

Universität Stuttgart
Institut für Sozialwissenschaften
Abteilung Soziologie und empirische Sozialforschung
Projektseminar: „Erhebung und Analyse xenophober Orientierungen und Verhaltensmuster. Möglichkeiten jenseits der standardisierten Befragung.“
Leitung: Thomas Krause, M.A.
Wintersemester 2015/2016

„Entschuldigung, können Sie mir bitte 5 € wechseln?“

**– Ein Feldexperiment zu den Auswirkungen von
Xenophobie auf die Hilfsbereitschaft –**

Vorgelegt von:

Nikolaus Hastreiter
Heinrich-Baumann-Straße 26
70190 Stuttgart
st100535@stud.uni-stuttgart.de
Matrikelnr. 2783655
MA. Empirische Politik- und
Sozialforschung

Fabienne Minn
Haußmannstr. 154
70188 Stuttgart
st101098@stud.uni-stuttgart.de
Matrikelnr. 2785705
MA. Empirische Politik- und
Sozialforschung

Alexander Schrage
Libanonstr. 71
70186 Stuttgart
st100641@stud.uni-stuttgart.de
Matrikelnr. 2789277
MA. Empirische Politik- und
Sozialforschung

Abgabedatum:

Stuttgart, 30. September 2016

Inhaltsverzeichnis

Tabellen- Abbildungs- und Abkürzungsverzeichnis.....	3
1. Einleitung	4
2. Theorie	5
2.1 Grundsätzliches zum Thema Fremdenfeindlichkeit.....	5
2.2 Gruppenbezogene Menschenfeindlichkeit	8
2.3 Die Aversive Racism Theorie	10
2.4 Hypothesenherleitung.....	13
3. Operationalisierung und Durchführung.....	15
3.1 Das Experiment in den Sozialwissenschaften	15
3.1.1 Feldexperiment.....	19
3.1.2 Ethische Aspekte experimenteller Untersuchungsdesigns	20
3.2. Experimentaldesign.....	21
4. Auswertung der Ergebnisse	31
4.1. Deskriptive statistische Analyse.....	31
4.1.1 Häufigkeiten nach LV	31
4.1.2 Häufigkeiten nach LV und nach Geschlecht.....	33
4.1.3. Deskriptive Kennzahlen	34
4.2. Inferenzstatistische Analyse.....	35
4.2.1 Multinominale logistische Regression	35
4.2.2 Binominale logistische Regression	37
5. Fazit.....	39
6. Literaturverzeichnis.....	40
7. Anhang	42
Erklärung über die Eigenständigkeit	53

Tabellen- Abbildungs- und Abkürzungsverzeichnis

Tabellenverzeichnis

Tabelle 1: Frauen in Stuttgart nach Altersgruppen (2015).....	24
Tabelle 2: Männer in Stuttgart nach Altersgruppen (2015)	24
Tabelle 3: Quotenplan für einen LV	24
Tabelle 4: Ordinale Zusammenhangsmaße für den Vergleich des Antwortverhaltens nach Lockvogel.....	34
Tabelle 5: Ordinale Zusammenhangsmaße für den Vergleich des Antwortverhaltens nach Geschlecht.....	34
Tabelle 6: Modelle der MLR.....	36
Tabelle 7: Likelihood ratio Tests von Modell 1	37
Tabelle 8: Parameterschätzungen von Modell 1	37
Tabelle 9: Parameterschätzungen der binominalen logistischen Regression.....	38
Tabelle 10: Tabelle zur Betrachtung der Störfaktoren nach Campbell und Stanley für das vorgestellte Experiment.....	42
Tabelle 11: Likelihood ratio Tests von Modell 2	45
Tabelle 12: Likelihood ratio Tests von Modell 3	45
Tabelle 13: Likelihood ratio Tests von Modell 4	45

Abbildungsverzeichnis

Abbildung 1: LV am 1. Erhebungstag (links die KP, rechts die EP)	27
Abbildung 2: schematische Darstellung der Experimentalsituation	29
Abbildung 3: Antwortverhalten insgesamt und pro LV in Prozent	32
Abbildung 4: Antwortverhalten pro LV nach Geschlecht	33

Abkürzungsverzeichnis

AR	Aversive Racism
EP	Experimentalperson
GMF	Gruppenbezogene Menschenfeindlichkeit
KP	Kontrollperson
LV	Lockvogel
VPN	Versuchsperson

1. Einleitung

In den Debatten um das Für und Wider von Kontingenten in der Aufnahme von Geflüchteten ist die Stimmung in Europa und der Bundesrepublik stark aufgeheizt. Nach den Ereignissen der Silvesternacht, den Anschlägen in Frankreich und nicht zuletzt in Deutschland selber scheint die Willkommenskultur nach und nach einer wachsenden Abwehrhaltung zu weichen. In deutschen Medien machen aber auch immer wieder Gewalt und Aggression gegen Asylsuchende Schlagzeilen, wenngleich 83% der deutschen Bevölkerung angeben, sich für die gewalttätigen Proteste gegen Flüchtlinge zu schämen (vgl. Infratest dimap 2016: o.S.). Auf politischer Ebene werden die Kontroversen durch populistische Äußerungen befeuert – und während über die Sicherung nationaler Grenzen, Kontingente und Abkommen mit Drittstaaten diskutiert wird, erfreut sich das rechte Parteienspektrum in- und außerhalb sozialer Netzwerke zunehmender Beliebtheit. In Anbetracht der Ergebnisse der Landtagswahlen in Baden-Württemberg, Rheinland-Pfalz, Sachsen-Anhalt und zuletzt nun auch Mecklenburg-Vorpommern die Frage, ob sich die gegenwärtige politische Situation alleine durch die aktuellen politischen Herausforderungen erklären lässt, oder ob nicht zuletzt auch die Existenz fremdenfeindlicher Einstellungen zur Erklärung beiträgt.

Doch auch abseits gegenwärtiger Geschehnisse stellt sich die Frage, wie sich das Vorhandensein xenophober Einstellungsmuster in einer Bevölkerung auf Gesellschaften auswirkt und für deren Persistenz bedeutet. Für die sozialwissenschaftliche Forschung sind selbstverständlich nicht nur die Auswirkungen von Einstellungen und die Frage nach deren handlungsleitendem Rahmen von Relevanz, sondern auch spezielle Möglichkeiten zur Erhebung xenophober Einstellungen. Diese Notwendigkeit erklärt sich insbesondere mit Hinblick auf die Problematik von Verzerrungen durch soziale Erwünschtheit bei sensiblen Themen.

Ausgehend von der Prämisse, dass xenophobe Einstellungen tatsächlich handlungsleitend werden können, wird im Rahmen dieser Arbeit ein experimentelles Erhebungsdesign vorgestellt, um die Frage zu untersuchen, inwiefern die augenscheinliche Fremdheit einer um Hilfe bittenden Person (unabhängige Variable) Auswirkungen auf die Hilfsbereitschaft von Stuttgarter BürgerInnen (abhängige Variable) hat. Hierfür werden jeweils ein in- und ausländischer Lockvogel in einer U-Bahnhaltestelle Passanten bitten, ihr einen Geldschein zum Erwerb eines Fahr-

scheins zu wechseln.

Im Rahmen dieser Arbeit wird daher zunächst einmal ein Überblick über den state-of-the-art in der Literatur zum Thema Xenophobie, Rassismus und gruppenbezogene Menschenfeindlichkeit gegeben, wobei ein besonderes Augenmerk auf die Theorie des Aversive Racism gelegt wird. Auf Grundlage dieser Analyse wird die Herleitung der für diese Arbeiten grundlegenden Hypothesen vorgenommen. Das folgende Kapitel beschäftigt sich mit der Operationalisierung, wobei die Rolle des Experiments als Erhebungsmethode in den Sozialwissenschaften diskutiert wird. Zudem wird in diesem Rahmen auf das Spezifikum des Feldexperiments gesondert eingegangen und das Experiment unter seinen ethischen Gesichtspunkten betrachtet. Im Anschluss daran wird das experimentelle Untersuchungsdesign en détail vorgestellt sowie im räumlichen und zeitlichen Kontext eingebettet. Zudem werden hier problematische Aspekte in der Forschung angeführt und weitere methodische und administrative Aspekte der Erhebung vertieft. Schlussendlich werden die gefundenen Ergebnisse statistisch ausgewertet. Hierzu werden deskriptive Interpretationen anhand von Häufigkeiten sowie inferenzstatistische Interpretationen anhand einer multinominalen logistischen Regression vorgenommen. Es zeigt sich, dass die Hypothese zur Fremdenfeindlichkeit zumindest in Teilen bestätigt wird, andere Faktoren wie das Geschlecht hingegen jedoch keine Auswirkungen zu haben scheinen. Im Anschluss an die Auswertungen wird ein abschließendes Fazit gezogen, das die Hauptideen der Arbeit zusammenfasst.

(FM)

2. Theorie

Im Folgenden werden zunächst grundlegende Begriffe zum Thema Fremdenfeindlichkeit erläutert. Anschließend wird eine Übersicht über verschiedenen Erklärungsansätze des Phänomens gegeben, bevor genauer auf die für diese Studie zentrale Theorie des Aversive Racism (AR) eingegangen wird.

2.1 Grundsätzliches zum Thema Fremdenfeindlichkeit

Im deutschen Sprachgebrauch wird häufig der Begriff der Ausländerfeindlichkeit anstelle von Rassismus verwendet. Letzterer dient meist der historischen Beschreibung des Nationalsozialismus oder des heutigen Rechtsextremismus. Dies

sei jedoch insofern problematisch, als dass dadurch das Problem als ein gesellschaftliches Randphänomen dargestellt und marginalisiert werde.¹

Rassismus ist in Deutschland jedoch keinesfalls ein Randphänomen: Im Jahr 1981 beklagten Professoren im *Heidelberger Manifest* “die Unterwanderung des deutschen Volkes durch Zuzug von vielen Millionen von Ausländern und ihren Familien, die Überfremdung unserer Sprache, unserer Kultur und unseres Volkstums’ und forderten ‘die Erhaltung des deutschen Volkes und seiner geistigen Identität auf der Grundlage unseres christlich-abendländischen Erbes’” (vgl. Shooman 2012: 54).

Die in dieser Aussage enthaltene Form des Rassismus wird als *kultureller* Rassismus bezeichnet. Diesem liegt die Vorstellung einer homogenen Gesellschaft zugrunde (vgl. Broden 2012: 9). Kombiniert wird dies mit dem “Glaube[n] an eine historisch gewachsene Inkompatibilität und Hierarchie der Kulturen (und der Religionen als Bestandteil von Kulturen)” (Shooman 2012: 54). Kulturelle Eigenschaften von Gesellschaften werden naturalisiert und als unwandelbar angesehen. Dies führe dazu, dass gewisse Eigenschaften oder kulturelle Verhaltensweisen auf alle Mitglieder der als homogen gedachten Gesellschaft verallgemeinert werden. Davon abweichendes soziales Verhalten werde mit der Zugehörigkeit zu einer anderen Gruppe begründet (vgl. ebd.). Der kulturelle Rassismus betrachtet ausschließlich die über dessen Abstammung definierte “Kultur” eines Individuums und übersieht dabei dessen übrige Identitäten und Zugehörigkeit zu anderen sozialen Gruppen.² Auch wird nicht berücksichtigt, zu welcher Gruppe, beziehungsweise zu welchem “Kulturkreis”, sich die betroffene Person selbst zuordnet (vgl. Shooman 2012: 54). Somit werde jede Person, die dem imaginären Leitbild dieser Gesellschaft nicht zu entsprechen vermag, diskreditiert (vgl. Broden 2012: 9). Kultureller Rassismus ist ein weit verbreitetes Phänomen: Umfragewerte aus dem Jahr 2011 zeigen in acht EU-Mitgliedsstaaten eine starke Zustimmung zu der

¹ Die Bezeichnung Ausländerfeindlichkeit sei jedoch irreführend, da sich das rassistische Verhalten nicht nur gegen Ausländer richte (vgl. Shooman 2012: 54). Auch deutsche Staatsangehörige können, beispielsweise aufgrund ihrer (vermuteten) Religionszugehörigkeit, betroffen sein.

² Es kommt in diesem Zusammenhang zum Phänomen, welches als „Rassifizierung“ bezeichnet wird. Ohne Rücksichtnahme auf individuelle Glaubensbekenntnisse wird die ethnische Fremdgruppe (beispielsweise Muslime) als homogene Gruppe wahrgenommen. Von ihrer Religion wird auf die fremde Herkunft der Mitglieder geschlossen und diese stigmatisiert (vgl. Shooman 2012: 55f).

Aussage, es lebten zu viele Musliminnen und Muslime im eigenen Land.³ Besonders hoch war die Zustimmung in Ländern mit einem Anteil von Muslimen in der Bevölkerung von unter einem Prozent. Dies zeigt, dass es sich bei der wahrgenommenen Überfremdung um einen subjektiven Eindruck handelt (vgl. Shooman 2012: 55).

Neben dem kulturellen Rassismus existiert der *biologische* Rassismus. Dieser diskriminiert⁴ Individuen aufgrund vermeintlicher biologischer Differenzen und Minderwertigkeit. Klassisches Beispiel für biologischen Rassismus ist der Nationalsozialismus in Deutschland. Nach dessen Ende fand der biologische Rassismus in der Bundesrepublik kaum noch Gebrauch. Doch auch wenn derartige Argumente heutzutage nicht mehr artikuliert würden, habe der biologische Rassismus nicht an Wirkung eingebüßt.⁵ Häufiger finden sich fremdenfeindliche Argumente, die dem kulturellen Rassismus zuzuordnen sind (vgl. Broden 2012: 9). Allerdings ist aktuell zu beobachten, dass der biologische Rassismus wieder salonfähig wird. Biologistische Argumente werden mit neoliberalen Auffassungen vermischt: Menschen mit Migrationshintergrund werden nicht nur aufgrund ihrer Abstammung diskriminiert, sondern ebenfalls aufgrund ihrer vermeintlichen wirtschaftlichen „Nichtverwertbarkeit“ (Broden 2012: 9).⁶

Auch wenn es nicht das Ziel dieser Studie ist, Hilfsbereitschaft gegenüber Muslimen zu messen, muss an dieser Stelle darauf hingewiesen werden, dass dies doch indirekt der Fall sein kann. So kommt es zu einer Ethnisierung der Religion: bei Personen werde anhand von Indikatoren wie Hautfarbe oder Sprache auf deren Religionszugehörigkeit geschlossen (vgl. Shooman 2012: 55). Somit kann es durchaus sein, dass der nicht-deutsche Lockvogel (LV) nicht aufgrund seiner Herkunft diskriminiert wird, sondern, da er aufgrund seiner Hautfarbe der Kategorie „Muslim“ zugeordnet wird. Womöglich ist diese vermutete Religionszugehörig-

³ Eine Zustimmung zur besagten Aussage ist dem kulturellen Rassismus zuzuordnen, da wahrgenommene Missstände in der Gesellschaft auf die Religionszugehörigkeit einer Bevölkerungsgruppe zurückgeführt werden und andere Faktoren nicht berücksichtigt werden.

⁴ Als Diskriminierung wird eine verweigerte Gleichbehandlung von Individuen oder Gruppen bezeichnet, obwohl diese Gleichbehandlung von ihnen erwünscht ist (vgl. Geschke 2012: 35).

⁵ Eine Ausnahme bildet die Äußerung des Thüringer AfD-Chefs Björn Höcke, welche für mediales Aufsehen sorgte. Im November 2015 sprach dieser bei einer Tagung des „Instituts für Staatspolitik“ vom „[...] Reproduktionsverhalten der Afrikaner“ (Hurtz 2015: o.S.). Ihm wurde anschließend vorgeworfen, wie die Nationalsozialisten biologische Theorien auf den Menschen zu übertragen (vgl. ebd.).

⁶ So auch in der aktuellen deutschen Flüchtlingsdebatte: <http://www.tagesspiegel.de/politik/fluechtlinge-als-fachkraefte-in-der-asylfrage-ist-nuetzlichkeit-kein-argument/11951852.html>

keit ausschlaggebend für die Diskriminierung und nicht etwa die Herkunft der Person. Das Problem, die Ursache der Diskriminierung zu ermitteln, kann zwar nicht gelöst werden, ist aber auch nicht Sinn und Zweck dieser Studie. Hier soll lediglich der Umfang der alltäglichen Diskriminierung gemessen werden, der eine Person mit offensichtlich fremder Herkunft in Deutschland ausgesetzt ist. Ob diese Diskriminierung nun aufgrund der Herkunft und/oder der vermuteten Religionszugehörigkeit geschieht, ist hierbei nicht relevant. Schließlich macht es auch für die betroffenen Personen im realen Alltag auch keinen Unterschied, aus welchem der genannten Gründe sie nun benachteiligt werden.

2.2 Gruppenbezogene Menschenfeindlichkeit

Neben dem Begriff des Rassismus findet häufig der Term der *gruppenbezogenen Menschenfeindlichkeit* (GMF) Verwendung. Im Gegensatz zum Begriff der Ausländerfeindlichkeit spielt in diesem Konzept nicht die (scheinbare) Herkunft einer Person die entscheidende Rolle. Wie die Bezeichnung bereits vermuten lässt, fallen unter GMF alle Formen von Diskriminierung von Individuen, welche aufgrund deren Zugehörigkeit zu einer (nicht zwingend von der Abstammung definierten) Gruppe geschieht.⁷ Aber es muss nicht erst faktisch zu einer Diskriminierung kommen, um von Menschenfeindlichkeit zu sprechen: die GMF umfasst „negative Stereotype, kognitive Vorurteile⁸, emotionale soziale Distanzierung [und die] Absicht der Schädigung“ (Groß et al. 2012: 11f). Wie beim kulturellen und biologischen Rassismus schafft die bloße Klassifizierung von Personen in Gruppen bereits einen Bias im Umgang mit ihnen (vgl. Gaertner/Dovidio 2005: 618). Grundlage für den Bias ist die Ideologie der Ungleichwertigkeit⁹ von Gruppen, die in Ungleichheit transformiert werden kann (vgl. Groß et al. 2012: 12f). Bei GMF handelt es sich um kein Randphänomen, vielmehr seien die verwendeten

⁷ Die Kategorisierung, das heißt die Zuordnung von Objekten (Menschen) zu Klassen, ist aufgrund der begrenzten kognitiven Fähigkeiten von Menschen ein notwendiges Verhalten, welches nicht zwingend abwertend ist (vgl. Geschke 2012: 34).

⁸ Während Stereotype Eigenschaften beschreiben, welche einer Gruppe (oft fälschlicherweise) zugeschrieben werden, enthalten Vorurteilen eine wertende, herabsetzende Eigenschaft. Sie sind „verzerrte Bewertungen eines sozialen Reizes, die kognitive (wie Stereotype), emotionale (wie Angst) und verhaltensmäßige Komponenten (wie Vermeidung) enthalten“ (Geschke 2012: 34).

⁹ Der Begriff Ungleichwertigkeit verdeutlicht die zuvor angesprochene Kopplung von Rassismus und neoliberaler Denkweise mit dem Fokus auf der Verwertbarkeit von Personen. „Die Ökonomisierung des Sozialen trägt dazu bei, insbesondere [...] als Konkurrenten und ‘Sozialschmarotzer’ empfundene Ausländer [und] Muslime [...] über die entsprechenden Vorurteile auf ihre statusniedrige Position zu verweisen. Diese fundamentale gesellschaftliche Entwicklung festigt und formt so indirekt und vermittelt über Menschenfeindlichkeit eine aktuelle Form der sozialen Ungleichheit.“ (Groß et al. 2012: 17)

Vorurteile in gesellschaftlichen Diskursen derart implementiert, dass sie „zum festen Bestandteil des vermeintlichen Wissens einer Gesellschaft“ (ebd.: 13) geworden seien. Es handle sich bei GMF um ein Phänomen, welches konjunkturellen Schwankungen unterliege (vgl. ebd.: 15).¹⁰

Untersucht wird das Konzept der GMF am Institut für interdisziplinäre Konflikt- und Gewaltforschung der Universität Bielefeld. Das Ergebnis einer durchgeführten Faktorenanalyse zeigt die enge Verknüpfung (Faktorladungen $> .75$) von Fremdenfeindlichkeit, Rassismus, Islamfeindlichkeit und der Abwertung von Obdachlosen und Asylbewerbern (vgl. ebd.). In den Bielefelder Studien konnte ebenfalls der bereits angesprochene Zusammenhang von Fremdenfeindlichkeit und einer ökonomischen Sichtweise nachgewiesen werden – besonders bei Personen, welche sich selbst „oben“ einordnen. „Wer eine ökonomische Sichtweise teilt – d.h. Menschen nach ihrem Nutzen beurteilt – neigt deutlich eher zur Abwertung schwacher Gruppen.“ (Broden 2012: 10).

Es gibt verschiedene Ansätze und Theorien zur Erklärung von GMF. Laut *realistischer Konflikttheorie* entsteht GMF bereits durch eine Konkurrenzsituation, wenn zwei oder mehrere Gruppen ein gemeinsames Ziel anstreben. Man spricht in diesem Kontext auch von *negativer Interdependenz*. Sie führt zu einer Aufwertung der Eigengruppe und Abwertung der Fremdgruppe.¹¹ Die eigene Gruppe wird aufgewertet, da so das eigene Selbstwertgefühl gestärkt wird. Ist die Interdependenz hingegen *positiv*, müssen beide Gruppen also zum Erreichen des Ziels zusammenarbeiten, führt dies zu sozialer Harmonie. Damit es zu einer Abwertung der Fremdgruppe kommt, muss die Interdependenz allerdings nicht tatsächlich negativ sein, die bloße Wahrnehmung einer Konkurrenzsituation ist ausreichend, damit sich entsprechende Effekte beobachten lassen (vgl. Geschke 2012: 35).

Eine zweite Theorie ist der *Soziale Identitätsansatz*. Laut diesem ergibt sich die Abwertung von Fremdgruppen auch, wenn die Kategorisierung künstlich erfolgt ist und die anderen Gruppen für die eigene Gruppe ohne Bedeutung sind (es also keine Konkurrenzsituation gibt und auch keine Kooperation notwendig ist). Aus diesem Ansatz lassen sich verschiedene Annahmen ableiten: Bei der Kategorisie-

¹⁰ Die Ergebnisse dieser Studie sind daher im Kontext der aktuellen Flüchtlingsdebatte in Deutschland und den Ereignissen in der Silvesternacht in Köln und anderen deutschen Städten zu betrachten.

¹¹ Diese Auf- und Abwertung wird auch als *positive Distinktheit* bezeichnet.

rung werden Unterschiede innerhalb der eigenen Gruppe als gering wahrgenommen, solche zu den Fremdgruppen als groß. So erscheint die eigene Gruppe als homogen. Auch kommt es im Prozess der Kategorisierung zu einer „Depersonalisierung“: Das Verhalten der Individuen wird durch die Normen und Werte der Gruppe gesteuert. Das Verhalten von Personen gegenüber anderen ergibt sich aus deren sozialer Identität, die durch die Gruppenzugehörigkeit bestimmt wird (vgl. ebd.: 36f.).

2.3 Die Aversive Racism Theorie

Nachdem nun verschiedene Begriffe zum Themenkomplex Fremdenfeindlichkeit geklärt und erste Erklärungsansätze genannt wurden, wird im folgenden Kapitel auf die für diese Studie zentrale Theorie des Aversive Racism (AR) eingegangen.

Grundlage für Diskriminierung, Rassismus und Fremdenfeindlichkeit sind Stereotype. Man kann zwischen *impliziten* und *expliziten* Stereotypen unterscheiden. Explizite Stereotype werden bewusst artikuliert. Implizite hingegen werden unbewusst und automatisch durch die Präsenz eines Stimulus aktiviert (vgl. Gaertner/Dovidio 2005: 623).

Offener – durch Umfragen gemessener – Rassismus hat in den letzten 40 Jahren abgenommen. Die Zustimmung zu Werten wie Gleichbehandlung von Minderheiten ist gestiegen. Daraus kann aber nicht geschlossen werden, dass es heutzutage weniger Rassismus gibt. Lediglich dessen offene Aussprache habe abgenommen (vgl. Pearson et al. 2009: 2). In der Tat ist eine Messung von Fremdenfeindlichkeit durch Befragungen schwer möglich, die Ergebnisse verschiedener Studien unterscheiden sich teils stark. Die Selbsteinschätzung der Befragten muss nicht immer den Tatsachen entsprechen (vgl. Saucier et al. 2005: 2).¹² Das gemeinsame Auftreten von Selbstbekenntnis zu Gleichheitswerten und rassistischem Alltagsverhalten auf individueller und gesellschaftlicher Ebene in den USA wurde von Gunnar Myrdal als „American dilemma“ bezeichnet (vgl. Gaertner/Dovidio 2005: 617). Dieses Paradox wird durch die Theorie des AR erklärt.

Der Begriff des Aversive Racism wurde von Joel Kovel (1970) geprägt, welcher ihn vom „dominative racism“ unterschied. Es gäbe eine qualitative Differenz zwi-

¹² Die Wahrnehmung von Fremdenfeindlichkeit unterscheidet sich, je nachdem ob ein Mitglied der Fremdgruppe (Schwarze in den US-amerikanischen Studien) oder ein Mitglied der Eigengruppe (Weiße) gefragt werden (vgl. Pearson et al. 2009: 2f; Gaertner/Dovidio 2005: 625).

schen AR und der herkömmlichen Form des Rassismus (→ explizite Stereotype): Erster sei subtiler und charakterisiere eher das Verhalten von politisch liberalen Individuen mit hoher Bildung (vgl. Gaertner/Dovidio 2005: 618). Theorien wie die des *modernen Rassismus* (McConahay 1986) oder des *symbolischen Rassismus* (Sears/Henry/Kosterman 2000) gehen bei betroffenen Personen von einem fundamentalen Konflikt zwischen dem Abstreiten von Vorurteilen und gleichzeitig vorhandenen, unterbewussten negativen Gefühlen und Einstellungen gegenüber der Fremdgruppe aus. Diese Formen des Rassismus beschreiben den „typischen Konservativen“ in den USA, welcher zwar Vorurteile hege, diese aber nicht öffentlich bekenne. Die AR-Theorie hingegen erklärt das Verhalten von Personen, welche Vorurteile nicht nur öffentlich abstritten, sondern politisch liberale Werte vertreten und sich selbst wahrhaftig als vorurteilsfrei ansehen. Bei diesen Personen kann es trotzdem zu diskriminierendem Verhalten kommen, wenn dies von unterbewusst vorhandenen Werten beeinflusst wird (→ implizite Stereotype) (vgl. Pearson et al. 2009: 3). Ein Aversive Racist spricht sich also für Gleichbehandlung aus, verspürt jedoch bei persönlichem Kontakt mit einem Mitglied der Fremdgruppe Unbehagen, Angst oder fehlende Gelassenheit. Dieses Unbehagen resultiere nicht in offen geäußertem Rassismus, sondern führe zu einer Meidung der Interaktion mit der Fremdgruppe, einer möglichst schnellen Beendigung einer Interaktion oder zu Verhalten, welches nicht offensichtlich fremdenfeindlich sei (vgl. Gaertner/Dovidio 2005: 619; Saucier et al. 2005: 3).

Rassistisches Verhalten entsteht daher häufig in Situationen, in welchen dieses nicht offensichtlich ist und leicht durch andere Faktoren erklärt werden kann (vgl. Saucier et al. 2005: 3). Laut *justification-suppression-model* wird ein Vorurteil dann nicht unterdrückt, wenn verweigerter Hilfe durch rationale Gründe gerechtfertigt werden kann, oder wenn das eigene Verhalten nicht als vorurteilsbelastet wahrgenommen wird (vgl. ebd.). So kann das eigene Selbstbild als vorurteilsfrei aufrechterhalten werden (vgl. Gaertner/Dovidio 2005: 620). Diskriminierung ist gering, wenn es sich um eine Situation mit hoher Wertgeladenheit handelt oder fremdenfeindliche Motive gegenüber anderen und sich selbst nicht gerechtfertigt werden können (vgl. ebd.). Aus diesem Grunde eignet sich die AR-Theorie besonders für das in Kapitel 3.2 genauer beschriebene Experiment, da sie die Unterschiede im Hilfsverhalten gegenüber Personen der Eigen- und Fremdgruppe am

besten erklären kann.¹³ Wenn man davon ausgeht, dass eine verweigerte Hilfe im Experiment (das Geldwechseln), leicht durch nicht-rassistische Motive erklärt werden kann (beispielsweise kein vorhandenes Bargeld, lediglich eine Bankkarte), so sollte sich laut AR das Unbehagen der helfenden Person besonders stark auf deren Verhalten auswirken. Somit würde die Interaktion mit der Person der Fremdgruppe entweder verweigert, oder möglichst schnell (zum Beispiel durch ein Kopfschütteln im Vorbeigehen) umgangen.

Laut *cost-reward-model* ist ein zweiter Faktor entscheidend für das Ergebnis: die Schwere der Hilfssituation. Nach diesem Modell wägt die Versuchsperson (VPN) die entstehenden Kosten der Hilfe (Zeit, Risiko, Anstrengung) gegen die Kosten einer Hilfsverweigerung (Schuldgefühl, Schaden des Opfers) ab. Wenn die Kosten einer Verweigerung überwiegen, wird Hilfe geleistet. Andersherum gilt: Ist die Hilfeleistung mit vergleichsweise hohen Kosten verbunden, wird diese verweigert (vgl. Saucier et al. 2005: 4).

Es ist jedoch anzumerken, dass sich geringe Veränderungen im Experimentdesign auf das Ergebnis auswirken können, da AR ein inkonsistentes Verhalten ist (vgl. Pearson et al. 2009: 5). So nimmt beispielsweise das Umfeld der Situation Einfluss auf das Ergebnis. Handelt es sich bei der Versuchsperson um den einzigen potentiellen Helfer, nimmt die Diskriminierung ab und es wird häufiger geholfen. Sind jedoch mehrere potentielle Helfer in der Nähe, wie in dem folgenden Experiment in der U-Bahn-Haltestelle, so nimmt die Wahrscheinlichkeit der Hilfe ab, da die Verantwortung zu helfen geteilt ist und verweigerte Hilfe leichter durch andere Faktoren erklärt werden kann (vgl. ebd.: 6; Gaertner/Dovidio 2005: 620). Diese Abhängigkeit von der Anwesenheit anderer Personen wird als „bystander effect“ bezeichnet (ebd.).

In empirischen Studien konnte die AR-Theorie jedoch nicht immer validiert werden: Saucier et al. 2005 haben eine Meta-Analyse von 48 Studien erstellt, die von 1967 bis 1986 durchgeführt wurden und sich inhaltlich mit der Diskriminierung bei Hilfsverhalten gegenüber Schwarzen befassen.¹⁴ Sie kommen dabei zu keinem eindeutigen Ergebnis: Das Ausmaß der geleisteten Hilfe werde von verschiedenen

¹³ So ergab eine Untersuchung der Kerner Commission, die sich mit den Ursachen der *race riots* in den USA im Jahr 1967 befasste, dass sich Rassismus gegenüber Schwarzen eher darin äußere, dass ihnen Hilfe verweigert werde, als dass ihnen aktiv geschadet würde (vgl. Pearson et al. 2009: 6).

¹⁴ Für eine Übersicht siehe Saucier et al. 2005: 7-9.

Moderatoren beeinflusst (vgl. Saucier et al. 2005: 9). Gegen die AR-Theorie spricht, dass in größeren Notsituationen auch die Diskriminierung gegenüber Schwarzen zunimmt – anders als von ihr vorausgesagt. Die Autoren erklären sich dieses Ergebnis mit der Theorie der *intergroup anxiety* nach Stephan/Stephan 1985. Diese besage, dass Individuen sich durch die Interaktion mit Personen der Fremdgruppe bedroht fühlen können. Diese negativen Erwartungen könnten dazu führen, dass die Interaktion vermieden werde (vgl. Saucier et al. 2005: 13f). Des Weiteren konnte das *justification-suppression model* nicht belegt werden, das *cost-reward model* hingegen schon (vgl. ebd.: 10).

2.4 Hypothesenherleitung

Wie in diesem Kapitel gezeigt wurde, gibt es unterschiedliche Ursachen für eine Diskriminierung der Fremdgruppe. Diese kann aufgrund von impliziten oder expliziten Vorurteilen entstehen. Motive können im Denken von kulturellem und/oder biologischem Rassismus verwurzelt sein. Bei dem im Folgenden beschriebenen Experiment, bei welchem Diskriminierung bei Hilfsverhalten gemessen werden soll, kommt besonders das von der AR-Theorie analysierte Verhalten zutage. Man kann nicht davon ausgehen, dass es sich bei allen Personen, welche der nicht-deutschen Testperson die Hilfe verweigern um Rassisten handelt. Vielmehr kommen bei der Interaktion mit einem Individuum der Fremdgruppe unterbewusste, implizite Vorurteile und Besorgnisse zutage, die sich möglicherweise auf das Verhalten auswirken und zu einem Bias führen. Wie bereits erwähnt, kann eine verweigerte Hilfeleistung leicht durch rationale Argumente gerechtfertigt werden, sodass es laut AR-Theorie zu diskriminierendem Verhalten kommen sollte. Auch das *cost-reward model* lässt dies erahnen. Bei dem in Kapitel 3.2 geschilderten Experimentaufbau sind die durch Hilfsverweigerung entstehenden Kosten vergleichsweise gering. Der LV kann problemlos andere Passanten um Hilfe bitten, einen erneuten Versuch mit glattgestrichenem Schein unternehmen, einen anderen Automaten auswählen oder notfalls ohne gelösten Fahrschein in die Bahn steigen. Die befürchteten Kosten einer Hilfeleistung sind dagegen groß, da die Versuchsperson neben Zeitverlust unter Umständen die Gefahr des Taschendiebstahls sieht. In der Tat ist Angst ein häufiger Grund, weshalb Hilfe verweigert wird (vgl. Pearson et al. 2009: 3f). Aus diesen Gründen wird von der folgenden Hypothese 1 ausgegangen:

H1: Es gibt einen Unterschied im Hilfsverhalten gegenüber einer deutschen und einer nicht-deutschen Person: Im Vergleich zu einer deutschen Person wird einer nicht-deutschen Person in einer alltäglichen Situation weniger Hilfsbereitschaft entgegengebracht.

Da bei der Durchführung des Experiments zur Einhaltung des Quotenplans auch das Geschlecht der Versuchspersonen berücksichtigt wird, bietet es sich an, dies in die statistische Analyse als erklärende Variable mit einfließen zu lassen.¹⁵

In der Literatur herrscht keine Eindeutigkeit bezüglich des Zusammenhangs von Geschlecht und Hilfsbereitschaft. So zeigt die 2016 durchgeführte Studie von Abdullahi/Kumar¹⁶, dass es einen signifikanten Unterschied zwischen Männern und Frauen gibt, was Bestandteile prosozialen Verhaltens betrifft. Frauen verfügen über signifikant mehr Einfühlvermögen und moralische Denkweisen. Dieser Unterschied in prosozialem Verhalten könnte nahelegen, dass Frauen auch in dem im Folgenden durchgeführten Experiment mehr Hilfe leisten.

Allerdings handelt es sich bei der Studie von Abdullahi/Kumar um einen Fragebogen, sodass es sich nur um die persönliche Einschätzung der Befragten handelt. Diese kann sich jedoch von den realen Verhaltensweisen unterscheiden. Tatsächlich ist die Möglichkeit eines Rückschlusses von antizipiertem auf reales Verhalten nicht empirisch belegt (vgl. Voigtländer 2008: 39-42). So hat in den von Voigtländer 2008 durchgeführten Experimenten das Geschlecht keinen Einfluss auf das Hilfsverhalten (vgl. ebd.: 56; 80).

Eine Metastudie von Toosi et al. 2012 untersucht 108 Studien aus 40 Jahren Forschung. Sie zeigen, dass das Geschlecht einen schwachen Effekt auf das Hilfsver-

¹⁵ Da die Versuchspersonen nach Durchführung des Experiments nicht aufgeklärt werden, kann auch ihre Geschlechtszugehörigkeit nicht erfragt werden, sondern muss durch die Beobachter anhand äußerer Kriterien in weiblich oder männlich eingeteilt werden. Die Autoren sind sich bewusst, dass es neben weiblichen und männlichen Personen auch solche gibt, die sich nicht ausschließlich mit einer dieser Kategorien identifizieren, oder sich trotz äußerer Merkmale mit dem anderen Geschlecht identifizieren. Die vom Beobachter notierte Geschlechtszugehörigkeit kann daher in Ausnahmefällen von der physischen oder der vom Individuum gefühlten Zugehörigkeit abweichen.)

¹⁶ Das Experiment wurde mit je 30 Frauen und Männern durchgeführt. In dem Index „prosoziales Verhalten“ fassen die Autoren die folgenden sieben Dimensionen zusammen: (1) social responsibility, (2) empathic concern, (3) perspective taking, (4) personal distress, (5) other oriented moral reasoning, (6) mutual concern moral reasoning und (7) self report altruism (vgl. Abdullahi/Kumar: 2016: 171). Für gewöhnlich ist Hilfsverhalten eine Unterkategorie von prosozialem Verhalten. Für eine genaue Abgrenzung von Hilfsverhalten, prosozialem Verhalten und Zivilcourage siehe Voigtländer 2008: 13-17.

halten einer Person anderer Hautfarbe hat. Waren beide Personen gleichen Geschlechts, wurde der Einfluss der Hautfarbe auf das Hilfsverhalten abgeschwächt. Allerdings waren die Effekte schwach und sollten im jeweiligen Kontext betrachtet werden. Im Laufe der Zeit haben nicht nur die durch Hautfarbe verursachten Unterschiede abgenommen, sondern auch die im Geschlecht begründeten Effekte (vgl. Toosi et al. 2012: 20).

Aufgrund dieser unterschiedlichen empirischen Befunde orientiert sich die im Folgenden ausgeführte Hypothese 2 an der theoretischen Literatur. Diese führt den Effekt des Geschlechts auf die Gruppenzugehörigkeit zurück. Die Lockvögel und die Testpersonen können sich im Hinblick auf zwei Gruppenzugehörigkeiten unterscheiden: ihre Hautfarbe und ihr Geschlecht. Wie bereits in Hypothese 1 vermutet, sollte die Hilfsbereitschaft gegenüber einer Person anderer Hautfarbe abnehmen, da diese der Fremdgruppe (outgroup) angehört. Dieser Effekt sollte abnehmen, wenn beide Personen gleichen Geschlechts sind, da sie somit auch über eine gemeinsame Gruppenzugehörigkeit (ingroup) verfügen. Andersrum sollte der Unterschied im Hilfsverhalten zunehmen, wenn LV und Testperson unterschiedlichen Geschlechts sind, da sie sich in beiden Merkmalen unterschiedlichen Gruppen zugehörig fühlen (vgl. Toosi 2012: 3). Da beide LV im Experiment männlich sind, wird folgende Hypothese angenommen:

H2: Das Hilfsverhalten unterscheidet sich je nach Geschlecht der Testperson: Die Hilfsbereitschaft weiblicher Testpersonen ist im Vergleich zu männlichen Testpersonen signifikant geringer.

(AS)

3. Operationalisierung und Durchführung

Zunächst soll dargelegt werden, welche Rolle Experimenten in den Sozialwissenschaften zukommt und die Frage behandelt werden, warum gerade ein experimentelles Design zur Untersuchung der vorliegenden Forschungsfrage geeignet ist. Hierzu wird insbesondere auf die Sonderform des Feldexperiments eingegangen.

3.1 Das Experiment in den Sozialwissenschaften

Im Forschungsprozess sind die experimentellen Methoden der Datenerhebung im Begründungszusammenhang verortet.

In einer vorläufigen Definition kann das Experiment als Beobachtung unter kontrollierten Bedingungen verstanden werden (vgl. Zimmermann 1972: 33). Ein Experiment ist immer auch die Überprüfung im Vorhinein festgelegter Aussagen nach festgelegten Bedingungen und keine willkürliche Variation von Faktoren (vgl. Atteslander 2010: 178). Weiterhin ist ein Höchstmaß der Kontrolle der sozialen Situation ein entscheidendes Kriterium in der Definition experimenteller Designs (vgl. ebd.: 177). Genauer gesagt dient aber das Experiment als Beweis für eine Kausalhypothese. Das Experiment ist eine „[w]iederholbare Beobachtung unter kontrollierten Bedingungen, wobei eine/mehrere unabhängige Variable/n derartig manipuliert wird /werden, daß eine Überprüfungsmöglichkeit der zugrunde liegenden Hypothese [...] in unterschiedlichen Situationen gegeben ist“ (Zimmermann 1972: 37). Nach Diekmann ist ein experimentelles Versuchsdesign unter Erfüllung folgender Bedingungen gegeben: Es werden mindestens zwei experimentelle Gruppen gebildet (1), denen die VPN nach einem Zufallsverfahren zugewiesen werden (2) und durch die Manipulation der unabhängigen Variable setzt der Forscher einen/mehrere experimentelle Stimuli (3) (vgl. Diekmann 2012: 337).¹⁷ Die Überprüfung mehrerer Variablen erlaubt demnach auch komplexe Versuchsanordnungen.

Doch wie findet das Experiment in den Sozialwissenschaften Anwendung? Gegenüber der Beobachtung bietet das Experiment den Vorteil, dass es über die Beschreibung eines Zustandes hinaus Kausalaussagen treffen kann (vgl. Zimmermann 1972: 35). Gegenüber Befragung und Beobachtung ist bei experimentellen Designs von Vorteil, dass VPN in einen künstlichen (oder beim Feldexperiment natürlichen) Prozess eingefügt werden. Dies erlaubt, Bedingungen konstant zu halten, sodass soziale Zusammenhänge unter (ständiger) Kontrolle dargestellt und reproduziert werden können. Positiv ist zudem, dass der experimentelle Stimulus der vermuteten Wirkung zeitlich vorausgeht, da er im Experiment geschaffen wird (vgl. Diekmann 2012: 349). So lassen sich nicht nur Hypothesen unter strengen Prüfbedingungen testen, sondern auch Extremsituationen konstruieren (vgl. Atteslander 2010: 179). Dies ist insbesondere für die hier zugrundeliegende Forschungsfrage von Bedeutung, bei der davon ausgegangen werden muss, dass es möglich ist, dass die Versuchspersonen sich ihrer Einstellungen gar nicht bewusst

¹⁷ Wobei die Gruppe mit Setzung des Stimulus als Versuchsgruppe, die Vergleichsgruppe als Kontrollgruppe bezeichnet wird. Der experimentelle Faktor kann dabei also in den Ausprägungen An-/Abwesenheit des gesetzten Stimulus auftreten (vgl. Diekmann 2012: 337f.).

sind (vgl. hierzu Kapitel 2.3).

Demnach ist das Experiment in der empirischen Sozialforschung eine wichtige Methode zur Überprüfung von Kausalbeziehungen im Bereich sozialer Phänomene. Diekmann bezeichnet das experimentelle Design gar als „ideales Versuchsaarrangement zum Test von Kausalhypothesen“ (ebd.: 330).

Dennoch ist das Experiment eine Forschungsmethode, die aus ethischen, theoretischen und forschungspraktischen Gründen selten in den Sozialwissenschaften zum Einsatz kommt.¹⁸ Wenngleich generell Fragen der Sozialforschung durch Experimente erforschbar sind, müssen auch die Nachteile des Experiments angeführt werden. Dazu zählen u.a. Probleme der externen und internen Validität der geprüften Zusammenhänge, Probleme der Reaktivität¹⁹ und ein hoher Aufwand bei (simultaner) Prüfung komplexer Zusammenhänge ebenso wie die Tatsache, dass viele soziale Zusammenhänge aus praktischen und/oder ethischen Gründen nicht randomisierbar sind (vgl. Diekmann 2012: 349). Zudem wird die soziale Realität nur unvollkommen berücksichtigt, da nur ein bestimmter Aspekt durch das Experiment isoliert wird und andere Faktoren keine Beachtung finden (vgl. Atteslander 2010: 189). Daher können auch experimentell bestätigte Aussagen ebenso wie Theorien keinen Anspruch auf Allgemeingültigkeit erheben und sind immer nur vorläufig verifiziert (vgl. ebd.). Es bleibt immer die Möglichkeit bestehen, dass ein Individuum nicht nach den rational vorausgesagten Kriterien handelt (vgl. ebd.: 191).²⁰

Um den durch experimentelle Designs gewonnenen Erkenntnissen daher eine möglichst hohe Genauigkeit und Gültigkeit zu verschaffen, sind neben der Einfüh-

¹⁸ Zimmermann konstatierte bereits in den 70er Jahren ebenso wie Atteslander und Diekmann zu Beginn der letzten Dekade eine vergleichsweise geringe Nutzung experimenteller Forschungsdesigns in den Sozialwissenschaften (vgl. Zimmermann 1972: 199; vgl. Atteslander 2010: 177; vgl. Diekmann 2012:349). Im Bereich der Vorurteils- und Diskriminierungsforschung mit experimentellen Designs ist in jüngerer Zeit insbesondere die Studie von Diekmann et al. zu erwähnen, bei welcher unterschiedliche Experimente durchgeführt wurden, um prosoziales und fremdenfeindliches Verhalten in der Schweiz zu erforschen (vgl. Diekmann et al 2014:185-199).

¹⁹ Reaktivität kann bspw. provoziert werden durch Interferenzen zwischen zeitlich versetzten Messungen, durch die Sensibilisierung aufgrund eines Pretests oder durch die Anwesenheit des Versuchsleiters.

²⁰ Campbell und Stanley haben in diesem Zusammenhang eine Liste von Störfaktoren aufgestellt, welche interne und externe Validität beeinträchtigen können (vgl. Campbell/Stanley zitiert nach Zimmermann 1972: 78-81). Dabei sind im hier vorliegenden Fall insbesondere auf Interferenzen und Reaktivität beruhende Problemquellen zu vernachlässigen, da ein verdecktes Feldexperiment durchgeführt wird, bei dem die jeweilige Versuchsperson nur einmal Teil des Untersuchungsdesigns ist. Auf die für den hier vorliegenden Versuchsaufbau problematisch erscheinenden Störfaktoren nach Campbell und Stanley verweist Tabelle 10 im Anhang.

rung einer Kontrollgruppe weitere Kontrolltechniken unabdingbar.²¹ Eine Vergrößerung der internen Validität (also die Ausblendung von Störvariablen) lässt sich zum einen dadurch erreichen, dass die Störquelle neutralisiert oder überlagert wird (vgl. Zimmermann 1972: 66f.). Zum anderen kann eine Zufallsaufteilung der Probanden auf die Versuchsgruppen vorgenommen werden, was die Kontrolle von Störfaktoren ermöglicht (vgl. Diekmann 2012: 339).²² Durch die Randomisierung wird der Einfluss der Drittvariablen neutralisiert: Von Zufallsfehlern abgesehen weisen die Drittvariablen in den Versuchsgruppen gleiche Verteilungen auf und sind statistisch berechenbar (vgl. Atteslander 2010: 186f.). Da bei Randomisierungen Selektionsverzerrungen nicht auftreten können, soll damit die Produktion verzerrender Scheineffekte durch unbekannte Drittfaktoren neutralisiert werden (vgl. Diekmann 2012: 339f.). Weiterhin gilt, dass mit steigender Zufallsstreuung umso eher eine Normalverteilung vorliegt (vgl. Zimmermann 1972: 70). Problematisch an diesem Verfahren ist jedoch, dass für die Untersuchung gewisser sozialer Zusammenhänge häufig keine Randomisierung möglich ist.

Eine weitere Möglichkeit der Kontrolle besteht im Matching: die Versuchs- und Kontrollgruppen werden bis auf die unabhängige Variable in allen relevanten Faktoren gleich zusammengesetzt, um Versuchseinheiten anzugleichen und den Einfluss unabhängiger Variablen im Experiment zu verringern (vgl. Atteslander 2010: 186f.). Beim Matching muss dem Forscher jedoch bekannt sein, welche Variablen kontrolliert werden sollen und bei welcher Kontrollvariable davon auszugehen ist, dass sie einen großen Einfluss auf die abhängige Variable hat (vgl. Zimmermann 1972: 73f.).

Wenngleich im Labor Störeffekte besonders gut zu kontrollieren sind, stellt sich bei Laborexperimenten die Frage nach der Generalisierbarkeit der Ergebnisse (vgl. Diekmann 2012: 344-346). Externe Validität lässt sich durch Feldexperimente in natürlichen sozialen Situationen schaffen (vgl. ebd.). Wichtiges Gütekriterium des Experiments ist daher auch die Wiederholbarkeit der Ergebnisse in

²¹ Gerade da die nicht-kontrollierbaren Faktoren im sozialwissenschaftlichen Experiment größer als in den Naturwissenschaften sind, kann ansonsten keine Erkenntnis über die Existenz des Kausalzusammenhangs und der Ursache der beobachteten Wirkung gewonnen werden. (vgl. Atteslander 2010: 186; Zimmermann 1972: 58f.).

²² Es existieren zwei Möglichkeiten, wie eine Zufallsverteilung vorgenommen werden kann: (1) Ziehen der Zufallsauswahl aus einer relevanten Grundgesamtheit und dann die Individuen nach dem Zufall auf die Kontroll-/Versuchsgruppe aufteilen. (2) Treffen zweier Zufallsauswahlen aus der Grundgesamtheit, wobei eine die Versuchsgruppe und die andere die Kontrollgruppe bildet (diese Strategie wird für den hier vorliegenden Fall angewendet) (vgl. Zimmermann 1972: 71).

Replikationen (vgl. Zimmermann 1972: 33).

Über diese gemeinsamen Ausgangskriterien hinaus existieren unterschiedliche Versuchsanordnungen sowie verschiedene Typen des Experiments, wobei im Folgenden auf das Feldexperiment näher eingegangen wird.²³

3.1.1 Feldexperiment

Im Gegensatz zum Laborexperiment, in dem in einer künstlichen Situation unter vereinfachten Bedingungen untersucht wird, ob ein Kausalfaktor die erwartete Wirkung provoziert, können im Feldexperiment (mind.) zwei unterschiedliche Gruppen in ihrer realen Umwelt untersucht werden, von denen eine dem (angenommenen) Kausalfaktor ausgesetzt ist (vgl. Atteslander 2010: 181-184).²⁴ Dabei müssen die Bedingungen so sorgfältig wie möglich kontrolliert werden, um die Milieuvorteile ausnutzen zu können. Auch für Feldexperimente gilt, dass gültige Messungen die Bildung von Versuchs- und Kontrollgruppen, Randomisierung, Replikation der Messung und einen Pretest benötigen. Beim letzterem ist ebenfalls darauf zu achten, dass keine reaktiven Effekte auftreten (vgl. ebd.).

Durch Blindtests²⁵ sollen Artefakte vermieden werden, welche durch die Kenntnis der Hypothese auftreten können. So z.B. Selbstsuggestion oder Verhaltensänderungen, welche self-fulfilling bzw. self-destroying prophecies provozieren können (vgl. Diekmann 2012: 338; vgl. Atteslander 2010: 188-190). Die Effekte durch die Präsenz des Forschers verschwinden beim Feldexperiment, da dieser nicht als solcher in der natürlichen Umgebung erkennbar ist. Dies erlaubt, Hypothesen zu testen, ohne die Situation zu modifizieren, da die Versuchsperson die Hypothese nicht kennt. Damit bietet das Feldexperiment den immensen Vorteil der Non-Reaktivität – was insbesondere bei heiklen und intimen Fragen, bei denen es durch die Normen sozialer Erwünschtheit zu Verzerrungen kommen kann, von großer Bedeutung ist.²⁶ Damit lassen sich bezüglich des tatsächlichen (prosozia-

²³ Zimmermann unterscheidet Labor-, Feld-, Natur-, Gedankenexperimente und Simulationen (vgl. Zimmermann 1972: 185-215); Atteslander unterscheidet neben Laboratoriums- und Feldexperimenten zudem zwischen projektiven und ex-post-facto Verfahren sowie simultanen und sukzessiven Experimenten ebenso wie Simulationen und Planspielen (vgl. Atteslander 2010: 181-184).

²⁴ Ist die Manipulation einer Variable durch den Versuchsleiter - Bedingung für experimentelle Tests von Hypothesen - unmöglich, handelt es sich um eine Feldstudie (vgl. Zimmermann 1972: 196).

²⁵ Unter einem Blindversuch versteht man Experimente, in denen Probanden weder ihre Zugehörigkeit zur Versuchsgruppe noch die zu überprüfende Hypothese kennen. In Doppelblindversuchen wird zudem der Versuchsleiter weitestgehend in Unkenntnis gelassen, um unbewusste Beeinflussungen zu vermeiden bzw. auf die Gruppen gleich zu verteilen (vgl. Diekmann 2012: 338).

²⁶ Neben dem Feldexperiment lassen sich weitere nichtreaktive Erhebungsmethoden nennen, wie z.B. unaufdringliche Beobachtung, Verhaltensspuren, Inhaltsanalysen, Sekundäranalysen prozess-

len) Verhaltens in der vom Forscher erzeugten Situation Zusammenhänge testen, die durch reaktive Erhebungsmethoden wie der Befragung nicht verlässlich überprüft werden könnten (vgl. Diekmann 2012: 361f.).²⁷ Damit haben Feldexperimente gegenüber Laborexperimenten einen Vorteil in Sachen externer Validität. Weitere Vorteile des Feldexperiments sind, dass sie bei einer Randomisierung zwischen Experimental- und Kontrollgruppe alle Vorteile des experimentellen Designs aufweisen und dadurch auch intern valide sind (vgl. ebd.: 633f.).

Problematisch ist jedoch, dass sich im Feldexperiment die Kontrolle alternativer Variablen schwierig darstellt und die Wechselbeziehungen mehrerer unabhängiger Variablen weniger gut zu kontrollieren sind. Da bei Feldexperimenten häufig mehrere Variablen auf einmal manipuliert werden, ist aber dabei oft unklar, auf welche spezifischen unabhängigen Variable der Effekt zurückzuführen ist. Daher bieten sich für komplexe Phänomene Feldexperimente an, für detaillierte Untersuchungen der einzelnen unabhängigen Variablen jedoch das Labor (vgl. Zimmermann 1972: 197). Bei Feldexperimenten muss jedoch kritisch hinterfragt werden, ob im Feld zwischen Versuchs- und Kontrollgruppe randomisiert werden kann und damit die externe Validität gewährleistet ist. Desweiteren muss untersucht werden, ob sich in der Realität gute Indikatoren für die Variablen finden lassen, oder ob sich dies auf die Validität der Messung auswirkt (vgl. ebd.: 634; 655f.).²⁸ Zudem bedeuten Feldexperimente für den Forscher einen erhöhten Zeit- und Kostenaufwand (vgl. Zimmermann 1972: 200f.). So können z.B. im Labor gemachte Erfahrungen nur bedingt übernommen werden und es müssen neue Strategien gefunden werden, um passende Kontrollbedingungen zu gewährleisten. Ebenfalls muss der Forscher in Betracht ziehen, dass die theoretisch leicht manipulierbare Variable in der Praxis nur schwer zu modifizieren ist (vgl. ebd.: 200f.).

3.1.2 Ethische Aspekte experimenteller Untersuchungsdesigns

Da Feldexperimente, insbesondere verdeckte und nichtreaktive - wie in der hier vorliegenden Studie angestrebt - nicht das Prinzip des Informed consent beachten, seien hier im Folgenden noch ethische Aspekte der experimentellen Sozialfor-

produzierter Daten (vgl. Diekmann 2012: 655).

²⁷ Hierzu zählen z.B. Hilfeleistungen (da sozial erwünscht).

²⁸ Hier kann die Kombination verschiedener reaktiver und nichtreaktiver Verfahren dazu beitragen, Informationslücken zu überbrücken, wobei die Kombination mehrerer Erhebungsverfahren aussagefähigere Schlüsse gibt, als die ausschließliche Verwendung einzelner Erhebungsmethoden (vgl. Diekmann 2012: 640-642).

schung dargelegt. Es versteht sich von selbst, dass Versuchspersonen nicht riskanten und/oder peinlichen Situationen ausgesetzt werden dürfen oder Situationen, die gar eine Beeinträchtigung der Psyche oder Gesundheit nach sich ziehen könnten (vgl. Diekmann 2012: 655f.).²⁹ Inwiefern jedoch im Rahmen eines Experiments getäuscht werden darf, ist strittig und variiert von forschungspragmatischen Argumenten bis hin zu dogmatischen Einwänden. Häufig wird argumentiert, dass Täuschungen dann gerechtfertigt sind, wenn sie der wissenschaftlichen Forschung dienen – wie die Grenze zwischen wissenschaftlichem Nutzen und Kosten für die Probanden zu ziehen oder wie diese Aspekte zu definieren sind, bleibt jedoch weiterhin diskutabel – insbesondere da die Fernwirkungen der Täuschung für den Forscher nicht einschätzbar oder kontrollierbar sind (vgl. Zimmermann 1972: 271). Hier ist die Faustregel interessant, dass Feldexperimente dann ethisch problematisch sind, wenn sie über die Simulation von Alltagssituationen hinausgehen und soziale Normen und Distanzregeln verletzen (vgl. Diekmann 2012: 361f.; Atteslander 2010: 190f.). Dieses Kriterium kann für das hier vorliegende Experiment jedoch vernachlässigt werden, da das Wechseln von Geld ein alltäglicher Vorgang ist.³⁰ Weiterhin wird auf die große Bedeutung der Aufklärung nach der Durchführung des Experiments hingewiesen, einerseits um den VPN aber auch der zukünftigen Wissenschaft so wenig wie möglich zu schaden (vgl. Zimmermann 1972: 271). Die Aufklärung der VPN kann jedoch im Rahmen des hier vorliegenden Experiments nicht geleistet werden, da die potenziellen Verzerrungen durch Reaktivität zu hoch wären und der damit verbundene Zeitfaktor das Projekt unmöglich machen würde.

(FM)

3.2. Experimentaldesign

Wie bereits angeführt, verfolgt das nachstehende nicht reaktive Feldexperiment das Ziel, xenophobe Einstellungen in der Stuttgarter Bevölkerung zu messen. Im Folgenden wird der Versuchsaufbau erläutert.

Der Ort der Versuchsdurchführung ist das U-Bahn Geschoss des Charlottenplatzes in Stuttgart. Hier werden Passanten darum gebeten, einen 5€-Schein für den Kauf

²⁹ Man denke hierbei an die Folgen des Milgram-Experiments für die Versuchspersonen.

³⁰ Weitere ethische Konflikte wie Auftraggebereffekte oder Zielkonflikte durch konfligierende Interessenlagen sind im Rahmen dieser Forschungsarbeit nicht gegeben und werden daher in den Ausführungen vernachlässigt.

einer Fahrkarte an einem Automaten in Münzgeld zu wechseln. Die Hilfsbereitschaft der Stuttgarter Bevölkerung wird demnach über die Bereitschaft, einer fremden Person Geld zu wechseln, operationalisiert. Die Operationalisierung der Xenophobie wiederum erfolgt über zwei unterschiedliche „Lockvögel“ (LV), welche die Passanten unabhängig voneinander ansprechen - eine ausländische „Experimentalperson“ (EP) und eine deutsche „Kontrollperson“ (KP). Xenophobe Einstellungen sollten sich demnach durch signifikante Unterschiede in der Hilfsbereitschaft gegenüber der EP im Vergleich zur KP ergeben. Neben der Variation der dichotomen Variablen „Lockvogel“ wurden außerdem das Geschlecht sowie das ungefähre Alter der Passanten (VPN) erfasst. Um ausschließen zu können, dass das Experiment an einer ungenügenden Teststärke scheitert, wurde im Vorfeld der Erhebung eine Mindestgröße für Stichprobe bestimmt. Des Weiteren wurde ein Quotenplan für die anzusprechenden VPN erstellt, um zu versuchen ein möglichst repräsentatives Bild der Stuttgarter Bevölkerung zu erhalten. Zunächst wird auf diese beiden Punkte näher eingegangen, bevor der Ort und der Zeitpunkt, die Wahl der Lockvögel sowie der detaillierte Ablauf des Experiments erläutert werden.

3.2.1 Bestimmung der Stichprobengröße

Die Bestimmung der Stichprobengröße erfolgte nach Anleitung von List et alii für Experimente mit dichotomen Treatment (KP vs. EP) und binominalem Outcome (Geholfen vs. Nicht-Geholfen). Hierfür wird eine erwartete Effektstärke benötigt (vgl. List et al. 2010: 13.f). Es wurde daher auf eine Studie von Klink und Wagner aus dem Jahr 1999 zurückgegriffen, in der 14 verschiedene Feldexperimente zur Messung von Diskriminierung von Ausländern in Deutschland durchgeführt wurden (vgl. Klink/Wagner 1999: 412f.). Insgesamt haben die Autoren eine signifikante ($p < 0.5$) overall Korrelation von $r = 0.23$ zwischen ethnischer Herkunft und Diskriminierung gefunden. Wenn unter den 14 Experimenten nach direktem und indirektem Kontakt mit den Lockvögeln unterschieden wird, zeigt sich bei direktem Kontakt sogar eine noch stärkere signifikante Korrelation von $r = 0,34$ (vs. 0.17 bei indirektem Kontakt) (vgl. Klink et al. 1999: 416f.). Im Vorfeld der vorliegenden Untersuchung wurde daher von einem ähnlich starken Zusammenhang mit einer Effektstärke von 0,2 ausgegangen. Bei einer gewünschten Power von 80% und auf einem Signifikanzniveau von 95% beträgt die benötigte Mindeststichpro-

bengröße somit pro Lockvogel jeweils 93 VPN und insgesamt 186 VPN.

In Bezug auf die abhängige Variable „Hilfsbereitschaft“ ergaben sich beim Pretest, welcher am Sonntag, den 28. Februar 2016, durchgeführt wurde, jedoch einige Änderungen. Es zeigte sich, dass für ein exaktes Erfassen der Reaktionen der VPN mehr als zwei Ausprägungen notwendig sind. In der Folge war es aber schwierig, die Konsequenzen dieser Änderung im Experimentaldesign auf die Bestimmung der Mindestgröße für die Stichprobe einschätzen zu können. Für den Fall, dass die genauere Ausdifferenzierung der abhängigen Variablen zu einer Abschwächung der Effektstärke führt, wäre eine Erhöhung der Stichprobengröße notwendig gewesen. Dies wäre allerdings aufgrund der begrenzten finanziellen Mittel der Forschergruppe ohnehin in keinem Fall möglich gewesen. Es wurde daher weiter mit der bereits berechneten Samplegröße gearbeitet, um zumindest für eine dichotome abhängige Variable eine ausreichende Teststärke zu gewährleisten.

3.2.2 Erstellung des Quotenplans

Der Quotenplan wurde nach Malhotras und Birks Anleitung zur Erstellung einer Quotenstichprobe entworfen. Diese schlagen vor, in einem ersten Schritt zentrale Charakteristika der Grundgesamtheit zu identifizieren. In einem zweiten Schritt werden die VPN auf Grundlage dieser Charakteristika ausgewählt, sodass die Stichprobe ein repräsentatives Abbild der Grundgesamtheit darstellt (vgl. Malhotra/Birks 2007: 412 ff.).

Für die vorliegende Studie wurde daher auf Statistiken über die Stuttgarter Gesamtbevölkerung zurückgegriffen. Als zentrale Charakteristika wurden das Geschlecht und das Alter determiniert, da diese bei den VPN für verdeckte Beobachter verhältnismäßig gut zu erkennen sind. Für die Bestimmung der genauen Quoten wurde erstens eine Statistik der weiblichen Einwohner nach Altersgruppen und zweitens eine Statistik der männlichen Einwohner nach Altersgruppen verwendet. Beide stammen von dem Statistischen Amt der Stadt Stuttgart und beziehen sich auf das Erhebungsjahr 2015. Während die Gesamtverteilung der Einwohner nach Geschlecht fast 50% beträgt - 50,4% weiblich, 49,6% männlich (vgl. Statistisches Amt Stuttgart 2016a und b: o.S.) - ist die Einteilung nach Altersgruppen erwartungsgemäß unausgeglichener. Tabelle 2 und 3 fassen diese zu-

sammen.³¹

Tabelle 1: Frauen in Stuttgart nach Altersgruppen (2015)

Altersgruppen	18 - 29 Jahre	30 - 64 Jahre	≥ 65 Jahre	
Einwohneranzahl	53 996	141 036	62 560	257 592
Anteil in Prozent	21 %	54,7 %	24,3 %	100 %

Quelle: Statistisches Amt Stuttgart 2016a: o.S.; Tabelle von den Autoren zusammengestellt

Tabelle 2: Männer in Stuttgart nach Altersgruppen (2015)

Altersgruppen	18 - 29 Jahre	30 - 64 Jahre	≥ 65 Jahre	
Einwohneranzahl	55 836	149 627	46 934	252 397
Anteil in Prozent	22,1 %	59,3 %	18,6 %	100 %

Quelle: Statistisches Amt Stuttgart 2016b: o.S.; Tabelle von den Autoren zusammengestellt

Wie bereits beschrieben, wurde für die Gesamterhebung die Mindestfallzahl auf 186 (je LV 93 VPN) festgelegt. Um den Quotenplan zu erfüllen und außerdem gut auf die drei Erhebungstage aufteilen zu können, wurde diese auf 192 (je LV 96 VPN) angehoben. Tabelle 4 zeigt den dazugehörigen Quotenplan für einen LV.³²

Tabelle 3: Quotenplan für einen LV

	Männlich	Weiblich	Gesamt
18 - 29 Jahre	12	9	21
30 - 64 Jahre	27	27	54
≥ 65 Jahre	9	12	21
Gesamt	48	48	96

Gruppen von Passanten, die gemeinsam am Ort der Versuchsdurchführung eintreffen, können wegen möglicher Verzerrungen in Bezug auf die Hilfsbereitschaft wie z.B. durch den Bystander- oder den Bühnen-Effekt nicht im Experiment berücksichtigt werden.³³ Da in dem vorgestellten Experiment vorrangig xenophobe

³¹ Die Altersgruppe 0 - 18 Jahre wurde nicht berücksichtigt, da aus ethischen Gründen nur volljährige Passanten als VPN in Frage kommen.

³² Die Prozentzahlen aus Tabelle 2 wurden für die erste und die letzte Altersgruppe leicht angehoben, damit diese nach Geschlecht und „Lockvogel“ zu 50% aufgeteilt werden konnten.

³³ Der Bystander-Effekt beschreibt das Phänomen, dass die Hilfsbereitschaft einer einzelnen Person in der Anwesenheit von Dritten aufgrund von Verantwortungsdiffusion abnimmt (vgl. Latané/Darley 1968: 90f.). Der Bühnen-Effekt beschreibt die gegenteilige Situation, wenn Personen in Anwesenheit von Dritten die Chance erkennen, sich mit prosozialem Verhalten gut zu präsentieren

Einstellungen gemessen werden sollen, die sowohl durch den Bystander- als auch durch den Bühnen-Effekt verzerrt werden könnten, kommen nur einzelne Passanten als VPN in Frage. Zudem würde die Berücksichtigung von Gruppen ein Einhalten des Quotenplans erschweren, da das Hilfsverhalten nicht mehr eindeutig auf eine VPN zurückgeführt werden könnte.³⁴

3.2.3 Ort und Zeitpunkt der Versuchsdurchführung

Das U-Bahn Geschoss am Charlottenplatz wurde primär aufgrund seiner sehr zentralen Lage gewählt. Der Charlottenplatz liegt in dem Stuttgarter Stadtteil „Stadt Mitte“. In diesem wohnen zwar nur 3,8% der Einwohner des gesamten Stadtgebiets (vgl. Statistisches Amt Stuttgart 2015: 23), jedoch ist die U-Bahn-Station Charlottenplatz mit dem Anschluss an neun Stadtbahnen der größte Knotenpunkt des Stuttgarter Liniennetzes und es herrscht dementsprechend viel Durchgangsverkehr (vgl. VVS 2016: o.S.). Der zweite Grund für die Wahl des U-Bahn Geschosses sind die für die Durchführung des Experiments sehr günstigen räumlichen Gegebenheiten. Der Fahrscheinautomat steht etwas abseits vom Bahnsteig und den auf die U-Bahn wartenden Fahrgästen, wodurch die Identifizierung einzelner VPN relativ einfach erscheint. Darüber hinaus ist es für eingeweihte Beobachter gut möglich, den Versuchsablauf zu verfolgen und zugleich eine gewisse Distanz zum Geschehen zu wahren.

Als Zeitpunkt für die Datenerhebung wurden drei Sonntage festgelegt, da sich während des Pretests zeigte, dass sonntags weniger Hektik herrscht als an den anderen Tagen der Woche und dadurch die Identifizierung von VPN erleichtert wird. Bei der Festlegung der Termine wurde außerdem darauf geachtet, dass nicht zugleich eine Großveranstaltung wie eine Demonstration in der Nähe des Charlottenplatzes stattfand, um eine Verzerrung der Selektion der VPN durch einen derart unkontrollierbaren Faktor zu vermeiden. Die Datenerhebung fand nach dem Pretest im Februar schließlich an drei Sonntagen, dem 22. Mai, dem 5. Juni und dem 12. Juni 2016 von 09:00 bis 18:00 Uhr statt. Das Experiment wurde stets den gesamten Vormittag über mit einem der beiden LV durchgeführt, am Nachmittag dann mit dem entsprechend anderen LV. Somit wurde versucht, Verzerrungen an

ren. Insbesondere bei Männern scheint der Bühnen-Effekt aufgrund von in der Gesellschaft verankerten Rollenerwartungen häufig aufzutreten (vgl. Alfermann 1996: 130).

³⁴ Im Pretest ergab sich die Situation, dass in Gruppen mehrere Personen gleichzeitig in ihren Geldbörsen nach Münzgeld suchten.

einem bestimmten Sonntag (beispielsweise durch ein zwischenzeitlich entstandenes mediales Ereignis) für beide LV konstant zu halten. Mit der KP wurde daher an zwei Vormittagen und einem Nachmittag Daten erhoben, mit der EP dementsprechend an einem Vormittag und zwei Nachmittagen.

3.2.4 Auswahl und Kleidung der „Lockvögel“

Die EP ist ein dreiundzwanzig jähriger Geflüchteter aus Eritrea, der Deutsch auf A2-Niveau spricht. Die KP ist ein ebenfalls dreiundzwanzig jähriger deutscher Studierender ohne Migrationshintergrund. Beide sind circa 1,80 m groß und tragen einen kurzen Bart. Ihre Kleidung sind ein unbedrucktes, weißes, langärmeliges T-Shirt sowie eine blaue Jeans³⁵. Howlett et alii stellen in einem Experiment fest, dass bereits Detailunterschiede in der Kleidung zu unterschiedlichen spontanen Einschätzungen von unbekannten Personen führen (vgl. Howlett u.a. 2013: 45). Aus diesem Grund wird in der vorliegenden Studie großen Wert daraufgelegt, den Kleidungsfaktor konstant zu halten. Die Wahl des weißen T-Shirts und der blauen Jeans beruht auf Ergebnissen einer Studie von Madden et alii über die Wahrnehmung von Farben in verschiedenen Kulturkreisen. Madden et alii stellten fest, dass die Farben blau, grün und weiß in allen untersuchten Ländern³⁶ mit den Adjektiven „peaceful“, „gentle“ und „calming“ assoziiert werden (vgl. Madden u.a. 2000: 97). Diese Assoziationen erscheinen für den Versuch vorteilhaft. Einerseits werden beiden LV die besten Chancen auf Hilfsbereitschaft eingeräumt und andererseits das Experiment nicht durch starke von anderen Farben hervorgerufene Emotionen verzerrt.³⁷

³⁵ Bei der Durchführung des Pretests im Februar wurde nicht auf die Wahl der Kleidung geachtet.

³⁶ Die untersuchten Länder waren Österreich, Brasilien, Kanada, Kolumbien, Hong Kong, die Vereinigten Staaten, Taiwan und die Volksrepublik China (vgl. Madden et al. 2000: 94).

³⁷ Beispielsweise wird die Farbe Rot als „active“, „hot“ und „vibrant“ empfunden (vgl. Madden et al. 2000: 98).

Abbildung 1: LV am 1. Erhebungstag (links die KP, rechts die EP)



3.2.5 Beschreibung des Versuchsaufbaus

Vor Beginn des Experiments platziert sich der LV vor dem Fahrscheinautomaten und täuscht vor, sich über den Kauf einer Fahrkarte informieren zu wollen. Der 5€-Schein befindet sich bereits in seiner Hand. Mögliche VPN können über zwei Zugangswege in die Nähe des LVs gelangen: entweder von oben über die ins Tiefgeschoss führende Treppe oder von dem Bahnsteig mit einer neu ankommenden U-Bahn. Beide Zugänge werden von drei verdeckten Beobachtern an unterschiedlichen Positionen im Untergeschoss observiert. Abbildung 2 veranschaulicht den räumlichen Versuchsaufbau.

Die Aufgaben der Beobachter umfassen zunächst das unabhängige Notieren der Merkmale der VPN („Alter“ und „Geschlecht“) und der „Hilfsbereitschaft“. Die „Hilfsbereitschaft“ wurde über vier Ausprägungen erfasst:

„Keine oder ablehnende Reaktion“	Die VPN beachtet den LV nicht, unterbricht nicht ihren Gang, schüttelt eventuell ablehnend den Kopf.
„Reaktion ohne Hilfeleistung“	Die VPN unterbricht, wenn auch nur kurz, ihren Gang und reagiert beispielsweise mit „Nein, leider nicht“.
„Andere Hilfeleistung“	Die VPN unterbricht ihren Gang und leistet Formen der Hilfe, ohne dabei die eigene Geldbörse hervor zu holen. Z.B. Verweis auf einen anderen Automaten im Tiefgeschoss oder Versuch, selbst mit dem 5€-Schein ein Ticket zu lösen.
„Volle Hilfeleistung“	Die VPN holt ihren Geldbeutel hervor, um nach Wechselgeld zu suchen.

Weiterhin achten die Beobachter auf Reaktionen von umstehenden Passanten an der U-Bahnstation, um erkennen zu können, ob diese das Experiment beobachteten und beispielsweise durch frühzeitiges Warnen der VPN eingreifen. Die Beobachtung des Umfelds erscheint insbesondere auch aus ethischen Gründen notwendig. Es wird die Gefahr gesehen, dass von Drittpersonen die Systematik, mit der die VPN von dem LV angesprochen werden, erkannt und daraus fälschlicherweise der Rückschluss gezogen wird, Zeugen einer kriminell motivierten Aktivität (z.B. einer neuen Masche von Taschendieben) zu werden. Dies ist vor allem in Bezug auf eventuell vorhandene Vorurteile gegenüber der EP nicht wünschenswert. Schlussendlich achtet der Beobachter B1 auf die Einhaltung des Quotenplans und gibt durch Husten das Signal für den Versuchsbeginn. Während ein einmaliges Husten ein Ankommen der VPN aus Richtung des Bahnsteigs ankündigt, bedeutet ein zweimaliges Husten ein Ankommen aus Richtung des Treppenaufgangs an. Das Husten erfolgt, sobald die identifizierte VPN einen zuvor festgelegten Radius um den Fahrscheinautomaten überschreitet. Sollten sich zwei für den Quotenplan relevante VPN aus den beiden entgegengesetzten Richtungen gleichzeitig nähern, soll die von B1 durchgeführte Zufallsauswahl durch eine vorher festgelegte Regel gewährleistet werden. An dem ersten und dritten Sonntag werden immer die von rechts ankommenden VPN bevorzugt, an dem zweiten immer die von links ankommenden VPN. Auf einen weiteren Zufallsgenerator für die Auswahl der VPN wurde aus praktischen Gründen verzichtet. Es wird davon ausgegangen, dass die Präsenz der VPN an dem Versuchsort bereits zufällig erfolgt und der Quotenplan ausreichend die Repräsentativität der Stichprobe garantiert.

Nach dem Startsignal orientiert sich der LV in Richtung der ankommenden VPN und spricht diese mit den Worten, „Entschuldigung, können Sie mir bitte 5€ wechseln?“ an. Fragt die VPN nach einer Begründung für das Wechseln, erklärt der LV, „Der Automat nimmt den Schein nicht.“ In regelmäßigen Abständen (jeweils nach 10 Fällen) wurde das Experiment unterbrochen und die Beobachtungen verglichen. Generell wurde immer die Meinung von B1 privilegiert gewertet, da dieser die beste Sicht auf den Versuchsablauf hatte. Falls sich die Reaktion der VPN jedoch hinter dem Fahrscheinautomaten abspielte, galt die Beobachtung von B2 als am korrektesten. Die Beobachtungen von B3 waren vernachlässigbar, da der Abstand zum Geschehen meistens zu groß für ein präzises Einschätzen der

Abbildung 2: Schematische Darstellung der Experimentalsituation



- Bedingt durch den Versuchsaufbau wurden Gruppen von mehreren Passanten nicht als VPN berücksichtigt. Dies diente, wie in 3.2 beschrieben, der Vermeidung des Bystander- und Bühneneffekts. In der Konsequenz können die Ergebnisse des Experiments allerdings auch nicht ohne weiteres auf eine reale für ausländische Personen auftretende Hilfssituation übertragen werden, da es möglich ist, dass diesen trotz tatsächlich vorhandener xenophober Einstellungen einzelner Personen in der Gruppe geholfen wird. Eine Untersuchung mit Einbeziehung von Gruppen als VPN bleibt weiteren Studien vorbehalten.

(NH)

3.2.7 Durchführung des Experimentaldesigns

Wie bereits beschrieben fand die Datenerhebung von 09 Uhr bis ca. 18 Uhr an drei Sonntagen in der U-Bahn Haltestelle des Charlottenplatzes statt. Trotz der getroffenen Vorkehrung, im Laufe eines Tages die Datenerhebung mit beiden LV durchzuführen, kann nicht gänzlich ausgeschlossen werden, dass aufgrund der Wetterlage Verzerrungen auftraten, da bei Regen generell weniger Personen anzutreffen waren. Da während unserer Durchführung jedoch nur an einem Tag leichte Unterschiede zwischen Vor- und Nachmittag auftraten kann jedoch hier davon ausgegangen werden, dass dies keine Auswirkungen auf die Ergebnisse der Erhebung hatte.³⁸ Zudem fanden während der Erhebungsphase keine medialen Ereignisse statt, die Verzerrungen während der Erhebungsphase befürchten lassen.

Jedoch traten bei der Auswahl der VPN einige generelle Probleme auf: So war es schwieriger, im Vormittag junge Leute anzutreffen und nachmittags hingegen wiederum ältere Personen. Generell war es den ganzen Tag über schwierig, Frauen (insbesondere der mittleren Alterskategorie) anzutreffen. Dennoch konnte an allen Tagen der Quotenplan erfüllt werden. Zudem kam es vor, dass einige Personen mehrmals am Tag die U-Bahn Station passierten. In zwei Fällen musste daher die Durchführung des Experiments aufgelöst werden, wobei nach jeder Unterbrechung eine längere Pause eingelegt wurde, um Probleme der Reaktivität zu vermeiden.³⁹ Weiterhin haben die beiden LV zum Teil unterschiedliche Reaktionen

³⁸ Der 22.05.2016 war den ganzen Tag sonnig und am 05.06.2016 wechselten sich Sonne und Regen über den Tag verteilt ab, alleine am 12.06.2016 war der Vormittag überwiegend verregnet und am Nachmittag zunehmend sonnig.

³⁹ Die Auflösung des Experiments schien in diesen Fällen sowohl aus ethischen als auch aus forschungspraktischen Gründen unabdingbar, um die Integrität von der LV und den erhobenen Daten nicht zu gefährden.

erfahren, die sich nicht im Beobachtungsbogen widerspiegeln: So wurde der KP mehrmals seitens der VPN initiativ Hilfe angeboten und es wurde sich seltener von ihm beim Herausholen des Geldbeutels abgewendet. Von der EP hingegen wurde sich häufiger beim Zücken des Geldbeutels weggedreht. Andererseits wurde der EP sogar ein Euro geschenkt, um den Betrag wechseln zu können und ihm sogar angeboten, das Ticket zu schenken.

Ungeachtet der oben geschilderten Vorkommnisse verlief die Erhebung jedoch wie im Pretest vorhergesehen und kann als problemlos bezeichnet werden.

(FM)

4. Auswertung der Ergebnisse

Im Folgenden wird die Auswertung der erhobenen Daten in Hinblick auf die gestellten Hypothesen vorgenommen. Dazu werden in einem ersten Teil beschreibende Auswertungen vorgenommen. In einem zweiten Teil der Analyse wird schließlich eine multinominale logistische Regression durchgeführt.

4.1. Deskriptive statistische Analyse

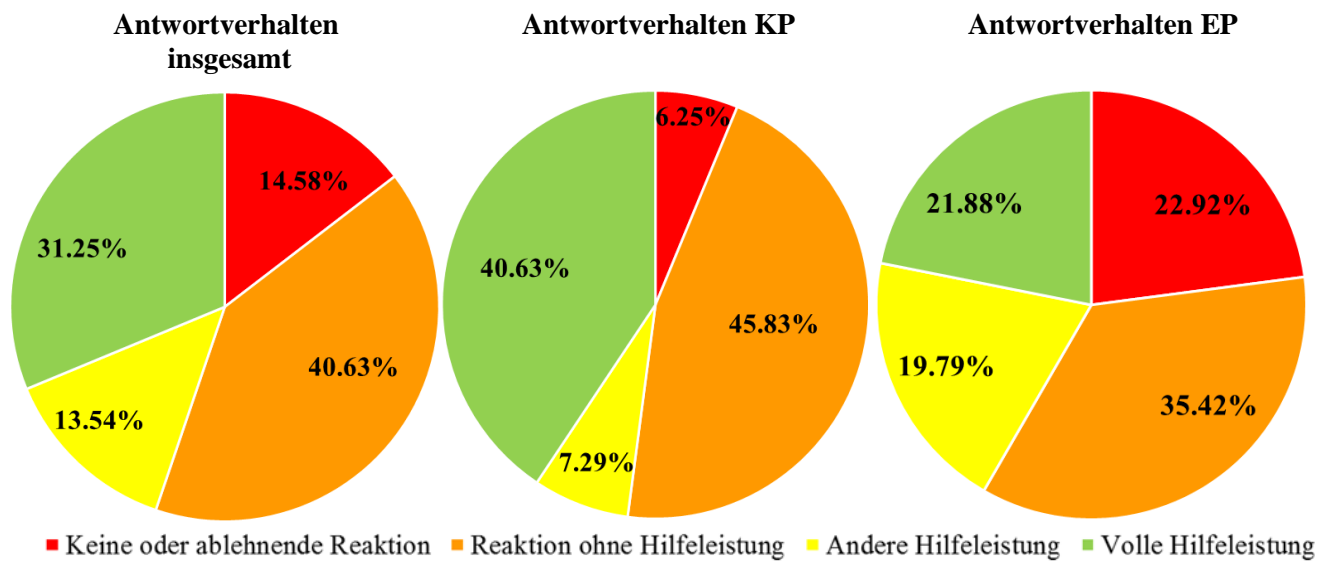
Im Rahmen der beschreibenden Analyse werden Aussagen über das Antwortverhalten mit Hinblick auf Lockvogel und Geschlecht auf Basis der Häufigkeiten getroffen und in einem weiteren Schritt die Korrelationen für die unabhängigen Variablen von Lockvogel und Geschlecht sowie der abhängigen ordinalen Variable des Antwortverhaltens berichtet.

4.1.1 Häufigkeiten nach LV

Die untenstehenden Kreisdiagramme geben einen Überblick über die Reaktionen, welche die LV auf ihre Bitte, den 5 Euro-Schein zu wechseln, erhielten.

Die Angaben werden hier für beide LV zusammen gemacht, um einen generellen Überblick über die Verteilung in den Antwortkategorien zu bekommen und dann für die jeweiligen LV einzeln dargestellt.

Abbildung 3: Antwortverhalten insgesamt und pro LV in Prozent



Betrachtet man in einem generellen Überblick das Antwortverhalten für die beiden LV insgesamt, stellt man eine Hauptverteilung in zwei Kategorien fest. So waren die überwiegenden Formen der Hilfeleistung der Kategorien *Reaktion ohne Hilfeleistung* (40,63%) und *volle Hilfeleistung* (31,25%). Betrachtet man zudem die beiden anderen Kategorien, lässt sich feststellen, dass 14,58% der angesprochenen VPN ablehnend oder gar nicht reagieren, immerhin aber noch 13,54% andere Formen der Hilfe anboten. Damit können insgesamt immerhin 44,79% der Reaktionen als Form der Hilfbereitschaft verbucht werden und etwas über die Hälfte des Antwortverhaltens (55,21%) fallen unter die Kategorie keiner Form der Hilfe.

Das Antwortverhalten gegenüber der KP weicht jedoch deutlich von diesem Muster ab. Auffällig ist hier, dass die Kategorien der Formen der Hilfbereitschaft *Andere Hilfe* und *Keine oder ablehnende Reaktion* deutlich unter dem Durchschnitt und weit unter den Häufigkeiten für die EP liegen. Vielmehr entfallen die Mehrheit der Reaktionen hier auf die Kategorien *Volle Hilfeleistung* (40,63%) und *Reaktion ohne Hilfeleistung* (45,83%). Festzuhalten ist hier insbesondere der hohe Anteil der Reaktionen, der auf die Kategorie *Volle Hilfeleistung* entfällt und den höchsten Wert für diese Kategorie in der Erhebung darstellt.

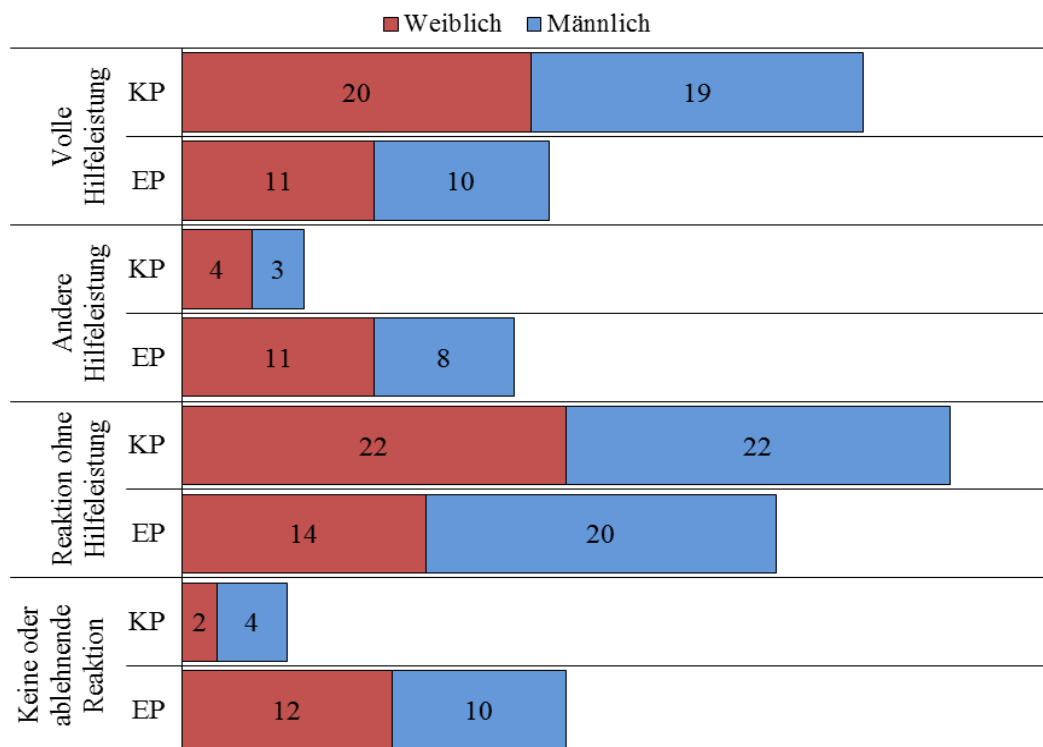
Das der EP entgegengebrachte Antwortverhalten wiederum stellt beinahe eine Gleichverteilung in allen Kategorien dar – auffällig sind hier insbesondere die relativ hohen Häufigkeiten für die Kategorien *Keine oder ablehnende Reaktion*

(22,92%) und *Andere Hilfeleistungen* (19,79%). Damit liegen die Werte dieser Kategorien weit über den Durchschnittswerten und denen der KP.

Mit Hinblick auf die Hypothese zur Hilfsbereitschaft lässt sich festhalten, dass der EP weniger häufig die volle Hilfeleistung entgegengebracht wird. Zudem sind insbesondere die Unterschiede in den Kategorien *Andere Hilfeleistung* und *Keine oder ablehnende Reaktion*. Denn wenngleich die Kategorie *Reaktion ohne Hilfeleistung* in der Untersuchung stets am häufigsten auftritt, wird deutlich, dass die EP deutlich häufiger ablehnende Reaktionen auf ihre Bitte erhält, als die KP. Es kann aber gleichfalls festgehalten werden, dass der EP deutlich häufiger andere Formen der Hilfe zugutekamen, als der KP. Es kann spekuliert werden, ob die VPN bei der EP eine Unvertrautheit mit dem Ticket-Automaten vermuteten, oder ob das Ausweichen auf eine andere Form der Hilfeleistung einen Mechanismus darstellt, um das tatsächliche Zücken des Geldbeutels zu vermeiden. Hier soll jedoch noch einmal ausdrücklich darauf verwiesen werden, dass das Ausbleiben einer Hilfeleistung oder Reaktion keinesfalls mit Xenophobie gleichgesetzt wird und nur die Werte der LV in Relation zueinander aussagekräftig sind.

4.1.2 Häufigkeiten nach LV und nach Geschlecht

Abbildung 4: Antwortverhalten pro LV nach Geschlecht



Generell lässt sich festhalten, dass das Antwortverhalten von Männern und Frauen

nicht eklatant voneinander abweicht. Vielmehr zeigt sich hier erneut die Dominanz der Kategorien *Reaktion ohne Hilfe* und *Volle Hilfeleistung* sowie bei der EP eine deutliche Zunahme der Kategorien *Andere Hilfeleistung* und *Keine oder ablehnende Reaktion*. Innerhalb dieser Kategorien sind die Häufigkeiten zwischen Männern und Frauen beinahe gleich verteilt. Das Geschlecht hat in Betrachtung der Häufigkeiten scheinbar keine größere Aussagekraft.

In Hinblick auf die Hypothese zu den Geschlechterunterschieden lässt sich feststellen, dass sich das Hilfsverhalten nicht eindeutig nach dem Geschlecht der VPN unterscheidet. Zudem kann des Weiteren nicht bestätigt werden, dass weibliche VPN wesentlich von denen der männlichen VPN abweichen, weder negativ, wie in H2 postuliert, noch positiv. Vielmehr zeigen die Frauen gegenüber denen der Männer sogar ein leicht hilfsbereiteres Verhalten – mit Ausnahme der Kategorie *Keine oder ablehnende Reaktion* für die EP.

(FM)

4.1.3. Deskriptive Kennzahlen

Um den Zusammenhang von Antwortverhalten und LV in statistische Kennzahlen zu fassen, werden im Folgenden sowohl Kendall's tau-b als auch Gamma wiedergegeben.⁴⁰

Tabelle 4: Ordinale Zusammenhangsmaße für den Vergleich des Antwortverhaltens nach Lockvogel

Zusammenhangsmaß		Wert
N = 192	Kendall's tau-b	-0,178**
	Gamma	-0,291**

Anmerkung: Die Signifikanz der Parameter ist folgendermaßen zu interpretieren:
*: $p < 0,05$; **: $p < 0,01$; ***: $p < 0,001$.

Tabelle 5: Ordinale Zusammenhangsmaße für den Vergleich des Antwortverhaltens nach Geschlecht

Zusammenhangsmaß		Wert
N = 192	Kendall's tau-b	-0,037
	Gamma	-0,062

Anmerkung: Die Signifikanz der Parameter ist folgendermaßen zu interpretieren:
*: $p < 0,05$; **: $p < 0,01$; ***: $p < 0,001$.

⁴⁰ Es werden sowohl Gamma als auch tau-b berichtet, da beide Maße sowohl Stärken als auch Schwächen aufweisen. So unterschätzt tau-b den wahren Zusammenhang, Gamma überschätzt diesen (siehe Grafik in Schulze 1978: 270). Da eine tiefergehende Diskussion beider Maßzahlen an dieser Stelle nicht möglich ist, wird für Interessierte lediglich auf den Text von Schulze 1978 verwiesen.

Wie Tabelle 4 zu entnehmen ist, weisen beide Werte auf einen schwachen negativen Zusammenhang von Antwortverhalten und LV hin, der trotzdem signifikant ist⁴¹. Wie bereits bei der Häufigkeitsverteilung zu sehen war, gibt es keinen signifikanten Zusammenhang in Bezug auf das Antwortverhalten und das Geschlecht der VPN, weshalb eine Falsifizierung der Hypothese H2 naheliegt (siehe *Tabelle 5*). Es ist allerdings anzumerken, dass die Werte nur eingeschränkt aussagekräftig sind, da sie den Zusammenhang von LV und Antwortverhalten über alle Kategorien hinweg schätzen und somit die Unterschiede zwischen einzelnen Kategorien nur unzureichend wiedergegeben werden. Somit bleiben die zuvor beschriebenen großen Unterschiede in den relativen Häufigkeiten von *keine oder ablehnende Reaktion* und *volle Hilfeleistung* unberücksichtigt.

(AS)

4.2. Inferenzstatistische Analyse

4.2.1 Multinominale logistische Regression

Um ebendiese Unterschiede zwischen den einzelnen Ausprägungen der abhängigen Variablen zu untersuchen, wurde eine multinominale logistische Regression (MLR) durchgeführt.⁴² Der Effekt des LV sowie des Geschlechts auf die einzelnen Kategorien des Antwortverhaltens kann somit statistisch geschätzt werden. Insgesamt wurden vier verschiedene Modelle berechnet, um den auspartialisierten Einfluss der jeweiligen Einflussfaktoren und Kontrollvariablen (Alter und Erhebungstag) schrittweise zu überprüfen. Tabelle 6 fasst die Variablen der einzelnen Modelle zusammen. Da jedoch nur in Modell 1 die unabhängige Variable LV über das Gesamtmodell signifikant war und in den anderen Modellen keine Variable über das Gesamtmodell einen signifikanten Wert erhielt, werden im Folgenden nur die Ergebnisse von Modell 1 interpretiert.⁴³

⁴¹ Da nur von einem einseitigen Effekt (der EP wird weniger Hilfe entgegengebracht als der KP) ausgegangen wird, wurden alle folgenden Signifikanztests nur einseitig berechnet.

⁴² Da die AV weniger als fünf Ausprägungen beinhaltet, die zudem nicht äquidistant sind, war für die hier vorliegende Untersuchung keine OLS durchführbar.

⁴³ Die Berechnungen von Modell 2 (7.3.1: Tabelle 11), Modell 3 (7.3.2 : Tabelle 12) und Modell 4 (7.3.3 : Tabelle 13) befinden sich im Anhang (siehe Punkt 7.3).

Tabelle 6: Modelle der MLR

Modell 1	Lockvogel
Modell 2	Lockvogel + Geschlecht
Modell 3	Lockvogel + Geschlecht + Lockvogel*Geschlecht
Modell 4	Lockvogel + Geschlecht + Lockvogel*Geschlecht + Alter + Erhebungsdatum

Als Referenzkategorie wurde in Modell 1 sowie auch in den folgenden Modellen *Keine oder ablehnende Reaktion* bestimmt. Tabelle 7 und Tabelle 8 fassen die Ergebnisse von Modell 1 zusammen. Zunächst sticht die sehr hohe Signifikanz, die die Variable LV über das Gesamtmodell erreicht ins Auge. Beim Betrachten der einzelnen Ausprägungen der abhängigen Variablen fällt jedoch auf, dass der Vergleich von *Keine oder ablehnende Reaktion* mit *Andere Hilfeleistung* aufgrund einer zu geringen Signifikanz nicht möglich ist. Über die verbleibenden zwei Vergleiche lassen sich aber Aussagen treffen. In beiden Fällen wirkt sich das Setzen des Stimulus negativ auf die erfahrene Hilfsbereitschaft aus. Die Odds ratios geben an, dass die KP im Vergleich zur EP eine 4,7-fache Chance hat, eine Reaktion ohne Hilfeleistung statt keiner oder eine ablehnende Reaktion zu erhalten. Noch frappierender ist der Unterschied in der Hilfeleistung bei Vergleich der ersten mit der letzten Ausprägung der abhängigen Variablen. Die KP hat im Vergleich zur EP eine 6,8-mal so große Chance eine volle Hilfsbereitschaft statt keine oder eine ablehnende Reaktion zu erfahren. Ausgedrückt in Wahrscheinlichkeiten bedeutet dieser Vergleich, dass die KP mit ca. 87 % eine volle Hilfsbereitschaft anstelle von keiner oder einer ablehnenden Reaktion erhält. Die EP hingegen erhält nur in ca. 51% der Fälle eine volle Hilfsbereitschaft anstelle von keiner oder einer ablehnenden Reaktion. Diese Ergebnisse bestätigen den ersten Eindruck der Betrachtung der Kreisdiagramme und die Hypothese H1 kann endgültig als bestätigt betrachtet werden. Für die Hypothesen H2 ist dies aufgrund der nicht ausreichenden Signifikanz der unabhängigen Variablen Geschlecht nicht der Fall.

Tabelle 7: Likelihood ratio Tests von Modell 1

Effekt	-2 LL	Chi ²	Freiheitsgrade	Signifikanz
Intercept	25,776	0	0	
Lockvogel	48,019	22,243	3	0,000

Tabelle 8: Parameterschätzungen von Modell 1

Hilfsbereitschaft		Logit	Odds ratio
Reaktion ohne Hilfeleistung	Intercept	0,435* (0,274)	4,745
	KP	1,557*** (0,514)	
	EP (Stimulus)	0	
Andere Hilfeleistung	Intercept	-0,147 (0,313)	1,351
	KP	0,301 (0,639)	
	EP (Stimulus)	0	
Volle Hilfeleistung	Intercept	-0,047 (0,305)	6,81
	KP	1,918 (0,534)	
	EP (Stimulus)	0	

Nagelkerkes R² = 0,119

Anmerkung: Die Eintragungen hinter den Logitkoeffizienten (in Klammern) sind die Standardfehler. Die Signifikanz der Parameter ist folgendermaßen zu interpretieren:

*: $p < 0,05$; **: $p < 0,01$; ***: $p < 0,001$.

Insgesamt beträgt die prozentuale Verbesserung der Schätzung der Hilfsbereitschaft Vergleich zum Nullmodell bei Kenntnis des Stimulus 11,9 % (betrachtet wurde Nagelkerkes R²). Dies ist zwar den Standardmaßen entsprechend kein besonders hoher Wert, allerdings durchaus beachtlich dafür, dass nur eine einzige Variable in das Modell aufgenommen wurde.

4.2.2 Binominale logistische Regression

Abschließend wurde eine binominale logistische Regression mit dichotomen Ausprägungen der abhängigen Variablen (1 = *Volle Hilfeleistung*, 0 = *Keine oder ablehnende Reaktion, Reaktion ohne Hilfeleistung, Andere Hilfeleistung*) gerechnet. Dies diente in erster Linie dazu, die Teststärke des Regressionsmodells im Nach-

hinein bestimmen zu können.⁴⁴ Somit wurde im Rückschluss ebenfalls kontrolliert, ob die im Vorfeld berechnete Stichprobengröße ausreichend war. Tabelle 9 zeigt die Ergebnisse. Die Parameterschätzungen bestätigen erwartungsgemäß nochmals die H1. Die EP hat im Vergleich zur KP nur eine 0,4-fache Chance, die volle Hilfeleistung zu erhalten. Im Umkehrschluss hat die KP also im Vergleich zur EP eine 2,44-mal so große Chance die volle Hilfeleistung statt jeglicher anderen Art von Reaktion zu erhalten als die KP.

Tabelle 9: Parameterschätzungen der binominalen logistischen Regression

	Logit	Odds Ratio
Intercept	-0,379* (0,208)	0,684
KP		
EP (Stimulus)	-0,893** (0,323)	0,409

Nagelkerkes R² = 0,057

Anmerkung: Die Eintragungen hinter den Logitkoeffizienten (in Klammern) sind die Standardfehler. Die Signifikanz der Parameter ist folgendermaßen zu interpretieren:

*: $p < 0,05$; **: $p < 0,01$; ***: $p < 0,001$.

Die prozentuale Verbesserung der Schätzung im binominalen Modell beträgt 5,7 %. Somit ergibt sich bei einem Stichprobenumfang von N=196, einem odds ratio von 0,409 und einem Signifikanzniveau von 0,034 eine Power von 72%. Die gewünschte Teststärke von 80% wurde demnach zwar knapp verfehlt, jedoch stellt dies nicht die Aussagekraft der gefundenen Ergebnisse in Frage.

(NH)

⁴⁴ Es wurde keine geeignete Methode für das berechnen der Teststärke über das Gesamtmodell der multinominalen logistischen Regression gefunden.

5. Fazit

Ziel dieser Studie war es, Fremdenfeindlichkeit mit einem eigens entwickelten Experimentaldesign zu erheben. Dazu wurden Testpersonen von zwei Lockvögeln – einem offensichtlich deutschen und einem (aufgrund von Hautfarbe und Aussprache) offensichtlich nicht-deutschen Lockvogel – in einer U-Bahnhaltestelle gefragt, ob sie einen 5-Euro-Schein wechseln könnten, da der Ticketautomat diesen nicht akzeptiere. Fremdenfeindlichkeit wurde also über die beobachtete Hilfebereitschaft operationalisiert, wobei zwischen vier Formen von Reaktionen unterschieden wurde. Aufgrund theoretischer Überlegungen und empirischer Erkenntnisse existenter Studien wurden zwei Hypothesen aufgestellt: H1 besagt, dass dem nicht-deutschen Lockvogel signifikant weniger Hilfe entgegengebracht wird als dem deutschen. H2 behauptet einen geschlechterspezifischen Unterschied im Hilfeverhalten, nämlich dass weibliche Testpersonen den männlichen Lockvögeln signifikant weniger helfen als männliche Testpersonen. Bereits bei der Betrachtung der Häufigkeiten zeigen sich Unterschiede im Hilfeverhalten gegenüber deutschem und nicht-deutschem Lockvogel, welche H1 zu bestätigen scheinen. Dem nicht-deutschen Lockvogel wird zwar etwa genauso häufig geholfen, wie seinem Pendant, allerdings wird dieser auch wesentlich häufiger ignoriert. Was das Geschlecht der Testpersonen betrifft, sind keine nennenswerten Unterschiede feststellbar. Die anschließend durchgeführte statistische Modellierung bestätigt diese Befunde und den von H1 postulierten signifikanten Unterschied im Hilfeverhalten. H2 hingegen kann nicht bestätigt werden, es gibt keinen signifikanten geschlechterspezifischen Unterschied.

6. Literaturverzeichnis

- Abdullahi, Isah Aliyu/Kumar, Pardeep 2016: Gender Differences in Prosocial Behaviour. In: The International Journal of Indian Psychology 3 (4), 171-175.
- Alfermann, D., 1996: Geschlechterrollen und geschlechtstypisches Verhalten. Stuttgart: Kohlhammer.
- Atteslander, Peter 2010: Methoden der empirischen Sozialforschung. 13., neu bearbeitete und erweiterte Auflage. Berlin: Erich Schmidt Verlag.
- Broden, Anne 2012: Anmerkungen zur Aktualität der Ungleichheit. In: Aus Politik und Zeitgeschichte 62 (16-17), 7-10.
- Diekmann, Andreas 2012: Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. 6. Auflage. Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verlag.
- Diekmann, Andreas/Jann, Ben/Näf, Matthias 2014: Wie fremdenfeindlich ist die Schweiz? In: Soziale Welt, 185-199.
- Gaertner, Samuel/Dovidio, John 2005: Understanding and Addressing Contemporary Racism: From Aversive Racism to the Common Ingroup Identity Model. In: Journal of Social Issues 61 (3), 615-639.
- Geschke, Daniel: Vorurteile, Differenzierung und Diskriminierung – sozialpsychologische Erklärungsansätze. In: Aus Politik und Zeitgeschichte 62 (16-17), 33-37.
- Groß, Eva/Zick, Andreas/Krause, Daniela 2012: Von der Ungleichwertigkeit zur Ungleichheit: Gruppenbezogene Menschenfeindlichkeit. In: Aus Politik und Zeitgeschichte 62 (16-17), 11-18.
- Howlett, Neil/ Pine, Karen/ Orakçioğlu, Ismail/ Fletcher, Ben 2013: The influence of clothing on first impressions: Rapid and positive responses to changes in male attires. In: Journal of Fashion Marketing and Management: An International Journal, Vol 17 Iss1, 38-48.
- Hurtz, Simon 2015: "Blanker Rassismus": Höcke und die Fortpflanzung der Afrikaner. In: Süddeutsche Zeitung. In: <http://www.sueddeutsche.de/politik/afd-thueringen-blanker-rassismus-hoecke-und-die-fortpflanzung-der-afrikaner-1.2780159>, aufgerufen am 29.03.2016
- Infratest dimap ARD DeutschlandTREND. Einstellungen zu fremdenfeindlichen Übergriffen und Protesten gegen Flüchtlinge. März 2016. In: <http://www.infratest-dimap.de/umfragen-analysen/bundesweit/ard-deutschlandtrend/2016/maerz/>, zugegriffen am: 25.03.2016
- Klink, Andreas/ Wagner, Ulrich 1999: Discrimination against ethnic minorities in Germany: Going back to the field. In: Journal of Applied Psychology, 29, 402-423.
- Kromrey, Helmut (2002): Empirische Sozialforschung. 10. überarbeitete Auflage. Opladen: Leske + Budrich.
- Latané, B./Darley, J.M., 1968: The Unresponsive Bystander: Why doesn't he

- help? New York: Appleton-Century-Crofts Educational Division Meredith Corporation.
- List, John A./Sadoff, Sally/Wagner, Mathis 2010: So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design. In: Economic Science Association 2011.
- Madden, Thomas J./Hewett, Kelly/Roth, Martin S. 2000: Managing Images in Different Cultures: A Cross-National Study of Color Meanings and Preferences. In: Journal of International Marketing, Vol. 8, No 4, 90-107.
- Malhotra, Maresh K./Birks, David F. 2007: Marketing Research. An Applied Approach. 3. Auflage. Pearson Education Limited: Harlow.
- Pearson, Adam R./Dovidio, John F./Gaertner, Samuel L. 2009: The Nature of Contemporary Prejudice: Insights from Aversive Racism. In: Social and Personality Psychology Compass 3, 1-25.
- Saucier, Donald A./Miller, Carol T./Doucet, Nicole 2005: Differences in Helping Whites and Blacks: A Meta-Analysis. In: Personality and Social Psychology Review 9 (1), 2-16.
- Schulze, Gerhard 1978: Rangkorrelation bei soziologischen Normaldaten – Ein Ansatz zur Überwindung der Schwächen von Tau und Gamma. In: Zeitschrift für Soziologie 7 (3), 267-272.
- Shooman, Yasemin 2012: Das Zusammenspiel von Kultur, Religion, Ethnizität und Geschlecht im antimuslimischen Rassismus. In: Aus Politik und Zeitgeschichte 62 (16-17), 53-57.
- Statistisches Amt Stuttgart 2015: Datenkompass Stadtbezirke Stuttgart 2014/2015. Stadtbezirk Mitte. In: https://service.stuttgart.de/lhs-services/komunis/documents/6433_1_01_Mitte_Stadtbezirksprofil.PDF, zugegriffen am 07.03.2016.
- Statistisches Amt Stuttgart 2016a: Frauen in Stuttgart seit 1986 nach Altersgruppen. In: <https://service.stuttgart.de/lhs-services/komunis/index.php?uid=99&objectid=13762>, zugegriffen am 05.03.2016
- Statistisches Amt Stuttgart 2016b: Männer in Stuttgart seit 1986 nach Altersgruppen. In: <https://service.stuttgart.de/lhs-services/komunis/index.php?uid=99&objectid=13761>, zugegriffen am 05.03.2016
- Toosi, Negin R./Ambady, Nalini/Babbitt, Laura G./Sommers, Samuel R. 2012: Dyadic Interracial Interactions: A Meta-Analysis. In: Psychological Bulletin 138 (1), 1-27.
- Voigtländer, Denise 2008: Hilfeverhalten und Zivilcourage: Ein Vergleich von antizipiertem und realem Verhalten. Diss., Georg-August-Universität Göttingen.
- VVS 2016: Stadtbahn-Liniennetz. In: http://www.vvs.de/download/-Stadtbahn_Liniennetz.pdf, zugegriffen am 07.03.2016
- Zimmermann, Ekkart 1972: Das Experiment in den Sozialwissenschaften. Stuttgart: Teubner.

7. Anhang

7.1

Tabelle 10: Tabelle zur Betrachtung der Störfaktoren nach Campbell und Stanley für das vorgestellte Experiment (vgl. Campbell/Stanley zitiert nach Zimmermann 1972: 78-81).

Einflussfaktoren auf die interne Validität	
1. Zeiteinflüsse (Ereignisse zwischen der ersten und zweiten Messung)	→ kann zu Problemen führen, wenn zwischen den Messungen/innerhalb des Messzeitraums sozial relevante Ereignisse geschehen (wie z.B. in der Silvesternacht)
2. biologisch-psychologische Veränderungen im Untersuchungsobjekt	→ unproblematisch, da nur einmalige Messung pro Versuchsperson
3. Messeffekte (Auswirkungen einer ersten Messung auf eine zweite)	→ unproblematisch, da nur einmalige Messung pro Versuchsperson (siehe auch Punkt 9)
4. Veränderungen in den Messinstrumenten	→ kann zu Problemen führen (z.B. durch Lern-/Ermüdungseffekte auch bei den Hilfspersonen; Ausfall des Lockvogels durch Krankheit)
5. statistische Regression (Gruppen werden wegen extremer Punktzahl ausgesucht)	→ unproblematisch, da Randomisierung durch Zufallsauswahl
6. Auswahlverzerrungen (Versuchseinheiten der Gruppen sind zu unterschiedlich)	→ unproblematisch, da Randomisierung durch Zufallsauswahl und Matching durch Quotenplan
7. Ausfälle unter den Versuchspersonen im Verlauf des Experiments	→ unproblematisch, da nur einmalige Messung pro Versuchsperson
8. Interaktion von Auswahlverzerrungen und biologisch-psychologischen Veränderungen	→ unproblematisch siehe Punkt 7 und Punkt 8
Einflussgrößen auf externe Validität:	
1. Reaktiver oder interaktiver Effekt von Messungen: (z.B. Sensibilisierung durch Pretest macht Generalisierung auf externe Population unmöglich)	→ potenziell problematisch; dem soll durch relativ langen zeitlichen Abstand zwischen Pretest und tatsächlicher Messung Sorge getragen werden
2. Interaktionseffekte durch die Auswahlverzerrungen und die experimentelle Variable	→ unproblematisch, da Randomisierung durch Zufallsauswahl und Matching durch Quotenplan
3. reaktive Effekte experimenteller Arrangements: Generalisierung außerhalb des experimentellen Rahmens nicht möglich	→ unproblematisch, da verdecktes Feldexperiment
4. Interferenzen durch mehrfache experimentelle Behandlungen derselben Versuchsperson (Effekte sind dann nicht immer zu trennen)	→ unproblematisch, da nur einmalige Messung pro Versuchsperson

7.2 Tabellarische Übersicht der Häufigkeitsverteilungen

7.2.1 Absolute und relative Häufigkeiten

		Beide LV			KP			EP		
		Weiblich	Männlich	Gesamt	Weiblich	Männlich	Gesamt	Weiblich	Männlich	Gesamt
<i>Keine oder ablehnende Reaktion</i>	absolut in %	14 (7,3)	14 (7,3)	28 (14,6)	2 (2,1)	4 (4,2)	6 (6,3)	12 (12,5)	10 (10,4)	22 (22,9)
<i>Reaktion ohne Hilfeleistung</i>	absolut in %	36 (18,8)	42 (21,9)	78 (40,6)	22 (22,9)	22 (22,9)	44 (45,8)	14 (14,6)	20 (20,8)	34 (35,4)
<i>Andere Hilfeleistung</i>	absolut in %	15 (7,8)	11 (5,7)	26 (13,5)	4 (4,1)	3 (3,1)	7 (7,3)	11 (11,5)	8 (8,3)	19 (19,8)
<i>Volle Hilfeleistung</i>	absolut in %	31 (16,1)	29 (15,1)	60 (31,3)	20 (20,8)	19 (19,8)	39 (40,6)	11 (11,5)	10 (10,4)	21 (21,9)
<i>Gesamt</i>	absolut in %	96 (50)	96 (50)	192 (100)	48 (50)	48 (50)	96 (100)	48 (50)	48 (50)	96 (100)

7.2.2 Erwartete und tatsächliche Häufigkeiten

7.2.2.1 Antwortverhalten nach LV: expected und absolute counts

		Keine oder ablehnende Reaktion	Reaktion ohne Hilfeleistung	Andere Hilfeleistung	Volle Hilfeleistung	Total
KP	Count	6	44	7	39	96
	Expected Count	14	39	13	30	96
	% of Total	(3,10)	(22,90)	(3,60)	(20,30)	(50,00)
EP	Count	22	34	19	21	96
	Expected Count	14	39	13	30	96
	% of Total	(11,50)	(17,70)	(9,90)	(10,90)	(50,00)
Total	Count	28	78	26	60	192
	Expected Count	28	78	26	60	192
	% of Total	(14,60)	(40,60)	(13,50)	(31,30)	(100,00)

7.2.2.2 Antwortverhalten nach Geschlecht: expected und absolute counts

		Keine oder ablehnende Reaktion	Reaktion ohne Hilfeleistung	Andere Hilfeleistung	Volle Hilfeleistung	Total
Weiblich	Count	14	36	15	31	96
	Expected Count	14	39	13	30	96
	% of Total	(7,30)	(18,80)	(7,80)	(16,10)	(50,00)
Männlich	Count	14	42	11	29	96
	Expected Count	14	39	13	30	96
	% of Total	(7,30)	(21,90)	(5,70)	(15,10)	(50,00)
Total	Count	28	78	26	60	192
	Expected Count	28	78	26	60	192
	% of Total	(14,60)	(40,60)	(13,50)	(31,30)	(100,00)

7.3 Übersicht der Modellvergleiche

7.3.1: Tabelle 11: Likelihood ratio Tests von Modell 2

Effekt	-2 Log Likelihood des reduzierten Modells	Chi- Quadrat	Freiheitsgrade	Signifikanz
Intercept	44,662	0	0	
Lockvogel	66,930	22,267	3	0,000
Geschlecht	45,833	1,170	3	0,380

7.3.2 : Tabelle 12: Likelihood ratio Tests von Modell 3

Effekt	-2 Log Like- lihood des reduzierten Modells	Chi- Quadrat	Freiheitsgrade	Signifikanz
Intercept	43,215	0	0	
Lockvogel	43,215	0	0	
Geschlecht	43,215	0	0	
Lockvogel*Geschlecht	44,662	1,448	3	0,347

7.3.3 : Tabelle 13: Likelihood ratio Tests von Modell 4

Effekt	-2 Log Like- lihood des reduzierten Modells	Chi- Quadrat	Freiheitsgrade	Signifikanz
Intercept	213,975	0	0	
Lockvogel	213,975	0	0	
Geschlecht	213,975	0	0	
Lockvogel*Geschlecht	215,459	1,484	3	0,343
Erhebungsdatum	222,858	8,883	6	0,09
Alter	220,063	6,088	6	0,207

7.4 SPSS-Syntax

7.4.1 Deskriptive Statistik

Frequencies Antwortverhalten insgesamt (beide LV) – inkl. Kreisdiagramm

USE ALL.

COMPUTE filter_\$=(Quotenplan = 1).

VARIABLE LABELS filter_\$ 'Quotenplan = 1 (FILTER)'.

VALUE LABELS filter_\$ 0 'Not Selected' 1 'Selected'.

FORMATS filter_\$ (f1.0).

FILTER BY filter_\$.

EXECUTE.

FREQUENCIES VARIABLES=Antwortverhalten

/PIECHART FREQ

/ORDER=ANALYSIS.

Frequencies Antwortverhalten KP – inkl. Kreisdiagramm

USE ALL.

COMPUTE filter_\$=(Quotenplan = 1 & Lockvogel = 0).

VARIABLE LABELS filter_\$ 'Quotenplan = 1 & Lockvogel = 0 (FILTER)'.

VALUE LABELS filter_\$ 0 'Not Selected' 1 'Selected'.

FORMATS filter_\$ (f1.0).

FILTER BY filter_\$.

EXECUTE.

FREQUENCIES VARIABLES=Antwortverhalten

/PIECHART FREQ

/ORDER=ANALYSIS.

Frequencies Antwortverhalten EP – inkl. Kreisdiagramm

USE ALL.

COMPUTE filter_\$=(Quotenplan = 1 & Lockvogel = 1).

VARIABLE LABELS filter_\$ 'Quotenplan = 1 & Lockvogel = 1 (FILTER)'.

VALUE LABELS filter_\$ 0 'Not Selected' 1 'Selected'.

```
FORMATS filter_$ (f1.0).  
FILTER BY filter_$.  
EXECUTE.  
FREQUENCIES VARIABLES=Antwortverhalten  
/PIECHART FREQ  
/ORDER=ANALYSIS.
```

Häufigkeiten Antwortverhalten nach LV und Geschlecht

```
USE ALL.  
COMPUTE filter_$=(Quotenplan = 1).  
VARIABLE LABELS filter_$ 'Quotenplan = 1 (FILTER)'.  
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.  
FORMATS filter_$ (f1.0).  
FILTER BY filter_$.  
EXECUTE.  
CROSSTABS  
/TABLES=Antwortverhalten BY Geschlecht BY Lockvogel  
/FORMAT=AVALUE TABLES  
/CELLS=COUNT TOTAL  
/COUNT ROUND CELL.
```

Zusammenhang von Antwortverhalten und LV (Kreuztabelle)

```
USE ALL.  
COMPUTE filter_$=(Quotenplan = 1).  
VARIABLE LABELS filter_$ 'Quotenplan = 1 (FILTER)'.  
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.  
FORMATS filter_$ (f1.0).  
FILTER BY filter_$.  
EXECUTE.  
CROSSTABS  
/TABLES=Lockvogel BY Antwortverhalten  
/FORMAT=AVALUE TABLES
```

```
/STATISTICS=GAMMA D BTAU
/CELLS=COUNT EXPECTED TOTAL
/COUNT ROUND CELL.
```

Zusammenhang von Geschlecht und Antwortverhalten insgesamt (Kreuztabelle)

```
USE ALL.
COMPUTE filter_$=(Quotenplan = 1).
VARIABLE LABELS filter_$ 'Quotenplan = 1 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.
CROSSTABS
  /TABLES=Geschlecht BY Antwortverhalten
  /FORMAT=AVALUE TABLES
  /STATISTICS=GAMMA D BTAU
  /CELLS=COUNT EXPECTED TOTAL
  /COUNT ROUND CELL.
```

Zusammenhang von Geschlecht und Antwortverhalten bei KP (Kreuztabelle)

```
USE ALL.
COMPUTE filter_$=(Quotenplan = 1 & Lockvogel = 0).
VARIABLE LABELS filter_$ 'Quotenplan = 1 & Lockvogel = 0 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.
CROSSTABS
  /TABLES=Geschlecht BY Antwortverhalten
  /FORMAT=AVALUE TABLES
  /STATISTICS=GAMMA D BTAU
  /CELLS=COUNT EXPECTED TOTAL
```


/COUNT ROUND CELL.

Zusammenhang von Geschlecht und Antwortverhalten bei EP (Kreuztabelle)

USE ALL.

COMPUTE filter_\$=(Quotenplan = 1 & Lockvogel = 1).

VARIABLE LABELS filter_\$ 'Quotenplan = 1 & Lockvogel = 1 (FILTER)'.

VALUE LABELS filter_\$ 0 'Not Selected' 1 'Selected'.

FORMATS filter_\$ (f1.0).

FILTER BY filter_\$.

EXECUTE.

CROSSTABS

/TABLES=Geschlecht BY Antwortverhalten

/FORMAT=AVALUE TABLES

/STATISTICS=GAMMA D BTAU

/CELLS=COUNT EXPECTED TOTAL

/COUNT ROUND CELL.

7.4.2 Inferenzstatistik

Multinominale logistische Regression

Modell I:

USE ALL.

COMPUTE filter_\$=(Quotenplan = 1).

VARIABLE LABELS filter_\$ 'Quotenplan = 1 (FILTER)'.

VALUE LABELS filter_\$ 0 'Not Selected' 1 'Selected'.

FORMATS filter_\$ (f1.0).

FILTER BY filter_\$.

EXECUTE.

NOMREG Antwortverhalten (BASE=FIRST ORDER=ASCENDING) BY Lockvogel

/CRITERIA CIN(95) DELTA(0) MXITER(100) MXSTEP(5) CHKSEP(20)
LCONVERGE(0) PCONVERGE(0.000001)

SINGULAR(0.00000001)

```

/MODEL=Lockvogel
/STEPWISE=PIN(.05)  POUT(0.1)  MINEFFECT(0)  RULE(SINGLE)  EN-
TRYMETHOD(LR) REMOVALMETHOD(LR)
/INTERCEPT=INCLUDE
/PRINT=PARAMETER SUMMARY LRT CPS STEP MFI.

```

Modell II

```

USE ALL.
COMPUTE filter_$=(Quotenplan = 1).
VARIABLE LABELS filter_$ 'Quotenplan = 1 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.
NOMREG Antwortverhalten (BASE=FIRST ORDER=ASCENDING) BY Lockvo-
gel Geschlecht
/CRITERIA CIN(95) DELTA(0) MXITER(100) MXSTEP(5) CHKSEP(20)
LCONVERGE(0) PCONVERGE(0.000001)
SINGULAR(0.00000001)
/MODEL=Lockvogel Geschlecht
/STEPWISE=PIN(.05)  POUT(0.1)  MINEFFECT(0)  RULE(SINGLE)  EN-
TRYMETHOD(LR) REMOVALMETHOD(LR)
/INTERCEPT=INCLUDE
/PRINT=PARAMETER SUMMARY LRT CPS STEP MFI.

```

Modell III

```

USE ALL.
COMPUTE filter_$=(Quotenplan = 1).
VARIABLE LABELS filter_$ 'Quotenplan = 1 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.

```

EXECUTE.

NOMREG Antwortverhalten (BASE=FIRST ORDER=ASCENDING) BY Lockvogel Geschlecht

/CRITERIA CIN(95) DELTA(0) MXITER(100) MXSTEP(5) CHKSEP(20)
LCONVERGE(0) PCONVERGE(0.000001)

SINGULAR(0.00000001)

/MODEL=Lockvogel Geschlecht Lockvogel*Geschlecht

/STEPWISE=PIN(.05) POUT(0.1) MINEFFECT(0) RULE(SINGLE) ENTRYMETHOD(LR) REMOVALMETHOD(LR)

/INTERCEPT=INCLUDE

/PRINT=PARAMETER SUMMARY LRT CPS STEP MFI.

Modell IV

USE ALL.

COMPUTE filter_\$=(Quotenplan = 1).

VARIABLE LABELS filter_\$ 'Quotenplan = 1 (FILTER)'.

VALUE LABELS filter_\$ 0 'Not Selected' 1 'Selected'.

FORMATS filter_\$ (f1.0).

FILTER BY filter_\$.

EXECUTE.

NOMREG Antwortverhalten (BASE=FIRST ORDER=ASCENDING) BY Lockvogel Geschlecht Erhebungsdatum Alter

/CRITERIA CIN(95) DELTA(0) MXITER(100) MXSTEP(5) CHKSEP(20)
LCONVERGE(0) PCONVERGE(0.000001)

SINGULAR(0.00000001)

/MODEL=Lockvogel Geschlecht Lockvogel*Geschlecht Erhebungsdatum Alter

/STEPWISE=PIN(.05) POUT(0.1) MINEFFECT(0) RULE(SINGLE) ENTRYMETHOD(LR) REMOVALMETHOD(LR)

/INTERCEPT=INCLUDE

/PRINT=PARAMETER SUMMARY LRT CPS STEP MFI.

Interaktionsvariable LV*Geschlecht

```
COMPUTE LVGeschlecht=Lockvogel * Geschlecht.
```

```
EXECUTE.
```

Umkodieren der UV und der AV

```
DATASET ACTIVATE DataSet1.
```

```
RECODE Antwortverhalten (3=1) (ELSE=0) INTO gewechselt.
```

```
EXECUTE.
```

```
RECODE Lockvogel (0=1) (1=0) INTO Lockvogel_andersrum.
```

```
EXECUTE.
```

Binominale logistische Regression

```
USE ALL.
```

```
COMPUTE filter_$=(Quotenplan = 1).
```

```
VARIABLE LABELS filter_$ 'Quotenplan = 1 (FILTER)'.  
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.  
FORMATS filter_$ (f1.0).  
FILTER BY filter_$.  
EXECUTE.
```

```
LOGISTIC REGRESSION VARIABLES gewechselt
```

```
  /METHOD=ENTER Lockvogel
```

```
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

Erklärung über die Eigenständigkeit

Wir erklären,

1. dass diese Arbeit (bzw. bei Gruppenarbeiten entsprechend gekennzeichnete Anteil der Arbeit) selbständig verfasst wurde,
2. dass keine anderen als die angegebenen Quellen benutzt und alle wörtlich oder sinngemäß aus anderen Werken übernommenen Aussagen als solche gekennzeichnet wurden,
3. dass die eingereichte Arbeit weder vollständig noch in wesentlichen Teilen Gegenstand eines anderen Prüfungsverfahrens gewesen ist,
4. dass die Arbeit weder vollständig noch in Teilen bereits veröffentlicht wurde und
5. dass – falls zutreffend – das elektronische Exemplar mit den gedruckten Exemplaren übereinstimmt.

Stuttgart, den 15.09.2016


Fabienne Minn


Nikolaus Hastreiter


Alexander Schrage