerst festgelegt werden, welche Items in die Auswertung überhaupt mit aufgenommen werden sollen. Die größte Eignung für das Messvorhaben konnte dabei bei einer Kombination beider Itembatterien aus den Fragestellungen 1.1 und 1.2 ermittelt werden. Hierfür wurde das sogenannte Kaiser-Meyer-Olkin-Kriterium (KMO) herangezogen, welches als Indikator angibt, ob die Durchführung einer Faktorenanalyse als sinnvoll erachtet wird (KAISER und RICE 1976: 112f.; BACKHAUS et al. 2008: 336). Hier konnte für den Fall, dass alle zwölf Items beider Fragenbatterien einbezogen werden, für das KMO-Kriterium ein Wert von 0,719 ermittelt werden. Dieser Wert wird von KAISER und RICE dabei als „*middling*“ und von BACKHAUS et al. als „*ziemlich gut*“ beschrieben. Weiterhin führen BACKHAUS et al. an, dass Daten mit einem solchen KMO-Wert für eine faktorenanalytische Auswertung geeignet sind (BACKHAUS et al. 2008: 336f.). Wurde hingegen nur jeweils eine Itembatterie in die Faktorenanalyse einbezogen, lagen die ermittelten Werte des KMO-Kriteriums mit 0,635 bzw. 0,619 deutlich darunter. Als weiteres Gütekriterium zur Überprüfung der Datengüte wurde die sogenannte Anti-Image-Kovarianz-Matrix betrachtet. Hier schlugen DZIUBAN und SHIRKLEY vor, die Daten für eine Faktorenanalyse als ungeeignet an-

zusehen, wenn der Anteil der nicht-diagonalen Elemente, die größer 0,09 sind, mehr als 25 % beträgt (DZIUBAN und SHIRKLEY 1974: 359). Der hier ermittelte Wert von 22,7 % bestätigt allerdings die Eignung des Datensatzes für eine Faktorenanalyse.

Nach Feststellung der Eignung des Datensatzes wurde anschließend eine Hauptkomponentenanalyse mit orthogonaler Rotation der Faktoren durchgeführt. Zur Rotation wurde dabei die Varimax-Methode mit Kaiser-Normalisierung ausgewählt. Zielsetzung der Hauptkomponentenanalyse ist es dabei, die Datenstruktur möglichst umfassend durch möglichst wenige Faktoren zu reproduzieren (BACKHAUS et al. 2008: 350). Es wird somit ein Trade-Off zwischen möglichst wenigen Faktoren sowie einer möglichst umfassenden Reproduktion der Datenstruktur vorgenommen.

Dem Kaiser-Kriterium entsprechend, wurden drei Faktoren extrahiert. Diese erklären in der Summe einen Anteil von 47,54 % an der Gesamtvarianz. Nach Durchführung der Rotation trägt der erste Faktor dabei mit einem Anteil von 18,32 %, der zweite mit einem Anteil von 15,19 % und der dritte Faktor mit einem Anteil von 14,03 % zur Erklärung der Gesamtvarianz bei (siehe hierzu Tabelle A 27 im Anhang).

**Tabelle 23: Rotierte Faktorladungsmatrix**

	Komponente		
	1	2	3
Maßnahmen zum Klimaschutz	0,706		
Höhere Steuern auf Benzin und Treibstoffe	0,675		
Bereitschaft Verzicht auf Flugreisen	0,643		
Schutz von Tier- und Pflanzenarten	0,629		
Verbesserung des Zusammenlebens mit Ausländern	0,449		
Bekämpfung der Arbeitslosigkeit		0,758	
Bekämpfung von Kriminalität		0,709	
Förderung der Wirtschaft		0,676	
Nur Änderung Konsumverhalten, wenn andere auch			0,722
Einzelner für Klimaschutz bringt nichts			0,670
Klimaschutz alleinige Sache der Politik			0,659
Kleine Dinge viel erreichen			-0,412

Extraktionsmethode: Hauptkomponentenanalyse.

Rotationsmethode: Varimax mit Kaiser-Normalisierung.

Quelle: Auswertung der eigenen Erhebung.

Zur Analyse der Einstellungsdimensionen ist nachfolgend die rotierte Faktorladungsmatrix relevant, die in Tabelle 23 dargestellt wird. Aufgrund der besseren Übersichtlichkeit und einer

einfacheren Interpretation werden dabei nur Faktorladungen dargestellt, die mit einem Wert von größer als 0,4 auf den entsprechenden Faktor bzw. die Komponente laden.

Wie anhand der Tabelle ersichtlich wird, ergibt sich nach der Faktorextraktion und der Rotation der Faktorladungsmatrix der Sachverhalt, dass fünf Items stark auf den ersten Faktor bzw. die erste Komponente laden. Aufgrund der Länge wurden die Item-Bezeichnungen dabei teilweise abgekürzt. Nachfolgend werden daher die vollständigen Item-Bezeichnungen sowie falls Abkürzungen verwendet wurden, diese in Klammern zusammen mit den jeweiligen Faktorladungen dargestellt:

- „Maßnahmen zum Klimaschutz“ (*keine Abkürzung*; 0,706),
- „Um das Klima zu schützen, sollten höhere Steuern auf Benzin und Treibstoff erhoben werden.“ (*„Höhere Steuern auf Benzin und Treibstoffe“*; 0,675),
- „Für den Klimaschutz bin ich bereit, auf Flugreisen zu verzichten.“ (*„Bereitschaft Verzicht auf Flugreisen“*; 0,643),
- „Schutz von Tier- und Pflanzenarten“ (*keine Abkürzung*, 0,629),
- „Verbesserung des Zusammenlebens mit Ausländern“ (*keine Abkürzung*; 0,449).

Dieser erste Faktor kann dabei möglicherweise als *zukunftsorientierte* Einstellungsdimension bzw. als *zukunftsorientiertes Handeln* der befragten Personen interpretiert werden. So gehen, wie im zweiten Kapitel der vorliegenden Arbeit erläutert wurde, vom Klimawandel starke Gefährdungen auf das alltägliche Leben und die Erde in ihrer jetzigen Form mit ihrer Tier- und Pflanzenvielfalt aus. Um diesen drohenden Gefahren begegnen zu können, ist ein verändertes menschliches Handeln in vielerlei Hinsicht nötig, was möglicherweise durch den Faktor ausgedrückt wird. So läßt die aus Befragtensicht starke Relevanz einer Verstärkung politischer Maßnahmen zum Klimaschutz mit einem Wert von 0,706 am stärksten auf den ersten Faktor. Zu der angeführten Interpretation passt ebenfalls, dass auch das Item *„Höhere Steuern auf Benzin und Treibstoffe“* mit einem Wert von 0,675 und das Item *„Bereitschaft Verzicht auf Flugreisen“* mit einem Wert von 0,643 auf diesen Faktor laden. So spielen, wie im zweiten Kapitel dargestellt wurde, fossile Brennstoffe bzw. deren Emissionen beim anthropogenen Klimawandel eine entscheidende Rolle. Im Bezug auf ein zukunftsorientiertes Handeln können somit höhere Steuern auf Treibstoffe und Benzin sowie der Verzicht auf Flugreisen eine wichtige Rolle einnehmen. Hierrüber kann ein Laden der Items gut erklärt werden. Daneben läßt weiterhin mit einem Wert von 0,629 das Item *„Schutz von Tier- und Pflanzenarten“* auch auf diesen ersten Faktor. So erwartet das IPCC, dass sich durch den anthropogenen Klimawandel deutliche Auswirkungen auf das Ökosystem ergeben werden (IPCC 2007c: 11ff.).

Hierdurch, aber auch aufgrund verschiedener Meldungen über vom Aussterben bedrohter Tiere durch Abholzung, Überfischung et cetera, sind die Befragten möglicherweise besorgt und sehen daher die Notwendigkeit, Tiere und Pflanzen besser schützen zu müssen, um das Ökosystem auch in Zukunft in Balance halten zu können. Deutlich schwächer hingegen lädt das Item „*Verbesserung des Zusammenlebens mit Ausländern*“ auf den Faktor. Dieses Item passt indirekt ebenfalls auf die Interpretation des Faktors als *zukunftsorientiertes Handeln*. So wird vom IPCC prognostiziert, dass durch den Klimawandel und den damit verbundenen Veränderungen der klimatischen Bedingungen in verschiedenen Regionen der Erde starke Migrationsbewegungen ausgelöst werden könnten. Diese würden eine Verstärkung des politischen Engagements zur Verbesserung des Zusammenlebens zwischen Einheimischen und Ausländern notwendig machen. Aber auch in der Gegenwart gibt es beispielsweise bereits Stimmen, die sich für eine stärkere Integration von Ausländern aussprechen, um den Zusammenhalt und das Gleichgewicht in der Gesellschaft zu stärken (KERSTING 2008: 107ff.; NUSSER 2008: 596). Anhand dieser Ausführungen lässt sich durchaus erkennen, dass dieses Item ebenfalls auf den ersten Faktor lädt. So spielt eine starke Gesellschaft, die Menschen integriert und absichert, für zukünftige Herausforderungen eine wichtige Rolle. Um diese Gesellschaft erreichen zu können, ist ein zukunftsorientiertes Handeln notwendig.

Auf den zweiten extrahierten Faktor laden nach der Faktorrotation weiterhin die folgenden drei Variablen:

- „Bekämpfung der Arbeitslosigkeit“ (*keine Abkürzung*; 0,758),
- „Bekämpfung von Kriminalität“ (*keine Abkürzung*, 0,709) und
- „Förderung der Wirtschaft“ (*keine Abkürzung*, 0,676).

Auf diesen zweiten Faktor laden nur Items der ersten Itematterie, mit deren Hilfe die Befragten zur Relevanz der Stärkung verschiedener politischer Aufgabenbereiche befragt wurden. Dieser Faktor kann dabei möglicherweise als *Stärkung der gesellschaftlichen Sicherheit* interpretiert bzw. als Sammelbegriff für die darauf ladenden Variablen herangezogen werden. Am stärksten lädt auf den Faktor dabei das Item „*Bekämpfung der Arbeitslosigkeit*“ mit einem Wert von 0,758. Im Kontext mit den beiden anderen, auf diesen Faktor ladenden Items „*Bekämpfung von Kriminalität*“ sowie „*Förderung der Wirtschaft*“ zeigt sich, dass bei diesem Faktor entsprechend der zuvor angeführten Interpretation höchstwahrscheinlich Sicherheitsaspekte im Vordergrund stehen. Diese Sicherheit kann sich dabei einmal aus einer materiellen Sicherheit ergeben, welche durch die Bekämpfung von Arbeitslosigkeit sowie einer Stärkung und Förderung der Wirtschaft erzielt werden kann. Ein weiterer Bestandteil dieses Sicherheitsaspektes kann dabei aber auch die physische, sowie kognitive Sicherheit sein. Diese Si-

cherheitsdimension wird dabei durch das Item „*Bekämpfung von Kriminalität*“, welches mit einem Wert von 0,709 auf den Faktor lädt, ausgedrückt. Während der erste Faktor sich tendenziell eher auf den Klimawandel und Klimaschutzaspekte bezog, lässt die Interpretation des letztgenannten Faktors den Schluss zu, dass dieser eher weniger auf den Klimawandel bezogen ist, sondern eher als allgemeine Einstellungsdimension der Befragten verstanden werden sollte.

Wie aus der Tabelle 23 hervorgeht, laden auf den dritten extrahierten Faktor die folgenden Items:

- „Ich bin nur dann bereit, mein Konsumverhalten für den Klimaschutz zu ändern, wenn alle anderen dies auch tun“ („*Nur Änderung Konsumverhalten, wenn andere auch*“, 0,722),
- „Wenn ich als Einzelner etwas für den Klimaschutz tue, ändert das auch nichts.“ („*Einzelner für den Klimaschutz bringt nichts*“, 0,670),
- „Klimaschutz ist alleinige Sache der Politik“ („*Klimaschutz ist alleinige Sache der Politik*“, 0,659), sowie
- „Auch mit kleinen Dingen wie sparsamer heizen und öfters mal das Auto stehen lassen kann schon viel erreicht werden.“ („*Kleine Dinge viel erreichen*“, -0,412).

Im Gegensatz zum zweiten Faktor beziehen sich hier alle auf diesen Faktor ladende Items auf den Themenkomplex Klimaschutz. Während im ersten Faktor jedoch ein zukunftsorientiertes Handeln hinsichtlich des Klimawandels bzw. dessen Folgen thematisiert wurde, laden auf diesen dritten Faktor nur negative Aspekte hinsichtlich des Klimawandels, weshalb dieser Faktor möglicherweise als *Desinteresse und mangelnde Motivation hinsichtlich des Klimawandels* interpretiert werden kann. So kann eine mangelnde Motivation bzw. eine mangelnde Bereitschaft, sein Verhalten für den Klimaschutz zu ändern, anhand des ersten Items „*Nur Änderung Konsumverhalten, wenn andere auch*“ erkannt werden, welches mit einem Wert von 0,722 auf den extrahierten Faktor lädt. Auch das zweite, mit einem Wert von 0,670 auf diesen Faktor ladende Item „*Einzelner für Klimaschutz bringt nichts*“ passt in diesen Kontext, wie ebenfalls das Item „*Klimaschutz ist alleinige Sache der Politik*“, welches mit einem Wert von 0,659 auf den Faktor lädt. Als viertes Item lädt abschließend die Variable „*Kleine Dinge viel erreichen*“ ebenfalls auf den Faktor. Da entsprechend der Faktorladungsmatrix in Tabelle 23 dieses Item mit einem negativen Wert von -0,412 auf den Faktor lädt, kann die oben angeführte Interpretation des Faktors als *Desinteresse und mangelnde Motivation hinsichtlich des Klimawandels* aufrecht erhalten bzw. sogar bestätigt werden, da aufgrund des negativen Vor-

zeichens das Item in negierter Form bei der Interpretation des Faktors berücksichtigt werden muss.

Das Vorhandensein einer solchen Einstellungsdimension ist dabei durchaus plausibel. So ist davon auszugehen, dass zumindest bei einem Teil der Befragten eine solche Einstellungsdimension existent ist und sich diese Befragten entsprechend nicht für den Klimawandel interessieren bzw. diese Befragten nicht davon ausgehen, dass ein verändertes Handeln ihrerseits zu einer Umkehr bzw. Reduktion des anthropogenen Klimawandels führen kann. Sie sehen sich stattdessen wahrscheinlich in einer tendenziell eher passiven Position, wie durch das auf den Faktor ladende Item „Nur Änderung Konsumverhalten, wenn andere auch“ aufgezeigt wird. Während diese Befragten scheinbar selbst nicht bereit sind, vorbildhaft voranzugehen und aktiv Maßnahmen zur Verhinderung oder Verminderung des Klimawandels zu ergreifen, erwarten sie dieses aktive Handeln hingegen von der Politik. So kann das auf den Faktor ladende Item „Klimaschutz ist alleinige Sache der Politik“ dahingehend interpretiert werden, dass diese Befragten sich selbst nicht in der Verantwortung für den Klimaschutz sehen, sondern eine derartige Verantwortung ausschließlich dem Staate zuordnen.

Weiterhin soll analysiert werden, ob zwischen den beiden Teilstichproben bzw. Befragungssamples F2F und Online Unterschiede hinsichtlich dieser Einstellungsdimensionen erkennbar sind. Hierfür wurden nach Durchführung der Faktorenanalyse zusätzlich die sogenannten Faktorwerte gebildet. Diese geben Informationen darüber, welche Werte die verschiedenen Objekte, im Falle der vorliegenden Arbeit die Befragten hinsichtlich der extrahierten Faktoren und somit der verschiedenen Einstellungsdimensionen annehmen (BACKHAUS et al. 2008: 358). Als verwendete Methodik zur Ermittlung der Faktorwerte wurde die Regressions-Methode in SPSS ausgewählt. Anschließend wurden für alle drei Einstellungsdimensionen und die beiden Teilstichproben jeweils die arithmetischen Mittel berechnet. Eine Darstellung dieser Mittelwerte wird in Tabelle 24 gegeben.

**Tabelle 24: Arithmetische Mittel der Einstellungsdimensionen in den Teilstichproben**

	Mittelwert			Spannweite	
	F2F (n=176)	Online (n=477)	Differenz	F2F (n=176)	Online (n=477)
Zukunftsorientiertes Handeln	0,1254	-0,0463	0,1717	5,2765	6,9412
Stärkung der gesellschaftlichen Sicherheit	-0,0412	0,0152	0,0564	4,3789	6,0873
Desinteresse und mangelnde Motivation hinsichtlich des Klimawandels	0,1659	-0,0612	0,2271	5,7575	5,5025

Quelle: Auswertung der eigenen Erhebung.

Wie aus der Tabelle hervorgeht, lassen sich bei den arithmetischen Mitteln der berechneten Faktorwerte gewisse Unterschiede zwischen den beiden Teilstichproben erkennen. So lässt

sich bei der ersten Einstellungsdimension „*Zukunftsorientiertes Handeln*“ in der F2F-Stichprobe ein positiver Mittelwert erkennen. Dieser kann so interpretiert werden, dass im Vergleich zur Gesamtstichprobe, zusammengesetzt aus beiden Teilstichproben, diese Einstellungsdimension in der F2F-Stichprobe überdurchschnittlich ausgeprägt ist<sup>31</sup>. Das negative arithmetische Mittel in der Online-Stichprobe weist hingegen darauf hin, dass hier diese Einstellungsdimension im Vergleich leicht unterdurchschnittlich ausgeprägt ist. Der umgekehrte Fall ist hingegen bei der zweiten extrahierten Einstellungsdimension „*Stärkung der gesellschaftlichen Sicherheit*“ ersichtlich. Während diese Einstellungsdimension bei den Befragten der F2F-Stichprobe leicht unterdurchschnittlich ausgeprägt ist, ist sie bei den Befragungsteilnehmern der Online-Stichprobe hingegen durchschnittlich stärker ausgeprägt. Im Bezug auf die dritte Einstellungsdimension „*Desinteresse und mangelnde Motivation hinsichtlich des Klimawandels*“ zeigt sich wiederum das Bild, dass diese Einstellungsdimension bei den Befragten in der F2F-Stichprobe überdurchschnittlich ausgeprägt ist, während sie bei den Befragten in der Online-Stichprobe nur eine unterdurchschnittliche Ausprägung annimmt. Hinsichtlich dieser Einstellungsdimension kann entsprechend ausgesagt werden, dass im Durchschnitt bei den Befragten der F2F-Stichprobe ein größeres Desinteresse hinsichtlich des Klimawandels besteht. Auch ist im Vergleich zur Online-Stichprobe die Motivation bei den Befragten, aktiv etwas gegen den Klimawandel zu unternehmen, geringer ausgeprägt.

In einem nächsten Schritt wurde mit Hilfe eines t-Tests überprüft, ob sich die ermittelten arithmetischen Mittel der Faktorwerte zwischen den beiden Teilstichproben signifikant voneinander unterscheiden. Eine Anwendung des t-Tests ist hier im Gegensatz zu den vorherigen Analysen möglich, da sowohl metrisch skalierte Daten vorliegen sowie diese auch einer Standardnormalverteilung folgen, wie anhand des arithmetischen Mittels der Gesamtstichprobe von null und der Standardabweichung von eins abgelesen werden kann.

Als Ergebnisse des t-Tests können festgehalten werden, dass sich die Befragten der beiden Teilstichproben hinsichtlich der Ausprägungen der Einstellungsdimensionen „*Zukunftsorientiertes Handeln*“ mit einer Irrtumswahrscheinlichkeit von 0,040 und „*Desinteresse und mangelnde Motivation hinsichtlich des Klimawandels*“ mit einer Irrtumswahrscheinlichkeit von 0,007 signifikant voneinander unterscheiden. Kein signifikanter Unterschied konnte hingegen bei den Mittelwerten der Einstellungsdimension „*Stärkung der gesellschaftlichen Sicherheit*“ ermittelt werden. Die Teststatistiken finden sich in Tabelle A 28 des Anhangs. Weiterhin wurde mit Hilfe des Programms G\*Power die Effektstärke  $d$  nach Cohen berechnet, die Auskunft

---

<sup>31</sup> Eine durchschnittliche Ausprägung der Einstellungsdimension wird durch ein arithmetisches Mittel von null ausgedrückt (BACKHAUS et al. 2008: 376).

über die Größe des ermittelten Effektes, nämlich den signifikanten Unterschieden der Teilstichproben bei zwei Einstellungsdimensionen, gibt. Hier konnte für die erste Einstellungsdimension mit signifikanten Unterschieden zwischen den Teilstichproben „*Zukunftsorientiertes Handeln*“ ein Wert von 0,1769 und für die zweite Einstellungsdimension mit signifikanten Unterschieden, „*Desinteresse und mangelnde Motivation hinsichtlich des Klimawandels*“ ein Wert von 0,2341 berechnet werden. Prinzipiell können diese Effektstärken von Cohens  $d$  nach COHEN als tendenziell eher klein interpretiert werden (COHEN 1992: 157). Im direkten Vergleich der Effektstärken zeigt sich, dass die ermittelten Unterschiede zwischen den Teilstichproben bei der Einstellungsdimension zum Desinteresse und zur mangelnden Motivation größer sind als die Unterschiede bei der Dimension zum zukunftsorientierten Handeln: Die hier ermittelte Effektstärke  $d$  wies einen kleineren Wert auf.

## 5.7 Regressionsanalytische Auswertung

Wie oben dargestellt wurde, konnten bei einigen Fragestellungen durch die zur Analyse herangezogenen nichtparametrischen Tests Antwortunterschiede zwischen den Teilstichproben Online und F2F ermittelt werden. Allerdings konnte nicht ausgeschlossen werden, dass die ermittelten signifikanten Antwortunterschiede auch auf anderen Ursachen, wie beispielsweise den im vorherigen Kapitel 5.6 analysierten Einstellungsunterschieden, beruhen – auch wenn im bisherigen Verlauf der Arbeit aufgrund des ausgewählten studentischen Samples von einer Homogenität hinsichtlich der Einstellungen ausgegangen wurde. Um nun herausarbeiten zu können, ob die ermittelten Antwortunterschiede zwischen den Teilstichproben wirklich auf die beiden unterschiedlichen Befragungsmethodiken zurückgeführt werden können, wurde für den fortfolgenden Analyseverlauf auf eine von GAUPP et al., HARTL sowie HEERSWEGH und LOOSVELDT herangezogene Vorgehensweise zurückgegriffen. Diese Autoren ziehen im Rahmen einer Regressionsanalyse jeweils zusätzlich zu den anderen erklärenden Variablen eine Dummyvariable für die verwendete Methodik heran, um deren Einfluss zu überprüfen. Die Analysen von HEERSWEGH und LOOSVELDT beziehen sich dabei ebenfalls, wie die Ausführungen der vorliegenden Arbeit, auf ein studentisches Sample (GAUPP et al. 2006: 17ff.; HARTL 2008: 70ff.; HEERSWEGH und LOOSVELDT 2008: 843ff.).

Als zu analysierende Variablen wurden einmal das Wissen der Befragten zu dem PCF-Label (siehe Kapitel 5.4.6) sowie die Mehrzahlungsbereitschaft für CO<sub>2</sub>-reduzierte Milch (siehe Kapitel 5.4.7) ausgewählt. Hintergrund dieser Auswahl ist dabei, dass bei der Mehrzahlungsbereitschaft über den Mann-Whitney U-Test kein signifikanter Unterschied zwischen den beiden Teilstichproben nachgewiesen werden konnte, während dies bei dem Wissen über das PCF-Zeichen der Fall war. Aufgrund ihrer metrischen Skalierung sind die Daten dabei für den

Analysezweck geeignet (BACKHAUS et al. 2008: 55). Anhand der herangezogenen multiplen linearen Regressionsanalyse soll entsprechend überprüft werden, ob diese unterschiedlichen Resultate auf die verwendete Befragungsmethodik zurückgeführt werden können.

HEERSWEGH und LOOSVELDT spezifizieren zur Erklärung der abhängigen Variablen ein grundsätzliches Modell, bei welchem neben einer Dummyvariablen für die angewendete Methodik auch allgemeine soziodemographische Variablen wie Alter und Geschlecht sowie eine studienbezogene Variable in Form des Fachbereichs zur Erklärung herangezogen werden:

$$Y = f(\text{Befragungsmethode}, \text{Alter}, \text{Geschlecht}, \text{Fachbereich}) \quad (12)$$

Auf ein ähnliches Grundmodell greifen ebenfalls GAUPP et al. zurück, um verschiedene abhängige Variablen zu erklären. In der vorliegenden Arbeit sollen neben den oben genannten Variablen weitere Variablen zur Erklärung herangezogen werden. So werden in die Modellspezifikation zur Erklärung der Mehrzahlungsbereitschaft für CO<sub>2</sub>-reduzierte Milch eine zusätzliche Dummyvariable für ein monatliches Haushaltsnettoeinkommen von über 1200 € („*Einkommen<sub>Hoch</sub>*“), eine Dummyvariable für ehrenamtliches Engagement („*Ehrenamt*“) sowie zwei Dummyvariablen für die Haupteinkaufsorte Bioladen („*Einkaufsort<sub>Bioladen</sub>*“) und Wochenmarkt („*Einkaufsort<sub>Wochenmarkt</sub>*“) integriert. Die Variablen zum Haupteinkaufsort wurden dabei im Rahmen der Fragestellung 3.2 des Fragebogens erfasst. Die Erhebung eines ehrenamtlichen Engagements der Studierenden erfolgte in Fragestellung 4.5 des Fragebogens bzw. der Online-Befragung. Bei diesen genannten Variablen wird ausgegangen, dass diese möglicherweise zu einer höheren Mehrzahlungsbereitschaft für die CO<sub>2</sub>-reduzierte Milch bei den Befragten führen könnten. Darüber hinaus finden die drei Einstellungsdimensionen „Zukunftsorientiertes Handeln“ („*Zukunftsorientierung*“), „Gesellschaftliche Sicherheit“ („*Gesellschaftliche\_Sicherheit*“) sowie „Desinteresse am Klimawandel“ („*Desinteresse*“), welche im vorherigen Abschnitt 5.6 gebildet wurden, ebenfalls Einzug in die Modellspezifikation. Um mögliche unterschiedliche Effekte zwischen den Fachbereichen herausarbeiten zu können, wurden diese jeweils als Dummyvariablen spezifiziert. Den Basis- bzw. Vergleichsfall stellt dabei der Fachbereich 09 („Agrarwissenschaften, Ökotrophologie und Umweltmanagement“) dar. Dieser Fachbereich wurde dabei ausgewählt, da hier ein besonderes Interesse bzw. ein höheres Wissen der Befragten erwartet wurde.

Letztlich wurde somit zur Erklärung der Zahlungsbereitschaftsdifferenz zwischen konventioneller und CO<sub>2</sub>-reduzierter Milch („*MZB*“) folgendes Modell spezifiziert:

$$MZB = \beta_0 + \beta_1 \cdot \text{Methode} + \beta_2 \cdot \text{Alter} + \beta_3 \cdot \text{Geschlecht} + \beta_4 \cdot \text{Ehrenamt} + \beta_5 \cdot \text{Einkommen}_{\text{Hoch}} + \beta_6 \cdot \text{Einkaufsort}_{\text{Bioladen}} + \beta_7 \cdot \text{Einkaufsort}_{\text{Wochenmarkt}} + \beta_8 \cdot \text{Zukunftsorientierung} + \beta_9 \cdot \text{Gesellschaftliche\_Sicherheit} + \beta_{10} \cdot \text{Desinteresse} + \beta_{11} \cdot \text{FB01} + \beta_{12} \cdot \text{FB02} + \beta_{13} \cdot \text{FB03} + \beta_{14} \cdot \text{FB04} + \beta_{15} \cdot \text{FB05} + \beta_{16} \cdot \text{FB06} + \beta_{17} \cdot \text{FB07} + \beta_{18} \cdot \text{FB08} + \beta_{19} \cdot \text{FB10} + \beta_{20} \cdot \text{FB11}.$$

Zur Erklärung der Anzahl der korrekt gegebenen Antworten zum PCF-Label („Anz\_PCF“) wurde ein gleichlautendes Modell spezifiziert. Die Regressionskoeffizienten, die jeweiligen Signifikanzniveaus sowie weitere Angaben zur Güte der Modelle werden in Tabelle 25 gegeben.

**Tabelle 25: Ergebnisse der Regression**

Variable	Modell I	Modell II
Abhängige Variable	MZB	Anz_PCF
Konstante	0,142	0,955 ***
Befragungsmethode (0=F2F)	-0,039	0,438 ***
Alter	0,005	-0,008
Geschlecht (0=männlich)	0,054 (*)	-0,227 *
Ehrenamt (0=kein Ehrenamt)	-0,017	-0,179 *
Einkommen <sub>Hoch</sub>	-0,132 (*)	0,231
Einkaufsort <sub>Bioladen</sub>	0,014	0,105
Einkaufsort <sub>Wochenmarkt</sub>	0,100	-0,358
Zukunftsorientierung	-0,065 ***	-0,019
Gesellschaftliche_Sicherheit	0,018	0,190 ***
Desinteresse	0,022 (*)	0,001
FB 01	-0,061	-0,230
FB 02	0,016	-0,243
FB 03	0,000	-0,328 **
FB 04	0,227 **	-0,085
FB 05	0,030	-0,394
FB 06	-0,040	-0,067
FB 07	-0,061	-0,120
FB 08	0,007	-0,247
FB 10	-0,039	-0,115
FB 11	0,056	-0,556 **
Korrigiertes R <sup>2</sup>	0,085	0,105
F-Wert	3,567 ***	4,919 ***

Anmerkung: \*\*\*, \*\*, \* und (\*) bedeuten signifikant auf dem

99,9 %-, 99 %-, 95 %- und 90 %-Niveau.

Quelle: Auswertung der eigenen Erhebung.

So zeigt sich beim ersten Modell, dass nur wenige Variablen einen signifikanten Einfluss auf die Mehrzahlungsbereitschaft für CO<sub>2</sub>-reduzierte Milch haben. So sind auf dem 90 %-Niveau die Variablen „Geschlecht“, „Einkommen<sub>Hoch</sub>“ sowie die Einstellungsdimension „Desinteresse“ signifikant. Während Frauen dabei gegenüber Männern eine Mehrzahlungsbereitschaft

von 0,054 € aufweisen, zeigt sich bei befragten Studierenden mit einem monatlichen Haushaltsnettoeinkommen von mehr als 1200 € hingegen eine negative Mehrzahlungsbereitschaft: Sie sind im Vergleich zu Studierenden mit einem niedrigeren Einkommen sogar nur bereit, weniger für die CO<sub>2</sub>-reduzierte Milch zu zahlen, wie das negative Vorzeichen des entsprechenden Regressionskoeffizienten zeigt. Ebenfalls nur eine geringere Zahlungsbereitschaft für die CO<sub>2</sub>-reduzierte Milch weisen Befragte auf, bei denen die Einstellungsdimension „*Zukunftsorientierung*“ überdurchschnittlich stark ausgeprägt ist. Dies kann möglicherweise so erklärt werden, dass diese Befragten aus ihrer Sicht bereits sehr viele Maßnahmen unternehmen, um ihren persönlichen Beitrag zum anthropogenen Klimawandel zu minimieren und die Wirkung einer solchen Milch nur sehr gering einschätzen. Gemäß theoretischen Überlegungen wäre bei diesem Koeffizienten allerdings ein positives Vorzeichen erwartet worden. Hinsichtlich der Fachbereiche zeigt sich, dass Angehörige des Fachbereichs 04 („Geschichts- und Kulturwissenschaften“) im Vergleich zur Referenzsituation mit Studierenden des Fachbereichs 09 über eine Mehrzahlungsbereitschaft von 0,22 € verfügen. Die Regressionsergebnisse des ersten Modells zeigen im Weiteren ebenfalls auf, dass kein signifikanter Einfluss der Befragungsmethode nachgewiesen werden konnte.

Während im ersten Modell kein signifikanter Einfluss der Befragungsmethodik nachgewiesen werden konnte, zeigt sich beim zweiten Modell hingegen, dass hier die Befragungsmethodik einen höchst signifikanten Einfluss auf die Anzahl der korrekt gegebenen Antworten zum PCF-Zeichen ausübt. Darüber hinaus kann im zweiten Modell auch ein signifikanter Einfluss der Einstellungsdimension „*Gesellschaftliche Sicherheit*“ ermittelt werden. Allerdings zeigten sich bei Anwendung des Mann-Whitney U-Tests in Kapitel 5.6 hinsichtlich dieser Einstellungsdimension zwischen den beiden Teilstichproben keine signifikanten Unterschiede, so dass die ermittelten Unterschiede wohl auf die beiden unterschiedlichen Befragungsmethodiken Online und F2F zurückgeführt werden können. Einen hoch signifikanten Einfluss haben des Weiteren die Fachbereiche 03 und 11 auf die Anzahl der korrekt gegebenen Antworten zum PCF-Zeichen. Im Vergleich zur Referenzsituation mit dem Fachbereich 09 weisen diese beiden Fachbereiche jedoch eine niedrigere Kenntnis des PCF-Zeichens auf, wie die negativen Koeffizienten ausdrücken. Ein ebenfalls signifikanter, negativer Einfluss zeigt sich auch bezüglich der Dummyvariablen „Geschlecht“ sowie „Ehrenamt“. Hier nimmt die Zahl der korrekt gegebenen Antworten ab, wenn Frauen statt Männern befragt wurden. Die Anzahl der richtig gegebenen Teilantworten geht ebenfalls zurück, wenn von den Befragten ein ehrenamtliches Engagement angegeben wurde.

Zumindest für die beiden betrachteten Fragestellungen zeigt sich somit, dass dort, wo kein signifikanter Unterschied durch den Mann-Whitney U-Test nachgewiesen werden konnte, ebenfalls kein signifikanter Einfluss der Befragungsmethodik ermittelt werden konnte. Im Unterschied dazu konnte bei der Fragestellung nach der Anzahl der korrekt gegebenen Teilantworten zum PCF-Zeichen, bei dem durch den U-Test nach Mann und Whitney ein signifikanter Unterschied in der zentralen Tendenz zwischen beiden Teilstichproben ermittelt wurde, durch die Regressionsanalyse ein signifikanter Einfluss der Befragungsmethode aufgedeckt werden. So zeigt der Regressionskoeffizient der Variablen „*Befragungsmethode*“ von 0,438, dass bei einem Wechsel der Befragungsmethode von F2F auf Online die Zahl der korrekt gegebenen Teilantworten um 0,438 steigt.

Bezüglich der dritten Forschungshypothese F3 *„Hinsichtlich der Beantwortung von Wissensfragen sind keine signifikanten Unterschiede zwischen den beiden Befragungssamples zu erwarten.“* ergibt sich somit, dass keine eindeutige Aussage getroffen werden kann. So konnten bei der Wissens- bzw. Überzeugungsfrage zum PCF-Zeichen signifikante Antwortunterschiede nachgewiesen werden. Diese konnten weiterhin anhand der regressionsanalytischen Auswertung auf die beiden Befragungsmethoden zurückgeführt werden. So lässt die Online-Befragung im Vergleich zu einer F2F-Befragung mit einem anwesenden Interviewer die Möglichkeit zu, unkontrolliert im Internet nach verschiedenen Fragestellungen zu suchen und somit ohne eigenes Wissen ebenfalls zur korrekten Antwort zu kommen, wie auch einer der Teilnehmer in der offen formulierten Fragestellung ausführte. Anhand dieser Ausführungen wird ersichtlich, dass bei Wissensfragen durchaus Antwortunterschiede zwischen mündlich bzw. online befragten Samples denkbar sind, die nicht auf einem unterschiedlichen Wissen der Befragten, sondern rein auf der herangezogenen Befragungsmethodik basieren. Relevant in diesem Zusammenhang ist vor allem die fehlende Kontrollmöglichkeit in Online-Befragungen.

In einer weiteren Wissens- bzw. Überzeugungsfrage in Form eines Quiz zum Themenkomplex „Klimaschutz“ in Kapitel 5.4.4 konnte ebenfalls bei einer von fünf Fragestellungen ein signifikanter Unterschied im Antwortverhalten zwischen den beiden Befragungsmethoden aufgedeckt werden. Zwar konnten in den anderen Fragestellungen des Quiz keine Antwortunterschiede durch den Mann-Whitney U-Test aufgedeckt werden, allerdings zeigen die nicht einheitlichen Ergebnisse auch, dass Antwortunterschiede nicht auszuschließen sind. Daher kann die dritte Forschungshypothese nicht verifiziert werden und muss abgelehnt werden:

**Hypothese F3:** Hinsichtlich der Beantwortung von Wissensfragen sind keine signifikanten Unterschiede zwischen den beiden Befragungssamples zu erwarten.



Ein ebenso uneindeutiges Bild zeigt sich auch bei anderen Fragetypen wie Einstellungs- oder Verhaltensfragen. Hier konnten im Analyseverlauf zwar immer wieder signifikante Antwortunterschiede durch den Mann-Whitney U-Test oder signifikante Zusammenhänge zwischen den Antworten und der Befragungsmethode durch den  $\chi^2$ -Test ermittelt werden. Ein einheitliches und konsistentes Bild lässt sich hier jedoch nicht erkennen. So zeigen sich beispielsweise innerhalb von Itembatterien Items mit Antwortunterschieden sowie Items, bei denen keine Antwortunterschiede ermittelt werden konnten, parallel nebeneinander. Dieser Sachverhalt gilt dabei sowohl für die als Einstellungsfragen konzipierten Itembatterien der Fragestellungen 1.1 und 1.2 des Fragebogens, für die Itembatterie der Fragestellung 1.3, welche als Verhaltensfrage konzipiert wurde sowie für weitere Fragestellungen des Fragebogens. Bei einem eindeutigen Vorliegen von Methodeneffekten, die sich durch signifikante Antwortunterschiede zwischen den beiden verglichenen Befragungssamples ausdrücken würden, wäre bezüglich der Antwortunterschiede jedoch ein konsistentes Bild dahingehend zu erwarten, dass bei sämtlichen Fragestellungen diese Methodeneffekte nachweisbar wären. Daher muss die Frage nach einem eindeutigen Vorliegen von Methodeneffekten, ausgedrückt durch Antwortunterschiede, verneint werden und entsprechend die vierte Forschungshypothese ebenfalls abgelehnt werden:

**Hypothese F4:** Auch bei anderen Fragetypen als Wissensfragen lassen sich eindeutig Methodeneffekte in Form von Antwortunterschieden zwischen Online- und F2F-Befragungen nachweisen.



## 6 Diskussion

Zielsetzung der vorliegenden Untersuchung war es, die beiden Befragungsmethoden Online und Face-to-Face hinsichtlich möglicher Unterschiede im Antwortverhalten zu analysieren. Als Analysemethode wurde dafür die sogenannte Split-Ballot-Technik gewählt, bei der eine Grundgesamtheit in zwei Teilstichproben aufgeteilt und anschließend mit unterschiedlichen Instrumenten bzw. Befragungsmethoden befragt wird. Bei dieser Methodik handelt es sich um ein weit verbreitetes Verfahren zur Analyse von Wirkungen unterschiedlicher Befragungsmethoden oder Frageformulierungen.

Ein Problem dieser Technik zeigte sich relativ schnell bei der Planung der beiden Befragungen. So musste, entsprechend der Theorie der Split-Ballot-Technik, verhindert werden, dass Teilnehmer sowohl an der mündlichen Befragung als auch der Online-Befragung teilnehmen, da sich hierdurch Verzerrungen in den Ergebnissen hätten ergeben können. Aufgrund der Beachtung von Datenschutzrichtlinien konnte hierbei nicht auf personenbezogene Daten wie etwa Namen, Telefonnummern oder ähnliches zurückgegriffen werden, um eventuelle Doppelteilnahmen von Personen an beiden Teilbefragungen zu verhindern. Daher musste hierfür eine alternative Lösung in Form von zusätzlichen Hinweisen der Interviewer an die Befragten sowie einer Filterfrage in der Online-Befragung gefunden werden, die in Kapitel 5.2.4 beschrieben wurden. Im Vergleich zu einer Lösung mit Filterung über personenbezogene Daten stellt diese Vorgehensweise natürlich die schwächere Alternative dar. Da sich jedoch keine weiteren umsetzbaren Lösungsmöglichkeiten finden ließen, wurde dieser Ansatz letztlich verwirklicht. Da zwischen den beiden Befragungen jedoch kein großer Zeitraum lag und aufgrund des Sachverhalts, dass zumindest bei einem Teil der Studierenden von Kenntnissen der empirischen Sozialforschung und dem davon abhängendem Erfolg eigener Untersuchungen gerechnet werden kann, muss in Verbindung mit den gegebenen Lösungsansätzen aus Sicht des Verfassers nicht mit negativen Auswirkungen auf die vorliegende Untersuchung ausgegangen werden.

Einen anderen Nachteil der Anwendung der Split-Ballot-Technik sieht TADDICKEN darin, dass mögliche Ungleichverteilungen in beiden Teilstichproben im Bezug auf relevante soziodemographische Variablen und Einstellungen zu Ergebnisverzerrungen führen können. Dies wird aus ihrer Sicht in Untersuchungen häufig nicht ausreichend beachtet (TADDICKEN 2008: 156ff.). Von einer derartigen Verzerrung ist aufgrund des angewendeten Forschungsdesigns in der vorliegenden Untersuchung allerdings nicht auszugehen. So wurde einerseits auf ein homogenes Studierenden-Sample zurückgegriffen, weshalb von keinen großen Unterschieden

in den Einstellungen auszugehen ist. Des Weiteren wurden die beiden Teilstichproben über eine Gewichtung anhand der beiden soziodemographischen Variablen „*Geschlecht*“ sowie „*Klassifizierter Fachbereich*“ aneinander angeglichen. Hinsichtlich dieser als relevant erachteter Merkmale stimmen die beiden Teilstichproben entsprechend überein. Bestätigung erfährt diese Vorgehensweise im Rahmen der gebildeten Einstellungsdimensionen in Kapitel 5.6. Hier konnte zwar bei zwei Einstellungsdimensionen („*Zukunftsorientiertes Handeln*“ und „*Desinteresse und mangelnde Motivation hinsichtlich des Klimawandels*“) ein signifikanter Unterschied nachgewiesen werden. Diese stellten sich jedoch bei der Berechnung der Effektstärke  $d$  als schwach heraus. Die relevanten soziodemographischen Variablen sowie die Einstellungen in den beiden befragten Teilstichproben stimmen, zumindest in der Tendenz, somit grundsätzlich überein. Daher zeigen sich keine Argumente, die im Nachhinein gegen die Verwendung der Split-Ballot-Technik sprechen.

Prinzipiell stellt sich in der Retrospektive allerdings die Frage, ob eine Gewichtung des Datensatzes überhaupt notwendig gewesen wäre. Verschiedene Studien haben sich mit dem Nutzen von Gewichtungsprozeduren bei der Angleichung von Teilstichproben beschäftigt. Während LOOSVELDT UND SONCK zu dem Ergebnis kommen, dass eine Gewichtung der Daten keinen Einfluss auf die Datengüte hat und starke Verzerrungen nicht ausgleichen kann (LOOSVELDT UND SONCK 2008: 144), kommen beispielsweise LEE sowie SCHONLAU et al. jeweils zu einem umgekehrten Ergebnis. Sie zeigen in ihren Untersuchungen auf, dass Gewichtungsfaktoren eine wichtige Rolle bei der Angleichung von Teilstichproben spielen können und so zur Reduktion von Verzerrungen, die aus unterschiedlich zusammengesetzten Teilstichproben her rühren, beitragen können (LEE 2006: 341ff.; SCHONLAU et al. 2009: 302ff.). Im Umkehrschluss kann ausgesagt werden, dass selbst bei Annahme einer Wirkungslosigkeit der Gewichtung nicht mit negativen Folgen oder Verzerrungen durch diese zu rechnen ist, womit sich im Nachhinein keine Einwände gegen die Gewichtungsprozedur zeigen.

Bezüglich der zu untersuchenden Fragestellung, ob zwischen den beiden zu vergleichenden Befragungsmethoden Online und Face-to-Face Methodeneffekte in Form von Antwortunterschieden feststellbar sind, ergibt sich in der Analyse kein konsistentes und eindeutiges Bild. So zeigen sich bei der Analyse unter Zuhilfenahme nichtparametrischer Tests wie dem  $\chi^2$ -Test nach Pearson oder dem U-Test nach Mann und Whitney bei verschiedenen Elementen des zur Untersuchung herangezogenen Fragebogens immer wieder Items bzw. Fragestellungen, bei denen signifikante Unterschiede im Antwortverhalten nachweisbar waren. Diese Unterschiede im Antwortverhalten folgen dabei jedoch keiner erkennbaren Systematik, beispielsweise im Bezug auf verschiedener Fragetypen wie etwa Einstellung-, Überzeugungs- oder Verhaltens-

fragen, so dass von der Annahme eines eindeutigen Vorliegens von Methodeneffekten abgerückt werden kann bzw. muss. Die erhaltenen Ergebnisse der Signifikanztests reichen nicht aus, um von einer eindeutigen Existenz von Methodeneffekten ausgehen zu können.

Dieses Resultat verwundert dabei nicht. So zeigte sich bereits in der Literaturübersicht in Kapitel 4, dass sich die Studienergebnisse zu Methodeneffekten stark voneinander unterscheiden und je nach Herangehensweise unterschiedlich ausfallen. Zusätzlich ist die Anzahl der Studien, die einen Methodenvergleich zwischen einer F2F-Befragung und einer Online-Befragung durchgeführt haben, noch sehr gering. Bei einer vergleichenden Betrachtung dieser Studien ergibt sich weiterhin, dass die Ergebnisse der Studien keiner eindeutigen Tendenz folgen. Während in einigen Studien, wie beispielsweise der qualitativen Studie von SCHOLL et al. keine signifikanten Antwortunterschiede zwischen den verglichenen Befragungssamples nachweisbar waren, kommen andere Studien hingegen zu gegenläufigen Ergebnissen, wie beispielsweise die Studien von SCHOEN et al. sowie DUFFY et al. (SCHOLL et al. 2002: 210ff.; SCHOEN et al. 2004: 274ff.; DUFFY et al. 2005: 615ff.). Da die wenigen Studien sich allesamt in ihrer Vorgehensweise und Methodik stark voneinander unterscheiden, ist ein Vergleich der unterschiedlichen Studienergebnisse darüber hinaus nur begrenzt möglich.

Die zusätzlich herangezogene multiple lineare Regressionsanalyse bestätigt bei zwei exemplarisch herangezogenen Fragestellungen die Resultate der zur Hauptanalyse herangezogenen nichtparametrischen Tests. So konnten bei der in einem ersten Regressionsmodell analysierten Fragestellung nach der Mehrzahlungsbereitschaft für CO<sub>2</sub>-reduzierte Milch durch den Mann-Whitney U-Test keine signifikanten Antwortunterschiede zwischen den beiden Teilstichproben ermittelt werden. Die Regressionsanalyse bestätigt dieses Ergebnis dahingehend, dass durch sie kein signifikanter Einfluss der Befragungsmethodik nachgewiesen werden konnte. Die zweite mittels der Regression analysierte Fragestellung beschäftigte sich mit den Determinanten des Wissens über das PCF-Zeichen. Während durch den nichtparametrischen Test hier ein höchst signifikanter Unterschied nachgewiesen werden konnte, zeigt im Gegensatz zur vorangegangenen Fragestellung die Regressionsanalyse hier einen signifikanten Einfluss der Befragungsmethodik auf. Bei der Besprechung der beiden Regressionsmodelle an dieser Stelle muss allerdings auch auf die relativ geringe Erklärungsgüte der beiden Modelle hingewiesen werden. So sind die geschätzten Modelle lediglich in der Lage, etwa 8,5 % bzw. 10,5 % der Gesamtvarianz zu erklären, wie die korrigierten Bestimmtheitsmaße aufzeigen. Vergleichsweise gut erscheint die Erklärungsgüte der Modelle hingegen, wenn als Vergleichsgrundlage die Studie von HEERSWEGH und LOOSVELDT herangezogen wird. Die Modelle der Autoren vermögen bei ähnlicher Spezifikation sogar nur drei bis fünf Prozent der Gesamtva-

rianz zu erklären (HEERSWEGH und LOOSVELDT 2008: 843). Ähnliche Bestimmtheitsmaße wie in der vorliegenden Arbeit konnten weiterhin GAUPP et al. ermitteln (GAUPP et al. 2006: 26f.). Aufgrund des Sachverhaltes, dass in Querschnittsanalysen eine verschieden große Zahl an erklärenden Faktoren für die Varianz bei der Modellbildung mangels Informationen nicht einbezogen werden kann, ist per se mit einem geringeren Wert für das Bestimmtheitsmaß zu rechnen. Die Werte für das Bestimmtheitsmaß in der vorliegenden Untersuchung ordnen sich dabei allerdings bei denen vorangegangener Studien ein oder liegen sogar darüber. Erwartungsgemäß werden beispielsweise bei der Erklärung der Mehrzahlungsbereitschaft für CO<sub>2</sub>-reduzierte Milch sehr viele Faktoren einen Einfluss haben, die im Rahmen empirischer Untersuchungen überhaupt nicht oder nur sehr aufwändig erfasst werden können. Dennoch sind die erhaltenen Ergebnisse wertvoll, da sie die durch die nichtparametrischen Tests erhaltenen Resultate bestätigen.

In der Tendenz der Ergebnisse ordnet sich die vorliegende Untersuchung inhaltlich zwischen den bisherigen Studien zur Thematik ein. So konnten bei den inhaltlich orientierten Fragestellungen des Kapitels 5.4 zwar vereinzelt Antwortunterschiede ermittelt werden, die aufgrund des unsystematischen und unregelmäßigen Auftretens jedoch gegen die generelle Existenz von Methodeneffekten sprechen, wie oben bereits angeführt werden konnte. Wurden bei Fragestellungen bzw. bei Items Antwortunterschiede zwischen den beiden Befragungsmethoden ermittelt, konnten diese weiterhin zumindest teilweise unter Zuhilfenahme des in Kapitel 3.3.3 erläuterten Konstrukts der sozialen Erwünschtheit als Methodeneffekte interpretiert werden. Offen bleibt jedoch, ob wirklich Konstrukte wie die soziale Erwünschtheit oder sonstige, unvorhersehbare Einflüsse zu dem unterschiedlichen Antwortverhalten in den beiden Teilstichproben geführt haben. Hierfür wäre die Integration spezieller Skalen zur Messung von sozialer Erwünschtheit notwendig gewesen. Derartige Skalen wurden beispielsweise von EDWARDS oder CROWNE und MARLOWE entwickelt (siehe hierzu EDWARDS 1957 sowie CROWNE und MALOWE 1960). Aufgrund des begrenzten Zeitfensters der vorliegenden Untersuchung wurde jedoch auf die Integration solcher Skalen verzichtet, da diese mit erheblichen Aufwendungen bei der Auswertung einher gehen. Diese zusätzlich notwendigen zeitlichen Ressourcen hätten letztlich zu einem starken Missverhältnis mit den verfügbaren (zeitlichen) Ressourcen gestanden, wodurch ein Verzicht gut und zweifelsfrei begründet erscheint.

Auf Basis der vorliegenden Untersuchungsergebnisse kann letztlich ausgesagt werden, dass grundsätzlich eine Substitution von mündlichen F2F-Befragungen durch online durchgeführte Befragungen möglich ist. Mit einem zwangsläufigen Auftreten von Methodeneffekten, die sich durch Antwortdifferenzen zwischen unterschiedlichen Befragungsmethoden kennzeich-

nen lassen, muss nicht gerechnet werden. Bei der Online-Befragung handelt es sich immer noch um eine relativ neue Befragungsmethode, die erst seit der Jahrtausendwende und einer zunehmenden Verbreitung von Computern mit Internetzugang in privaten Haushalten Anhänger findet. Während sie zu Beginn noch mit vielen Schwächen wie etwa der Repräsentativität oder der erreichbaren Datengüte gekennzeichnet war, scheinen diese Probleme zumindest für Populationen mit hohen Anteilen an Internetnutzern nicht mehr zu gelten. Allerdings haftet der Online-Befragung aufgrund dieser ehemaligen Probleme bis zum aktuellen Zeitpunkt immer noch ein in der Tendenz schlechter Ruf an. Ähnliche Bedenken zeigten sich in den Sozialwissenschaften auch bei dem Aufkommen der telefonischen Befragung. Zu Beginn mit ähnlichen Problemen behaftet wie anfangs die Online-Befragung, hing ihr relativ lange ein negatives Image bzw. ein negativer Beigeschmack an.

Heute gehört die telefonische Befragung von Personen zum Usus der empirischen Sozialforschung und wird beispielsweise im Rahmen der Wahl- bzw. Meinungsforschung zur Ermittlung der politischen Stimmung in der Bundesrepublik Deutschland eingesetzt (INFRATEST DIMAP 2010; FORSCHUNGSGRUPPE WAHLEN 2010). Verantwortlich hierfür ist allerdings auch, dass im Prinzip heute jede Person in Deutschland über einen Telefonanschluss verfügt bzw. telefonisch erreichbar ist, so dass die Ziehung bevölkerungsrepräsentativer Stichproben möglich ist. Bezüglich der Online-Befragung ist ebenfalls zu erwarten, dass diese aufgrund ihrer Vorteile weiter an Bedeutung zunehmen wird. Unter der Annahme, dass zukünftig noch mehr Menschen Zugang zum Internet haben werden, sind hier auch bevölkerungsrepräsentative Online-Befragungen denkbar. Wichtig wäre in diesem Zusammenhang allerdings die Zugangsregulierung zu derartigen Befragungen, um die Repräsentativität solcher Befragungen sicherstellen zu können, da zentrale Register für E-Mail-Adressen, ähnlich den Telefonbüchern, aus denen Zufallsstichproben gezogen werden könnten, für die Zukunft nicht zu erwarten sind. Ebenfalls ist nicht zu erwarten, dass die F2F-Befragung in der nahen Zukunft vollständig an Bedeutung verlieren wird. So sind mit ihr immer noch viele Vorteile, wie etwa die gute Möglichkeit der Kontrolle der Interviewsituation, verbunden, die auch zukünftig für eine Durchführung mündlicher Befragungen sprechen. Die vorliegende Untersuchung konnte allerdings zeigen, dass grundsätzlich nichts gegen eine Substitution von F2F-Befragungen durch online durchgeführte Befragungen spricht.

## 7 Zusammenfassung und Ausblick

Bei der mündlichen oder auch Face-to-Face-Befragung handelte es sich lange um das meistverwendete Erhebungsverfahren der empirischen Sozialforschung, weshalb die F2F-Befragung häufig auch als „Königsweg“ der empirischen Sozialforschung bezeichnet wurde bzw. wird. Seit etwa der Jahrtausendwende und der zunehmenden Verbreitung des Internets über alle gesellschaftlichen Gruppen und Milieus hinweg, zeigen sich jedoch zunehmend Tendenzen dahingehend, dass F2F-Befragungen durch online durchgeführte Befragungen substituiert werden. Hierbei stellt sich jedoch die Frage, ob eine derartige Substitution problemlos möglich ist oder ob möglicherweise sogar Verzerrungen in den Antworten zu erwarten sind. Um diese Fragestellung beantworten zu können, lautete die Zielsetzung der vorliegenden Arbeit, einen Methodenvergleich zwischen einer F2F-Befragung und einer Online-Befragung durchzuführen, um mögliche Unterschiede im Antwortverhalten, welche auf der Befragungsmethode basieren, identifizieren zu können.

Bei einem diesbezüglich vorgenommenen Literaturüberblick ergibt sich, dass es zwar eine gewisse Zahl an Untersuchungen gibt, die Antwortunterschiede zwischen verschiedenen Befragungsmethoden thematisieren, häufig hierbei jedoch schriftliche Befragungen mit online durchgeführten Befragungen verglichen werden. Vergleiche von Online-Befragungen mit F2F-Befragungen stellen hingegen die Ausnahme dar. Hier ist die Zahl der Studien, auch aufgrund des Sachverhaltes, dass es sich bei Online-Befragungen noch um eine relativ neue Befragungsmethode handelt, sehr gering. Hierdurch ergibt sich auch die Notwendigkeit nach weiteren Forschungsergebnissen.

Kapitel 3 thematisiert die beiden Befragungsmethoden F2F und Online. Hierbei zeigt sich, dass im Rahmen der F2F-Befragung grundsätzlich die alltägliche Befragung von der hier interessierenden, wissenschaftlichen Befragung abgegrenzt werden muss. Die Zielsetzung der wissenschaftlichen F2F-Befragung ist es dabei, Zusammenhänge aufzudecken, diese zu messen sowie weiterhin statistische Beziehungen zu interpretieren sowie zu verfeinern. Unter dem Oberbegriff der mündlichen Befragung können dabei weitere Unterarten voneinander abgegrenzt werden, die sich beispielsweise anhand ihres Standardisierungs- und Strukturierungsgrades, der Zahl der zur gleichen Zeit befragten Personen, der Art der Befragungssituation sowie der Frequenz bzw. der Häufigkeit der Befragung unterscheiden.

Während F2F-Befragungen dabei schon sehr lange im Bereich der empirischen Sozialforschung eingesetzt werden, erfreuen sich Online-Befragungen auch in der Wissenschaft zuneh-

mender Beliebtheit. Ursächlich hierfür kann sein, dass mit ihnen verschiedene Vorteile im Vergleich zu F2F-Befragungen beziehungsweise traditionellen Befragungsmethoden wie etwa schriftlicher oder telefonischer Befragung, verbunden werden. So ist im Allgemeinen die Durchführungsgeschwindigkeit von Online-Befragungen höher, da mehr Befragungsteilnehmer zur gleichen Zeit befragt werden können. Dadurch, dass im Rahmen von Online-Befragungen die Dateneingabe wie bei anderen Befragungsmethoden entfällt, sind ebenfalls Zeitvorteile zu verwirklichen. Auch wird mit Online-Befragungen der Aspekt der niedrigeren Kosten verbunden. So fallen im Rahmen von online durchgeführten Befragungen beispielsweise Ausgaben für Interviewer weg, wodurch die variablen Kosten stark reduziert werden können. Allerdings lassen sich auch Meinungen finden, die mit den niedrigen Kosten von Online-Befragungen auch die Gefahr von Qualitätsverlusten sehen. Weitere Vorteile von Online-Befragungen sind die besseren Gestaltungsmöglichkeiten im Vergleich zu anderen Befragungsmethoden, da multimediale Inhalte eingebunden werden können, sowie auch der Sachverhalt, dass der Interviewer als mögliche Fehlerquelle wegfällt.

Zur Unterscheidung verschiedener Typen von Online-Befragungen können verschiedene Zugangsarten voneinander abgegrenzt werden. Das mögliche Spektrum reicht dabei von einer öffentlichen Online-Befragung, an der jeder Interessierte teilnehmen kann, bis zu Befragungen, die auf vorab rekrutierte Panels zurückgreifen. Zur Ziehung repräsentativer Stichproben sind dabei nur wenige Zugangsarten geeignet.

Im Rahmen der Befragung können verschiedene Fehlerarten voneinander abgegrenzt werden, die sich auf einer ersten Stufe in Zufallsfehler sowie systematische Fehler aufteilen lassen. Die systematischen Fehler lassen sich weiterhin in Coveragefehler, Samplingfehler, Nonresponsefehler sowie Messfehler unterteilen. Coveragefehler entstehen, wenn die gezogenen Grund- und Auswahlgesamtheiten nicht miteinander übereinstimmen. Hier wird auch ein Problem von Online-Befragungen gesehen. So waren es bisher vor allem Männer, jüngere sowie höhergebildete Personen, die an Online-Befragungen teilnahmen. Während Samplingfehler bei der Stichprobenziehung entstehen können, ergibt sich der sogenannte Nonresponsefehler aus der Nichtteilnahme von ausgewählten Befragungspersonen an der gesamten Befragung oder durch die Nichtbeantwortung einzelner Fragestellungen. Neben diesen Fehlern kann zusätzlich noch die Gruppe der Messfehler abgegrenzt werden, deren Ursachen beim Interviewer, beim angewendeten Instrument, beim Befragten oder auch bei der angewendeten Befragungsmethodik liegen können. Diese verschiedenen Störeinflüsse können dabei unter dem Sammelbegriff des Methodeneffektes vereint werden.

Zielsetzung der empirischen Analyse war, Methodeneffekte zwischen einer F2F-Befragung sowie einer Online-Befragung aufzudecken, die sich durch signifikante Antwortunterschiede zeigen. Hierfür wurden im Zeitraum vom 01.07.2009 bis zum 31.08.2009 Studierende der Justus-Liebig-Universität unter Anwendung der Split-Ballot-Technik sowohl durch Interviewer als auch online befragt. Bei der Split-Ballot-Technik werden aus einer definierten Grundgesamtheit zwei Teilstichproben gezogen, die mit einem grundsätzlich gleichen, jedoch in einem Sachverhalt abgewandelten Instrument befragt werden. Im Rahmen der vorliegenden Untersuchung ist dabei entsprechend die Befragungsmethodik variiert worden. Daher wurde eine Teilstichprobe F2F mit Hilfe von Interviewern befragt, während die zweite Teilstichprobe online befragt wurde. Der herangezogene Fragebogen war dabei gleich und wurde nur an wenigen, notwendigen Punkten an die Befragungsmethodik angepasst. Thematik des Fragebogens war der anthropogene Klimawandel und mögliche Carbon-Labeling-Strategien des Lebensmitteleinzelhandels. Den Befragten war dabei zu keiner Zeit die wahre Intention der Teilbefragungen, nämlich das Methodenexperiment, bewusst. Nach Abzug unvollständig ausgefüllter Fragebogen blieben letztlich 700 Datensätze übrig. Da sich die beiden Teilstichproben hinsichtlich demographischer Variablen voneinander unterschieden, sind die Teilstichproben über eine Gewichtungsprozedur aneinander angeglichen worden. Die eigentliche Testauswertung erfolgte mit Hilfe nichtparametrischer Testverfahren, da die Daten den Anforderungen des t-Tests nicht genügten.

Hinsichtlich möglicher Methodeneffekte konnten über den gesamten Fragebogen hinweg immer wieder signifikante Antwortunterschiede zwischen den beiden Teilstichproben über die nichtparametrischen Tests aufgedeckt werden. Allerdings folgten diese keiner Systematik, weshalb hieraus nicht automatisch auf Methodeneffekte geschlossen werden konnte. So wurde zu Beginn der Analyse zwar davon ausgegangen, dass es sich hinsichtlich der Einstellungen bei dem Studierendensample um ein recht homogenes Sample handelt, aufgrund der Ergebnisse musste jedoch davon ausgegangen werden, dass möglicherweise auch andere Einflussfaktoren, wie beispielsweise unterschiedliche Einstellungen in den Teilstichproben zu diesen signifikanten Antwortunterschieden geführt haben.

Daher sind mit Hilfe der Faktorenanalyse verschiedene Einstellungsdimensionen zur Politik sowie zum Klimawandel extrahiert worden. Bei Überprüfung mit Hilfe des t-Tests konnte hier bei zwei Einstellungsdimensionen ein signifikanter Unterschied zwischen den beiden Teilstichproben aufgedeckt werden. Bei der Berechnung der Effektstärke  $d$  nach Cohen stellten sich diese allerdings als schwach heraus.

Um letztlich die Determinanten des Antwortverhaltes ermitteln zu können, wurde bei zwei exemplarisch herangezogenen Fragestellungen zur Mehrzahlungsbereitschaft für CO<sub>2</sub>-reduzierte Milch sowie zum Wissen zum Product Carbon Footprint-Zeichen eine multiple lineare Regressionsanalyse durchgeführt. Die Ergebnisse der beiden Regressionsmodelle bestätigen dabei die Ergebnisse der Signifikanztests: Bei der Fragestellung nach der Mehrzahlungsbereitschaft, bei der durch die nichtparametrischen Tests kein signifikanter Antwortunterschied nachgewiesen werden konnte, zeigt sich auch im Regressionsmodell kein signifikanter Einfluss der Befragungsmethodik. Bei der Fragestellung nach dem Wissen der Befragten zum Product Carbon Footprint-Zeichen zeigt sich hingegen ein signifikanter Einfluss der Befragungsmethode im Regressionsmodell. Hier konnte bereits mit Hilfe des Mann-Whitney U-Tests ein signifikanter Antwortunterschied nachgewiesen werden. Insgesamt bestätigt die Regressionsanalyse somit die Ergebnisse. Weiterhin zeigt sie allerdings auch signifikante Einflüsse der ermittelten Einstellungsdimensionen auf das Antwortverhalten der Befragten. Möglicherweise sind daher die ermittelten Antwortunterschiede auch auf die Einstellungsunterschiede in den Teilstichproben zurückzuführen.

Insgesamt ergibt sich jedoch, dass bei einer Substitution einer F2F-Befragung durch eine online durchgeführte Befragung nicht zwangsläufig mit dem Auftreten von Methodeneffekten gerechnet werden muss, da nur vereinzelt Antwortunterschiede nachgewiesen werden konnten. Hier kommt es immer auf den Individualfall an, wodurch sich auch der weitere Forschungsbedarf ergibt.

So könnten im Rahmen weiterführender Untersuchungen alle Items und Fragestellungen des Fragebogens, bei denen durch die nichtparametrischen Tests signifikante Antwortunterschiede nachgewiesen werden konnten, mit Hilfe der multiplen linearen Regressionsanalyse analysiert werden. Hierdurch könnten weitere Erkenntnisse gewonnen werden.

Des Weiteren wurden im Fragebogen zwar verschiedene Fragetypen verwendet und in der Analyse auf das Vorliegen von Methodeneffekten hin überprüft. Allerdings könnte im Rahmen einer zukünftigen Untersuchung weiterführend analysiert werden, ob sich zwischen verschiedenen Fragetypen unterschiedliche Anfälligkeiten für Methodeneffekte ergeben. So ist es denkbar, dass Methodeneffekte neben anderen Determinanten auch vom gewählten Fragentyp abhängen, wie die nachfolgende Gleichung präzisiert:

$$\text{Methodeneffekte} = f(\text{Fragentyp}, \dots) \quad (13)$$

Auch bezieht sich die vorliegende Untersuchung auf ein rein studentisches Sample. Hier stellt sich abschließend die Frage, inwiefern die gefundenen Ergebnisse auch auf ein bevölkerungs-

repräsentatives Sample übertragbar sind. Aufgrund des ungleich höheren Erhebungsaufwandes könnte ein derartig formulierter Forschungsgegenstand in der Zukunft aufgegriffen werden. Diese Frage erscheint unter dem Sachverhalt, dass auch zukünftig mit einer Substitution von F2F-Befragungen durch online durchgeführte Befragungen bei bevölkerungsrepräsentativen Stichproben zu rechnen ist, nicht uninteressant.

## 8 Literaturverzeichnis

- ADM ARBEITSKREIS DEUTSCHER MARKT- UND SOZIALFORSCHUNGSINSTITUTE (2008): Jahresbericht 2008. Internet: [http://www.adm-ev.de/fileadmin/user\\_upload/PDFS/Jahresbericht\\_08.pdf](http://www.adm-ev.de/fileadmin/user_upload/PDFS/Jahresbericht_08.pdf). Abgerufen am: 16.10.2009.
- AGOF ARBEITSGEMEINSCHAFT ONLINE FORSCHUNG (2009): Berichtsband zur internet facts 2009-II. Internet: <http://www.agof.de/berichtsband-if2009-ii.download.848e0b18fb-27fa58bbe4541db0354a5c.pdf>. Abgerufen am: 17.09.2009.
- AKERLOF, G. A. (1970): The Market for „Lemons“: Quality uncertainty and the Market Mechanism. In: The Quarterly Journal of Economics, Volume 84, No. 3 (August). S. 488-500.
- ALEMANN, H. VON (1977): Der Forschungsprozeß: Eine Einführung in die Praxis der empirischen Sozialforschung. (Studienskripte zur Soziologie, Band 30). Stuttgart: Teubner.
- ALTOBELLI, C. F. (2007): Marktforschung: Methoden - Anwendungen - Praxisbeispiele. Stuttgart: Lucius & Lucius.
- ANONYMUS (2008): Methode mit Effekt. In: a3-boom, Nr. 07-08/08 vom 07.08.2008. Internet: [http://www.wiso-net.de/webcgi?START=A60&DOKV\\_DB=ZGEN&DOKV\\_NO=AAA0653066079079077095200808+071636560009&DOKV\\_HS=0&PP=1](http://www.wiso-net.de/webcgi?START=A60&DOKV_DB=ZGEN&DOKV_NO=AAA0653066079079077095200808+071636560009&DOKV_HS=0&PP=1). Abgerufen am: 16. 12.2009.
- ATTESLANDER, P. und H.-U. KNEUBÜHLER (1975): Verzerrungen im Interview: Zu einer Fehlertheorie der Befragung. (Studien zur Sozialwissenschaft, Band 32). Opladen: Westdeutscher Verlag.
- ATTESLANDER, P. und M. KOPP (1999): Befragung. In: ROTH, E., K. HEIDENREICH und H. HOLLING (Hrsg.): Sozialwissenschaftliche Methoden: Lehr- und Handbuch für Forschung und Praxis. 5., durchgesehene Auflage. München, Wien: Oldenbourg. S. 146-174.
- ATTESLANDER, P. (2008): Methoden der empirischen Sozialforschung. 12., durchgesehene Auflage. Berlin: Erich Schmidt Verlag.
- AUSPURG, K. und M. ABRAHAM (2007): Die Umzugsentscheidung von Paaren als Verhandlungsproblem. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie, Jahrgang 59, Heft 1. S. 271-293.

- BÄCHLE, M. (2006): Social Software. In: Informatik-Spektrum, Volume 29, Nr. 2 (April 2006). S. 121-124.
- BACKHAUS, K., B. ERICHSON, W. PLINKE und R. WEIBER (2008): Multivariate Analysemethoden: Eine anwendungsorientierte Einführung. 12., vollständig überarbeitete Auflage. Berlin, Heidelberg: Springer.
- BAGOZZI, R. P. (1984): A Prospectus for Theory Construction in Marketing. In: The Journal of Marketing, Volume 48, No. 1 (Winter 1984). S. 11-29.
- BANDILLA, W. und M. BOSNJAK (2000): Perspektiven der Online-Forschung. In: STATISTISCHES BUNDESAMT (Hrsg.): Neue Erhebungsinstrumente und Methodeneffekte. (Spektrum Bundesstatistik, Band 15). S. 166-174.
- BANDILLA, W., M. BOSNJAK und P. ALTDORFER (2001): Effekte des Erhebungsverfahrens? Ein Vergleich zwischen einer Web-basierten und einer schriftlichen Befragung zum ISSP-Modul Umwelt. In: ZUMA-Nachrichten, Jahrgang 25, Heft 49. S. 7-28.
- BANDILLA, W., L. KACZMIREK, M. BLOHM und W. NEUBARTH (2009): Coverage- und Non-response-Effekte bei Online-Bevölkerungsumfragen. In: JAKOB, N., H. SCHOEN und T. ZERBACK (Hrsg.): Sozialforschung im Internet: Methodologie und Praxis der Online-Befragung. Wiesbaden: VS Verlag für Sozialwissenschaften. S. 129-143.
- BASS, B. M. (1955): Authoritarianism or Acquiescence? In: Journal of Abnormal and Social Psychology, Volume 51, No. 3 (November). S. 616-623.
- BASS, B. M. (1956): Development and Evaluation of a Scale for Measuring Social Acquiescence. In: Journal of Abnormal and Social Psychology, Volume 53, No. 3 (November). S. 296-299.
- BAUR, N. (2008): Das Ordinalskalnenproblem. In: BAUR, N. und S. FROMM (Hrsg.): Datenanalyse mit SPSS für Fortgeschrittene: Ein Arbeitsbuch. 2., überarbeitete und erweiterte Auflage. Wiesbaden: VS Verlag für Sozialwissenschaften. S. 279-289.
- BAUR, N. und M. J. FLORIAN (2009): Stichprobenprobleme bei Online-Umfragen. In: JAKOB, N., H. SCHOEN und T. ZERBACK (Hrsg.): Sozialforschung im Internet: Methodologie und Praxis der Online-Befragung. Wiesbaden: VS Verlag für Sozialwissenschaften. S. 109-128.
- BECH, M. und M. B. KRISTENSEN (2009): Differential Response Rates in Postall and Web-based Surveys among older Respondents. In: Survey Research Methods, Volume 3, No. 1. S. 1-6.

- BENNINGHAUS, H. (2001): Einführung in die sozialwissenschaftliche Datenanalyse. 6., überarbeitete Auflage. München, Wien: Oldenbourg.
- BEREKOVEN, L., W. ECKERT und P. ELLENRIEDER (2009): Marktforschung: Methodische Grundlagen und praktische Anwendung. 12., überarbeitete und erweiterte Auflage. Wiesbaden: Gabler.
- BERRENS, R. P., A. K. BOHARA, H. JERKINS-SMITH, C. SILVA und D. L. WEIMER (2003): The Advent of Internet Surveys for Political Research: A Comparison of Telephone and Internet Samples. In: Political Analysis, Volume 11, Number 1. S. 1-22.
- BILHARZ, M. und F.-M. BELZ (2008): Öko als Luxus-Trend: Rosige Zeiten für die Vermarktung „grüner“ Produkte? In: Marketing Review St. Gallen, Volume 25, Heft 4 (April 2008). S. 4-10.
- BODEN, T.A., G. MARLAND und R.J. ANDRES (2009): Global, Regional, and National Fossil-Fuel CO<sub>2</sub> Emissions. Internet: [http://cdiac.ornl.gov/trends/emis/tre\\_glob.html](http://cdiac.ornl.gov/trends/emis/tre_glob.html). Abgerufen am: 17.01.2010.
- BOLWIG, S. und P. GIBBON (2009a): Overview of Product Carbon Footprinting Schemes and Standards: Report prepared for the OECD Joint Working Party on Trade and Environment, Trade and Agriculture Directorate. Internet: [http://www.se2009.eu/polopoly\\_fs/1.24631!menu/standard/file/Simon%20Bolwig.pdf](http://www.se2009.eu/polopoly_fs/1.24631!menu/standard/file/Simon%20Bolwig.pdf). Abgerufen am: 10.02.2010.
- BOLWIG, S. und P. GIBBON (2009b): Emerging Product Carbon Footprint Standards and Schemes and their possible Trade Impacts. (Risø-R-Report Nr. 1719(EN)). Internet: <http://orbit.dtu.dk/getResource?recordId=254090&objectId=1&versionId=1>. Abgerufen am: 10.02.2010.
- BORTZ, J. (2005): Statistik für Human- und Sozialwissenschaftler. 6., vollständig überarbeitete und aktualisierte Auflage. Heidelberg: Springer Medizin.
- BORTZ, J. und N. DÖRING (2006): Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. 4., überarbeitete Auflage. Heidelberg: Springer Medizin.
- BORTZ, J., G. A. LIENERT und K. BOEHNKE (2008): Verteilungsfreie Methoden in der Biostatistik. 3., korrigierte Auflage. Heidelberg: Springer Medizin.
- BOSNJAK, M., T. L. TUTEN und W. BANDILLA (2001): Participation in Web Surveys: A Typology. In: ZUMA-Nachrichten, Jahrgang 25, Heft 48. S. 7-17.

- BOSNJAK, M. (2003): Teilnahmeverhalten bei Web-Befragungen: Nonresponse und Selbstselektion. In: THEOBALD, A., M. DREYER und T. STARSETZKI (Hrsg.): Online-Marktforschung: Theoretische Grundlagen und praktische Erfahrungen. 2., vollständig überarbeitete und erweiterte Auflage. Wiesbaden: Gabler. S. 55-72.
- BRENTON, P., G. EDWARDS-JONES und M. F. JENSEN (2009): Carbon Labelling and Low-income Country Exports: A Review of the Development Issues. In: Development Policy Review, Volume 27, No. 3. S. 243-267.
- BROSIUS, H.-B., F. KOSCHEL und A. HAAS (2008): Methoden der empirischen Kommunikationsforschung: Eine Einführung. 4., überarbeitete und erweiterte Auflage. Wiesbaden: VS Verlag für Sozialwissenschaften.
- CANTRIL, H. (1947): Gauging Public Opinion. Princeton: Princeton University Press.
- CARBON TRUST (2008): Working with Tesco: Product Carbon Footprinting in Practice. Internet: <http://www.carbon-label.com/casestudies/Tesco.pdf>. Abgerufen am: 11.02.2010.
- CARINI, R. M., J. C. HAYEK, G. D. KUH, J. M. KENNEDY und J. A. OUMET (2003): College Student Responses to Web and Paper Surveys: Does Mode Matter? In: Research in Higher Education, Volume 44, No. 1 (February). S. 1-19.
- CHEN, R., K. L. KRAEMER und P. SHARMA (2009): Google: Das weltweit erste „Information Utility“? In: Wirtschaftsinformatik, Volume 51, Nr. 1 (Februar 2009). S. 61-71.
- CHRISTL, A. (2005): Cookies, Web-Logs, Location Based Services, eMail, Webbugs, Spyware: Datenschutz im Internet. München, Ravensburg: Grin. Zugleich: Dissertation, Universität Salzburg.
- CLEFF, T. (2008): Deskriptive Statistik und moderne Datenanalyse: Eine computergestützte Einführung mit Excel, SPSS und STATA. Wiesbaden: Gabler.
- COBANOGLU, C., B. WARDE und P. J. MOREO (2001): A Comparison of Mail, Fax and Web-based Survey Methods. In: International Journal of Market Research, Volume 43, Quarter 4. S. 441-452.
- COHEN, J. (1992): A Power Primer. In: Psychological Bulletin, Volume 112, No. 1 (January 1992). S. 155-159.
- CONVERSE, J. M. und S. PRESSER (1986): Survey Questions: Handcrafting the Standardized Questionnaire. (Sage University Paper: Quantitative Applications in the Social Sciences, Number 07-063). Beverly Hills, Newbury Park, London, New Delhi: Sage Publications.

- COOK, C., F. HEATH und R. L. THOMSON (2000): A Meta-Analysis of Response Rates in Web- or Internet-Based Surveys. In: Educational and Psychological Measurement, Volume 60, No. 6 (December). S. 821-836.
- COUPER, M. P., J. BLAIR und T. TRIPLETT (1999): A Comparison of Mail and E-Mail for a Survey of Employees in U.S. Statistical Agencies. In: Journal of Official Statistics, Volume 15, No. 1. S. 39-56.
- COUPER, M. P. (2000): Web Surveys: A Review of Issues and Approaches. In: Public Opinion Quarterly, Volume 64. S. 464-494.
- COUPER, M. P. und E. COUTTS (2006): Online-Befragung: Probleme und Chancen verschiedener Arten von Online-Erhebungen. In: DIEKMANN, A. (Hrsg.): Methoden der Sozialforschung. (Sonderhefte der Kölner Zeitschrift für Soziologie und Sozialpsychologie). Wiesbaden: VS Verlag für Sozialwissenschaften. S. 217-243.
- COX, D. R. (1961): Design of Experiments: The Control of Error. In: Journal of the Royal Statistics Society, Series A (General), Volume 124, No. 1. S. 44-48. Zitiert in: ATTESLANDER, P. und H.-U. KNEUBÜHLER (1975): Verzerrungen im Interview: Zu einer Fehlertheorie der Befragung. (Studien zur Sozialwissenschaft, Band 32). Opladen: Westdeutscher Verlag.
- CRONBACH, L. J. (1946): Response Sets and Test Validity. In: Educational and Psychological Measurement, Volume 6, Issue 4. S. 475-494.
- CROWNE, D. P. und D. MARLOWE (1960): A new Scale of Social Desirability Independent of Psychopathology. In: Journal of Consulting Psychology, Volume 24, No. 4. S. 349-354.
- DIEKMANN, A. (2008): Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen. 19. Auflage Januar 2008. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag.
- DIENER, J. (1999): Ökologischer Vergleich von NMT- und GMT-Bauteilen. Internet: [http://www.uni-kassel.de/fb15/ifw/kutech/zu%20downloaden/Tagung\\_1999\\_PDF/33%20Diener.pdf](http://www.uni-kassel.de/fb15/ifw/kutech/zu%20downloaden/Tagung_1999_PDF/33%20Diener.pdf). Abgerufen am: 11.02.2010.
- DILLMAN, D. A. (1978): Mail and Telephone Surveys: The Total Design Method. New York: Wiley.
- DILLMAN, D. A. (2007): Mail and Internet Surveys: The Tailored Design Method. Hoboken: Wiley.

- DUFFY, B., K. SMITH, G. TERHANIAN und J. BREMER (2005): Comparing data from online and face-to-face-surveys. In: *International Journal of Market Research*, Volume 47, Issue 6. S. 615-639.
- DYKEMA, J., D. BASSON und N. C. SCHAEFFER (2008): Face-to-Face-Surveys. In: DONSBACH, W. und M. W. TRAUGOTT (Hrsg.): *The SAGE Handbook of Public Opinion Research*. London, New Delhi, Singapur, Thousand Oaks: Sage Publications. S. 240-248.
- DZIUBAN, C. D. und E. C. SHIRKLEY (1974): When is a Correlation Matrix appropriate for Factor Analysis?: Some Decision Rules. In: *Psychological Bulletin*, Volume 81, No. 6 (June 1974). S. 358-361.
- EDWARDS, A. L. (1957): *The Social Desirability Variable in Personality Assessment and Research*. New York: Holt.
- EIA U.S. ENERGY INFORMATION ADMINISTRATION (2009): *International Energy Outlook 2009*. Internet: <http://www.eia.doe.gov/oiaf/ieo/pdf/0484%282009%29.pdf>. Abgerufen am: 10.02. 2010.
- ESSER, H. (1986): Können Befragte lügen? Zum Konzept des „wahren Wertes“ im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, Jahrgang 38, Heft 2. S. 314-336.
- EUROPEAN COMMISSION (2009): *Europeans' Attitudes towards the Issue of sustainable Consumption and Production: Analytical Report*. (Flash Eurobarometer, Publication TBD 2009). Internet: [http://ec.europa.eu/public\\_opinion/flash/fl\\_256\\_en.pdf](http://ec.europa.eu/public_opinion/flash/fl_256_en.pdf). Abgerufen am: 11.02.2010.
- EVANS, J. R. und A. MATHUR (2005): The Value of Online Surveys. In: *Internet Research*, Volume 15, No. 2. S. 195-219.
- FAAS, T. (2003): Umfragen im Umfeld der Bundestagswahl 2002: Offline und Online im Vergleich. In: *ZA-Information*, Heft 52 (Mai). S. 120–135.
- FAAS, T. und H. SCHOEN (2009): Fallen Gewichte ins Gewicht? Eine Analyse am Beispiel dreier Umfragen zur Bundestagswahl 2002. In: JAKOB, N., H. SCHOEN und T. ZERBACK (Hrsg.): *Sozialforschung im Internet: Methodologie und Praxis der Online-Befragung*. Wiesbaden: VS Verlag für Sozialwissenschaften. S. 145-157.

- FALLOWS, D. (2005): How Women and Men use the Internet. Internet: [http://www.pewinternet.org/~media/Files/Reports/2005/ PIP\\_Women\\_and\\_Men\\_online.pdf](http://www.pewinternet.org/~media/Files/Reports/2005/PIP_Women_and_Men_online.pdf). Abgerufen am: 16.09.2009.
- FAUL, F., E. ERDFELDER, A.-G. LANG und A. BUCHNER (2007): G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. In: Behavior Research Methods, Volume 39, Issue 2. S. 175-191.
- FORSCHUNGSGRUPPE WAHLEN (2010): Politbarometer Februar 2010. Internet: <http://www.forschungsgruppe.de/Aktuelles/Politbarometer/>. Abgerufen am: 08.03.2010.
- FOWLER, F. J. (1984): Survey Research Methods. (Applied Social Research Methods Series). Thousand Oaks: Sage Publications.
- FREY, B. und D.-S. MÜHLBACH (2009): Betriebliche Klimastrategie und Carbon Footprint: Gesellschaftliche Verantwortung im Einzelhandel. In: ABLÄNDER, M. S. und K. SENGE (Hrsg.): Corporate Social Responsibility im Einzelhandel. (Ethik und Ökonomie, Band 8). Marburg: Metropolis. S. 179-202.
- FRIEDRICHS, J. (1973): Methoden empirischer Sozialforschung. (Rororo-Studium, Band 28). Reinbeck: Rowohlt.
- FUCHS, M. (2003): Kognitive Prozesse und Antwortverhalten in einer Internet-Befragung. In: Österreichische Zeitschrift für Soziologie, Heft 4. S. 19-45.
- GABLER, S. (2006): Gewichtungprobleme in der Datenanalyse. In: DIEKMANN, A. (Hrsg.): Methoden der Sozialforschung. (Sonderhefte der Kölner Zeitschrift für Soziologie und Sozialpsychologie). Wiesbaden: VS Verlag für Sozialwissenschaften. S. 128-147.
- GADEIB, A. (2005): Online-Stichproben: Anspruch und Wirklichkeit. In: Planung & Analyse, Nr. 1. S. Internet: [http://www2.dialego.de/uploads/media/050322D\\_\\_pua\\_Flie\\_text\\_Online\\_Stichproben\\_03.pdf](http://www2.dialego.de/uploads/media/050322D__pua_Flie_text_Online_Stichproben_03.pdf). Abgerufen am: 04.11.2009.
- GAUPP, N., R. KUHNKE und E. SCHWEIGARD (2006): Vergleich unterschiedlicher Erhebungsmethoden: Arbeitsbericht im Rahmen der Dokumentationsreihe: Methodische Erträge aus dem „DJI-Übergangspanel“. Internet: [http://www.cgi.dji.de/bibs/276\\_5912\\_WT\\_1\\_2006\\_gauppua.pdf](http://www.cgi.dji.de/bibs/276_5912_WT_1_2006_gauppua.pdf). Abgerufen am: 28.12.2009.
- GOOGLE (2010): Suche nach Carbon Labelling. Internet: <http://www.google.de/#hl=de&q=Carbon+Labelling&meta=&aq=f&oq=Carbon+Labelling&fp=749e75dee4f1c52f>. Abgerufen am 14.01.2010.

- GRABL, H. (2009): Wissenschaftliche Basis globaler Klimaänderungen durch den Menschen. In: WALLACHER, J. und K. SCHARPENSEEL (Hrsg.): Klimawandel und globale Armut. (Globale Solidarität - Schritte zu einer neuen Weltkultur, Band 18). Stuttgart: Kohlhammer. S. 5-17.
- GROUPE CASINO (2010): The Casino Carbon Index: A green Label for Products. Internet: <http://www.groupe-casino.fr/en/The-Casino-Carbon-Index-a-green.html>. Abgerufen am: 12.02.2010.
- GROVES, R. M., F. J. FOWLER, M. P. COUPER, J. M. LEPKOWSKI, E. SINGER und R. TOURANGEAU (2004): Survey Methodology. (Wiley series in survey methodology). Hoboken: Wiley.
- GÜNTHER, M., U. VOSSEBEIN und R. WILDNER (2006): Marktforschung mit Panels: Arten - Erhebung - Analyse - Anwendung. 2., vollständig überarbeitete Auflage. Wiesbaden: Gabler.
- HABERMEHL, W. (1992): Angewandte Sozialforschung. München, Wien: Oldenbourg.
- HAMBURGER BILDUNGSSERVER (2007): Der natürliche Treibhauseffekt. Internet: [http://www.hamburger-bildungsserver.de/welcome.phtml?unten=/klima/klimawandel/atmosphaere/treibhauseffekt\\_nat.html](http://www.hamburger-bildungsserver.de/welcome.phtml?unten=/klima/klimawandel/atmosphaere/treibhauseffekt_nat.html). Abgerufen am: 17.01.10.
- HARTL, J. (2008): Die Nachfrage nach genetisch veränderten Lebensmitteln: Anwendung neuerer Entwicklungen der Discrete-Choice-Analyse zur Bewertung genetisch veränderter Lebensmittel mit Output-Traits. Frankfurt: DLG-Verlag. Zugleich: Dissertation, Universität Gießen.
- HAYSLETT, M. M. und B. M. WILDEMUTH (2004): Pixels or Pencils? The relative Effectiveness of Web-based versus Paper Surveys. In: Library and Information Science Research, Volume 26. S. 73-93.
- HDE HAUPTVERBAND DES DEUTSCHEN EINZELHANDELS (2008): Positionspapier zur Kennzeichnung von Produkten mit dem jeweiligen CO<sub>2</sub>-Verbrauch (produktbezogener CO<sub>2</sub>-Fußabdruck bzw. „carbon footprint of products“). Internet: [http://www.einzelhandel.de/pb/site/hde/get/params\\_Dattachment/75454/2008\\_01\\_18\\_%20HDE-Position\\_%20carbon\\_footprint.pdf](http://www.einzelhandel.de/pb/site/hde/get/params_Dattachment/75454/2008_01_18_%20HDE-Position_%20carbon_footprint.pdf). Abgerufen am: 15.02.2010.
- HEERWEGH, D. und G. LOOSVELDT (2008): Face-to-Face versus Web Surveying in a High-Internet-Coverage Population: Differences in Response Quality. In: Public Opinion Quarterly, Volume 72, No 5. S. 836-846.

- HEIMANN, M. (2010): How stable ist he Methane Cycle? In: Science, Volume 327, No. 5970 (March 2010). S. 1211-1212.
- HENKEL, T. (2008): Die Bedeutung der Regionalität beim Lebensmittelkauf: Empirische Ergebnisse am Beispiel „Landmarkt“. (Arbeitsberichte des Instituts für Agrarpolitik und Marktforschung der Justus-Liebig-Universität Gießen, Nr. 45). Gießen: Institut für Agrarpolitik und Marktforschung.
- HUDSON, D., L.-H. SEAH, D. HITE und T. HAAB (2004): Telephone Presurveys, Self-Selection, and Non-Resonse Bias to Mail and Internet Surveys in Economic Research. In: Applied Economic Letters, Volume 11, Issue 4. S. 237-240).
- HUI, C. H. und H. C. TRIANDIS (1985): The Instability of Response Sets. In: Public Opinion Quarterly, Volume 49, Issue 2. S. 253-260.
- INFRAEST DIMAP (2010): ARD-DeutschlandTREND März 2010. Internet: <http://www.infratest-dimap.de/umfragen-analysen/bundesweit/ard-deutschlandtrend/2010/maerz/>. Abgerufen am: 08.03.2010.
- INITIATIVE D21 (2009a): (N)Onliner Atlas 2009: Eine Topographie des digitalen Grabens durch Deutschland: Nutzung und Nichtnutzung des Internets, Strukturen und regionale Verteilung. Internet: <http://www.initiaved21.de/wp-content/uploads/2009/06/NONLINER2009.pdf>. Abgerufen am: 17.09.2009.
- INITIATIVE D21 (2009b): Struktur und Organisation. Internet: <http://www.initiaved21.de/ueber/struktur>. Abgerufen am: 17.09. 2009.
- IPCC INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (2007a): Summary for Policymakers: A Report of Working Group I of the Intergovernmental Panel on Climate Change. In: SOLOMON, S., D. QIN, M. MANNING, Z. CHEN, M. MARQUIS, K. B. AVERYT, M. TIGNOR und H. L. MILLER (Hrsg.): Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, New York: Cambridge University Press. S. 1-18
- IPCC INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (2007b): Changes in Atmospheric Constituents and Radiative Forcing. In: SOLOMON, S., D. QIN, M. MANNING, Z. CHEN, M. MARQUIS, K. B. AVERYT, M. TIGNOR und H. L. MILLER (Hrsg.): Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, New York: Cambridge University Press. S. 129-234.

- IPCC INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (2007c): Summary for Policymakers: In: PARRY, M. L., O. F. CANZIANI, J. P. PALUTIKOF, P. J. VAN DER LINDEN und C. E. HANSON (Hrsg.): Climate Change 2007: Impacts, Adaption and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press. S. 7-22.
- IPCC INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (2007d): New Assessment Methods and the Characteristics of Future Conditions. In: PARRY, M., O. CANZIANI, J. PALUTIKOF, P. VAN DER LINDEN und C. HANSON (2007): Climate Change 2007: Impacts, Adaption and Vulnerability: Working Group II Contribution to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press. S. 133-171.
- IPCC INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (2010a): Organization. Internet: [http:// www.ipcc.ch/organization/organization.htm](http://www.ipcc.ch/organization/organization.htm). Abgerufen am: 10.02.2010.
- IPCC INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (2010b): Figures and Tables. Internet: <http://www.ipcc.ch/graphics/ar4-wg1/jpg/spm5.jpg>. Abgerufen am: 10.02.2010.
- ISO INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (2010): ISO/WD 14067: Carbon Footprint of Products - Part I. Internet: [http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=43278](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43278). Abgerufen am: 15.02.2010.
- JACKMAN, M. R. (1973): Education and Prejudice or Education and Response-Set? In: American Sociological Review, Volume 38, No. 3 (June). S. 327-339.
- JANSSEN, J. und W. LAATZ (2007): Statistische Datenanalyse mit SPSS für Windows: Eine anwendungsorientierte Einführung in das Basissystem. 6., neu bearbeitete und erweiterte Auflage. Berlin, Heidelberg, New York: Springer.
- JAROS, D. und G. L. MASON (1969): Party Choice and Support for Demagogues. In: American Political Science Review, Volume 63, Number 1 (March). S. 100-110.
- JLU JUSTUS-LIEBIG-UNIVERSITÄT GIEßEN (2002): Studienordnung des Fachbereichs Agrarwissenschaften, Ökotropologie und Umweltmanagement der Justus-Liebig-Universität Gießen vom 20. Juni 2001. Internet: [http://www.uni-giessen.de/cms/fbz/fb09/studiumlehre/bsc/BSc\\_alt/pdf-bsc-alt/pa-po.pdf](http://www.uni-giessen.de/cms/fbz/fb09/studiumlehre/bsc/BSc_alt/pdf-bsc-alt/pa-po.pdf). Abgerufen am: 05.10.2009.
- JLU JUSTUS-LIEBIG-UNIVERSITÄT GIEßEN (2007): E-Mail-Adresse, Mailbox. Internet: <http://www.uni-giessen.de/hrz/service/email/em-adresse.html#k1>. Abgerufen am: 30.09.2009.

- JLU JUSTUS-LIEBIG-UNIVERSITÄT GIEßEN (2009a): Stud.IP. Internet: <http://studip.uni-giessen.de>. Abgerufen am: 21.09.2009.
- JLU JUSTUS-LIEBIG-UNIVERSITÄT GIEßEN (2009b): FlexNow!. Internet: <https://flexnow.uni-giessen.de>. Abgerufen am: 21.09.2009.
- JLU JUSTUS-LIEBIG-UNIVERSITÄT GIEßEN (2009c): Students' Personal Information Center SPIC. Internet: <http://wiwi.uni-giessen.de/home/fb02/spic-login/>. Abgerufen am: 21.09.2009.
- JLU JUSTUS-LIEBIG-UNIVERSITÄT GIEßEN (2009d): k-MED. Internet: <http://www.k-med.uni-giessen.de/>. Abgerufen am: 21.09.2009.
- JLU JUSTUS-LIEBIG-UNIVERSITÄT GIEßEN (2009e): Statistik der Studierenden, SS 2009. Internet: [http://www.uni-giessen.de/cms/org/admin/kb/stat/stat\\_publ/studstat\\_jlu\\_intern/aktuell-1/jlu-studierendenstatistik-gesamt-ss-2009.pdf](http://www.uni-giessen.de/cms/org/admin/kb/stat/stat_publ/studstat_jlu_intern/aktuell-1/jlu-studierendenstatistik-gesamt-ss-2009.pdf). Abgerufen am: 25.09.2009.
- JLU JUSTUS-LIEBIG-UNIVERSITÄT GIEßEN (2009f): Master-Studiengänge. Internet: <http://www.uni-giessen.de/cms/studium/studienangebot/master>. Abgerufen am: 05.10.2009.
- JLU JUSTUS-LIEBIG-UNIVERSITÄT GIEßEN (2009g): Fachbereiche, Zentren, Einrichtungen. Internet: <http://www.uni-giessen.de/cms/fbz>. Abgerufen am: 18.12.2009.
- JOINSON, A. (1999): Social Desirability, Anonymity, and Internet-based Questionnaires. In: Behavior Research Methods, Instruments, and Computers, Volume 31, No. 3. S. 433-438.
- JOINSON, A. N. (2001): Knowing Me, Knowingf You: Reciprocal Self-Disclosure in Internet-Based Surveys. In: CyperPsychology and Behavior, Volume 4, No. 5. S. 587-591.
- KÄHLER, W.-M. (2008): Statistische Datenanalyse: Verfahren verstehen und mit SPSS gekonnt einsetzen. 4., völlig neubearbeitete Auflage 2008. Wiesbaden: Vieweg.
- KAISER, H. F. und J. RICE (1976): Little Jiffy, Mark IV. In: Educational and Psychological Measurement, Volume 34, No. 1 (April 1974). S. 111-117.
- KALTON, G. und C. F. CITRO (1993): Panel Surveys: Adding the Fourth Dimension. In: Survey Methodology, Volume 19, No. 2. S. 205-215. Zitiert in: COUPER, M. P. und E. COUTTS (2006): Online-Befragung: Probleme und Chancen verschiedener Arten von Online-Erhebungen. In: DIEKMANN, A. (Hrsg.): Methoden der Sozialforschung. (Sonderhefte der Kölner Zeitschrift für Soziologie und Sozialpsychologie). Wiesbaden: VS Verlag für Sozialwissenschaften. S. 217-243.

- KAPLOWITZ, M. D., T. D. HADLOCK und R. LEVINE (2004): A Comparison of Web and Mail Survey Response Rates. In: Public Opinion Quarterly, Volume 68, No. 1. S. 94-101.
- KAYA, M. (2007): Verfahren der Datenerhebung. In: ALBERS, S., D. KLAPPER, U. KONRADT, A. WALTER und J. WOLF (Hrsg.): Methodik der empirischen Forschung. 2., überarbeitete und erweiterte Auflage. Wiesbaden: Gabler. S. 49-64.
- KEPPER, G. (2008): Methoden der qualitativen Marktforschung. In: HERRMANN, A., C. HOMBURG und M. KLARMANN (Hrsg.): Handbuch Marktforschung. 3. Auflage. Wiesbaden: Gabler. S. 175-212.
- KERSTING, N. (2008): Beiräte und Kommissionen: Integration von Partikularinteressen. In: KERSTING, N. (Hrsg.): Politische Beteiligung: Einführung in dialogorientierte Instrumente politischer und gesellschaftlicher Partizipation. Wiesbaden: VS Verlag für Sozialwissenschaften. S. 107-122.
- KEUSCH, F. (2007): Methodenvergleich Online-Offline. Vortrag im Rahmen des Expertenforums Onlineforschung am 20.06.2007 in Wien. Internet: [http://www.expertenforum.at/files/keusch\\_charts\\_2008\\_06.pdf](http://www.expertenforum.at/files/keusch_charts_2008_06.pdf). Abgerufen am: 16.12.2009.
- KHLAVNA, V. (2005): Nationale Identität: Ost- und westdeutsche Studierende im Vergleich. (Diskussionspapiere aus der Fakultät für Sozialwissenschaft, Ruhr-Universität Bochum, Nr. 05-3). Internet: <http://www.sowi.ruhr-uni-bochum.de/mam/content/fakultaet/diskuss/dp05-3.pdf>. Abgerufen am: 02.09.2009.
- KIERNAN, N. E., M. KIERNAN, M. A. OYLER und C. GILLES (2005): Is a Web Survey as Effective as a Mail Survey? A Field Experiment Among Computer Users. In: American Journal of Evaluations, Volume 26, No. 2 (June). S. 245-252.
- KLIMATMÄRKNING (2010): Climate certification for food: Information in English. Internet: <http://www.klimatmarkningen.se/in-english/>. Abgerufen am: 12.02.2010.
- KÖHLER, W., G. SCHACHTEL und P. VOLESKE (2007): Biostatistik: Eine Einführung für Biologen und Agrarwissenschaftler. 4., aktualisierte und erweiterte Auflage. Berlin, Heidelberg: Springer.
- KÖLTRINGER, R. (1997): Richtig fragen heißt besser messen: Optimale Formulierungstechniken für Umfragen. Mannheim: Forschung Raum und Gesellschaft e. V..
- KOMM KOORDINATIONSSTELLE FÜR MULTIMEDIA DER JUSTUS-LIEBIG-UNIVERSITÄT GIEßEN (2009): Limesurvey. Internet: <http://www.uni-giessen.de/hrz/komm/onlinebefragungen/limesurvey.htm>. Abgerufen am: 22.09.2009.

- KOOLWIJK, J. VAN (1969): Unangenehme Fragen: Paradigma für die Reaktionen des Befragten im Interview. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie, Jahrgang 21. S. 864-875.
- KREUTER, F., S. PRESSER und R. TOURANGEAU (2008): Social Desirability Bias in CATI, IVR, and Web Surveys: The Effect of Mode and Question Sensitivity. In: Public Opinion Quarterly, Volume 72, No. 5. S. 847-865.
- KROMREY, H. (2006): Empirische Sozialforschung: Modelle und Methoden der standardisierten Datenerhebung und Datenauswertung. 11., überarbeitete Auflage. Stuttgart: Lucius & Lucius.
- KUß, A. (2007): Marktforschung: Grundlagen der Datenerhebung und Datenanalyse. 2., überarbeitete und erweiterte Auflage. Wiesbaden: Gabler.
- KÜTH, M. (2009): Schriftliche Auskunft per E-Mail von Herrn Küth, Postmaster des Hochschulrechenzentrums der Justus-Liebig-Universität Gießen, vom 28.08.2009.
- KWAK, N. und B. RADLER (2002): A Comparison between Mail and Web Surveys: Response Pattern, Respondent Profile, and Data Quality. In: Journal of Official Statistics, Volume 18, No. 2. S. 257-273.
- LEE, S. (2006): Propensity Score Adjustment as a Weighting Scheme for Volunteer Panel Web Surveys. In: Journal of Official Statistics, Volume 22, No. 2. S. 329-349.
- LIMESURVEY (2009): Features. Internet: <http://www.limesurvey.org/de/ueber-limesurvey/features.html>. Abgerufen am: 22.09.2009.
- LÖFFLER, U. (1999): Ein Kurz-Überblick über die gebräuchlichsten Stichproben-Verfahren in der Marktforschung. In: ARBEITSKREIS DEUTSCHER MARKT- UND SOZIALFORSCHUNGSINSTITUTE ADM und ARBEITSGEMEINSCHAFT MEDIA ANALYSE AG.MA (Hrsg.): Stichproben-Verfahren in der Umfrageforschung: Eine Darstellung für die Praxis. Opladen: Leske und Budrich. S. 17-21.
- LOOSVELDT, G. und N. SONCK (2008): An Evaluation of the Weighting Procedures for an Online Access Panel Survey. In: Survey Research Methods, Volume 2, No. 2. S. 93-105.
- LORD, F. M. und M. R. NOVICK (1968): Statistical Theories of Mental Test Scores. Reading, Menlo Park, London, Don Mills: Addison-Wesley Publishing.

- LOZAR MANFREDA, K., Z. BATAGELI und V. VEHOVAR (2002): Design of Web Survey Questionnaires - Three Basic Experiments. In: Journal of Computer-Mediated Communications, Volume 7, No. 3. Internet: <http://jcmc.indiana.edu/vol7/issue3/vehovar.html>. Abgerufen am: 01.12.2009.
- MAYERHOFER, W. (2006): Die Beobachtung als Instrument der Werbewirkungsmessung. In: STREBINGER, A., W. MAYERHOFER und H. KURZ (Hrsg.): Werbe- und Markenforschung: Meilensteine - State of the Art - Perspektiven. Wiesbaden: Gabler. S. 465-486.
- MICHAELOWA, A. und K. KRAUSE (2000): International Maritime Transport and Climate Policy. In: Intereconomics, Volume 35, No. 3 (May/June 2000). S. 127-136.
- MIGROS (2010): Klimafreundliche Produkte deklarieren. Internet: <http://www.migros.ch/DE/Sortiment/Engagement/Climatop/Seiten/Produktedeklaration.aspx>. Abgerufen am: 15.02.2010.
- MSC MARINE STEWARDSHIP COUNCIL (2010a): Über uns. Internet: <http://www.msc.org/de/uber-den-msc>. Abgerufen am: 13.10.2010.
- MSC MARINE STEWARDSHIP COUNCIL (2010b): MSC-Zertifizierung so funktioniert's. Internet: <http://www.msc.org/de/msc-zertifizieren/msc-zertifizierung>. Abgerufen am: 13.10.2010.
- NUSSER, M. (2008): Internationale Wettbewerbsfähigkeit forschungs- und wissensintensiver Unternehmen. In: Wirtschaftsdienst, Volume 88, Nr. 9 (September). S. 594-603.
- OECD ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (2010): About OECD. Internet: [http://www.oecd.org/pages/0,3417,en\\_36734052\\_36734103\\_1\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/pages/0,3417,en_36734052_36734103_1_1_1_1_1,00.html). Abgerufen am: 01.03.2010.
- PCF PRODUCT CARBON FOOTPRINT PILOTPROJEKT DEUTSCHLAND (2009): Product Carbon Footprinting - Ein geeigneter Weg zu klimaverträglichen Produkten und deren Konsum?: Erfahrungen, Erkenntnisse und Empfehlungen aus dem Product Carbon Footprint Pilotprojekt Deutschland. Internet: [http://www.pcf-projekt.de/files/1241099725/ergebnisbericht\\_2009.pdf](http://www.pcf-projekt.de/files/1241099725/ergebnisbericht_2009.pdf). Abgerufen am: 03.10.2009.
- PEPELS, W. (2008): Computergestützte Befragung. In: Pepels, W. (Hrsg.): Marktforschung: Verfahren, Datenauswertung, Ergebnisdarstellung. 2., überarbeitete und erweiterte Auflage. Düsseldorf: Symposion Publishing. S. 189-202.

- PETERSEN, T. (2008): Split Ballots as an Experimental Approach to Public Opinion Research. In: DONSBACH, W. und M. W. TRAUGOTT (Hrsg.): The SAGE Handbook of Public Opinion Research. London, New Delhi, Singapur, Thousand Oaks: Sage Publications. S. 322-330.
- PISTORIUS, T., J. ZELL und C. HARTEBRODT (2006): Untersuchungen zur Rolle des Waldes und der Forstwirtschaft im Kohlenstoffhaushalt des Landes Baden-Württemberg. Internet: <http://www.bwplus.fzk.de/berichte/SBer/ZO3K23004SBer.pdf>. Abgerufen am: 17.01.2010.
- POHL, S., R. STEYER und K. KRAUS (2008): Modelling Method Effects as Individual Causal Effects. In: Journal of Royal Statistics - Series A: Statistics in Society, Volume 171, Part 1. S. 41-63.
- PÖTSCHKE, M. (2009): Potentiale von Online-Befragungen: Erfahrungen aus der Hochschulforschung. In: JAKOB, N., H. SCHOEN und T. ZERBACK (Hrsg.): Sozialforschung im Internet: Methodologie und Praxis der Online-Befragung. Wiesbaden: VS Verlag für Sozialwissenschaften. S. 75-89.
- RAITHEL, J. (2008): Quantitative Forschung: Ein Praxiskurs. 2., durchgesehene Auflage. Wiesbaden: VS Verlag für Sozialwissenschaften.
- REIPS, U.-D. und L. FRANEK (2004): Mitarbeiterbefragungen per Internet oder Papier? Der Einfluss von Anonymität, Freiwilligkeit und Alter auf das Antwortverhalten. In: Wirtschaftspsychologie, Heft 1. S. 67-83.
- RÖSCH, G. (1994): Kriterien der Gewichtung einer nationalen Bevölkerungstichprobe. In: GABLER, S., J. H. P. HOFFMEYER-ZLOTNIK und D. KREBS (Hrsg.): Gewichtung in der Umfragepraxis. Opladen: Westdeutscher Verlag. S. 7-26.
- ROSENTHAL, D. (1998): Infopool Internet: Methoden, Tricks und Quellen der Profils zur effizienten Recherche. Zürich: Orell Füssli.
- ROWE, G., W. POORTINGA und N. PIDGEON (2006): A Comparison of Responses to Internet and Postal Surveys in a Public Engagement Context. In: Science Communication, Volume 27, No. 3 (March 2006). S. 352-375.
- RUBIN, D. B. und N. THOMAS (1996): Matching Using Estimated Propensity Scores: Relating Theory to Practice. In: Biometrics, Volume 52, Issue 1 (March). S. 249-264.
- SACHS, L. und J. HEDDERICH (2006): Angewandte Statistik: Methodensammlung mit R. 12., vollständig neu bearbeitete Auflage. Berlin, Heidelberg, New York: Springer.

- SAUER, C., K. AUSPURG, T. HINZ, S. LIEBIG und J. SCHUPP (2009): Die Bewertung von Erwerbseinkommen: Methodische und inhaltliche Analysen zu einer Vignettenstudie im Rahmen des SOEP-Pretest 2008. (Data Documentation, Band 42).
- SCHAEFER, D. R und D. A. DILLMAN (1998): Development of a standard E-Mail Methodology: Results of an Experiment. In: Public Opinion Quarterly, Volume 62. S. 378-397.
- SCHEUCH, E. K. (1973): Das Interview in der Sozialforschung. In: KÖNIG, R. (Hrsg.): Handbuch der empirischen Sozialforschung, Band 2. 3. Auflage. Stuttgart: Enke. S. 66-190.
- SCHLICH, M., D. THOMAE und E. H. SCHLICH (2009a): Public Perception of Carbon Footprint(s) as additional Food Label. Posterpräsentation auf der Life Cycle Assessment XI, 29.11.2009-02.10.2009 in Boston. Internet: [http://lcacenter.org/LCA9/presentations/Schlich\\_Poster\\_Boston\\_2009.pdf](http://lcacenter.org/LCA9/presentations/Schlich_Poster_Boston_2009.pdf). Abgerufen am: 15.02.2010.
- SCHLICH, E., B. HARDTERT und F. KRAUSE (2009b): Rindfleisch aus Sicht der Ecology of Scale: Rindfleisch aus regionaler und globaler Herkunft: Von der Primärproduktion bis zum PoS - Ein Vergleich vollständiger Prozessketten hinsichtlich Energieumsatz und Kohlenstoffdioxidemissionen. In: Fleischwirtschaft, Nr. 9. S. 114-118.
- SCHNEEKLOTH, U. und I. LEVEN (2003): Woran bemisst sich eine „gute“ allgemeine Bevölkerungsumfrage? Analysen zu Ausmaß, Bedeutung und zu den Hintergründen von Non-response in zufallsbasierten Stichprobenerhebungen am Beispiel des ALLBUS. In: ZUMA-Nachrichten, Jahrgang 27, Heft 53. S. 16–57.
- SCHNELL, R., P. B. HILL und E. ESSER (2005): Methoden der empirischen Sozialforschung. 7., völlig überarbeitete und erweiterte Auflage. München, Wien: Oldenbourg.
- SCHOEN, H. (2004): Online-Umfragen-schnell, billig, aber auch valide? Ein Vergleich zweier Internetbefragungen mit persönlichen Interviews zur Bundestagswahl 2002. In: ZA-Information, Heft 54 (Mai). S. 27–52.
- SCHOLL, A. (1993): Die Befragung als Kommunikationssituation: Zur Reaktivität im Forschungsinterview. (Studien zur Sozialwissenschaft, Band 109). Opladen: Westdeutscher Verlag.
- SCHOLL, N., S. MULDER und R. DRENT (2002): On-line Qualitative Market Research: Interviewing the World at a Fingertip. In: Qualitative Market Research: An International Journal, Volume 5, Number 3. S. 210-223.

- SCHONLAU, M., A. VAN SOEST, A. KAPTEYN und M. COUPER (2009): Selection Bias in Web Surveys and the Use of Propensity Scores. In: *Sociological Methods and Research*, Volume 37, No. 3 (February 2009). S. 291-318.
- SCHÖNWIESE, C.-D. (1996): Anthropogene Verstärkung des Treibhauseffektes. In: *Geographie und Schule*, Nr. 101. S. 14-24.
- SCHULDT, B. A. und J. W. TOTTEN (1994): Electronic Mail vs. Mail Survey Response Rates. In: *Marketing Research*, Volume 6, No. 1. S. 36-39.
- SCHWARZ, N., F. STRACK, H.-J. HIPPLER und G. BISHOP (1991): The Impact of Administration Mode on Response Effects in Survey Measurement. In: *Applied Cognitive Psychology*, Volume 5. S. 193-212.
- SCHWARZ, S. und U.-D. REIPS (2001): CGI versus JavaScript - A Web Experiment on the Reverse Hindsight. In: REIPS, U.-D. und M. BOSNJAK (Hrsg.): *Dimensions of Internet Science*. Lengerich: Pabst Science Publishers. S. 75-90.
- SHIH, T.-H. und X. FAN (2007): Response Rates and Mode Preferences in Web-Mail Mixed-Mode Surveys: A Meta-Analysis. In: *International Journal of Internet Science*, Volume 2, No. 1. S. 59-82.
- SINGER, E., M. R. FRANKEL und M. B. GLASSMAN (1983): The Effect of Interviewer Characteristics and Expectations on Response. In: *Public Opinion Quarterly*, Volume 47, No. 1 (Spring). S. 68-83.
- SINN, H. W. (2008): Das grüne Paradoxon: Warum man das Angebot bei der Klimapolitik nicht vergessen darf. (Ifo Working Paper, No. 54). Internet: [http://www.cesifo-group.de/pls/guest/download/Ifo%20Working%20Papers%20\(seit%202005\)/IfoWorkingPaper-54.pdf](http://www.cesifo-group.de/pls/guest/download/Ifo%20Working%20Papers%20(seit%202005)/IfoWorkingPaper-54.pdf). Abgerufen am: 11.01.2009.
- SPSS (2009): Product Naming Guide. Internet: <http://www.spss.com/de/software/product-name-guide/>. Abgerufen am: 23.09.2009.
- STATISTISCHES BUNDESAMT (2009): *Statistisches Jahrbuch 2009 für die Bundesrepublik Deutschland*. Wiesbaden: Statistisches Bundesamt.
- STERN, N. (2008): *The Economics of Climate Change: The Stern Review on the Economics of Climate Change*. Fifth Printing. Cambridge: Cambridge University Press.
- STEVENS, S. S. (1959): Measurement, Psychophysics, and Utility. In: CHURCHMAN, C. W. und P. RATOOSH (Hrsg.): *Measurement: Definitions and Theories*. New York: Wiley. S. 18-63.

- SUDMAN, S. und N. M. BRADBURN (1982): Asking Questions: A Practical Guide to Questionnaire Design. San Francisco, Washington, London: Jossey-Bass Publishers.
- SUE, V. M. und L. A. RITTER (2007): Conducting Online Surveys. Los Angeles, London, New Delhi, Singapore: Sage Publications.
- SVENSKT SIGILL (2010): Who's behind. Internet: <http://www.svensksigill.se/website2/1.0.2.0/446/1/>. Abgerufen am: 12.02.2010.
- TADDICKEN, M. (2008): Methodeneffekte bei Web-Befragungen: Einschränkungen der Datengüte durch ein >reduziertes Kommunikationsmedium<? (Neue Schriften zur Online-Forschung, Band 5). Köln: Halem. Zugleich: Dissertation, Universität Stuttgart-Hohenheim.
- TADDICKEN, M. (2009): Methodeneffekte von Web-Befragungen: Soziale Erwünschtheit vs. Soziale Entkontextualisierung. In: WEICHBOLD, M., J. BACHER und C. WOLF (Hrsg.): Umfrageforschung: Herausforderungen und Grenzen. (Österreichische Zeitschrift für Soziologie, Sonderheft 9/2009). Wiesbaden: VS Verlag für Sozialwissenschaften. S. 85-103.
- TESCO (2009): Carbon Labelling. Internet: [http://www.tescopl.com/plc/corporate\\_responsibility\\_09/environment/climate\\_change/leading\\_a\\_revolution/carbon\\_labelling](http://www.tescopl.com/plc/corporate_responsibility_09/environment/climate_change/leading_a_revolution/carbon_labelling). Abgerufen am: 29.12.2009.
- THEOBALD, A. (2007): Zur Gestaltung von Online-Fragebögen. In: WELKER, M. und O. WENZEL (Hrsg.): Online-Forschung 2007: Grundlagen und Fallstudien. Köln: Herbert von Halem. S. 103-118.
- TOEPOEL, V., M. DAS und A. van SOEST (2009): Relating Question Type to Panel Conditioning: Comparing Trained and Fresh Respondents. In: Survey Research Methods, Volume 3, No. 2. S. 73-80.
- TOUTENBERG, H. und C. HEUMANN (2008): Deskriptive Statistik: Eine Einführung in Methoden und Anwendungen mit R und SPSS. 6., aktualisierte und erweiterte Auflage. Berlin, Heidelberg: Springer.
- TUTEN, T. L (1997): Getting a Foot in the Electronic Door: Understanding why People read or delete Electronic Mail. (ZUMA-Arbeitsbericht, Nr. 97/08). Internet: [http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/zuma\\_arbeitsberichte/97\\_08.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/zuma_arbeitsberichte/97_08.pdf). Abgerufen am: 08.11.2009.

- VEHOVAR, V., K. LOZAR MANFREDA und G. KOREN (2008): Internet Surveys. In: DONSBACH, W. und M. W. TRAUGOTT (Hrsg.): The SAGE Handbook of Public Opinion Research. London, New Delhi, Singapur, Thousand Oaks: Sage Publications. S. 271-283.
- WACKERNAGEL, M. und W. E. REES (1997): Unser ökologischer Fußabdruck: Wie der Mensch Einfluss auf die Umwelt nimmt. Basel, Boston, Berlin: Birkhäuser.
- WEBHITS (2010): Web-Barometer. Internet: <http://www.webhits.de/deutsch/index.shtml?deutsch/webstats.html>. Abgerufen am: 14.01.2010.
- WEIBLE, R. und J. WALLACE (1998): Cyber Research: The Impact of the Internet on Data Collection. In: Marketing Research, Volume 10, Issue 3 (Fall). S. 18-24.
- WELKER, M., A. WERNER und J. SCHOLZ (2005): Online-Research: Markt- und Sozialforschung mit dem Internet. Heidelberg: dpunkt.verlag.
- WELKER, M. und U. MATZAT (2009): Online-Forschung: Gegenstände, Entwicklung, Institutionalisierung und Ausdifferenzierung eines neuen Forschungszweiges. In: JAKOB, N., H. SCHOEN und T. ZERBACK (Hrsg.): Sozialforschung im Internet: Methodologie und Praxis der Online-Befragung. Wiesbaden: VS Verlag für Sozialwissenschaften. S. 33-47.
- WIEDMANN, T. und J. MINX (2007): A Definitino of ‚Carbon Footprint‘. (ISA<sup>UK</sup> Research Report 07-01). Internet: [http://www.censa.org.uk/docs/ISA-UK\\_Report\\_07-01\\_carbon\\_footprint.pdf](http://www.censa.org.uk/docs/ISA-UK_Report_07-01_carbon_footprint.pdf). Abgerufen am: 15.11.2009.
- WIEGAND, E. (2000): Chancen und Risiken neuer Erhebungstechniken in der Umfrageforschung. In: STATISTISCHES BUNDESAMT (Hrsg.): Neue Erhebungsinstrumente und Methodeneffekte. (Spektrum Bundesstatistik, Band 15). Stuttgart: Metzler-Poeschel. S. 12-21.
- WORLD BANK (2008): Development and Climate Change: A Strategic Framework for the World Bank Group. Internet: <http://siteresources.worldbank.org/EXTCC/Resources/407863-1219339233881/DevelopmentandClimateChange.pdf>. Abgerufen am: 10.02.2010.
- WTO WORLD TRADE ORGANIZATION und UNEP UNITED NATIONS ENVIRONMENT PROGRAMME (2009): Trade and Climate Change. Genf: WTO Publications.

ZERBACK, T., H. SCHOEN, N. JACKOB und S. SCHLERETH (2009): Zehn Jahre Sozialforschung mit dem Internet: Eine Analyse zur Nutzung von Online-Befragungen in den Sozialwissenschaften. In: JACKOB, N., H. SCHOEN und T. ZERBACK (Hrsg.): Sozialforschung im Internet: Methodologie und Praxis der Online-Befragung. Wiesbaden: VS Verlag für Sozialwissenschaften. S. 15-31.

## Anhang

Fragebogen zur Thematik Carbon Labelling.....	165
Verwendete Skalen sowie Labels.....	179
Tabelle A 1: Kolmogorov-Smirnov-Test auf Normalverteilung der Variablen "Alter" in der Gesamtstichprobe.....	182
Tabelle A 2: Kolmogorov-Smirnov-Test auf Normalverteilung der Variablen "Alter" in den beiden Teilstichproben.....	182
Tabelle A 3: Kolmogorov-Smirnov Z-Test auf unterschiedliche Verteilungen der Variablen "Alter" in den beiden Teilstichproben „F2F“ sowie „Online“.....	182
Tabelle A 4: Genaue Bezeichnungen der Fachbereiche (Legende zu den Abkürzungen der Tabelle 7).....	182
Tabelle A 5: Ergebnisse des Mann-Whitney U-Tests zur klassifizierten Einkommensvariablen (ungewichtet).....	183
Tabelle A 6: Kolmogorov-Smirnov-Test auf Normalverteilung der Variablen "Interviewdauer" in der Gesamtstichprobe.....	183
Tabelle A 7: Ergebnisse des Mann-Whitney U-Tests zur Interviewdauer.....	183
Tabelle A 8: Ergebnisse der Kontingenzanalyse zur Ermittlung des Zusammenhangs zwischen der Befragungsmethode und demographischen Merkmalen.....	183
Tabelle A 9: Mann-Whitney U-Test zur Altersvariablen.....	183
Tabelle A 10: Ergebnisse des Mann-Whitney U-Tests zur klassifizierten Einkommensvariablen (gewichtet).....	183
Tabelle A 11: Ergebnisse des Kolmogorov-Smirnov-Tests zur Überprüfung der Items der Fragestellung zur Relevanz der Stärkung verschiedener politischer Aufgabenbereiche auf Normalverteilung.....	184
Tabelle A 12: Ergebnisse des Mann-Whitney U-Tests zu den Items der Fragestellung zur Relevanz der Stärkung verschiedener politischer Aufgabenbereiche.....	184
Tabelle A 13: Ergebnisse des Mann-Whitney U-Tests zu den Items der Einstellungsfrage zum Themenkomplex „Klimaschutz“.....	184
Tabelle A 14: Ergebnisse des Mann-Whitney U-Tests zu den Items der Verhaltensfrage zum Themenkomplex „Klimaschutz“. Genaue Fragestellung: „Wie oft hast Du bereits eine der folgenden Maßnahmen unternommen?“.....	185
Tabelle A 15: Ergebnisse der Kontingenzanalyse zur Ermittlung des Zusammenhangs zwischen dem Bezug von „Grünem“ Strom und der Befragungsmethode.....	185

Tabelle A 16: Ergebnisse der Kontingenzanalyse zur Ermittlung des Zusammenhangs zwischen der Beantwortung der einzelnen Quiz-Items und der Befragungsmethode.....	185
Tabelle A 17: Ergebnisse des Kolmogorov-Smirnov-Tests zur Überprüfung der korrekt gegebenen Antworten des Klimaschutz-Quiz auf Normalverteilung.....	185
Tabelle A 18: Ergebnisse des Mann-Whitney U-Tests zu der Anzahl der korrekt gegebenen Antworten des Klimaschutz-Quiz.....	186
Tabelle A 19: Ergebnisse des Mann-Whitney U-Tests für die verschiedenen Zeichen. Genauer Wortlaut der Fragestellung: „Hast Du dieses Zeichen (bzw. Siegel) schon mal gesehen? Kennst Du diese Zeichen?“ .....	186
Tabelle A 20: Ergebnisse des Mann-Whitney U-Tests zu den korrekt gegebenen Teilantworten hinsichtlich des Product Carbon Footprints (PCF).....	186
Tabelle A 21: Ergebnisse des Mann-Whitney U-Tests zum semantischen Differentials des Product Carbon Footprints. Genauer Wortlaut der Fragestellung „Hältst Du ein solches ‚CO2-Zeichen‘ für...“, gefolgt von den Eigenschaftspaaren .....	186
Tabelle A 22: Ergebnisse der Kontingenzanalyse zur Ermittlung des Zusammenhangs zwischen der Frage nach der rechtlichen Regelung eines solchen PCFs und der Befragungsmethode.....	187
Tabelle A 23: Ergebnisse der Kontingenzanalyse zur Ermittlung des Zusammenhangs zwischen der Frage nach Ausgestaltung eines solchen PCFs und der Befragungsmethode .....	187
Tabelle A 24: Ergebnisse des Kolmogorov-Smirnov-Tests zur Überprüfung der gebildeten Variablen „Zahlungsbereitschaftsdifferenz“ auf Normalverteilung (Gesamtstichprobe).....	187
Tabelle A 25: Ergebnisse des Kolmogorov-Smirnov-Tests zur Überprüfung der gebildeten Variablen „Zahlungsbereitschaftsdifferenz“ auf Normalverteilung (Teilstichproben) .....	187
Tabelle A 26: Ergebnisse des Mann-Whitney U-Tests zu der Variablen „Zahlungsbereitschaftsdifferenz“.....	187
Tabelle A 27: Erklärte Gesamtvarianz durch die Faktorenanalyse .....	188
Tabelle A 28: Ergebnisse des t-Tests zum Mittelwertvergleich der Einstellungsdimensionen aus der Faktorenanalyse (F2F n=176; Online n=477) .....	188

## Fragebogen zur Thematik Carbon Labelling

Justus-Liebig-Universität Gießen  
 Institut für Agrarpolitik und Marktforschung  
 Professur für Marktlehre der Agrar- und Ernährungswirtschaft

Interviewer:			1
Ort:	Datum:	Uhrzeit (Start):	

Hallo!

Wir studieren hier auch an der Uni Gießen am Fachbereich 09 und führen im Rahmen einer Masterarbeit am Institut für Agrarpolitik und Marktforschung eine Befragung zum Thema **Klimaschutz und Lebensmittel** durch.

Wärst Du bereit, Dir etwa 10-15 Minuten Zeit zu nehmen, um mir ein paar Fragen zu beantworten?

(Deine Meinung ist uns sehr wichtig. Außerdem werden Deine Daten streng vertraulich behandelt und es kann kein Zusammenhang zu Deiner Person hergestellt werden.)

Zusätzlich verlosen wir unter allen Teilnehmern zwei Amazon-Gutscheine in Höhe von 10 Euro!

### **FILTERFRAGE:**

#### **Darf ich fragen, ob Du an der Uni Gießen Student bist?**

Wenn **ja**: Interview fortsetzen,  
 wenn **nein**: Interview abbrechen.

## TEIL 1

### 1.1

**Ich lese Dir nun verschiedene politische Aufgabenbereiche in Deutschland vor.**

**Bitte sage mir anhand der Skala, wie wichtig Dir persönlich eine zusätzliche Stärkung dieser Aufgabenbereiche jeweils ist.**

**Eine 1 bedeutet sehr wichtig, eine 5 bedeutet unwichtig, dazwischen kannst Du Deine Meinung abstimmen.**

*Bitte Skala 1 übergeben und erklären; Items vor Beantwortung erst einmal insgesamt, dann zur Beantwortung jeweils nochmal einzeln vorlesen; je Aufgabe nur eine Nennung möglich! Pro Interview Startpunkt wechseln!*

Start		1	2	3	4	5	
	Bekämpfung der Arbeitslosigkeit						A1_1
	Schutz von Tier- und Pflanzenarten						A1_2
	Verbesserung des Zusammenlebens mit Ausländern						A1_3
	Maßnahmen zum Klimaschutz						A1_4
	Bekämpfung von Kriminalität						A1_5
	Förderung der Wirtschaft						A1_6



**1.4 Bezieht Du momentan „Grünen“ Strom oder hast Du schon mal „Grünen“ Strom bezogen? (A4)**

Ja, momentan.	<input type="checkbox"/>	1
Ja, habe ich schon, momentan aber nicht mehr.	<input type="checkbox"/>	2
Nein.	<input type="checkbox"/>	0
Weiß nicht.	<input type="checkbox"/>	9

**TEIL 2**

**2.1 Im Folgenden würde ich mit Dir gerne ein kleines Quiz rund um das Thema Klimaschutz machen. Da wir uns hier nicht in einer Prüfungssituation befinden, kannst Du ganz entspannt bleiben! Bitte sage mir, ob Du diese Aussagen eher für richtig oder eher für falsch hältst.**

*Pro Interview Startpunkt wechseln; Maßnahmen einzeln vorlesen!*

Start		Richtig (1)	Falsch (0)	weiß nicht (9)	
	CO <sub>2</sub> , auch Kohlendioxid genannt, ist ein Gas, das bei der Verbrennung von Erdöl, Erdgas, Benzin, Kohle oder auch Holz entsteht.				B1_1
	Wenn CO <sub>2</sub> in die Atmosphäre gelangt, wirkt es dort wie eine Isolierung. Dadurch steigt die Durchschnittstemperatur an und das Klima verändert sich.				B1_2
	Es gibt keinen von Menschen verursachten Klimawandel.				B1_3
	Der Ausstoß an Treibhausgasen ist bei der Erzeugung von tierischen Lebensmitteln wie Fleisch, Milch und Eiern wesentlich höher als bei pflanzlichen Erzeugnissen.				B1_4
	Die menschliche Ernährung trägt nur minimal zum Klimawandel bei.				B1_5

## 2.2

**Hast Du diese Zeichen (bzw. Siegel) schon mal gesehen? Kennst Du diese Zeichen?**

Dem Befragten das Extrablatt mit den Zeichen zeigen und Skala 4 übergeben. Nicht den Namen, sondern die Nummer nennen.

		Kenne ich gut <b>1</b>	Schon mal gesehen, Vorstellung, was es bedeutet <b>2</b>	Schon mal gesehen, keine Vorstellung, was es bedeutet <b>3</b>	Noch nie gesehen <b>4</b>	
Zeichen 1:						B2_1
Zeichen 2:						B2_2
Zeichen 3:						B2_3
Zeichen 4:						B2_4
Zeichen 5:						B2_5

**2.3**

**Ich möchte nun noch genauer über das Zeichen 5 sprechen. Was könnte es Deiner Meinung nach bedeuten bzw. was bedeutet es? Wichtig: Hier gibt es keine richtigen oder falschen Antworten - es geht uns rein um Deine persönliche Meinung!**

*Label auf Extrablatt nochmals zeigen. Alle Antworten detailliert aufschreiben! (B3)*

---

---

---

---

*Erklärung:*

**Das Zeichen gibt an, welche Menge an klimaschädigenden Emissionen (zum Beispiel CO<sub>2</sub>) im gesamten Lebenszyklus des Produktes, also bei Herstellung, Verarbeitung, Verpackung, Transport etc. freigesetzt werden.**

**Man nennt dies den CO<sub>2</sub>-Fußabdruck oder „Carbon Footprint“ des Produktes.**

**Die Mengenangabe (hier: 125g) bezieht sich auf 100g Produktmasse.**

**Ich nenne dieses Zeichen im weiteren Verlauf des Interviews „CO<sub>2</sub>-Zeichen“.**

## 2.4

**Hältst Du ein solches „CO<sub>2</sub>-Zeichen“ für...**

*Pro Interview Startpunkt wechseln! Befragten in Fragebogen schauen lassen. Jedes Item einzeln vorlesen!*

Start		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>		<b>Weiß nicht (9)</b>	
	<b>notwendig</b>						<b>überflüssig</b>		B4_1
	<b>effektiv</b>						<b>ineffektiv</b>		B4_2
	<b>nicht umsetzbar</b>						<b>umsetzbar</b>		B4_3
	<b>verständlich</b>						<b>unverständlich</b>		B4_4
	<b>kompliziert</b>						<b>einfach</b>		B4_5

## 2.5

Welche Regelungen für dieses „CO<sub>2</sub>-Zeichen“ würdest Du bevorzugen?

a)

Ein solches „CO<sub>2</sub>-Zeichen“ sollte für alle Lebensmittel... (B5\_1)

...freiwillig sein.	<input type="checkbox"/>	1
...gesetzlich vorgeschrieben sein.	<input type="checkbox"/>	2
weiß nicht	<input type="checkbox"/>	9

b)

Ein solches „CO<sub>2</sub>-Zeichen“ sollte für alle Lebensmittel... (B5\_2)

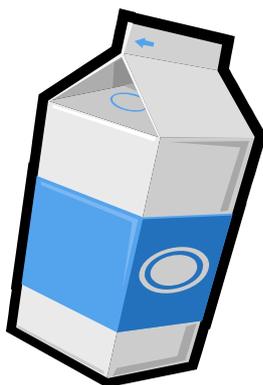
...eine konkrete Mengenangabe über den CO <sub>2</sub> -Ausstoß je Produktmenge enthalten.	<input type="checkbox"/>	1
...ohne konkrete Mengenangabe jeweils als roter, gelber oder grüner Fußabdruck eine <u>schlechte</u> , <u>durchschnittliche</u> oder <u>gute</u> CO <sub>2</sub> -Bilanz symbolisieren.	<input type="checkbox"/>	2
weiß nicht	<input type="checkbox"/>	9

## 2.6

Nun geht es konkret um ein Lebensmittel:

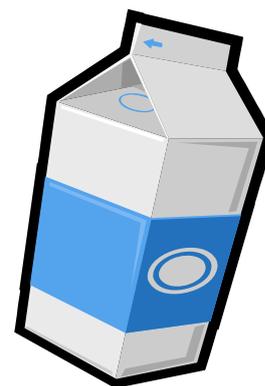
*Befragten in dieser Frage wieder mit reinschauen lassen!*

Wie viel zahlst Du normalerweise für  
1 Liter Milch?



\_\_\_\_\_ €/Liter (B6\_1)

Wie viel wärest Du maximal bereit für  
1 Liter Milch zu bezahlen, bei dem in der  
Produktion 30% weniger  
CO<sub>2</sub>-Emissionen freigesetzt werden?



-30% CO<sub>2</sub>

\_\_\_\_\_ €/Liter (B6\_2)

**TEIL 3: Zum Abschluss noch einige wenige Fragen zur Statistik:****3.1 Bist Du in Deinem Haushalt für die Lebensmittel-Einkäufe zuständig? (C1)**

Ja, ausschließlich.	<input type="checkbox"/>	<b>1</b>
Ja, überwiegend.	<input type="checkbox"/>	<b>2</b>
Ja, teilweise.	<input type="checkbox"/>	<b>3</b>
Nein.	<input type="checkbox"/>	<b>0</b>

**3.2 Wo kaufst Du Lebensmittel ein?**

		<b>überwiegend</b> <b>(1)</b>	<b>gelegentlich</b> <b>(2)</b>	<b>selten</b> <b>(3)</b>	<b>gar nicht</b> <b>(4)</b>
<b>C2_1</b>	Supermarkt (REWE, Edeka, real, ...)				
<b>C2_2</b>	Discounter (Aldi, Lidl, Penny, ...)				
<b>C2_3</b>	Wochenmarkt				
<b>C2_4</b>	Bioladen/Reformhaus				
<b>C2_5</b>	Fachgeschäft (Bäcker, Metzger, ...)				
<b>C2_6</b>	Sonstiges:				

**TEIL 4****4.1 Geschlecht (D1)**

männlich	<input type="checkbox"/>	(0)
weiblich	<input type="checkbox"/>	(1)

**4.2 Wie alt bist Du? (D2)**

\_\_\_\_\_ Jahre

**4.3 An welchem Fachbereich studierst Du hauptsächlich? (D3)**

(FB 01) Rechtswissenschaft	<input type="checkbox"/>	1
(FB 02) Wirtschaftswissenschaften	<input type="checkbox"/>	2
(FB 03) Sozial- und Kulturwissenschaften (inkl. LEHRAMT)	<input type="checkbox"/>	3
(FB 04) Geschichts- und Kulturwissenschaften	<input type="checkbox"/>	4
(FB 05) Sprache, Literatur, Kultur	<input type="checkbox"/>	5
(FB 06) Psychologie und Sportwissenschaft	<input type="checkbox"/>	6
(FB 07) Mathematik und Informatik, Physik, Geographie	<input type="checkbox"/>	7
(FB 08) Biologie und Chemie	<input type="checkbox"/>	8
(FB 09) Agrarwissenschaften, Ökotoxikologie und Umweltmanagement	<input type="checkbox"/>	9
(FB 10) Veterinärmedizin	<input type="checkbox"/>	10
(FB 11) Medizin	<input type="checkbox"/>	11

**4.4 Was ist der Abschluss Deines momentanen Studiengangs (D4)**

Bachelor	<input type="checkbox"/>	1
Master	<input type="checkbox"/>	2
Lehramt	<input type="checkbox"/>	3
Staatsexamen (Jura, Medizin, Vetmed)	<input type="checkbox"/>	4
Diplom	<input type="checkbox"/>	5
Magister	<input type="checkbox"/>	6
sonstige: _____	<input type="checkbox"/>	7

**4.5 Bist Du ehrenamtlich engagiert?**

Nein	<input type="checkbox"/>	D5_1
Ja, an der Uni	<input type="checkbox"/>	D5_2
Ja, außerhalb der Uni	<input type="checkbox"/>	D5_3

**4.6 Bitte ordne die Summe aller Deiner monatlichen Nettoeinkünfte in eine der Kategorien ein: (D6)**

unter 400 €	<input type="checkbox"/>	1
400 - 600 €	<input type="checkbox"/>	2
601 - 800 €	<input type="checkbox"/>	3
801 - 1200 €	<input type="checkbox"/>	4
1200 - 2500 €	<input type="checkbox"/>	5
über 2500 €	<input type="checkbox"/>	6
keine Angabe	<input type="checkbox"/>	9

Wenn Du am Gewinnspiel für den Amazon-Gutschein teilnehmen willst, gebe bitte hier Deine E-Mail-Adresse an:

---

Vor Auswertung der Fragebögen wird dieser Teil mit Deiner E-Mail-Adresse entfernt, entsprechend bleibt alles anonym. Da die Stichprobe circa 100 Befragte umfassen soll, beträgt deine Gewinnchance bei zwei Gutscheinen á 10 Euro 1/50. Viel Erfolg!

Ein wichtiger Hinweis noch: Eine identische Umfrage zu der Thematik „Klimaschutz und Lebensmittel“ wird in der nächsten Zeit auch online noch einmal durchgeführt werden. Für den Erfolg der Untersuchung und der Masterarbeit wäre es notwendig, dass Du an dieser Online-Befragung NICHT teilnimmst!

Vielen Dank für Deine Mitarbeit.

*Schokolade überreichen.*

Tschüß!

**TEIL 5 (Erst nach Beendigung des Interviews ausfüllen!!!)****5.1 Ende Interview (Uhrzeit) (E1)**

--

**5.2 Der Interviewte hat alleine geantwortet (→Keine Gruppe, die geantwortet hat) (E2)**

Ja	<input type="checkbox"/>	<b>1</b>
Nein	<input type="checkbox"/>	<b>0</b>

**5.3 Gab es Verständnisprobleme im Fragebogen? (E3)**

Ja	<input type="checkbox"/>	<b>1</b>
Nein	<input type="checkbox"/>	<b>0</b>

**5.4 Wenn in 5.3 „ja“ angekreuzt wurde: Bei welchen Fragen gab es Verständnisprobleme? (E4)**

Frage	Item	<input checked="" type="checkbox"/>	Welche Probleme?
1.1	A1_1		
	A1_2		
	A1_3		
	A1_4		
	A1_5		
	A1_6		
1.2	A2_1		
	A2_2		
	A2_3		
	A2_4		
	A2_5		
	A2_6		
	A2_7		
	A2_8		
1.3	A3_1		
	A3_2		
	A3_3		
	A3_4		
	A3_5		
Frage	Item	<input checked="" type="checkbox"/>	Welche Probleme?
1.3	A3_6		
	A3_7		
	A3_8		
2.1	B1_1		
	B1_2		
	B1_3		
	B1_4		
	B1_5		
2.2	B2_1		
	B2_2		
	B2_3		
	B2_4		
	B2_5		
2.3	B3		
2.4	B4_1		
	B4_2		
	B4_3		

	B4_4		
	B4_5		
2.5 a)	B5_1		
2.5 b)	B5_2		
2.6	B6_1		
	B6_2		
3.1	C1		
3.2	C2_1		
	C2_2		
	C2_3		
	C2_4		
	C2_5		
	C2_6		
4.1	D1		
4.2	D2		
4.3	D3		
4.4	D4		
4.5	D5_1		
	D5_2		
	D5_3		
4.6	D6		

## Verwendete Skalen sowie Labels

### Skala 1

sehr wichtig	wichtig	etwas wichtig	weniger wichtig	unwichtig
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>

### Skala 2

stimme vollkommen zu	stimme eher zu	bin unentschieden	lehne eher ab	lehne vollkommen ab
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>

### Skala 3

Tue ich oft	Tue ich manchmal	Tue ich eher selten	Bislang nicht, aber vielleicht in Zukunft	Tue ich nicht und habe es auch nicht vor
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>

### Skala 4

Kenne ich gut	Schon mal gesehen; Vorstellung, was es bedeutet	Schon mal gesehen; <u>keine</u> Vorstellung, was es bedeutet	Noch nie gesehen
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>

## Zeichen 1:



## Zeichen 3:



## Zeichen 2:



## Zeichen 4:



## Zeichen 5:



## Tabellen im Anhang

**Tabelle A 1: Kolmogorov-Smirnov-Test auf Normalverteilung der Variablen "Alter" in der Gesamtstichprobe**

Variable	Kolmogorov-Smirnov <sup>a</sup>		
	Statistik	df	Signifikanz
Alter (n=697)	0,179	6971	0,000

a. Signifikanzkorrektur nach Lilliefors

Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 2: Kolmogorov-Smirnov-Test auf Normalverteilung der Variablen "Alter" in den beiden Teilstichproben**

Variable „Alter“ in der Teilstichprobe	Kolmogorov-Smirnov <sup>a</sup>		
	Statistik	df	Signifikanz
Face-to-Face (n=186)	0,188	186	0,000
Online (n=511)	0,186	511	0,000

a. Signifikanzkorrektur nach Lilliefors

Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 3: Kolmogorov-Smirnov Z-Test auf unterschiedliche Verteilungen der Variablen "Alter" in den beiden Teilstichproben „F2F“ sowie „Online“**

	Alter <sup>a</sup>
Extremste Differenzen Absolut	0,107
Positiv	0,021
Negativ	-0,107
Kolmogorov-Smirnov-Z	1,246
Asymptotische Signifikanz (2-seitig)	0,089
Exakte Signifikanz (2-seitig)	0,028
Punkt-Wahrscheinlichkeit	0,002

a. Gruppenvariable: Befragungssample

Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 4: Genaue Bezeichnungen der Fachbereiche (Legende zu den Abkürzungen der Tabelle 7)**

Abkürzung	Genaue Fachbereichsbezeichnung
FB 01	Rechtswissenschaft
FB 02	Wirtschaftswissenschaften
FB 03	Sozial- und Kulturwissenschaften
FB 04	Geschichts- und Kulturwissenschaften
FB 05	Sprache, Literatur, Kultur
FB 06	Psychologie und Sportwissenschaft
FB 07	Mathematik und Informatik, Physik, Geographie
FB 08	Biologie und Chemie
FB 09	Agrarwissenschaften, Ökotoxikologie und Umweltmanagement
FB 10	Veterinärmedizin
FB 11	Medizin

Quelle: Eigene Darstellung nach JLU 2009g.

**Tabelle A 5: Ergebnisse des Mann-Whitney U-Tests zur klassifizierten Einkommensvariablen (ungeachtet)**

	Mann-Whitney U-Test <sup>a</sup>	
	Mann-Whitney-U	Asympt. Signifikanz <sup>b</sup>
Monatliches verfügbares Nettoeinkommen (n=594)	34006,500	0,321

a. Gruppenvariable: Befragungssample; b. Zweiseitige asymptotische Signifikanz  
Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 6: Kolmogorov-Smirnov-Test auf Normalverteilung der Variablen "Interviewdauer" in der Gesamtstichprobe**

Variable	Kolmogorov-Smirnov <sup>a</sup>		
	Statistik	df	Signifikanz
Interviewdauer (n=698)	0,135	698	0,000

a. Signifikanzkorrektur nach Lilliefors  
Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 7: Ergebnisse des Mann-Whitney U-Tests zur Interviewdauer**

	Mann-Whitney U-Test <sup>a</sup>	
	Mann-Whitney-U	Asympt. Signifikanz <sup>b</sup>
Interviewdauer (n=698)	19122,500	0,000

a. Gruppenvariable: Befragungssample; b. Zweiseitige asymptotische Signifikanz  
Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 8: Ergebnisse der Kontingenzanalyse zur Ermittlung des Zusammenhangs zwischen der Befragungsmethode und demographischen Merkmalen**

	$\chi^2$ -Wert	Freiheitsgrade	Signifikanz
Geschlecht (n=700)	38,607	1	0,000
Fachbereich (n=700)	111,955	10	0,000
Bildungsabschluss (n=280)	37,547	5	0,000

Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 9: Mann-Whitney U-Test zur Altersvariablen**

	Mann-Whitney U-Test <sup>a</sup>	
	Mann-Whitney-U	Asympt. Signifikanz <sup>b</sup>
Alter (n=697)	43473,000	0,083

a. Gruppenvariable: Befragungssample; b. Zweiseitige asymptotische Signifikanz  
Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 10: Ergebnisse des Mann-Whitney U-Tests zur klassifizierten Einkommensvariablen (gewichtet)**

	Mann-Whitney U-Test <sup>a</sup>	
	Mann-Whitney-U	Asympt. Signifikanz <sup>b</sup>
Monatliches verfügbares Nettoeinkommen (n=594)	32683,500	0,583

a. Gruppenvariable: Befragungssample; b. Zweiseitige asymptotische Signifikanz  
Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 11: Ergebnisse des Kolmogorov-Smirnov-Tests zur Überprüfung der Items der Fragestellung zur Relevanz der Stärkung verschiedener politischer Aufgabenbereiche auf Normalverteilung**

	Kolmogorov-Smirnov <sup>a</sup>		
	Statistik	df	Signifikanz
Bekämpfung der Arbeitslosigkeit (n=699)	0,252	699	0,000
Schutz von Tier- und Pflanzenarten (n=699)	0,218	699	0,000
Verbesserung des Zusammenlebens mit Ausländern (n=699)	0,268	699	0,000
Maßnahmen zum Klimaschutz (n=699)	0,299	699	0,000
Bekämpfung von Kriminalität (n=699)	0,215	699	0,000
Förderung der Wirtschaft (n=699)	0,249	699	0,000

a. Signifikanzkorrektur nach Lilliefors

Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 12: Ergebnisse des Mann-Whitney U-Tests zu den Items der Fragestellung zur Relevanz der Stärkung verschiedener politischer Aufgabenbereiche**

	Mann-Whitney U-Test <sup>a</sup>	
	Mann-Whitney-U	Exakte Signifikanz
Bekämpfung der Arbeitslosigkeit (n=638)	43115,000	0,421
Schutz von Tier- und Pflanzenarten (n=640)	42942,000	0,291
Verbesserung des Zusammenlebens mit Ausländern (n=640)	40961,000	0,041
Maßnahmen zum Klimaschutz (n=640)	44777,000	0,852
Bekämpfung von Kriminalität (n=640)	44490,500	0,754
Förderung der Wirtschaft (n=640)	44756,000	0,849

a. Gruppenvariable: Befragungssample

Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 13: Ergebnisse des Mann-Whitney U-Tests zu den Items der Einstellungsfrage zum Themenkomplex „Klimaschutz“**

	Mann-Whitney U-Test <sup>a</sup>	
	Mann-Whitney-U	Exakte Signifikanz
Auch mit kleinen Dingen wie sparsamer heizen und öfters mal das Auto stehen lassen kann schon viel erreicht werden. (n=637)	42242,500	0,224
Wenn ich als Einzelner etwas für den Klimaschutz tue, ändert das auch nichts. (n=623)	41874,000	0,485
Für den Klimaschutz bin ich bereit, auf Flugreisen zu verzichten. (n=623)	31595,000	0,000
Ich bin nur dann bereit, mein Konsumverhalten für den Klimaschutz zu ändern, wenn alle anderen dies auch tun. (n=626)	42760,000	0,723
Klimaschutz ist alleinige Sache der Politik. (n=635)	40277,000	0,026
Um das Klima zu schützen, sollten höhere Steuern auf Benzin und Treibstoffe erhoben werden. (n=635)	43949,500	0,750

a. Gruppenvariable: Befragungssample

Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 14: Ergebnisse des Mann-Whitney U-Tests zu den Items der Verhaltensfrage zum Themenkomplex „Klimaschutz“. Genaue Fragestellung: „Wie oft hast Du bereits eine der folgenden Maßnahmen unternommen?“**

	Mann-Whitney U-Test <sup>a</sup>	
	Mann-Whitney-U	Exakte Signifikanz
Häufiger die Heizung runterdrehen, weniger warmes Wasser verbrauchen (n=635)	41016,500	0,126
Energiesparlampen kaufen (n=633)	42191,000	0,324
Beim Kauf von Haushaltsgeräten auf Energieverbrauch achten (n=609)	30207,500	0,000
Produkte aus der näheren Umgebung kaufen (n=638)	41009,000	0,065
Das Auto stehen lassen und zu Fuß gehen bzw. Fahrrad oder öffentliche Verkehrsmittel nutzen (n=634)	42401,500	0,353
Mit der Familie oder Freunden über das Thema Globale Erwärmung sprechen (n=638)	38670,500	0,003
Aus Klimaschutzgründen auf Fernreisen verzichten (n=570)	28448,500	0,000

a. Gruppenvariable: Befragungssample  
Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 15: Ergebnisse der Kontingenzanalyse zur Ermittlung des Zusammenhangs zwischen dem Bezug von „Grünem“ Strom und der Befragungsmethode**

	$\chi^2$ -Wert	Freiheitsgrade	Signifikanz
Befragungsmethode (n=700)	7,979	3	0,046

Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 16: Ergebnisse der Kontingenzanalyse zur Ermittlung des Zusammenhangs zwischen der Beantwortung der einzelnen Quiz-Items und der Befragungsmethode**

	$\chi^2$ -Wert	Freiheitsgrade	Signifikanz
Gegebene Antwort Quizfrage 1 (n=700)	0,574	1	0,502
Gegebene Antwort Quizfrage 2 (n=700)	2,344	1	0,134
Gegebene Antwort Quizfrage 3 (n=700)	0,767	1	0,422
Gegebene Antwort Quizfrage 4 (n=700)	9,635	1	0,002
Gegebene Antwort Quizfrage 5 (n=700)	0,166	1	0,753

Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 17: Ergebnisse des Kolmogorov-Smirnov-Tests zur Überprüfung der korrekt gegebenen Antworten des Klimaschutz-Quiz auf Normalverteilung**

	Kolmogorov-Smirnov <sup>a</sup>		
	Statistik	df	Signifikanz
Anzahl korrekt gegebener Quiz-Antworten (n=699)	0,242	699	0,000

a. Signifikanzkorrektur nach Lilliefors  
Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 18: Ergebnisse des Mann-Whitney U-Tests zu der Anzahl der korrekt gegebenen Antworten des Klimaschutz-Quiz**

	Mann-Whitney U-Test <sup>a</sup>	
	Mann-Whitney-U	Exakte Signifikanz
Anzahl korrekt gegebener Quizantworten (n=639)	43658,000	0,501

a. Gruppenvariable: Befragungssample  
Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 19: Ergebnisse des Mann-Whitney U-Tests für die verschiedenen Zeichen. Genauer Wortlaut der Fragestellung: „Hast Du dieses Zeichen (bzw. Siegel) schon mal gesehen? Kennst Du diese Zeichen?“**

	Mann-Whitney U-Test <sup>a</sup>	
	Mann-Whitney-U	Exakte Signifikanz
Zeichen 1: Bio-Zeichen (n=640)	44213,000	0,552
Zeichen 2: Markenqualität aus deutschen Landen (n=640)	44977,000	0,934
Zeichen 3: Fairtrade-Zeichen (n=640)	42571,500	0,179
Zeichen 4: MSC-Zeichen (n=640)	35464,500	0,000
Zeichen 5: Product Carbon Footprint (n=640)	41661,500	0,019

a. Gruppenvariable: Befragungssample  
Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 20: Ergebnisse des Mann-Whitney U-Tests zu den korrekt gegebenen Teilantworten hinsichtlich des Product Carbon Footprints (PCF)**

	Mann-Whitney U-Test <sup>a</sup>	
	Mann-Whitney-U	Exakte Signifikanz
Korrekt gegebene Teilantworten zum PCF (n=640)	38460,000	0,000

a. Gruppenvariable: Befragungssample  
Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 21: Ergebnisse des Mann-Whitney U-Tests zum semantischen Differentials des Product Carbon Footprints. Genauer Wortlaut der Fragestellung „Hältst Du ein solches ‚CO2-Zeichen‘ für...“, gefolgt von den Eigenschaftspaaren**

	Mann-Whitney U-Test <sup>a</sup>	
	Mann-Whitney-U	Exakte Signifikanz
Eigenschaftspaar „notwendig/überflüssig“ (n=609)	40370,000	0,788
Eigenschaftspaar „effektiv/ineffektiv“ (n=597)	36903,500	0,212
Eigenschaftspaar „nicht umsetzbar/umsetzbar (n=596)“	33419,000	0,001
Eigenschaftspaar „verständlich/unverständlich“ (n=621)	36380,500	0,001
Eigenschaftspaar „kompliziert/einfach“ (n=614)	39331,000	0,154

a. Gruppenvariable: Befragungssample  
Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 22: Ergebnisse der Kontingenzanalyse zur Ermittlung des Zusammenhangs zwischen der Frage nach der rechtlichen Regelung eines solchen PCFs und der Befragungsmethode**

	$\chi^2$ -Wert	Freiheitsgrade	Signifikanz
Befragungsmethode (n=698)	10,459	2	0,005

Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 23: Ergebnisse der Kontingenzanalyse zur Ermittlung des Zusammenhangs zwischen der Frage nach Ausgestaltung eines solchen PCFs und der Befragungsmethode**

	$\chi^2$ -Wert	Freiheitsgrade	Signifikanz
Befragungsmethode (n=698)	17,559	2	0,000

Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 24: Ergebnisse des Kolmogorov-Smirnov-Tests zur Überprüfung der gebildeten Variablen „Zahlungsbereitschaftsdifferenz“ auf Normalverteilung (Gesamtstichprobe)**

	Kolmogorov-Smirnov <sup>a</sup>		
	Statistik	df	Signifikanz
Zahlungsbereitschaftsdifferenz (n=591)	0,165	591	0,000

a. Signifikanzkorrektur nach Lilliefors

Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 25: Ergebnisse des Kolmogorov-Smirnov-Tests zur Überprüfung der gebildeten Variablen „Zahlungsbereitschaftsdifferenz“ auf Normalverteilung (Teilstichproben)**

	Kolmogorov-Smirnov <sup>a</sup>		
	Statistik	df	Signifikanz
F2F-Stichprobe (n=167)	0,119	167	0,000
Online-Stichprobe (n=424)	0,177	424	0,000

a. Signifikanzkorrektur nach Lilliefors

Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 26: Ergebnisse des Mann-Whitney U-Tests zu der Variablen „Zahlungsbereitschaftsdifferenz“**

	Mann-Whitney U-Test <sup>a</sup>	
	Mann-Whitney-U	Exakte Signifikanz
Zahlungsbereitschaftsdifferenz (n=545)	32683,500	0,583

a. Gruppenvariable: Befragungssample

Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 27: Erklärte Gesamtvarianz durch die Faktorenanalyse**

Komponente	Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion			Rotierte Summe der quadrierten Ladungen		
	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	2,641	22,010	22,010	2,641	22,010	22,010	2,198	18,318	18,318
2	1,812	15,102	37,111	1,812	15,102	37,111	1,823	15,189	33,507
3	1,252	10,430	47,541	1,252	10,430	47,541	1,684	14,034	47,541
4	0,923	7,693	55,235						
5	0,895	7,458	62,692						
6	0,853	7,112	69,805						
7	0,700	5,832	75,636						
8	0,675	5,623	81,259						
9	0,641	5,345	86,605						
10	0,610	5,083	91,688						
11	0,517	4,312	96,000						
12	0,480	4,000	100,000						

Extraktionsmethode: Hauptkomponentenanalyse.

Quelle: Auswertung der eigenen Erhebung.

**Tabelle A 28: Ergebnisse des t-Tests zum Mittelwertvergleich der Einstellungsdimensionen aus der Faktorenanalyse (F2F n=176; Online n=477)**

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit		
		F	Signifikanz	T	df	Sig. (2-seitig)
Zukunftsorientierung	Varianzen sind gleich	4,176	0,041	1,951	651	0,051
	Varianzen sind nicht gleich			2,062	349,608	0,040
Stärkung_Gesellschaft	Varianzen sind gleich	0,233	0,630	-0,639	651	0,523
	Varianzen sind nicht gleich			-0,645	317,986	0,519
Desinteresse_Klimawandel	Varianzen sind gleich	05,411	0,020	2,586	651	0,010
	Varianzen sind nicht gleich			2,729	348,444	0,007

Quelle: Auswertung der eigenen Erhebung.

## **Arbeitsberichte des Instituts für Agrarpolitik und Marktforschung der Justus-Liebig-Universität Gießen\***

- 1 bis 23 siehe Verzeichnis in Nr. 24
- 24 SCHUMACHER, Silke, Quantitative Erfassung des Anbaus und der Verwertung nachwachsender Rohstoffe in der Europäischen Union.  
1997, 119 S. und Anhang.
- 25 ECKERT, Sabine, Ökonomische Effekte von Lebensmittelskandalen. Das Beispiel BSE.  
1998, 104 S. und Anhang.
- 26 GÄRTNER, Susanne, Freizeit und Nahrungsmittelnachfrage: Theoretische Überlegungen und empirische Auswertung der Nationalen Verzehrsstudie.  
1999, 105 S. und Anhang.
- 27 KROLL, Steffi, Der Einfluß von Verkaufsförderung auf den Absatz von Markenartikeln – Eine empirische Analyse für den Cerealienmarkt.  
2000, 119 S. und Anhang.
- 28 WERNER, Elke, Marktstruktur und –entwicklung des deutschen Konfitüremarktes: Beschreibung, Analyse, Determinanten des Konsumentenverhaltens.  
2000, 109 S. und Anhang.
- 29 ANDERS, Sven, Quantitative Analyse der Entwicklung des Fleischverbrauchs in Hessen: Ursachen von Verbrauchsstrukturänderungen und Folgen für das hessische Gemeinschaftsmarketing.  
2000, 101 S. und Anhang.
- 30 GAST, Michael, Nichttarifäre Handelshemmnisse bei heterogenen Gütern der Agrar- und Ernährungswirtschaft – Theoretische Grundlagen und das Beispiel US-amerikanischer Käseimporte.  
2001, 82 S.
- 31 SCHRÖTER, Christiane, Consumer perceptions of three innovations related to meat processing.  
2001, 87 S.
- 32 WENZEL, Montserrat, Hedonistische Preisanalyse zum Einfluß von Qualität auf den Preis von Fruchtsaft: Das Beispiel Apfelsaft.  
2001, 157 S.
- 33 ROGGENKAMP, Liz, Erfolgreiche Innovationen in der Ernährungswirtschaft – Messung und Determinanten –.  
2002, 101 S.
- 34 RÖSE, Stefan, Marktanalyse über Soja-Lebensmittel in Deutschland.  
2002, 127 S.
- 35 GAST, Michael, Der Importmarkt für ausgewählte pflanzliche Drogen in Deutschland.  
2003, 45 S.
- 36 KUBITZKI, Sabine, Innovationsaktivitäten im Ernährungsgewerbe – Eine branchenspezifische Untersuchung des Mannheimer Innovationspanels 1999 -.  
2003, 105 S.
- 37 HARTL, Jochen, Estimating the Demand for Risk Reduction from Foodborne Pathogens.  
2004, 120 S.
- 38 EGENOLF, Petra, Ökonomische Konsequenzen von BSE: Stand der Forschung und empirische Analyse des Verbraucherverhaltens in der deutschen BSE-Krise.  
2004, 106 S.
- 39 FAUST, Ulrike, Gemeinschaftsmarketing für Lebensmittel unter dem Einfluss von EU-Recht und Verbraucherverhalten – das Beispiel „Geprüfte Qualität – HESSEN“.  
2005, 118 S.
- 40 TÖNNIGES, Stefan, Die Determinanten der Nachfrage nach Fisch und Fischwaren.  
2005, 117 S. und Anhang.

\* Die Arbeitsberichte können für eine Schutzgebühr von 25,-- €(15,-- €bis Nr. 38) erworben werden beim:  
Institut für Agrarpolitik und Marktforschung, Justus-Liebig-Universität Gießen, Senckenbergstr. 3, 35390 Gießen,  
Tel.: (06 41) 99-3 70 20, Fax: (06 41) 99-3 70 29, e-mail: Sekretariat.Marktlehre@agr.uni-giessen.de

- 41 WETTNER, Christoph, Wirkungsanalyse negativer Agrarprotektion: Quantifizierungsansätze und Möglichkeiten der Implementierung in GTAP. 2006, 84 S. und Anhang.
- 42 SCHÜTZ, Daniel, Verbraucherpräferenz für regionale Lebensmittel – Eine Untersuchung der Einflussfaktoren mit Hilfe multivariater Analysemethoden. 2006, 120 S.
- 43 HEINZE, Daniela, Experimentelle Analyse von Zahlungsbereitschaften für das Qualitäts- und Herkunftssiegel „Geprüfte Qualität – HESSEN“. 2006, 103 S.
- 44 OBERBECK, Corinna, Produktdifferenzierung im deutschen Kaffeemarkt. 2008, 90 S.
- 45 HENKEL, Tobias, Die Bedeutung der Regionalität beim Lebensmittelkauf: Empirische Ergebnisse am Beispiel „Landmarkt“. 2008, 83 S.
- 46 MAAS, Sarah, What Explains EU Food Aid? 2008, 110 S.
- 47 ROEBEN, Anna, Erzielen Produkte mit höherer Qualität auch einen höheren Preis im Lebensmitteleinzelhandel? - Eine empirische Untersuchung am Beispiel Fruchtsaft. 2008, 95 S. und Anhang.
- 48 STAUDIGEL, Matthias, Der Einfluss institutioneller Rahmenbedingungen auf Ernährung und Gesundheit am Beispiel Russlands. 2009, 109 S. und Anhang.
- 49 KUBITZKI, Sabine, Dr. HENSELEIT, Meike, Prof. Dr. HERRMANN, Roland, HENKEL, Tobias, Lebensmittelkennzeichnung „ohne Gentechnik“: Verbraucherwahrnehmung und –verhalten. 2009, 77 S. und Anhang mit Fragebogen.
- 50 SCHRÖCK, Rebecca, Determinanten der Nachfrage nach Biomilch – eine ökonometrische Analyse. 2010, 142 S. und Anhang.
- 51 HENKEL, Tobias, Die Wirkung der Befragungsform auf das Antwortverhalten: Eine vergleichende Untersuchung am Beispiel des Carbon Footprints bei Lebensmitteln. 2010, 162 S. und Anhang.