

# Vertrauen in autonome Technik

Modellierung und Simulation von Mensch-Maschine-Interaktion  
in experimentell-soziologischer Perspektive

## **Dissertation**

zur Erlangung des akademischen Grades

Dr. rer. pol.

vorgelegt an der

Wirtschafts- und Sozialwissenschaftlichen Fakultät der  
Technischen Universität Dortmund

von

Herrn Dipl.-Inform. Robin David Fink

aus

Mülheim an der Ruhr

1. Gutachter: Prof. Dr. Johannes Weyer
2. Gutachter: JProf. Dr. Maximiliane Wilkesmann



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>11</b>
<b>1</b>	<b>Vertrauen und Automation</b>	<b>15</b>
<b>2</b>	<b>Definitiorische Klärung des Vertrauensbegriffs</b>	<b>17</b>
2.1	Das Klassifizierungsschema von McKnight und Chervany . . . . .	18
2.1.1	Vertrauensdisposition . . . . .	19
2.1.2	Institutionelles Vertrauen . . . . .	19
2.1.3	Vertrauende Überzeugung . . . . .	21
2.1.4	Vertrauende Absichten . . . . .	22
2.1.5	Auf Vertrauen bezogenes Verhalten . . . . .	23
2.1.6	Zusammenhänge zwischen den Vertrauenskonstrukten . . . . .	24
2.1.7	Fazit und kritische Würdigung . . . . .	27
2.2	Facetten und Objekte des Vertrauens nach Nooteboom . . . . .	28
2.2.1	Vertrauen als vierstelliges Prädikat . . . . .	29
2.2.2	Exkurs: Multiple Kausalität . . . . .	30
2.2.3	Vertrauen: Eine auf Verhalten bezogene Disposition . . . . .	32
2.3	Vertrauensbegriff und Implikationen . . . . .	35
2.3.1	Vertrauen als eine Beziehung zwischen Subjekt und Objekt . . . . .	37
2.3.2	Vertrauen als Zuschreibung von Gründen und Verhalten . . . . .	37
2.3.3	Abgabe von Kontrolle und Inkaufnahme Verletzlichkeit . . . . .	40
<b>3</b>	<b>Vertrauen in der Sozialtheorie</b>	<b>43</b>
3.1	Probabilistischer Vertrauensbegriff . . . . .	44
3.1.1	Deutsch: Vertrauen als Abwägung negativer und positiver Konsequenzen . . . . .	45
3.1.2	Coleman: Vertrauen als Entscheidung unter Risiko . . . . .	47
3.1.3	Kritische Würdigung der probabilistischen Vertrauensbegriffe . . . . .	52
3.2	Spieltheoretischer Vertrauensbegriff . . . . .	55
3.2.1	Das Vertrauensspiel . . . . .	56
3.2.2	Experimentelle Untersuchung von Vertrauen . . . . .	58
3.2.3	Iterierte bzw. reputationsbasierte Vertrauensspiele . . . . .	59

3.2.4	Buskens: Synthese spieltheoretischer Vertrauensmodelle . . . . .	65
3.2.5	Kritische Würdigung . . . . .	68
3.3	Implikationen . . . . .	70
<b>4</b>	<b>Vertrauen in der Automationsforschung</b>	<b>73</b>
4.1	Vertrauen und Funktionsallokation . . . . .	75
4.2	Das Vertrauensproblem in der Automationsforschung . . . . .	76
4.3	Automationsstudien zu Vertrauen . . . . .	79
4.3.1	Steuerung einer Pasteurierungsanlage . . . . .	80
4.3.2	Klassifizierung feindlicher Flugzeuge . . . . .	84
4.3.3	Erkennung versteckter Soldaten . . . . .	87
4.3.4	Langzeitstudie an einem Realsystem . . . . .	91
4.4	Implikationen . . . . .	94
<b>II</b>	<b>Verbindende Konzepte zwischen Computersimulation, Handlungstheorie und autonomer Technik</b>	<b>99</b>
<b>5</b>	<b>Autonome Technik als Gegenstand der Soziologie</b>	<b>101</b>
5.1	Die Provokation der Akteur-Netzwerk-Theorie . . . . .	101
5.1.1	Generalisierte Symmetrie . . . . .	102
5.1.2	Der Hybrid-Akteur . . . . .	103
5.1.3	Fazit ANT . . . . .	104
5.2	Praxisbeispiele und Handlungswirksamkeit autonomer Technik . . . . .	107
5.2.1	Autonomous Trading Agent . . . . .	107
5.2.2	(Teil-)autonomes Fahren . . . . .	109
5.2.3	Wikipedia Bots – Nicht-menschliche Mitglieder der Wikipedia-Gemeinschaft . . . . .	118
5.3	Verteilte Handlungsträgerschaft und hybride Systeme . . . . .	137
5.3.1	Einschub: Technik und Intentionen . . . . .	139
5.3.2	Ein gradualisiertes Modell verteilter Handlungsträgerschaft . . . . .	142
5.3.3	Schlussüberlegungen und kritische Würdigung . . . . .	146
<b>6</b>	<b>Handlungstheoretische Mikro-Makro-Modelle</b>	<b>151</b>
6.1	Das Coleman'sche Erklärungsmodell . . . . .	152
6.1.1	Die drei Übergänge . . . . .	153
6.1.2	Erkenntnistheoretische Überlegungen zum Makro-Mikro-Makro-Modell . . . . .	154
6.2	MSE und SEU-Theorie . . . . .	160
6.3	Weiterentwicklungen . . . . .	164
6.4	Implikationen . . . . .	167

---

<b>7</b>	<b>Computersimulation und Experimente</b>	<b>169</b>
7.1	Experimente in den Sozialwissenschaften . . . . .	170
7.2	Computersimulation als Methode der Sozialwissenschaften . . . . .	175
7.3	Agentenbasierte Modellierung und Simulation . . . . .	177
7.3.1	Struktur einer ABMS . . . . .	178
7.3.2	ABMS-Anwendungen . . . . .	182
7.3.3	Das KISS-Prinzip und seine Tücken . . . . .	184
7.3.4	Für eine mikrofundierte ABMS . . . . .	185
<b>8</b>	<b>Vorschlag einer integrierenden Methodologie</b>	<b>189</b>
8.1	Modell soziologischer Erklärung hybrider Systeme (HMSE) . . . . .	190
8.1.1	Symmetrische Modellierung . . . . .	194
8.1.2	Illustration des HMSE am Beispiel von Adaptive Cruise Control	197
8.2	Modellzuschreibungen . . . . .	200
8.2.1	Systematischer Vergleich von SEU-Modellen . . . . .	201
8.2.2	Das Konzept der wechselseitigen Modellzuschreibungen . . . . .	203
8.3	Erklärungsanspruch und methodologischer Nutzen des HMSE . . . . .	208
<b>III</b>	<b>Empirie</b>	<b>211</b>
<b>9</b>	<b>Vorstellung der Experimente und Hypothesen</b>	<b>213</b>
9.1	Überblick über die bisherigen forschungsleitenden Annahmen . . . . .	213
9.2	Kurzvorstellung der Experimente und zugehörigen Hypothesen . . . . .	214
9.2.1	MensaAssist . . . . .	215
9.2.2	Simhybs . . . . .	216
<b>10</b>	<b>Experiment: MensaAssist</b>	<b>221</b>
10.1	Modellierungsgrundlage: Zwei empirische Erhebungen . . . . .	222
10.1.1	Papierbefragung in der Mensa . . . . .	222
10.1.2	Online-Befragung zur Essenswahl in der Mensa . . . . .	227
10.1.3	Fazit aus beiden Befragungen und Implikationen für MensaAssist	230
10.2	Design von MensaAssist-Experiment und Fragebogenelementen . . . . .	231
10.2.1	Elemente des Fragebogens . . . . .	231
10.2.2	Arbeitsweise von MensaAssist und methodische Einschränkungen	234
10.2.3	Entscheidungsgrundlage der Befragten / Erfassung von U-Werten	235
10.2.4	Vertrauensmessung . . . . .	235
10.3	Auswertung . . . . .	237
10.3.1	Überblick über die Stichprobe . . . . .	239
10.3.2	Vertrauensmessung . . . . .	240
10.3.3	SEU-Modelle und Essenstypen . . . . .	242
10.3.4	SEU-Modelle und Modelldifferenzen . . . . .	245

10.3.5	Zusammenhang von Vertrauen und SEU-Modellstruktur . . . . .	247
10.3.6	Inhaltliche Interpretation unterschiedlicher Vertrauensstärke . . .	249
10.4	Zwischenfazit: MensaAssist . . . . .	253
<b>11</b>	<b>Experiment: Simhybs</b>	<b>255</b>
11.1	Simhybs-Oberfläche und Handlungsmodell . . . . .	256
11.1.1	Oberfläche . . . . .	256
11.1.2	Handlungsmodell . . . . .	258
11.2	Design des Simhybs-Experiments . . . . .	259
11.3	Auswertung . . . . .	262
11.3.1	Überblick über die Stichprobe . . . . .	262
11.3.2	Vertrauensmessung . . . . .	263
11.3.3	Automationsmodi im Vergleich . . . . .	264
11.3.4	Zusammenhang von Vertrauen und Zielintensitäten . . . . .	270
11.3.5	Entwicklung von Vertrauen . . . . .	273
11.3.6	Vertrauensentwicklungstypen . . . . .	277
11.4	Zwischenfazit . . . . .	284
<b>12</b>	<b>Fazit</b>	<b>287</b>
<b>IV</b>	<b>Literaturverzeichnis und Anhang</b>	<b>293</b>
	<b>Literaturverzeichnis</b>	<b>295</b>
<b>A</b>	<b>Anhang</b>	<b>315</b>
A.1	Übersicht über Wikipedia-Namensräume . . . . .	315
A.2	Wikipedia-Datenquellen . . . . .	315
A.2.1	Datengrundlage zu Abbildung 5.4 . . . . .	315
A.2.2	Datengrundlage zu Abbildung 5.5 . . . . .	315
A.3	Mensa-Befragung Winkler . . . . .	317
A.3.1	Durchschnittliches SEU-Modell für häufigste Wahl . . . . .	317
A.4	Mensa-Befragung Yilmaz . . . . .	319
A.4.1	Durchschnittliche SEU-Modelle . . . . .	319
A.5	MensaAssist . . . . .	321
A.5.1	Fragebogen zu MensaAssist . . . . .	321
A.5.2	Entscheidungsgrundlage MensaAssist / Übereinstimmungen . . .	335
A.5.3	Auswertung MensaAssist . . . . .	338
A.6	Simhybs . . . . .	342
A.6.1	Fragebogen . . . . .	342
A.6.2	Faktoranalyse Vertrauen . . . . .	347
A.6.3	Faktorenanalyse Semantisches Differenzial . . . . .	355

A.6.4	Mittelwertdifferenzen und Konfidenzintervalle . . . . .	357
A.6.5	Clusteranalyse Vertrauensentwicklungstypen . . . . .	358
A.6.6	Korrelationen Vertrauensentwicklungstypen . . . . .	363
A.6.7	Performancedaten Vertrauensentwicklungstypen . . . . .	365

## Abbildungsverzeichnis

2.1	Zusammenhänge zwischen interdisziplinären Vertrauenskonstrukten . .	25
3.1	Vertrauensspiel in extensiver Form . . . . .	56
3.2	Spielbaum des iterierten Vertrauensspiel . . . . .	62
4.1	Higher orders of automaticity . . . . .	73
4.2	Das Problem des Vertrauens in der Automationsforschung . . . . .	76
4.3	Steuerung einer Pasteurierungsanlage . . . . .	81
4.4	Klassifizierung feindlicher Flugzeuge . . . . .	84
4.5	Erkennung versteckter Soldaten . . . . .	87
4.6	Grafische Anzeige des Shipboard Piloting Expert System (SPES) . . .	92
5.1	Selbstbeschreibungsseite eines Bots . . . . .	119
5.2	Versionshistorie eines Bildes . . . . .	124
5.3	IWLC: Beispiel für die hybride Konzeption eines Bots . . . . .	126
5.4	Aktivität von DrTrigonBot . . . . .	131
5.5	Beispiel für Inter-Bot-Kooperation . . . . .	132
6.1	Coleman'sche Erklärung im H-O-Schema . . . . .	158
6.2	Das Modell soziologischer Erklärung (MSE) . . . . .	161
6.3	MFS-Schaubild . . . . .	165
7.1	Grundstruktur einer ABMS . . . . .	179
7.2	Segregation in Netlogo . . . . .	184
8.1	Hybrides Interaktionssystem . . . . .	191
8.2	Hybrid-Akteur im Makro-Mikro-Makro-Modell . . . . .	193
8.3	Modell soziologischer Erklärung hybrider Systeme . . . . .	196
8.4	ACC-Cockpit . . . . .	198
10.1	Visuelle Analogskala im Mensa-Fragebogen . . . . .	224
10.2	Fragesequenz im Online-Fragebogen zu MensaAssist . . . . .	233

10.3	MensaAssist: Abfrage von U-Werten . . . . .	235
10.4	Boxplot für kumulierte SEU-Differenzen . . . . .	246
10.5	Zusammenhang von kumulierten SEU-Differenzen und Vertrauen . . . . .	248
10.6	MensaAssist – Mittelwerte der Zielintensitäten . . . . .	251
10.7	MensaAssist – Mittelwerte der Zielintensitäten stark Vertrauender . . . . .	252
11.1	Elemente der Simhybs-Oberfläche . . . . .	257
11.2	Simhybs Experimentaldesign . . . . .	260
11.3	Fragebatterie zum semantischen Differenzial . . . . .	267
11.4	Grafische Darstellung der Vertrauensentwicklungstypen . . . . .	278
11.5	SIMHYBS– U-Werte für Vertrauensentwicklungscluster . . . . .	281
11.6	Vertrauensentwicklungstypen und Tastaturanschläge . . . . .	282

## Tabellenverzeichnis

2.1	Elemente des Vertrauens in Anlehnung an Nooteboom . . . . .	33
3.1	Auszahlungsmatrix der Grundversion des Vertrauensspiels . . . . .	57
3.2	Auszahlungen im experimentellen Vertrauensspiel . . . . .	58
3.3	Zusammentreffen Typ 1- / Typ 2-Spieler . . . . .	61
3.4	Auszahlungsmatrix Typ 1- / Typ 2-Spieler . . . . .	61
3.5	Auszahlungsmatrix Typ 2- / Typ 2-Spieler . . . . .	61
3.6	Auszahlungsmatrix für das heterogene Vertrauensspiel . . . . .	65
5.1	Tätigkeitsfelder der Bots . . . . .	121
5.2	Anteil Bot-Editierungen 2010 vs. 2012/13 . . . . .	123
5.3	Kommentierte Versionshistorie . . . . .	125
5.4	Aktivitäten ausgewählter Bots . . . . .	128
6.1	H-O-Erklärung . . . . .	155
10.1	Mensa-Befragung: SEU-Modelle für häufigste Wahl . . . . .	225
10.2	Unterschiede laut t-Test „häufigste Wahl“ . . . . .	227
10.3	Mensa-Befragung Yilmaz – SEU-Modelle für häufigste Wahl . . . . .	229
10.4	Vergleich der SEU-Modelle für häufigste Wahl . . . . .	230
10.5	Essenstypen aus dem MensaAssist-Fragebogen . . . . .	232
10.6	MensaAssist – Operationalisierung von Essensbewertungen . . . . .	234

---

10.7	Umsetzung Vertrauensskala von Jian u. a. . . . .	236
10.8	Soziodemografika zum Experiment MensaAssist . . . . .	239
10.9	Faktorladungen MensaAssist 2 Faktoren . . . . .	241
10.10	MensaAssist – Faktorladungen für den Faktor Vertrauen . . . . .	242
10.11	Zielbewertungen für drei Essenstypcluster . . . . .	243
11.1	Simhybs Fahrmodi . . . . .	259
11.2	Umsetzung Vertrauensskala von Jian u. a. . . . .	261
11.3	Simhybs-Fallzahlen . . . . .	263
11.5	Simhybs – Faktorladungen für den Faktor Vertrauen . . . . .	263
11.4	Soziodemografika zum Experiment SIMHYBS . . . . .	264
11.6	Mittelwerte und 95%-Konfidenzintervalle je Modus . . . . .	265
11.7	Semantisches Differenzial ja Modus . . . . .	267
11.8	Dem FAS zugeschriebene U-Werte . . . . .	269
11.9	Korrelation zwischen Vertrauensmessungen und U-Werten . . . . .	271
11.10	Simhybs – Korrelationen zwischen Vertrauensmessungen . . . . .	274
A.1	Wikipedia-Namensräume . . . . .	316



# 1 Einleitung

Autonome Technik durchzieht inzwischen nahezu jeden gesellschaftlichen und wirtschaftlichen Lebensbereich. Fahrerassistenzsysteme, intelligente WCs, Pflege- und Operationsroboter, smarte Stromnetze und autonome Softwareagenten sind nur einige Beispiele, die verdeutlichen, in welchem Umfang die Technisierung des Alltags bereits vorangeschritten ist. Autonome Technik ist längst keine exklusive Spielerei von Computer-Nerds mehr, oder solchen Personen, die sich wissenschaftlich mit autonomer Technik auseinandersetzen (wollen) – im Gegenteil.

Als besonders prägnantes Beispiel dafür, dass *autonome* Technik in der Mitte der Gesellschaft angekommen ist, kann die publizistische Reaktion auf den Dow Jones-Kursverfall um über 1 000 Punkte am 06.06.2010, mitverursacht durch autonome technische Systeme, gedeutet werden. Vermutlich durch eine fehlerhafte Zahleneingabe initiiert, entschieden sich zahlreiche autonome High Frequency Trader für massenhafte Verkäufe und trugen damit zu einem sich selbst verstärkenden Kursverfall bei. In dem Zuge waren Formulierungen wie „die Maschinen haben einfach die Kontrolle übernommen“ (Kuls und Mohr 2010) oder „der Kampf der Algorithmen ist [...] nicht zu durchschauen“ (Spiegel Online 2010) vielfach in den Zeitungen zu lesen.

Sobald autonome Technik beginnt, sich nicht einfach nur auf bloßes „Funktionieren“ zu beschränken, sondern aktiv in gesellschaftliche Interaktionszusammenhänge eingreift, Handlungsmöglichkeiten menschlicher Akteure einschränkt und die Gefahr des Fehllaufens besteht, stellen sich Fragen nach dem Vertrauen in ebendiese Technik.

Autonome Technik dient im zweifachen Sinne *nicht* dem Selbstzweck: Erstens ist sie das Produkt eines wie auch immer gearteten intentionalen Konstruktionsprozesses. Daraus folgt zweitens, dass „Technik in Aktion“ (Rammert 2003) ihr Verhalten auf ein oder mehrere Ziele hin orientiert. Über den Einsatzzweck und die Justierung der Ziele bestimmt, trotz aller Autonomie, nicht die Technik allein. Vielmehr ist sie eingebunden in ein soziotechnisches Geflecht der Mensch-Maschine-Interaktion.

Vor dem Hintergrund möchte sich die vorliegende Arbeit verstanden wissen als eine Analyse der Mensch-Maschine-Interaktion (MMI) mit dem Ziel, ein *soziologisches* Verständnis von Elementen und Aspekten des Vertrauens in autonome Technik zu entwickeln. Vertrauensfragestellungen sind auch Gegenstand der Automationsforschung, werden dort allerdings weitgehend ohne Bezug zu genuin soziologischen Theorien als ingenieurwissenschaftliche Herausforderung behandelt. Deswegen ist eine tiefergehende soziologische Analyse, die anschlussfähig ist an ingenieurwissenschaftliche Denkweisen, mehr denn je erforderlich.

Das Zitat „There are as many forms of trust as there are [...] causes of action“ (Nooteboom 2002: 49) deutet an, dass zahlreiche Vertrauensformen und -konzepte im Hinblick auf ihre Relevanz für den Gegenstandsbereich der autonomen Technik zu berücksichtigen sind. Es weist auch auf einen engen Zusammenhang zwischen Vertrauen, Handlungen und Gründen hin, der systematisch untersucht werden wird. Anstatt Technik lediglich als „Infrastruktur der Gesellschaft“ (Esser 2000a: 452) aufzufassen, sind die *zunehmende Eigenaktivität* und das *Mithandeln* autonomer Technik (Rammert und Schulz-Schaeffer 2002a; Weyer 2006) im Rahmen einer soziologischen Analyse konstitutiv zu berücksichtigen. Dazu werden Überlegungen angestellt, auf welche theoretischen Konzepte eine solche Analyse rekurrieren kann. Es wird sich herausstellen, dass für diesen Zweck insbesondere an handlungstheoretische Makro-Mikro-Makro-Modelle Esser'scher Provenienz (Esser 1993: 83ff.) angeschlossen werden kann.

Das Zusammenhandeln von Mensch und autonomer Technik, so die übergeordnete These, lässt sich im Rahmen einer symmetrisch-teleologischen und handlungstheoretisch fundierten Methodologie untersuchen. Die Begründung dafür liegt in der Überlegung, dass sich das Zusammenspiel von Mensch und autonomer Technik stets im Kontext einer gemeinsam zu erfüllenden Aufgabe abspielt und vor dem Hintergrund symmetrisch modelliert werden kann. In den Fokus der Betrachtung rücken dabei insbesondere wechselseitige Modellzuschreibungen und operationalisierbare Modell-differenzen. Dabei wird sich herausstellen, dass die Mensch-Maschine-Interaktion gerade für diejenigen Fälle problematisch ist, in denen Mensch und Nicht-Mensch je unterschiedliche Erwartungen aneinander stellen und damit ein Vertrauensproblem evozieren.

Der Empirie-Teil der vorliegenden Arbeit nimmt derartige Unterschiede in experimenteller Perspektive in den Blick und analysiert ihre Auswirkungen auf Vertrauen. Im Rahmen von zwei Experimentalstudien, MensaAssist und SIMHYBS, erfolgt eine systematische Verbindung unterschiedlicher Überlegungen zu Vertrauen und der handlungstheoretisch fundierten Methodologie zur Mensch-Maschine-Interaktion.

Um mit Mayntz (1967: 27) zu sprechen sind der durch „Modellkonstruktion ausgeübte Zwang zur Präzisierung“ und die Notwendigkeit, „*ausdrückliche* Annahmen zu machen“ diejenigen Aspekte eines *formalen* Modells, die seinen heuristischen Mehrwert ausmachen.

Der Zwang zu expliziten Annahmen ist für die entwickelten Experimente unbedingt erforderlich. Bei SIMHYBS wird Vertrauen in autonome Technik mit Hilfe einer interaktiven und (soziologisch) mikrofundierten Computersimulation untersucht, die an eine Fahrsimulation angelehnt ist. Die Untersuchung zu MensaAssist bildet mittels einer Online-Befragung einen Entscheidungshelfer nach, der die Befragten bei ihrer individuellen Essenswahl unterstützen soll. Die beiden Experimente sind so angelegt, dass sie jeweils unterschiedliche Aspekte des Vertrauens abdecken und untersuchbar machen.

---

## Kapitelübersicht und Argumentationsgang

Die Arbeit gliedert sich grob in drei Teile: *Vertrauen und Automation* (Teil I), *Verbindende Konzepte zwischen Computersimulation, Handlungstheorie und autonomer Technik* (Teil II) und schließlich der *empirische* Teil III mit den beiden Experimentalstudien. Aus den ersten beiden Teilen werden, wenn möglich, *forschungsleitende Annahmen* abgeleitet, die jeweils mit (FLA 1), (FLA 2) usw. bezeichnet sind. Die Kapitel werden jeweils auf vorangegangene FLAs Bezug nehmen und insbesondere im Empirieteil als Grundlage für das Design der Experimente sowie zur Ableitung der experimentbezogenen Hypothesen dienen. Eine Übersicht über alle forschungsleitenden Annahmen findet sich in Abschnitt 9.1 auf Seite 213.

In Teil I erfolgen definitorische Klärungen (Kapitel 2) des Vertrauensbegriffs. Auf Grundlage von Meta-Analysen zu Vertrauensbegriffen werden relevante Aspekte des Vertrauens in autonome Technik identifiziert. Es schließt sich mit Kapitel 3 eine Würdigung unterschiedlicher Forschungstraditionen, die sich mit Vertrauen beschäftigen, an. Dabei werden u. a. probabilistische und spieltheoretische Vertrauensmodelle dahingehend untersucht, ob sie für den Gegenstandsbereich autonome Technik geeignet sind. Insbesondere wird gezeigt, dass Vertrauen ein multidimensionales Konzept ist. In Bezug auf Charakteristika autonomer Technik lässt sich Vertrauen in interaktivprozessförmiger Perspektive als Zuschreibung auf Verhalten und Motive auffassen. Schließlich umfasst Kapitel 4 eine gründliche Beschäftigung mit Vertrauensstudien aus der Automationsforschung. Aus diesem Kapitel werden Anregungen für die Experimentaldesigns im Empirieteil gewonnen sowie inhaltliche Ergebnisse in Bezug auf Vertrauen und Automation, z. B. den Einfluss von Ausgangsvertrauen, zusammengestellt. Einige der Ergebnisse gehen dann in die Hypothesen im Empirieteil ein.

Teil II befasst sich schwerpunktmäßig mit autonomer Technik und mit Möglichkeiten, wie diese soziologisch fundiert im Rahmen einer integrierenden Methodologie, die Gegenstand von Kapitel 8 ist, untersucht werden kann. In Vorbereitung auf diese Integration sind in Kapitel 5 einige Vorüberlegungen hinsichtlich einer soziologischen Sichtweise auf autonome Technik anzustellen. Es wird u. a. geprüft, welche soziologischen Theorieangebote für die Berücksichtigung des Mithandelns von autonomer Technik bereitstehen und welche Lücken diese Ansätze aufweisen. Es werden dort auch Fallbeispiele, wie Wikipedia-Bots und Fahrerassistenzsysteme angeführt, die zeigen, dass autonome Technik auch in einem soziologischen Verständnis *handlungswirksam* ist und aktiv in Interaktionszusammenhänge eingreift. In Kapitel 6 erfolgt mit Blick auf die theoretische Grundlage der integrierenden Methodologie eine ausführliche Würdigung des Modells soziologischer Erklärung, gekoppelt mit erkenntnistheoretischen Überlegungen zum methodologischen Individualismus. Kapitel 7 befasst sich mit dem Einsatz von Computersimulation und Aspekten der Modellierung und des Experimentierens in der Soziologie. Das Kapitel läuft argumentativ darauf hinaus, dass die Methode der agentenbasierten Modellierung und Simulation, entsprechende

soziologische Mikrofundierung vorausgesetzt, das Modell soziologischer Erklärung im Computer implementieren kann. Die integrierende Methodologie spezifiziert ein Modell soziologischer Erklärung hybrider Systeme (HMSE), mit dem sich wechselseitige Modellzuschreibungen und Modelldifferenzen, auch mit Hilfe von interaktiven Computersimulationen, untersuchen lassen. Diese haben, so die Vermutung, Einfluss auf Vertrauen.

Teil III startet zunächst mit Kapitel 9, in dem alle bis dahin formulierten forschungsleitenden Annahmen systematisch zusammengeführt und für die beiden Experimente, gekoppelt mit Überlegungen aus den vorangegangenen Kapiteln, zu Hypothesen verdichtet werden. Diese werden dann für MensaAssist (Kapitel 10) und SIMHYBS (Kapitel 11) überprüft. Die Experimente sind so aufgebaut, dass sich unterschiedliche Vertrauensaspekte und verschiedene Modellzuschreibungsrichtungen untersuchen lassen.

Das abschließende Kapitel 12 fasst schließlich die inhaltlichen, forschungsheuristischen und theoretischen Ergebnisse zusammen und gibt einen Ausblick auf weitere Forschungsperspektiven.

**Teil I**

# **Vertrauen und Automation**



## 2 Definitiorische Klärung des Vertrauensbegriffs

Folgt man der Definition des Duden Universalwörterbuchs, so bedeutet Vertrauen „festes Überzeugtsein von der Verlässlichkeit, Zuverlässigkeit einer Person, Sache“. Der so definierte Vertrauensbegriff erstreckt sich interessanterweise nicht nur auf Personen, sondern auch explizit auf „Sachen“ und damit auch auf Technik. Weiter wird aus der Definition deutlich, dass Vertrauen ausgehend von einem Subjekt (in der Definition nicht näher genannt), dem Treugeber, einem Objekt (Person, Sache), dem Treuhänder zugeschrieben wird und sich zusammensetzt aus den Eigenschaften Verlässlichkeit und Zuverlässigkeit.

Als erste Annäherung an den Vertrauensbegriff mag diese Definition genügen. Bei einer näheren Beschäftigung mit Vertrauen ist schnell festzustellen, dass im wissenschaftlichen Kontext der Begriff des Vertrauens mannigfaltig verwendet wird und eine klare Verortung der eigenen (Vertrauens-)forschung und des zugrundeliegenden Vertrauensbegriffs von großer Wichtigkeit ist. Die vorliegende Arbeit stützt sich bei der Einordnung auf zwei elaborierte Metaanalysen zum Thema Vertrauen: Zunächst wird der Aufsatz „Trust and Distrust Definitions: One Bite at a Time“ von McKnight und Chervany (2001) ausführlich dargelegt und kritisch diskutiert. Die Autoren haben insgesamt 65 Vertrauensdefinitionen untersucht und daraus ein abstrahiertes Klassifizierungsschema abgeleitet. Als zweiter definitiorischer Überblick dienen die von Bart Nooteboom (2002) vorgeschlagenen Klassifizierungen, die im Rahmen seiner Monografie „Trust. Forms, Foundations, Functions, Failures and Figures“ entstanden sind.

Bei der Würdigung beider Metaanalysen werden an Stellen, an denen es inhaltlich sinnvoll erscheint, Querbezüge zu weiteren Forschungsarbeiten und Begriffen hergestellt. Es wird zu überprüfen sein, ob sich die Erkenntnisse der Autoren auf den Untersuchungsgegenstand dieser Arbeit, auf autonome Technik und damit auf das Feld nicht-menschlicher Akteure, übertragen lassen. In Abschnitt 2.3 werden die zentralen Ergebnisse zu einem eigenen, für diese Arbeit adäquaten, Vertrauensbegriff kondensiert und forschungsleitende Annahmen formuliert, auf die im weiteren Verlauf der Arbeit Bezug genommen wird.

## 2.1 Das Klassifizierungsschema von McKnight und Chervany

McKnight und Chervany motivieren ihre Metaanalyse zu Vertrauenskonzepten mit der Erwartung, dadurch vergleichbare wissenschaftliche Ergebnisse zu erhalten: „To enable trust researchers to more easily compare empirical results“ (McKnight und Chervany 2001: 27).

Den Ausgangspunkt der Autoren bildet die Feststellung, dass sich unterschiedliche wissenschaftliche Disziplinen mit unterschiedlichen Vertrauensaspekten beschäftigen (ebd.: 31). So befasse sich die Psychologie mit der persönlichen Seite („personality site“), die ökonomische Vertrauensforschung mit den Rational Choice-Aspekten des Vertrauens und die Soziologie mit dem Zusammenhang von Vertrauen und sozialen Strukturen. Ob diese strikte disziplinäre Zuordnung für die Soziologie und damit insbesondere für die vorliegende Arbeit vor dem Hintergrund soziologischer Tiefenerklärungen und den damit einhergehenden Makro-Mikro-Makro-Erklärungen berechtigt ist, wird in Abschnitt 2.1.6 kritisch hinterfragt werden.

Die Autoren untersuchen zwar nur auf Personen (bzw. Institutionen) bezogenes Vertrauen, schließen aber nicht aus, dass sich der Vertrauensbegriff auch grundsätzlich auf Technik übertragen lässt: „Although we define trust in terms of people, these definitions may be adapted for trust of people in computers or trust between computer agents“ (vgl. ebd.: 28).

Die Grobeinordnung der Definitionen erfolgt zunächst in drei Großkategorien: interpersonelles, dispositionelles und institutionelles Vertrauen. Innerhalb dieser Großkategorien werden Subkategorien, sogenannte konzeptionelle Level gebildet, die ihrerseits wieder operationalisierbare Subkonstrukte enthalten und Vertrauen messbar machen können<sup>1</sup>.

Zur Entwicklung ihres Klassifizierungsschemas haben McKnight und Chervany (2001: 31) Vertrauensdefinitionen aus Monografien und wissenschaftlichen Fachartikeln untersucht. Davon entstammen 25 der Psychologie, 23 der Management-Literatur und 19 den Bereichen Soziologie, Ökonomie und Politikwissenschaften. Herausgekommen sind bei der Analyse fünf *konzeptionelle Vertrauentypen* (ebd.: 33ff.) zu denen jeweils eine eigene Übersetzung angegeben wird:

- Disposition to Trust – Vertrauensdisposition
- Institution-Based Trust – Institutionelles Vertrauen
- Trusting Beliefs – Vertrauende Überzeugung
- Trusting Intentions – Vertrauende Absichten
- Trust-Related Behaviour – Auf Vertrauen bezogenes Verhalten

---

<sup>1</sup>Die Großkategorien und Subkonstrukte beziehen sich nicht auf die direkte Messung von Vertrauen, sondern auf die gefundenen Zusammenhänge in der analysierten Vertrauensliteratur (vgl. auch McKnight u. a. 1998).

Für jeden der fünf Typen erfolgt einer Operationalisierung auf Basis zugeschriebener Treuhänder-Charakteristika, z. B. Kompetenz („competence“), Gutmütigkeit („benevolence“) oder Vorhersagbarkeit („predictability“). Die je Typ resultierenden Konstrukte sind empirisch messbar und können auf diese Weise in vergleichenden Vertrauensstudien verwendet werden. Zwischen den Konstrukten bestehen Kausalzusammenhänge: Beispielsweise hängt auf Vertrauen bezogenes Verhalten mit vertrauenden Absichten und vertrauender Überzeugung zusammen (vgl. ebd.: 33 / Abschnitt 2.1.6). In den folgenden Abschnitten werden die fünf von McKnight und Chervany identifizierten Vertrauenskonstrukte näher beschrieben.

### 2.1.1 Vertrauensdisposition

„Disposition to Trust means the extent to which one displays a consistent tendency to be willing to depend on general others across a broad spectrum of situations and persons“ (McKnight und Chervany 2001: 38). Das Konstrukt Vertrauensdisposition fällt gemäß der Autoren in den Bereich der Psychologie und bezieht sich auf „generalisierte Andere“ und damit auf nicht angebbare Personen oder Institutionen. Die Disposition ist relativ unspezifisch und stellt lediglich eine Tendenz dar, ob ein Treugeber grundsätzlich gewillt ist, zu vertrauen. Folgt man McKnight und Chervany (ebd.) wirkt die Vertrauensdisposition insbesondere auf die Einordnung einer bislang unbekanntem Situation oder Person durch den Treugeber.

Das Konstrukt „Disposition to Trust“ setzt sich zusammen aus (vgl. McKnight und Chervany 2001: 38f.):

- *Faith in Humanity* – Genereller Glaube an Menschlichkeit
- *Trusting Stance* – Vertrauensvolle Grundhaltung (als persönliche Strategie)

*Faith in Humanity* bezeichnet eine persönliche Einstellung zur angenommenen Menschlichkeit generalisierter Anderer, z. B. bei einem Unfall von einem Fremden echte Hilfe zu erhalten.

*Trusting Stance* ist eine persönliche Strategie, die sich im Sinne einer subjektiven Kosten-Nutzen-Kalkulation im Lichte vorheriger Erfahrungen (ebd.) als recht erfolgreich erwiesen hat. Vertrauen als Strategie, so lässt sich folgern, ist die Antwort auf die von Diego Gambetta (1988a) formulierte Frage: „can we trust trust?“, die er eindeutig bejaht (vgl. auch Abschnitt 3.2.5) und argumentiert, dass es rational sei, Vertrauen zu vertrauen (234).

### 2.1.2 Institutionelles Vertrauen

„Institution based Trust means one believes, with feelings of relative security, that favorable conditions are in place that are conducive to situational success in a risky endeavor or aspect of one’s life“ (McKnight und Chervany 2001: 37).

Institutionelles Vertrauen bezieht sich auf (gesellschaftliche) Strukturen und nicht auf konkrete Personen: „Institution-based trust refers to beliefs about those protective structures, not about the people involved“ (37). Das operationalisierte Konstrukt umfasst:

- *Structural Assurance / Strukturelle Sicherheit*
- *Situational Normality / Situative Normalität*

*Structural Assurance* garantiert Strukturen, in denen Verträge und Prozessabläufe Gültigkeit besitzen und entsprechend ihrer Vorgabe funktionieren bzw. rechtsbindend sind. Als Beispiel nennen die Autoren u. a. die von verschlüsselter Datenübertragung im Web ausgehende strukturelle Sicherheit, die gewährleistet, dass sensible Informationen nicht von Dritten mitgelesen werden können (37). Der (überzogene) Glaube in Sicherheitsprozeduren bei der NASA im Zuge des Challenger-Unglücks ist als weiteres Beispiel für strukturelle Sicherheit angeführt (ebd.). Die Gefahr, überzogenes Vertrauen als Resultat an den Glauben struktureller Sicherheit zu betrachten, wird noch im Abschnitt 4.1 zu „Vertrauen als Problem der Automation“ thematisiert werden. So wäre denkbar, dass ein überzogenes strukturelles Vertrauen z. B. in die Fähigkeiten der deutschen Automobil-Ingenieure<sup>2</sup>, perfekte Fahrerassistenzsysteme für ihre Autos zu konstruieren, u. a. zum Problem des „overtrust“ führt, das ein wichtiger Untersuchungsgegenstand der Automationsforschung ist.

*Situational Normality* stellt einen Zusammenhang her zwischen Vertrauen und dem Vorliegen einer geordneten Situation, die in vorhandene Muster passt (vgl. McKnight und Chervany 2001: 37f.). Die Autoren beziehen sich dabei auf interaktionistische Konzepte von Garfinkel (2005), der in seinen sozialpsychologischen Krisenexperimenten u. a. herausgefunden hat, dass Interaktionssituationen von impliziten und meist unhinterfragten Vorbedingungen geprägt sind. Um diese Vorbedingungen sichtbar zu machen, konfrontierte er in alltäglichen Interaktionssituationen die jeweiligen Interaktionspartner (die nicht wussten, dass sie Bestandteil eines sozialpsychologischen Experiments waren) mit vollkommen inadäquaten Reaktionen: Nachfragen zu scheinbaren Selbstverständlichkeiten verursachten Verunsicherung und Verwirrung (Garfinkel 2005: 371f.). Auf die Aussage „Hurry or we will be late“ fragte der Versuchsleiter nach, „[...] what did he mean by late and from what point of view did it have reference“ (372). Der Proband reagierte darauf mit Ablehnung und großer Irritation. Weitere Experimente, etwa des Behandeln eines normalen Kunden in einem Geschäft als wäre er Verkäufer (372ff.), oder eine völlig verquere Interpretation einer scheinbar offensichtlichen Situation durch Dritte (376ff.) führten immer wieder zu Irritation und stellten die situative Normalität in Frage.

---

<sup>2</sup>Aus Gründen der Vereinfachung wird ausschließlich die männliche Form verwendet. Personen weiblichen wie männlichen Geschlechts sind darin gleichermaßen eingeschlossen.

Garfinkel konnte mit seinen Experimenten zeigen, dass die interaktive Herstellung abnormaler Situationszustände zur Zerstörung des interpersonellen Vertrauens führen kann und im schlimmsten Fall eine Interaktion abrupt enden lässt.

Welche Implikationen sich aus der Hinterfragung der situativen Normalität für autonome Technik ergeben, wird in Abschnitt 2.3 diskutiert.

### 2.1.3 Vertrauende Überzeugung

Während es sich bei institutionellem und dispositionellem Vertrauen um unpersönliche Vertrauenskonzepte handelt, die sich auf keine spezifische Sache oder Person beziehen, rekurren *Trusting Beliefs*, *Trusting Intentions* und *Trust-Related Behaviour* auf spezifische Personen.

„Trusting Beliefs means the extent to which one believes, with feelings of relative security, that the other person has characteristics beneficial to one“ (McKnight und Chervany 2001: 36).

Die Definition verdeutlicht zwei Eigenschaften des hier gemeinten Vertrauensbegriffs: Erstens sieht der Treugeber eine Chance („feeling of relative security“), dass der Treuhänder das Vertrauen tatsächlich bestätigt. Zweitens hat der Treuhänder charakterliche Eigenschaften, die für den Treugeber in irgendeiner Weise nützlich („beneficial to one“) sind. Bei der Nützlichkeit geht es nicht um monetären Nutzen, sondern vor allem um für den Treugeber nützliche Eigenschaften, wie das Konstrukt „Trusting Beliefs“ verdeutlicht. McKnight und Chervany (2001: 36) setzen ihr Konstrukt aus folgenden Sub-Konstrukten zusammen:

- *Trusting Belief-Competence / Kompetenzüberzeugung*  
Die feste Überzeugung des Treugebers, dass ein Treuhänder die Fähigkeit besitzt, die vom Treugeber erwünschte Veränderung herbeizuführen.
- *Trusting Belief-Benevolence / Gutmütigkeitsüberzeugung*  
Die feste Überzeugung, dass dem Treuhänder die Interessen des Treugebers wichtig sind und er im treugeberischem Sinne handelt.
- *Trusting Belief-Integrity / Integritätsüberzeugung*  
Die feste Überzeugung des Treugebers, dass der Treuhänder Zusicherungen einhält und in guter Absicht handelt.
- *Trusting Belief-Predictability / Vorhersagbarkeitsüberzeugung*  
Unabhängig vom positiven oder negativen Effekt einer Handlung des Treuhänders bezieht sich die Vorhersagbarkeitsüberzeugung darauf, dass der Treugeber die zukünftigen Handlungen des Treuhänders vorhersagen kann. McKnight und Chervany betonen (ebd.), dass die Vorhersagbarkeitsüberzeugung im Gegensatz zur Integritätsüberzeugung wertneutral ist.

Trusting Beliefs beziehen sich auf die Überzeugung des Treugebers hinsichtlich der Eigenschaften Kompetenz, Gutmütigkeit, Integrität und Vorhersagbarkeit des Treu-

händers. Dabei sind zwei Sub-Konstrukte dediziert normativ konstruiert. Die Überzeugung in Gutmütigkeit und Integrität des Treuhänders erfordert, dass der Treugeber eine generelle Vorstellung von Gut und Böse hat und charakterliche Eigenschaften des Treuhänders nach diesen Kriterien beurteilen kann. Kompetenz und Vorhersagbarkeit sind hinsichtlich ihrer Tragweite deutlich eingeschränkter, da sie sich auf ein erwartetes Verhalten des Treuhänders in einer spezifischen Situation beziehen. Charakteristisch, so lässt sich folgern, für das Gesamtkonzept von Trusting Beliefs ist dessen *positive* Definition, die Nebenwirkungen von Vertrauen, etwa Risiko, eigene Verletzbarkeit und mangelnde Kontrolle über den Treuhänder, ausblendet. Diese Nebenwirkungen werden im folgenden Konstrukt der „Trusting Intentions“ aufgegriffen.

#### 2.1.4 Vertrauende Absichten

Ebenso wie bei den im vorherigen Abschnitt erläuterten „Trusting Beliefs“ handelt es sich bei „Trusting Intentions“ um ein *personenspezifisches* Konstrukt:

„Trusting Intentions means one is willing to depend, or intends to depend, on the other party with a feeling of relative security, in spite of lack of control over that party, and even though negative consequences are possible“ (McKnight und Chervany 2001: 34).

Die Definition der vertrauenden Absichten bezieht explizit die Möglichkeit negativer Konsequenzen („even though negative consequences are possible“) mit ein. Trusting Intentions bedeutet nicht nur eine (folgenlose) Zuschreibung von Eigenschaften, sondern beinhaltet bereits eine Absicht des sich Verlassens („willing to depend, or intends to depend“) bzw. die Absicht einer Vertrauenshandlung, trotz mangelnder Kontrolle („lack of control over that party“) über den Treuhänder.

Der beschriebene Zusammenhang wird als Konstrukt wie folgt operationalisiert:

- *Willingness to Depend*: der Wille, sich auf den Treuhänder zu verlassen
- *Subjective Probability of Depending*: subjektive Wahrscheinlichkeit sich zu verlassen

Der Unterschied zwischen beiden Eigenschaften besteht laut McKnight und Chervany (2001: 34) insbesondere darin, dass die „Subjective Probability of Depending“ konkreter, spezifischer und handlungswirksamer ist als die bloße „Willingness to Depend“.

Kritisch lässt sich anmerken, dass die Operationalisierung der inhaltlich reichhaltigen Definition von „Trusting Intentions“ nicht gerecht wird. „Willingness to Depend“ bedeutet, wie die Autoren selbst anmerken (34), Verwundbarkeit in Kauf zu nehmen, indem man sich auf eine andere Person verlässt. Folglich umfasst die „Willingness to Depend“ mindestens zwei Unterkonstrukte, nämlich Risikoaffinität und geschätztes Risiko im Falle eines Vertrauensmissbrauchs, die leichter zu messen sind als die diffuse Operationalisierung „Willingness to Depend“.

Noch problematischer ist die „Subjective Probability of Depending“, da der Referent ein anderer ist als bei der vorherigen Eigenschaft. Ein Dritter oder ggf. auch der Treuhänder selbst „[...] forecasts or predicts that one will depend on the other person [...]“ (ebd.). Es handelt sich dabei nicht mehr um eine Eigenschaft des Treugebers, sondern um die Eigenschaft einer Zuschreibung durch Dritte. Schließlich fehlt der Operationalisierung auch der Aspekt der mangelnden Kontrolle über den Treuhänder.

### 2.1.5 Auf Vertrauen bezogenes Verhalten

Als letztes Konstrukt beschreiben die Autoren schließlich auf Vertrauen bezogenes Verhalten („Trust-Related Behaviour“):

„Trust-Related Behavior means that a person voluntarily depends on another person with a feeling of relative security, even though negative consequences are possible“ (McKnight und Chervany 2001: 34).

Im Gegensatz zu „Trusting Intentions“ bezieht sich „Trust-Related Behaviour“ auf die bereits vollzogene Vertrauenshandlung(en), die trotz möglicher Risiken („even though negative consequences are possible“) gewählt wurde(n). Trusting Intentions umfasst lediglich die *Absicht*, zukünftig eine Vertrauenshandlung zu wählen. Die Erscheinungsformen von Vertrauenshandlungen sind vielfältig und lassen sich nicht erschöpfend operationalisieren. Daraus lässt sich folgern, dass auf Vertrauen bezogenes Verhalten Chance und Gefahr zugleich ist und einen ambivalenten Charakter hat. Als mögliche Sub-Konstrukte, die sich jeweils mit einem eigenen kurzen Beispiel erläutern lassen, benennen McKnight und Chervany (2001: 34f.):

- *Cooperation / Kooperation*

Auf der einen Seite sind Kooperationsgewinne möglich, z. B. in Form von Reputation, im Falle eines erfolgreich gemeinsam publizierten Fachartikels, auf der anderen Seite muss sich der Treugeber darauf verlassen, dass er vom Treuhänder nicht betrogen wird. Kümmert sich der Treuhänder etwa um den empirischen Teil des Artikels, muss sich der Treugeber darauf verlassen, dass die Daten nicht manipuliert wurden.

- *Information sharing / Teilen von Informationen*

Dieses Verhalten macht den Treugeber dahingehend verwundbar, dass der Treuhänder die geteilten Informationen ausnutzen kann, z. B. bei Bereitstellung von persönlichen Informationen auf Social Networking Sites (SNS). Das Teilen von Informationen erfordert Vertrauen gegenüber der SNS und gegenüber den Personen, mit denen die Informationen geteilt werden.

- *Informal agreements / Informelle Absprachen*

Bei dieser Art von Verhalten geht der Treugeber davon aus, dass der Treuhänder sich an eine Absprache hält, ohne dass diese in Form eines schriftlichen Vertrags fixiert wäre.

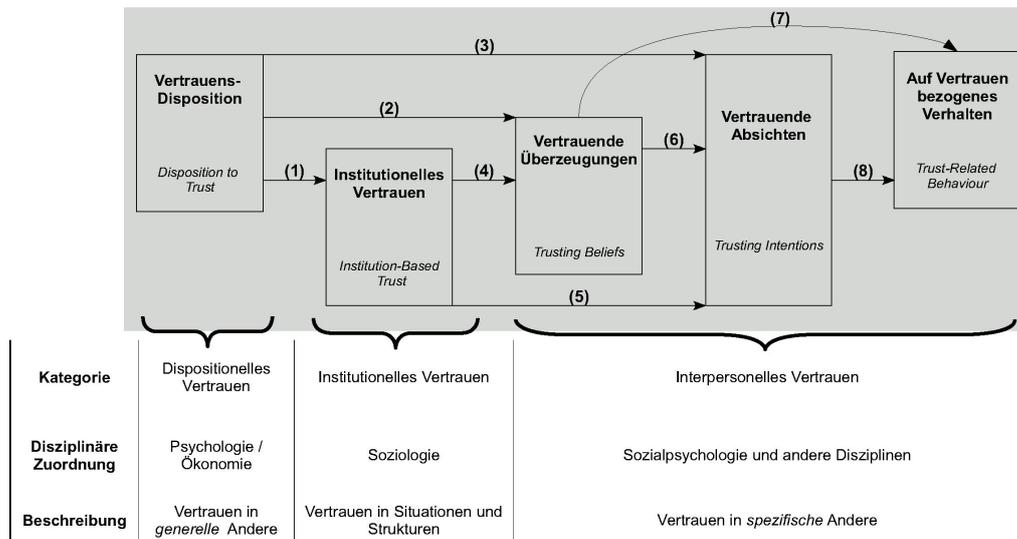
- *Reducing the controls / Kontrollreduktion*  
Ein Treugeber verringert seine Kontrolle über den Treuhänder und setzt sich dadurch einem größeren Risiko aus. Die abnehmende Kontrolle eines Fahrerassistenzsystems durch den Fahrer beinhaltet das Risiko, einen eventuellen Fehler des FAS nicht zu erkennen, gleichzeitig aber die Chance die mentale Belastung beim Autofahren zu senken.
- *Accepting influence / Akzeptieren von Einflussnahme*  
Das Annehmen fremder Einflussnahme erfordert das Eingehen des Risikos, einer eventuell falschen Meinung zu folgen, z. B. im Falle einer falschen Routenempfehlung.
- *Granting autonomy and more decision-making power / Zulassen von Autonomie und höherer Entscheidungsgewalt*  
McKnight und Chervany (2001: 34) beziehen ihre Erläuterung vor allem darauf, dass „[...] the boss must depend on the employee to make the right decisions“. Dies bedeutet als Verantwortlicher Entscheidungsgewalt abzutreten. Das Zulassen von Autonomie tangiert einen wichtigen Bereich der vorliegenden Arbeit: die teilweise Abgabe von Autonomie an autonome Technik. Ein Autofahrer muss sich z. B. darauf verlassen, dass ein Notbremsassistent in einer Gefahrensituation die richtige Entscheidung trifft.
- *Transaction business / Durchführen eines Geschäfts*  
Lässt sich ein Treugeber auf ein Geschäft ein, muss er dem Treuhänder dahingehend vertrauen, dass das Geschäft gemäß der getroffenen Absprachen korrekt durchgeführt wird.

Nach Vorstellung der unterschiedlichen Vertrauenskonstrukte schließt sich die Frage an, wie diese untereinander zusammenhängen. McKnight und Chervany (2001: 33) entwickeln eine zusammenfassende interdisziplinäre Vertrauentypologie, die Gegenstand des nächsten Abschnitts ist.

### 2.1.6 Zusammenhänge zwischen den Vertrauenskonstrukten

Wie McKnight und Chervany (2001: 48f.) trefflich bemerken, ist Vertrauen ein multidimensionales Konstrukt, das viele Facetten umfasst: „Like an elephant, trust is so large that it needs to be digested a bite at a time in order to make orderly progress.“ Die unterschiedlichen Dimensionen des Vertrauens, die im vorherigen Abschnitt bereits besprochen wurden, und ihre jeweilige disziplinäre Zuordnung sind in Abbildung 2.1 dargestellt.

Die eingezeichneten Pfeile sind als Einflusspfeile zu verstehen, z. B. beeinflusst die Vertrauensdisposition das institutionelle Vertrauen. Die Informationen in der Grafik umfassen zu jedem Konstrukt eine Vertrauenskategorie, eine *disziplinären Zuordnung* sowie eine kurze *Beschreibung*. Die disziplinäre Zuordnung ist kritisch zu hinterfragen.



**Abbildung 2.1:** Zusammenhänge zwischen interdisziplinären Vertrauenskonstrukten. Ein Pfeil von A nach B ( $A \rightarrow B$ ) bedeutet, dass A Einfluss auf B hat. Das Originalschaubild von McKnight und Chervany (2001: 33) ist um die Zeilenbeschriftungen (Kategorie, Beschreibung etc.), Pfeilnummerierungen und deutsche Übersetzungen ergänzt.

gen (vgl. Abschnitt 2.1.7), wird für die Erläuterung des Schaubildes jedoch zunächst als gegeben hingenommen. Die geschweiften Klammern deuten die Zuordnung der Konstrukte zu den zugehörigen erläuternden Tabellenspalten an.

Der logische Aufbau des Schaubildes, auch wenn die Autoren dies nicht explizit benennen, zielt darauf ab, auf Vertrauen bezogenes Verhalten mit je vorgelagerten beeinflussenden Konstrukten zu erklären. Überdies enthält das Schaubild implizit unterschiedliche Ausgangssituationen: eine Situation vor der Vertrauenswahl und eine nach der Vertrauenswahl. Das letzte Konstrukt, das auf Vertrauen bezogene Verhalten, ist das einzige Konstrukt, das sich auf einen Zeitpunkt *nach* der Vertrauenswahl bezieht. Hat sich ein Akteur oder eine Institution für die Vergabe von Vertrauen entschieden<sup>3</sup>, folgt auf Vertrauen bezogenes Verhalten. Die weiteren vier Konstrukte verweisen auf einen Zeitpunkt vor der Vertrauenswahl.

Im Folgenden werden die einzelnen Zusammenhänge zwischen den Konstrukten beschrieben:

**Vertrauensdisposition  $\rightarrow$  Institutionelles Vertrauen (1):** McKnight und Chervany (2001: 40) bemerken, dass die generelle Vertrauensdisposition, also die Grundeinstel-

<sup>3</sup>Der Aspekt der *expliziten* Vertrauenswahl, eine Entscheidung zwischen den Handlungsalternativen *vertrauen* und *nicht vertrauen*, spielt bei McKnight und Chervany (2001) keine Rolle. Dieser nimmt hingegen einen zentralen Stellenwert bei den probabilistischen und spieltheoretischen Vertrauensmodellen ein (vgl. Kapitel 3).

lung gegenüber Anderen, automatisch auf die Institution bzw. die Struktur und damit auch auf Personen innerhalb der Struktur abfärbt.

**Vertrauensdisposition → Vertrauende Überzeugungen (2):** Die generelle Vertrauensdisposition kann auch die vertrauenden Überzeugungen beeinflussen. Der direkte Einfluss findet allerdings nur in völlig neuen, bisher unbekanntem, Situationen statt. In der Regel wird dieser Einfluss über das institutionelle Vertrauen (Pfeile 1 und 4) vermittelt (ebd.).

**Vertrauensdisposition → Vertrauende Absichten (3):** Eine genauere Erläuterung zu diesem Zusammenhang findet sich in einem älteren Text von McKnight u. a. (1998: 478). Dort wird argumentiert, dass vor allem die vertrauende Grundhaltung aus dem Vertrauensdispositionskonstrukt Einfluss auf die vertrauenden Absichten nimmt und damit auf den Bereitschaft, sich auf eine andere Person zu verlassen. Besonders wirksam wird dieser Zusammenhang ebenso wie die Vertrauensdisposition in neuen, bisher unbekanntem, Situationen.

**Institutionelles Vertrauen → Vertrauende Überzeugung (4):** Der Aspekt der situativen Normalität als Teil des institutionellen Vertrauens wirkt sich auf die vertrauende Überzeugung aus. Eine analytische Begründung für den Zusammenhang liefern die Autoren nicht, sondern erläutern den Zusammenhang anhand eines Beispiels (McKnight und Chervany 2001: 38): Sofern ein Beschäftigter sein aktuelles berufliches Umfeld als situative Normalität wahrnimmt, überträgt er die Normalität auf die Personen, die sich in diesem Umfeld bewegen und damit auf seine vertrauende Grundüberzeugung in diese Personen. Analytisch könnte man, ganz im interaktionistischen Sinne, argumentieren, dass die situative Normalität im Wesentlichen von der Normalität der im situativen Kontext stattfindenden Interaktionen ausgeht. Träger von Interaktionen sind in aller Regel Personen (allgemein: Aktanten), so dass sich die situative Normalität auf zukünftige Interaktionen extrapolieren lässt und sich damit auf die vertrauenden Überzeugungen in ebendiese Personen auswirkt.

**Institutionelles Vertrauen → Vertrauende Absichten (5):** Die strukturelle Sicherheit („structural assurance“) als Teil des institutionellen Vertrauens wirkt sich auf vertrauende Absichten aus. Sie stellt die Umgebung bzw. den strukturellen Kontext bereit, in dem sich vertrauende Absichten entwickeln können (ebd.: 37).

**Vertrauende Überzeugungen → Vertrauende Absichten (6):** Die vertrauenden Überzeugungen sind eine wesentliche Quelle („foundation“) für vertrauende Intentionen (vgl. McKnight und Chervany 2001: 36). Entscheidend, so die Autoren, sei die Kombination aller im Konstrukt vorkommender Einzeleinflüsse (vgl. Abschnitt 2.1.3).

So sei z. B. der gute Wille des Treuhänders („benevolence“) nicht ausreichend, um vertrauende Intentionen auszubilden, da nicht klar ist, ob dieser überhaupt kompetent genug ist, um zu helfen (ebd.).

**Vertrauende Überzeugungen → Auf Vertrauen bezogenes Verhalten (7):** Die Autoren bemerken, dass Personen dazu neigen, „[...] to translate their beliefs and intentions into actions“ (ebd.: 37). Dieser Übersetzungsprozess kann im Rahmen einer soziologischen Handlungstheorie, etwas des Esser’schen Modells soziologischer Erklärung (vgl. Abschnitt 6.2) erfolgen, bei der situative Opportunitäten und Intentionen in die Variablen der Handlungstheorie übersetzt werden und letztlich zur Selektion einer bestimmten Handlung führen.

**Vertrauende Absichten → Auf Vertrauen bezogenes Verhalten (8):** Die Begründung ist analog zu Pfeil (7) und ergibt sich alleine aus der Tatsache, dass eine teleologische<sup>4</sup> Handlungstheorie für die Auswahl einer adäquaten Handlungsalternative *immer* auf die Intentionen des betrachteten Akteurs zurückgreift. Hat der Akteur (aus guten Gründen) vertrauende Intentionen ausgebildet, wirken diese Intentionen auf die Selektion einer intentional passenden Vertrauenshandlung.

### 2.1.7 Fazit und kritische Würdigung

Zusammenfassend lässt sich festhalten, dass das Klassifizierungsschema bzw. die Meta-Analyse zur Vertrauensforschung von McKnight und Chervany (2001) eine sinnvolle Möglichkeit bietet, um eigene Forschungsarbeiten, die sich mit Vertrauen beschäftigen, einzuordnen. Besonders hervorzuheben ist der Anspruch der Autoren, die destillierten Vertrauensbegriffe anhand von Konstrukten und Sub-Konstrukten zu operationalisieren und Zusammenhänge zwischen ebendiesen herzustellen. Details darüber, wie sich die Konstrukte statistisch hinsichtlich ihrer Konsistenz verhalten, finden sich leider nicht. Zwar wird auf eine empirische Untersuchung zur Quantifizierung dieser Zusammenhänge verwiesen (ebd.: 33), der dort referenzierte unveröffentlichte Aufsatz ist allerdings nach wie vor nicht auffindbar (Stand: 31. Januar 2014). Aus den Details zur Ermittlung der Konstrukte und Sub-Konstrukte hätten sich ggf. interessante Implikationen für die Konstruktion einer Vertrauensskala für die vorliegende Arbeit ergeben.

Neben den logischen Zusammenhängen zwischen den Konstrukten ordnen die Autoren diese jeweiligen Fachdisziplinen zu: Psychologie / Ökonomie, Soziologie und Sozialpsychologie. Kritisch lässt sich dazu bemerken, dass eine solch strikte Zuordnung, zumindest für die Soziologie, die bei McKnight u. Chervany ausschließlich auf Institutionen bezogen wird, nicht gerechtfertigt erscheint. Eine *erklärende* Soziologie im Sinne von Makro-Mikro-Makro-Erklärungen (vgl. Coleman 1991) soll zwar

<sup>4</sup>Von *télos* (griech.), Zweck / Ziel.

primär (Makro-)Strukturen und damit auch Institutionen erklären, diese entstehen, folgt man der Logik der Tiefenerklärung, allerdings erst durch Handlungen auf der Mikro-Ebene. Für die Mikro-Ebene können sowohl sozialpsychologische Aspekte Berücksichtigung finden, wie etwa die SEU-Theorie (vgl. Esser (1993) / Abschnitt 6.2), als auch psychologisch-dispositionelle, so geschehen beim Modell der Frame-Selektion (Kroneberg 2011). Ebenso wenig einsichtig ist, warum interpersonelles Vertrauen, das vertrauende Überzeugungen, vertrauende Absichten und schließlich sogar auf Vertrauen bezogenes Verhalten umfasst, der Sozialpsychologie und nicht auch der Soziologie zugeordnet wird.

Weiterhin ist kritisch anzumerken, dass dem Schema eine zeitliche Dimension fehlt. Die spezifizierten Konstrukte beziehen sich auf unterschiedliche Ebenen des Vertrauensproblems: Auf Vertrauen bezogenes Verhalten ist nur möglich, wenn sich ein Treugeber für die Vergabe von Vertrauen entschieden hat. Alle vorgeschalteten Konstrukte, die vor allem institutionelle und dispositionelle Faktoren umfassen, sind auf einen Zeitpunkt *vor* der Vertrauenswahl zurückzuführen.

Im zusammenfassenden Abschnitt 2.3 werden einige Aspekte aus dem Klassifizierungsschema zusammen mit dem im nächsten Abschnitt vorgestellten Vertrauensbegriff von Bart Nooteboom (2002) verwendet und in Form von forschungsleitenden Annahmen verdichtet werden.

## 2.2 Facetten und Objekte des Vertrauens nach Nooteboom

Bart Nooteboom (2002) hat sich in seiner Monografie ausführlich mit Vertrauen auseinandergesetzt und dabei zunächst eine sehr präzise Begriffsbildung betrieben, die im folgenden Abschnitt in Ergänzung zu McKnight und Chervany referiert wird. Sein Entwurf einer Vertrauensontologie wird hilfreich sein, um die relevanten Aspekte des Vertrauensbegriffs für die vorliegende Arbeit näher zu spezifizieren.

Eine Beschäftigung mit Vertrauen sei notwendig, folgt man Nooteboom (2002: 5), da menschliches Verhalten grundsätzlich geprägt sei von Unsicherheit und Unvorhersagbarkeit. Sofern man Vertrauen ignoriere, so Nooteboom weiter, ignoriere man ebenso Unsicherheit:

„As I will argue at length in this book, the importance and nature of trust arise from the unpredictability, or radical uncertainty, of human behaviour. Trust can be ignored only at the price of ignoring uncertainty.“  
(Nooteboom 2002: 5)

Nooteboom (2002: 8) hebt zur Klärung seines spezifischen Vertrauensbegriffs die doppelte Bedeutung von Vertrauen hervor, die insbesondere für die je spezifische Fundierung des Vertrauens bedeutsam ist: Vertrauen, so Nooteboom (ebd.), ist eine Eigenschaft des Treugebers, Vertrauenswürdigkeit eine des Treuhänders. Den Zusammen-

hang zwischen den beiden Konzepten Vertrauen und Vertrauenswürdigkeit erklärt Nootboom über Zuschreibungen: „Rational trust is based on an attribution by the trustor of reasons for trustworthiness to the trustee“ (8). Rationales Vertrauen ist demnach eine Zuschreibung von Gründen für Vertrauenswürdigkeit, die vom Treugeber für seine Einschätzung des Treuhänders vorgenommen wird. Darüber hinaus gibt es noch weitere (weniger rationale) Quellen des Vertrauens (ebd.), z. B. psychologische Quellen des Vertrauens.

### 2.2.1 Vertrauen als vierstelliges Prädikat

Nach den ersten definitiven Klärungen präzisiert Nootboom den Vertrauensbegriff weiter und führt die Vertrauens*beziehung* zwischen Treugeber und Treuhänder als vierstelliges Prädikat<sup>5</sup> und zentrale Analyseeinheit ein (8): „Someone [Subjekt des Vertrauens / Treugeber] has trust in something [Objekt des Vertrauens / Treuhänder] in some respect under some conditions“. Als Subjekte des Vertrauens kommen Agenten, die nach Verständnis von Nootboom Personen oder korporative Akteure sein können, in Frage. Als Objekte des Vertrauens eignen sich Personen, Organisationen, Institutionen, Sozioökonomische Systeme (allgemein: Dinge) (ebd.). Interessant an dieser Unterscheidung zwischen Objekten und Subjekten des Vertrauens ist die Tatsache, dass Nootboom als Treugeber ausschließlich menschliche Akteure (oder entsprechend übergeordnete aggregierte Akteureinheiten) vorsieht, als Treuhänder jedoch allgemein Dinge und damit auch autonome Technik (8). In Abschnitt 2.3 wird dieser Gedanke noch einmal aufgegriffen und für eine zwar attributionstheoretische, aber doch symmetrische Sichtweise plädiert.

Nootboom (2002: 10) fasst seine Überlegungen zu einer Vertrauensontologie zusammen, die sich nur teilweise aus der Definition von Vertrauen als vierstelliges Prädikat erschließt. Dabei unterscheidet er zwischen folgenden *Dimensionen des Vertrauens*, die jeweils mit eigenen Erläuterungen versehen sind:

- Subjekt des Vertrauens / Treugeber:  
Als Subjekte kommen menschliche Akteure oder auch korporative Akteure wie Organisationen in Frage.
- Objekte, auf die sich Vertrauen bezieht:  
Dabei unterscheidet Nootboom eher unspezifische Adressaten des Vertrauens und konkrete Aspekte des Verhaltens, auf die sich Vertrauen beziehen kann:
  - Treuhänder: Vertraute Dinge oder Akteure auf unterschiedlichen Ebenen, Organisationen, Institutionen und Systeme

---

<sup>5</sup>Der Begriff *Prädikat* ist im Sinne der mathematischen Prädikatenlogik (vgl. Hartmann 2012: 43) zu verstehen. Für diesen hiesigen Fall würde das Prädikat Vertrauen (genauer: Vertrauensbeziehung) bestehen aus  $Vertrauen(a, b, c, d)$  mit z. B.  $Vertrauen(Student, Professor, Wissen, Fachgebiet)$  mit der inhaltlichen Interpretation: Der Student hat Vertrauen in das Wissen des Professor bezogen auf dessen Fachgebiet.

- Aspekte des Verhaltens, auf die sich Vertrauen bezieht: Kompetenz, Intentionen und andere
- Grenzen oder Bedingungen des Vertrauens:  
Gemeint sich insbesondere institutionelle Strukturen und Rahmenbedingungen, in denen sich Vertrauen abspielt.
- Ressourcen:  
Größe der potenziellen Belohnungen (bei Bestätigung des Vertrauens) und Bestrafungen (bei Missbrauch des Vertrauens), die auf dem Spiel stehen
- Basis oder Quellen der Vertrauenswürdigkeit:  
Nooteboom bezieht sich dabei auf rationale Motive wie Angst vor Sanktionen durch eine übergeordnete Autorität (z. B. ein Gericht) und Reputationsverluste (vgl. ebd.: 64f.).
- Mentale und psychologische Quellen des Vertrauens:  
Dabei spielen eher irrationale Aspekte wie z. B. Liebe und Empathie eine Rolle (ebd. 77f.).

Wie bereits genannt, erfolgt unter partiellem Rückgriff auf die hier vorgestellte Vertrauensontologie in Abschnitt 2.3 eine Einordnung des Vertrauensbegriffs der vorliegenden Arbeit. Von besonderer Bedeutung ist die Eingrenzung auf bestimmte Objekte des Vertrauens. Aus forschungspragmatischen Gesichtspunkten, die mit der eigenen theoretischen Position und der Empirie zusammenhängen, sind bestimmte Aspekte des Verhaltens als Objekt des Vertrauens von exponiertem Interesse. Nooteboom (2002: 30) hat mit seinem Konzept der „multiplen Kausalität“, das auf Kausalitätsaspekte von Aristoteles zurückgeht, einen Vorschlag präsentiert, der Formen des Vertrauens aus *Gründen* für Handlungen herleitet: „There are as many forms of trust as there are (failures of) causes of action“ (49). Vertrauen über zugeschriebene Handlungsgründe zu modellieren ist anschlussfähig an deduktiv-nomologische Erklärungsmodelle, wie etwa das HMSE (Modell soziologischer Erklärung hybrider Systeme, vgl. Abschnitt 8.1), das eine theoretisch-soziologische Basis der Arbeit darstellt. Dort nimmt die Zuschreibung von Erwartungen und Gründen eine wichtige Funktion ein. Im folgenden Exkurs wird das Konzept der multiplen Kausalität etwas ausführlicher vorgestellt.

### 2.2.2 Exkurs: Multiple Kausalität

Der Lexikonartikel von Falcon (2011) fasst das Konzept der multiplen Kausalität von Aristoteles, das er insbesondere in dem Buch *Physik II 3* beschrieben hat, zusammen. Aristoteles geht es dabei um Erklärungen von Bewegungen bzw. Veränderungen in der Natur. Am Anfang seiner Überlegungen steht die Feststellung, dass die Suche

nach Gründen zentral für die Erklärung der uns umgebenden Welt ist (Falcon 2011: Abs. 0). Laut Falcon (2011: Abs. 2) besteht die zentrale These von Artistoteles darin, dass sich alles Erklärungsbedürftige durch vier Typen von Gründen erklären lässt.

Am Beispiel einer Bronze-Statue lassen sich die vier Kausalitätstypen erklären: Der sogenannte *material cause* beschreibt das, woraus etwas besteht. Im Falle der Bronze-Statue das Material Bronze. Der Rohmaterial Bronze bildet die Grundlage für weitere Veränderungen, die im Endergebnis zu einer fertigen Bronze-Statue werden. Der *formal cause* erklärt die äußere Form der Statue, den Archetyp, in dem sich das Designprinzip der Statue widerspiegelt, z. B. ein bestimmtes harmonisches Verhältnis zwischen Körper und Gliedmaßen. Der *efficient cause* erklärt handwerklich-künstlerischen Schritte, die notwendig sind, um die Bronze-Statue herzustellen. Gemeint ist dabei nicht ein bestimmter Handwerker oder eine an einen Akteur gebundene Fertigkeit, sondern allgemein das *Wissen*, das notwendig ist, um die gewünschte Materialumwandlung zu vollziehen: „This knowledge, not the artisan who has mastered it, is the salient explanatory factor that one should pick as the most accurate specification of the efficient cause“ (Abs. 2). Der letzte Kausalitätstyp ist der *final cause*, der das Endergebnis beschreibt, das Ziel, auf das mit den vorangegangenen Gründen hingearbeitet wurde. Das Endergebnis, die fertige Bronze-Statue, ist der Grund für die vorangegangenen Schritte im Produktionsprozess: „[...] that for the sake of which everything in the production process is done“ (ebd.).

Entscheidend für die aristotelische Kausalitätstypologie ist, dass die vier Gründe jeweils aufeinander aufbauen. Die ersten drei Typen (*material cause*, *formal cause* und *efficient cause*) sind stets im Lichte eines bestimmten Ziels bzw. erwünschten Endzustands zu betrachten. Es handelt sich dabei um ein teleologisches Erklärungsmodell. Der Einsatz eines bestimmten Materials, der Rekurs auf ein bestimmtes Konstruktionsprinzip und die konkrete handwerkliche Umsetzung sind nie Selbstzweck, sondern werden immer auf ein Ziel hin, im Beispiel die Herstellung einer Statue, durchgeführt.

### Der Erweiterung der aristotelischen Kausalitätstypen durch Nooteboom

Neben den vier bereits beschriebenen Kausalitätstypen ergänzt Nooteboom (2002: 30) zwei weitere: den *exemplary cause* und den *conditional cause*. Der *exemplary cause* umfasst Prototypen und Rollenmodelle, ist also konkreter als der *formal cause*, den Nooteboom eher mit Wissen, Fertigkeiten und Methoden (ebd.) beschreibt. Der *conditional cause* als weitere Ergänzung ist auf die Tatsache bezogen, dass Handlungen nie *ex nihilo* vollzogen werden, sondern im Rahmen gesellschaftlicher Strukturen bzw. gesellschaftlicher Regelsysteme. Nooteboom führt beispielhaft „law and market structure[s]“ sowie „conditions for ideas and initiatives to arise“ (ebd.) an. Dieser Vertrauentyp entspricht, so lässt sich folgern, der „Structural Assurance“ als einer Ausprägung des institutionellen Vertrauens bei McKnight und Chervany (2001).

Zur Erläuterung der sechs Typen nennt er das Beispiel eines Schreiners: Der arbeitende Schreiner (efficient cause) nutzt Holz (material cause) gemäß professioneller Regeln und gelerntem Wissen (formal cause) oder anhand eines Mustermodells (exemplary cause), um Möbel herzustellen und seinen Lebensunterhalt zu verdienen (final cause) im Kontext einer gewissen Markt- und Handwerkskammerstruktur (conditional cause) (ebd.)

### 2.2.3 Vertrauen: Eine auf Verhalten bezogene Disposition

Folgt man Nooteboom (2002: 37), so ist unter Vertrauen eine auf vertrauendes Verhalten („trusting behaviour“) eines Treuhänders bezogene Disposition des Treugebers zu verstehen. Die jedem Verhalten zugrundeliegende Disposition ergibt sich aus den Gründen für das Verhalten und bezieht sich damit auf die sechs zuvor vorgestellten Kausalitätstypen (ebd.: 49). Die Disposition beinhaltet die Inkaufnahme von Verletzbarkeit mit der Erwartung, dass dieses Risiko beschränkt ist:

„Summing up, trust is a disposition towards trusting behaviour, that is behaviour with limited safeguards, accepting vulnerability, based on the expectation that this risk is limited“ (ebd.:37).

Nooteboom unterscheidet zwischen dem weiten und engen Vertrauensbegriff (ebd.: 45ff.), der sich bei beiden Definitionen sowohl auf Personen als auch auf Dinge („things“) beziehen kann. In beiden Fällen handelt es sich nicht um Vertrauenshandlungen, sondern um Erwartungen bzw. Dispositionen, die sich auf vertrauensbedürftige bzw. riskante Handlungen des Treuhänders beziehen, die spätere Vertrauenshandlungen ermöglichen können. Der *weite* Vertrauensbegriff umfasst die *Bereitschaft*, das Risiko eines möglichen Vertrauensmissbrauchs in Kauf zu nehmen, mit der Erwartung, dass dieses Risiko nicht eintreten wird:

„Trust in things or people entails the *willingness* to submit to the risk that they *may* fail us, with the expectation that they will not, or the neglect or lack of awareness of the possibility that they might.“ (45; Hervorhebung R. F.)

Einige Seiten später schreibt Nooteboom (2002: 49), dass als weiter Vertrauensbegriff eigentlich „reliance“, also Verlässlichkeit gemeint sei: „I use the term 'reliance' for the wide notion of trust [...] as an expectation that 'things will not go wrong', regardless of the basis for that expectation [...]“.

Im Gegensatz zum weiten Vertrauensbegriff umfasst der *enge* Vertrauensbegriff zusätzlich den Aspekt, dass sich der Treugeber über den Anreiz eines Vertrauensmissbrauchs des Treuhänders im Klaren ist, dieses Risiko jedoch billigend in Kauf nimmt:

„'Real' trust, or trust in the strong sense, is an expectation that things or people will not fail us, or the neglect or lack of awareness of the possibility

of failure, even if there are perceived opportunities and incentives for it.“  
(48)

In weiteren Differenzierungen bezieht sich Nootboom nur noch auf „real trust“, das sich (obwohl dieser Aspekt in der Definition nicht recht herauskommt) auf Motive von Vertrauenswürdigkeit bezieht, die auch über bloßes Eigeninteresse hinausgehen können (49).

### Elemente des Vertrauens

Der Zweck des Exkurses zur multiplen Kausalität (vgl. Abschnitt 2.2.2) wird deutlich, wenn man sich das Zitat „There are as many forms of trust as there are (failures of) causes of action“ (Nootboom 2002: 49) noch einmal verdeutlicht. Nootboom argumentiert hier, dass es so viele Formen des Vertrauens gibt wie Handlungsgründe. Mit dieser Formulierung geht er über seine vorherige Diktion des Vertrauens als „disposition towards trusting behaviour“ (37) deutlich hinaus. Elemente des Vertrauens können, so Nootbooms Argument, anhand der multiplen Kausalität von Handlungen unterschieden werden (s. Tabelle 2.1).

Die unterschiedlichen Elemente des Vertrauens sollen im Folgenden erläutert werden und dabei auf ein Beispiel für autonome Technik angewendet werden. Da sich Bart Nootboom als Organisationsforscher vor allem für Vertrauen innerhalb und zwischen

Vertrauensform	Objekt des Vertrauens	Kausalitätstyp
Verhaltensvertrauen	Akteure bzw. ihre Handlungen	Effektiver Grund
Materialvertrauen	Beiträge, Mittel	Materieller Grund
Kompetenzvertrauen	Fähigkeit, Fertigkeit, Wissen über den Gebrauch einer Technologie, Methoden, Sprache etc.	Formaler Grund
Intentionales Vertrauen	Ziele, Intentionen	Finaler Grund
Zuneigungsvertrauen	Zuneigung, Fürsorge	
Gutmütigkeitsvertrauen	Gutmütigkeit, Wohlwollen, kein Opportunismus	
Konditionalvertrauen	äußere Rahmenbedingungen, Grenzen	Konditionaler Grund
Vorbildvertrauen	Rollenmodelle	Exemplarischer Grund
Informationsvertrauen	Informationen über Gründe	Alle Gründe
Ehrlichkeitsvertrauen	Aufrichtigkeit, Wahrhaftigkeit	

**Tabelle 2.1:** Elemente des Vertrauens, übersetzt und angelehnt an Nootboom (2002: 50)

Organisationen interessiert, ist eine Anwendung auf Technik nicht selbstverständlich. In seiner Vertrauensmonografie wendet er seine theoretischen Überlegungen auf Organisationen der Fotokopierindustrie, Elektrokostenindustrie und Automobilindustrie an (ebd.: 157ff.). Für die vorliegende Arbeit stellt sich die Frage, ob das Vertrauensmodell auch auf den Untersuchungsgegenstand autonome Technik anwendbar ist.

Als Beispiel dient Adaptive Cruise Control (ACC), ein Fahrerassistenzsystem, das je nach Voreinstellung des Autofahrers eine bestimmte Geschwindigkeit einhält, im Falle von vorausfahrenden Fahrzeugen jedoch abbremst, um einen bestimmten Mindestabstand einzuhalten<sup>6</sup>. Im Ergebnis wird sich zeigen, dass sich die Nooteboom'schen Vertrauentypen auch auf nicht-menschliche Vertrauensobjekte anwenden lassen. Die Wahl des Beispiels ist kein Zufall, da im empirischen Teil der Arbeit ein ACC ähnliches System untersucht wird (vgl. Abschnitt 11). Zur Anwendung:

*Verhaltensvertrauen* bedeutet, dass ein Treugeber Vertrauen in das Verhalten eines Treuhänders hat. Objekt des Vertrauens ist dabei das konkrete Verhalten, z. B. dass ACC bei Einstellung auf 100 km/h auch genau 100 km/h fährt und nicht etwa 105 km/h. *Materialvertrauen* bezieht sich auf das bei der Handlung eingesetzte Material. Im Falle von ACC könnte ein Treugeber Vertrauen in die physikalisch reibungslose Übertragung der Sensordaten zur Abstandsermittlung an die ACC-Steuerung haben.

*Kompetenzvertrauen* bezieht sich auf die Fähigkeit eines Treuhänders bzw. auf sein Wissen. Kompetenzvertrauen adressiert keine konkrete Handlung, sondern bewegt sich auf formaler Ebene hinsichtlich der potenziell anwendbaren, abstrakten Methoden. Nooteboom (2002: 50) nennt „communicate and collaborate in teams“ und „plan and coordinate activities“ als Beispiele. Im Falle von ACC könnte sich das Kompetenzvertrauen z. B. darauf beziehen, dass aus den verschiedenen Sensorinformationen der korrekte Abstand zum vorausfahrenden Fahrzeug berechnet wird.

*Intentionales Vertrauen* bezieht sich auf Dispositionen und Intentionen, die dem Verhalten zugrunde liegen. Als Unterkategorien des intentionalen Vertrauens nennt Nooteboom Zuneigungsvertrauen und Gutmütigkeitsvertrauen (s. Tabelle 2.1), das sich jeweils auf bestimmte dispositionale Aspekte bezieht. Da Nooteboom die verschiedenen Verhaltenstypen ohnehin als Zuschreibungen konzeptualisiert, läßt sich die Zuschreibung von Intentionen bzw. Zielen prinzipiell auch auf autonome Technik übertragen (vgl. Fink und Weyer (2011); Abschnitt 8.1 und 8.2). Im Falle von ACC hat der Fahrer (Treugeber) z. B. Vertrauen darin, dass das Assistenzsystem eine eingestellte Geschwindigkeit einhält und Auffahrunfälle vermeidet.

*Konditionalvertrauen* bezieht sich darauf, dass sich Verhalten innerhalb bestimmter Grenzen abspielt, „[...] that no freak accident will occur that affects outcomes [...]“ (51). Bezogen auf ACC würde dies z. B. bedeuten, dass das System nicht plötzlich auf die Höchstgeschwindigkeit beschleunigt oder auf einmal ausfällt.

<sup>6</sup>Für weitere Informationen zu ACC vgl. Winner u. a. (2002).

*Vorbildvertrauen* bezieht sich auf die Imitation von Rollenmodellen. Die Orientierung an Vorbildern, so bemerkt Nooteboom, spielt insbesondere in der Unternehmenskultur eine große Rolle (51). ACC, das Beispiel ist zugegebenermaßen sehr konstruiert, könnte z. B. in einer speziellen Hochgeschwindigkeits-Einstellung die Rolle von Formel 1-Fahrern imitieren und besonders schnell beschleunigen.

*Informationsvertrauen* als letztgenannte Vertrauensform bezieht sich auf Informationen über Gründe, auf denen das Vertrauen basiert. Im Falle von ACC würde z. B. eine Kontrollanzeige über den aktuellen Zustand des Assistenzsystems Auskunft geben und anzeigen, dass gerade abgebremst wurde, weil sich ein vorausfahrendes Fahrzeug gefährlich genähert hat. Adressat des Vertrauens ist in diesem Falle nicht das beobachtbare Verhalten bzw. die zugeschriebenen Gründe für das beobachtbare Verhalten, sondern Informationen über mögliche Gründe des beobachtbaren Verhaltens. Kritisch anzumerken ist, dass Informationsvertrauen in der Typologie von Nooteboom gegenüber den anderen Vertrauensformen eine Sonderstellung einnimmt, da sich damit das Vertrauensproblem auf eine weitere Ebene verlagert: Kann den Informationen *über* mögliche Eigenschaften des Treuhänders vertraut werden? Weyer (2007: 39) beschreibt dieses Problem für hochautomatisierte Systeme als „Automatisierungs-Paradox“, bei dem ein menschlicher Entscheider „[...] Entscheidungen zweiter Ordnung treffen muss, nämlich unter hohem Zeitdruck feststellen, ob die Anzeigen der automatischen Systeme korrekt sind und ob er sich auf die von diesem Systemen generierten Anweisungen verlassen kann“. Gewissermaßen wird, sofern eine korrekte Wiedergabe der Informationen nicht gewährleistet ist, ein zusätzliches Vertrauensproblem verursacht.

Resümierend lässt sich für die Überprüfung der Anwendbarkeit des Nooteboom'schen Vertrauensschemas auf autonome Technik feststellen, dass diese grundsätzlich möglich erscheint und nicht zu Widersprüchen führt.

## 2.3 Vertrauensbegriff und Implikationen

Die vorangegangenen Abschnitte hatten die Funktion, die im folgenden vorgenommene Ableitung forschungsleitender Annahmen vorzubereiten.

Die Annahmen werden die eigene theoretische Position (Kapitel 8) anleiten und als Grundlage für die empirischen Hypothesen (Kapitel 9) dienen.

Neben der Ableitung forschungsleitender Annahmen wird sich der folgende Abschnitt auch mit grundsätzlicheren Fragen beschäftigen, etwa damit, welche Objekte und Subjekte des Vertrauens in Frage kommen (Abschnitt 2.3.1) und inwiefern der Risiko- und Kontrollbegriff, der bei den bisher vorgestellten Definitionen einen wesentlichen Teilaspekt darstellte, auch für den Gegenstandsbereich dieser Arbeit – Vertrauen in autonome Technik – anwendbar ist. Die empirische Nutzbarkeit soll dabei ebenso Be-

rücksichtigung finden wie Überlegungen, ob die vorgestellten Konzepte auf autonome Technik übertragbar sind.

Insbesondere bei McKnight und Chervany (2001: 330) findet sich der Vorschlag, unterschiedliche Aspekte des Vertrauens den jeweiligen Forschungsdisziplinen zuzuordnen. Diesem Vorschlag wird, wie schon zuvor angedeutet, *nicht* gefolgt, da die strikte Zuordnung zu starr erscheint und dies die eigene Forschungsarbeit zu stark einschränken würde. Von den Autoren wurde z. B. vorgeschlagen, den Zuständigkeitsbereich der Soziologie ausschließlich bei den Institutionen zu verorten – dies erscheint als zu starke Engführung.

Gefolgt wird allerdings der vorgeschlagenen Multidimensionalität von Vertrauen, die beide Metaanalysen gemeinsam haben. Beim Klassifizierungsschema von McKnight und Chervany spiegelt sich die Multidimensionalität von Vertrauen in den miteinander in Beziehung stehenden Konstrukten wider. Bei Nooteboom ergibt sich die Multidimensionalität aus der Ableitung der Vertrauentypen aus dem aristotelischen Konzept der multiplen Kausalität. Für die vorliegende Arbeit wird den Gemeinsamkeiten Rechnung getragen und als forschungsleitende Annahme (FLA 1) festgehalten:

#### **FLA 1 (Multidimensionalität)**

Vertrauen ist als multidimensionales Konzept aufzufassen.

Weiterhin ist beiden Konzepten gemeinsam, dass Verhalten und Vertrauen in einem interdependenten Zusammenhang stehen. Bei McKnight u. a. (1998: 34) ist von „trust-related behaviour“, auf Vertrauen bezogenem Verhalten, die Rede, während Nooteboom (2002: 49) von „behavioural trust“ spricht, was mit Verhaltensvertrauen übersetzt werden kann. Zwischen den beiden Verhaltens-Vertrauens-Konzepten gibt es einen wichtigen Unterschied: Auf Vertrauen bezogenes Verhalten ist eine Aktivität des Treugebers, die *nach* Vergabe von Vertrauen eine bestimmte risikobehaftete Aktion beinhaltet. Verhaltensvertrauen ist hingegen eine Zuschreibung, die der Treugeber gegenüber dem Treuhänder vornimmt – er schreibt dem Treuhänder Vertrauen hinsichtlich seines Verhaltens zu. Welche Schlussfolgerungen er aus der Zuschreibung zieht, ist keine Frage des Verhaltensvertrauens. An einem Beispiel erläutert, umfasst die Entscheidung, ein Fahrerassistenzsystem einzuschalten, auf Vertrauen bezogenes Verhalten. Verhaltensvertrauen wird demgegenüber wirksam, wenn der Fahrer davon ausgeht, dass sich das Assistenzsystem korrekt verhält – dass es z. B. die eingestellte Wunschgeschwindigkeit korrekt einhält.

Bezogen auf autonome Technik sind beide Konzepte von Bedeutung: Auf Vertrauen bezogenes Verhalten kann die Frage der Nutzung autonomer Technik in den Blick rücken, Verhaltensvertrauen die Erwartungen an das Verhalten von autonomer Technik.

**FLA 2 (Verhalten und Vertrauen)**

Verhalten und Vertrauen stehen in einem interdependenten Zusammenhang. Beim Zusammenspiel von Mensch und autonomer Technik sind sowohl auf Vertrauen bezogenes Verhalten wie auch Verhaltensvertrauen von Bedeutung.

**2.3.1 Vertrauen als eine Beziehung zwischen Subjekt und Objekt**

Vertrauen als eine *Beziehung* zwischen einem *Subjekt* und einem *Objekt* zu beschreiben, erscheint zunächst trivial und völlig unstrittig. Vertrauenssubjekt wird weitgehend synonym mit Treugeber, Vertrauensobjekt mit Treuhänder verwendet. Entscheidend ist jedoch die Frage, wer oder was als Objekt bzw. Subjekt des Vertrauens überhaupt in Frage kommt. Wie oben beschrieben, beschränkt Bart Nooteboom (2002) Vertrauenssubjekte auf menschliche Akteure, sieht als Vertrauensobjekte hingegen allgemein Dinge vor. Bei McKnight und Chervany (2001) sind als Treugeber und Treuhänder auch nicht-menschliche Akteure möglich – Vertrauen zwischen Computer-Agenten ist explizit erwähnt, obwohl sie Vertrauen im Sinne von menschlichen Akteuren definieren: „[...] [a]lthough we define trust in terms of people [...]“ (ebd.: 28).

Unabhängig von der Frage, wer Vertrauen in wen haben kann, findet sich in beiden Analysen die Gemeinsamkeit einer Subjekt-Objekt-Unterscheidung.

Für die vorliegende Arbeit wird diese Unterscheidung aus analytischen Gesichtspunkten beibehalten, allerdings soll nicht von vornherein eine Zuordnung erfolgen, die menschliche Akteure als Vertrauens*subjekte* und autonome Technik als Vertrauens*objekte* definiert. Ein besonderes Charakteristikum autonomer Technik ist, dass sie die Grenzen der Handlungsträgerschaft verschwimmen lässt (vgl. dazu Abschnitte 5.1 und 5.3). Übertragen auf eine Interaktions„beziehung“ zwischen Mensch und autonomer Technik folgt daraus, dass sich die Handlungslogik autonomer Technik (in Form komplexer Algorithmen) ebenso auf Erwartungen gegenüber einem menschlichen Akteur bzw. Bediener stützen kann, um eine gemeinsam intendierte Gesamthandlung zu bewirken. Die Zuordnung von Mensch und Technik zu Vertrauensobjekt bzw. Vertrauenssubjekt soll daher als variabel aufgefasst werden und kann je nach Untersuchungsfrage und betrachtetem Interaktionszeitpunkt unterschiedlich sein.

**2.3.2 Vertrauen als Zuschreibung von Gründen und Verhalten**

Der Idee von Nooteboom (2002), Vertrauen als eine auf Verhalten bezogene Disposition zu konzeptualisieren, folgt auch die vorliegende Arbeit. Entscheidend ist die Klärung, welche Aspekte der Disposition mit einbezogen werden. Bei McKnight und Chervany (2001) umfassen Dispositionen vor allem psychologische Zustände und sind

eines von vielen Vertrauenskonstrukten. Bei Nootboom hingegen ist der Dispositionsbegriff deutlich breiter gefasst und enthält verschiedene auf vertrauendes Verhalten bezogene Gründe. Vertrauen und Disposition werden dort weitgehend synonym verwendet. Erkenntnistheoretisch, im Sinne einer deduktiv-nomologischen Erklärung (vgl. Hempel und Oppenheim 1948), ist vertrauendes Verhalten das zu erklärende Phänomen, das Explanandum, das durch die zuvor beschriebenen Explananda, durch die unterschiedlichen Elemente und Facetten des Vertrauens erklärt wird.

Mit der zuvor vorgenommenen Unterscheidung zwischen Subjekt und Objekt des Vertrauens lässt sich Vertrauen als eine vom Treugeber ausgehende Zuschreibung auffassen, die sich auf Verhalten oder Gründe des Verhalten des Treuhänders bezieht:

### FLA 3 (Zuschreibung)

Vertrauen wird hier aufgefasst als eine gerichtete Zuschreibung, ausgehend vom Treugeber, die sich auf (zukünftiges) Verhalten bzw. dem Verhalten zugrundeliegender Motive des Treuhänders bezieht.

Auf welche Aspekte des Verhaltens oder auf welche spezifischen zugeschriebenen Gründe des Verhaltens sich die Zuschreibung bezieht, hängt stark vom empirischen Fall ab.

In Bezug auf die unterschiedlichen Elemente des Vertrauens bei Nootboom lassen sich einige Typen des Vertrauens benennen, die für den Untersuchungsgegenstand der autonomen Technik von Interesse sind: Der Fokus wird gelegt auf *Verhaltensvertrauen*, *Kompetenzvertrauen* und *intentionales Vertrauen* (vgl. Tabelle 2.1 auf Seite 33). Die drei genannten Vertrauensformen sollen im Folgenden anhand von Beispielen erläutert werden. Dabei kommt es vor allem auf die Richtung der Vertrauenszuschreibung an. Der Pfeil “ $\rightarrow$ ” ist dabei als Zuschreibungspfeil zu deuten, wobei auf der linken Seite des Pfeils der Treugeber steht, der dem auf der rechten Seite befindlichen Treuhänder Vertrauen zuschreibt. Als Vertrauenssubjekte und -objekte kommen dabei Mensch (H) und Nicht-Mensch (NH, engl. Non-human) in Frage. Für die drei als besonders relevant identifizierten Vertrauentypen lassen sich die Vertrauenszuschreibungen näher beschreiben, zunächst abstrakt, dann am Beispiel ACC, das schon mehrfach zur Erklärung verwendet wurde:

- **Verhaltensvertrauen**

- H  $\rightarrow$  NH: Ein menschlicher Akteur vertraut darauf, dass ein nicht-menschlicher Akteur ein erwartetes Verhalten zeigt. Bsp.: Der Fahrer eines Autos vertraut darauf, dass sein Assistenzsystem die Geschwindigkeit verringert, da sich in unmittelbarer Nähe ein weiteres Fahrzeug befindet.
- NH  $\rightarrow$  H: Ein nicht-menschlicher Akteur erwartet von einem menschlichen Akteur ein bestimmtes Verhalten. Eine solche Erwartung ist insbesondere

dann von Bedeutung, wenn H und NH in einem wechselseitigen Interaktionszusammenhang stehen, also die Aktionen von H und NH aufeinander aufbauen bzw. voneinander abhängen. Denkbar wäre, dass NH mit eigenen Aktionen erst fortfahren kann, wenn H eine erwartete Aktion durchführt. Bsp.: Die Benutzerschnittstelle des ACC erwartet die Eingabe einer Wunschgeschwindigkeit bevor das ACC aktiviert wird. Unterbleibt die Eingabe, unterbleibt auch die Aktivierung. In dem Fall hat der menschliche Akteur die Erwartung, die vom nicht-menschlichen Akteur an sein Verhalten gestellt wurde, nicht erfüllt.

- **Kompetenzvertrauen**

- H  $\rightarrow$  NH: Der menschliche Akteur erwartet, dass der nicht-menschliche Akteur hinsichtlich eines erwarteten End- oder Zwischenzustandes den Weg dorthin gemäß üblicher Verfahren und Methoden fehlerfrei beschreitet. Bsp.: Der Fahrer erwartet, dass ACC durch „Gasgeben“ korrekt auf die eingestellte Wunschgeschwindigkeit beschleunigt.
- NH  $\rightarrow$  H: Im Kontext eines gemeinsamen Interaktionszusammenhangs erwartet NH, dass H über bestimmte Fertigkeiten und über Wissen verfügt, die für ein gemeinsam zu lösendes Problem von Bedeutung sind. Bsp.: ACC erwartet von dem Fahrer ein gewisses Verständnis physikalischer Bewegungsgesetze. Befindet sich auf der Straße ein unmittelbar vorausfahrendes Fahrzeug, so sollte dem Fahrer bekannt sein, dass eine zusätzliche Beschleunigung unvermeidlich zu einem Zusammenstoß führt.

- **Intentionales Vertrauen**

- H  $\rightarrow$  NH: Der menschliche Akteur schreibt dem nicht-menschlichen Akteur bestimmte Intentionen zu und vertraut darauf, dass diese Intentionen auch tatsächlich verfolgt werden. Intentionen sind dabei abstrakter als konkretes Verhalten. Bsp.: Der Fahrer erwartet, dass ACC die Geschwindigkeit einhält und im Falle eines vorausfahrenden Fahrzeugs abbremst.
- NH  $\rightarrow$  H: Der nicht-menschliche Akteur unterstellt dem menschlichen Akteur bestimmte Intentionen. Ohne die Unterstellung bestimmter Intentionen, wäre jeder NH, der mit einem H interagiert, bloß Selbstzweck. Fasst man autonome Technik vor allem als assistierende Technik auf, so „enthält“ jede Technik eine grobe „Vorstellung“ darüber, welche Intentionen ihr Interaktionspartner verfolgt. Bsp.: ACC erwartet bzw. vertraut darauf, dass der Fahrer die eingestellte Wunschgeschwindigkeit auch tatsächlich fahren möchte.

Die wechselseitige Zuschreibung im Kontext bestimmter Arten des Vertrauens kann nicht nur für *eine* statische Situation betrachtet werden, sondern erfordert eine in-

teraktive (m. a. W. prozessuale) Betrachtung. Weder bei Nooteboom noch bei McKnight und Charvany finden sich Ansätze zu einer solchen, interaktiven Sichtweise auf Vertrauen, die für den Gegenstandsbereich autonomer Technik allerdings von großer Bedeutung wäre.

Allenfalls eine kurze Referenz in dem Aufsatz von McKnight und Charvany auf die Interaktionsexperimente von Garfinkel (s. meine Ausführungen dazu auf S. 20) im Zuge der Infragestellung der situativen Normalität deutet die Relevanz einer interaktiven Sichtweise auf Vertrauen an. Übertragen auf die Interaktion mit autonomer Technik bedeuten die Erkenntnisse von Garfinkel, dass bestimmte Erwartungen an Folgeinteraktionen evoziert werden, die bei Nichterfüllung die situative Normalität in Frage stellen. Die Folge könnte ein Zusammenbruch des Vertrauens in die nicht *interaktionserwartungskonforme* Technik sein – so ließe sich diese bezeichnen. Solche Vertrauenszusammenbrüche zu identifizieren und auf die genannten Vertrauentypen zu spiegeln, stellt aus Sichtweise des Autors eine für die vorliegende Arbeit fruchtbare Forschungsheuristik dar.

### 2.3.3 Abgabe von Kontrolle und Inkaufnahme Verletzlichkeit

Es findet sich in fast allen hier behandelten Ansätzen zu Vertrauen ein enger Zusammenhang zwischen Vertrauen und Risiko. So spricht etwa Nooteboom (2002: 37) von beschränkten Garantien und Inkaufnahme von Verletzlichkeit, McKnight und Charvany (2001: 34ff.) davon, dass auch „negative consequences are possible[...]“ sind. Und auch Luhmann (1988: 97f.), der hier nicht näher behandelt wird, spricht darüber, dass Vertrauen durch Risikosituationen gekennzeichnet sei.

Risiko und Verletzlichkeit sind keine zentralen Untersuchungsgegenstände dieser Arbeit, dennoch sollen einige kursorischen Überlegungen zum Verhältnis von Vertrauen und Verletzlichkeit in Bezug auf autonome Technik angestellt werden.

Entscheidend für einen adäquaten Risikobegriff<sup>7</sup> ist die Vorbedingung, dass die vorliegende Arbeit sogenannte „assistierende Systeme“ in den Blick nimmt. Die primäre Funktion dieser Systeme ist die Unterstützung eines menschlichen Akteurs (oder derer mehrere) bei der Erfüllung einer spezifischen Aufgabe.

Das Risiko oder die Verwundbarkeit kommt bei Interaktion mit dieser Art Technik nicht daher, dass sich ebendiese strategisch<sup>8</sup> gegen den Menschen wendet und Vertrau-

<sup>7</sup>Risiko ist hier nicht im Sinne von Systemrisiko mit Katastrophenpotential (Perrow 1987) oder als Gesellschaftsdiagnose (Beck 2012) zu verstehen, sondern bezieht sich auf riskante Entscheidungen und mögliche Verluste, die dadurch in Kauf genommen werden.

<sup>8</sup>Bei probabilistischen Vertrauensmodelle (vgl. z. B. Coleman 1991: 126) wird Vertrauenswürdigkeit und Vertrauensunwürdigkeit in Abwägung zu Gewinn und Verlust explizit modelliert (vgl. dazu ausführlich Abschnitt 3.1). Auch spieltheoretische Modelle modellieren Vertrauen als Prozess, bei dem die Möglichkeit besteht, vom Treuhänder betrogen zu werden (vgl. dazu Abschnitt 3.2). Dieser strategische Aspekt des Vertrauensmissbrauchs spielt für die vorliegende Arbeit allenfalls eine untergeordnete Rolle.

en bewusst missbraucht, sondern vielmehr daher, dass eine Vertrauenswahl automatisch die Abgabe von Kontrolle beinhaltet. Weyer (1997: 244; 2008b: 244; 2008a: 192) diagnostiziert als zentrales Charakteristikum autonomer Technik ihre partielle Undurchschaubarkeit und folgert, dass die Funktionsweise häufig nur noch als Blackbox aufgefasst wird und häufig keine direkten Eingriffe mehr ermöglicht. Daher, so eine naheliegende Folgerung, bedeutet Vertrauen in autonome Technik immer auch eine Abgabe von Kontrolle mit dem Wissen, dass die Eingriffsmöglichkeiten dadurch mitunter eingeschränkt werden. Zudem kann der Treugeber (meist der menschliche Akteur) nicht mehr bis auf die operative Ebene eingreifen – entweder weil ein solcher Eingriff nicht vorgesehen ist oder weil er das System nur bedingt versteht. Das Risiko besteht darin, dass Kontrolle abgegeben und Aufgaben delegiert werden, obwohl nicht klar ist, ob die Erwartungen an die autonome Technik tatsächlich in der gewünschten Weise erfüllt werden.

Insgesamt lässt sich zu den vorgestellten Vertrauensmetaanalysen und den bisher daraus abgeleiteten forschungsleitenden Annahmen einschränkend anmerken, dass interaktive bzw. prozessualen Zusammenhänge in den bisher vorgestellten Arbeiten keine wesentliche Rolle spielen. Für den Gegenstand autonomer Technik erscheint dieser Zusammenhang jedoch essenziell. Aus dem Grund wird im folgenden Kapitel 3 ein spezieller Zweig der Vertrauensforschung näher eruiert: die probabilistische Vertrauensforschung, die mit einem Vertrauensbegriff operiert, der die Grundlage für eine spieltheoretische, und damit interaktive Sichtweise auf Vertrauen bietet.



### 3 Vertrauen in der Sozialtheorie

Nachdem im vorherigen Kapitel zwei elaborierte Meta-Analysen zu Vertrauen vorgestellt wurden, verfolgt dieses Kapitel das Ziel, einige konkrete Ansätze aus der Sozialtheorie vorzustellen. Zweifelsohne erfassen die Vertrauenskonzepte von Nooteboom (2002) und McKnight und Chervany (2001) eine enorme Bandbreite von Facetten des Vertrauens, von denen auch einige für die vorliegende Arbeit nutzbar sind. In Hinblick auf eine laborexperimentelle empirische Untersuchung von Vertrauen in autonome Technik (ab Kapitel 11) ist allerdings zu prüfen, ob sich ggf. auch Ansätze eignen, die mit deutlich weniger Variablen und damit insgesamt mit einer sparsameren Modellierung auskommen. Den behandelten Meta-Analysen fehlen zudem Überlegungen zu einem *interaktiven* Vertrauensbegriff: Zwar wird Vertrauen als facettenreiches und multidimensionales Konzept vorgestellt, letztlich werden die dortigen Konstrukte allerdings zu einer eher *statischen* Zustandsbeschreibung des Vertrauens als Disposition zugespitzt. Der interaktive Charakter von Vertrauen, der für die Entwicklung eines Vertrauenskonzepts für autonome Technik von Bedeutung ist, findet in spieltheoretischen Überlegungen seinen Niederschlag, die Gegenstand des vorliegenden Kapitels sind.

Kennzeichnend für die zuvor vorgestellten Vertrauenskonzepte ist zudem, dass Vertrauen durch seine Modellierung als Disposition kein dyadisches Konzept ist und keine *Beziehung* zwischen zwei Akteuren abbildet. Die in dem folgenden Kapitel vorgestellten Ansätze fußen explizit auf Vertrauens*beziehungen*, so dass für die Frage nach Vertrauen in autonome Technik zu prüfen ist, inwieweit eine dyadische Konzeption von Vertrauen dem Untersuchungsgegenstand gerecht wird. Um mit Preisendörfer (1995: 264) zu sprechen, ist Vertrauen als soziologische Kategorie ein wichtiges Element einer Sozialtheorie, da es „sich nicht um ein Merkmal handelt, das an irgendeine Einzelakteure gebunden ist, sondern um ein Merkmal von sozialen Beziehungen, so daß also stets mindestens zwei Akteure [individuelle oder korporative] involviert sind.“ Folgt man Preisendörfer weiter, entsteht Vertrauen aus zwei grundlegenden Problemen, die charakteristisch für soziale Beziehungen sind: dem Zeitproblem und dem Informationsproblem. Das Zeitproblem ist eine Konsequenz aus dem meist sequentiellen Charakter sozialer Interaktionen, da „Leistungen in sozialen Beziehungen zumeist nur mit gewissen zeitlichen Verzögerungen und sequentiell ausgetauscht werden können“ (264). Aus dem Zeitproblem ergibt sich ein Informationsproblem, nämlich die Frage, ob sich „die Interaktionspartner tatsächlich an die impliziten oder auch expliziten Vereinbarungen halten werden“ (ebd.).

In Kapitel 8 wird ein eigener handlungstheoretischer Ansatz vorgestellt, der die Interaktion zwischen Mensch und autonomer Technik modellieren kann. Im Kern basiert dieser Ansatz auf subjektiven Wahlhandlungen, die in einen interaktiven Prozess eingebettet sind, der sich zwischen Mensch und autonomer Technik abspielt. Vor dem Hintergrund ist zu prüfen, ob es bereits existierende Vertrauensansätze gibt, die in besonderer Weise anschlussfähig sind an subjektive Wahlhandlungen und geschätzte Eintrittswahrscheinlichkeiten. Dabei werden zwei Stoßrichtungen der sozialtheoretischen Vertrauensansätze detailliert vorgestellt: der probabilistische Vertrauensansatz (Abschnitt 3.1) und der spieltheoretische Vertrauensansatz (Abschnitt 3.2).

Probabilistische Ansätze modellieren Vertrauen als Entscheidung unter Risiko. Das Maß für die subjektive Einschätzung der Vertrauenswürdigkeit des Gegenübers bestimmt den Ausgang solcher Entscheidungen (vgl. Preisendörfer 1995: 264). Zum anderen behandelt das Kapitel unterschiedliche spieltheoretische Ansätze zur Modellierung von Vertrauen, bei denen vor allem strategische Interdependenzsituationen im Modellierungsfokus stehen.

Spieltheoretische Modelle eignen sich besonders gut für eine interaktive Sichtweise auf Vertrauen. Derartige Ansätze dominieren beispielsweise den Sammelband von Diego Gambetta (1988b) „Trust: making and breaking cooperative relations“, der den damaligen Stand der Vertrauensforschung zusammenfasst. Ebenso gibt es eine lange Tradition probabilistischer Vertrauensmodelle, etwa bei Deutsch (1958) und Coleman (1991), die als Vorläufer der späteren, auf Rational-Choice-Annahmen basierenden Vertrauensmodelle anzusehen sind.

Der zentrale Unterschied zwischen dem probabilistischen Vertrauensbegriff und dem spieltheoretischen Vertrauensbegriff besteht im zugrundeliegenden Entscheidungskalkül: Im ersten Fall wird ausschließlich die Entscheidung des Treugebers modelliert, der auf Basis der situativen Opportunitäten rational darüber entscheidet, ob er vertraut oder misstraut. Dabei handelt es sich um eine zustandsorientierte bzw. situative Entscheidung. Im zweiten Fall, dem spieltheoretischen Vertrauensbegriff, ist die Entscheidungssituation zwischen Treugeber und Treuhänder sequentiell-prozessförmig angelegt.

Es wird sich allerdings zeigen, dass die probabilistischen und spieltheoretischen Ansätze Annahmen treffen, die eine direkte Anwendung auf autonome Technik weitgehend verhindern.

### 3.1 Probabilistischer Vertrauensbegriff

Der Abschnitt beginnt mit einer ausführlichen Würdigung des von Morton Deutsch (1958; 1976) entwickelten Vertrauensbegriffs (Abschnitt 3.1.1), der Vertrauen als Entscheidungsproblem ansieht, das eine Abwägung zwischen positiven und negativen Konsequenzen umfasst.

James Coleman (1991: 123ff.) hat die Ideen von Deutsch später aufgegriffen und dahingehend weiterentwickelt (s. Abschnitt 3.1.2), dass er Vertrauen als rationale Wahlhandlung rekonstruiert und unterschiedliche Typen von Vertrauensbeziehungen in den Blick nimmt. Der Modellierung von Coleman hat sich Hartmut Esser (2000b: 87f.), wie gezeigt wird, zwar auf formaler Ebene weitgehend angeschlossen. Dennoch unterscheidet sich diese inhaltlich von den Überlegungen bei Coleman.

Der abschließende Abschnitt 3.1.3 beschäftigt sich mit der Frage, ob der probabilistische Vertrauensbegriff auch für autonome Technik anwendbar ist. Er wird zu dem Ergebnis kommen, dass die Modellierung von Vertrauen als bivalentes Entscheidungsproblem mit nur zwei Handlungsalternativen (vertrauen vs. nicht vertrauen) am eigentlichen Kern des Vertrauensproblems vorbeigeht.

Der probabilistische Vertrauensbegriff verschiebt etwa im Vergleich zu Nooteboom's Ansatz (Abschnitt 2.2.3) den Fokus auf die Konsequenzen des Vertrauens hinsichtlich eventuell eintretender Gewinne oder Verluste und nicht auf die Zusammenhänge, die vor Eintreten einer Vertrauenshandlung wirksam werden.

### 3.1.1 Deutsch: Vertrauen als Abwägung negativer und positiver Konsequenzen

Morton Deutsch hat sich nicht nur als einer der ersten Sozialwissenschaftler mit einer experimentell-spieltheoretischen Modellierung von Vertrauen (vgl. Deutsch 1958) beschäftigt (vgl. Abschnitt 3.2), sondern legte auch den Grundstein für einen probabilistischen Vertrauensbegriff (Deutsch 1976).

#### Definitiver Aufbau des Vertrauensbegriffs

Deutsch baut seinen probabilistischen Vertrauensbegriff definitiv/systematisch auf. Zentral sind dabei die Definitionen 1 bis 5 (vgl. Deutsch 1976: 139f.):

*Definition 1–2* entsprechend, hat ein Ereignis für einen Akteur eine positiv motivierende Bedeutung  $Va^+$ , wenn „sein Auftreten das Gelingen einer Sache oder das Wohlergehen einer Person (einschließlich der eigenen) verbessert [...]“ (139). Im Gegensatz dazu hat ein Ereignis eine negativ motivierende Bedeutung, wenn „das Individuum erwartet, dass sein Auftreten das Gelingen einer Sache oder das Wohlbefinden einer Person (einschließlich der eigenen) vermindert oder ein Anwachsen verhindert“ (ebd.). Positiv und negativ motivierende Ereignisse sind genau gegensätzlich konstruiert und lassen sich in die Sprache von Rational Choice-Ansätzen übersetzen als positiver bzw. negativer Nutzen.

In *Definition 3* verknüpft Deutsch die Eintrittswahrscheinlichkeit eines Ereignisses mit seiner Handlungswirksamkeit, denn „[je] möglicher ein erhofftes Ereignis [ist], um so wirksamer ist dieses Ereignis, ein Verhalten zu erzeugen, das auf die Förderung oder Verhinderung eines Zustandekommens gerichtet ist“ (140).

Schließlich führt Deutsch in *Definition 4* den zentralen Begriff des ambiguenten Weges ein, dessen Existenz eine Vertrauenswahl erst notwendig macht. Ein ambigüenter Weg ist dabei ein Weg, „der genau so gut zu einem Ereignis mit positiv motivierender Bedeutung ( $Va^+$ ) wie zu einem Ereignis mit negativ motivierender Bedeutung ( $Va^-$ ) führen kann“ (140). Die Formulierung „genau so gut“ bedeutet dabei nicht, dass die beiden Ereignisse mit gleicher Wahrscheinlichkeit eintreten, sondern nur, dass überhaupt die Möglichkeit besteht, dass entweder das eine oder das andere Ereignis eintritt.

Es handelt sich bei der Wahl des ambiguenten Weges gemäß *Definition 5* um eine Vertrauenswahl, wenn „ $Va^+$  geringer ist als  $Va^-$ “, wenn die negativen Konsequenzen also die positiven überwiegen und um eine Misstrauenswahl, wenn ein solcher Weg vermieden wird (ebd.).

Ist umgekehrt  $Va^-$  geringer als  $Va^+$ , handelt es sich nur um eine Wahl unter „Risiko und Wagnis“.

### Vertrauen als rationale Abwägung

Nach einigem definatorischen Vorlauf kommt Deutsch zum Kern seines Vertrauensbegriffs: der Beantwortung der Frage, wann eine Entscheidung auf Vertrauensbasis erfolgt. Entscheidend ist dabei die subjektiv geschätzte Wahrscheinlichkeit für den Eintritt des negativ motivierten Ereignis im Vergleich zum Eintritt des positiv motivierten Ereignis. Unter der Voraussetzung, dass  $Va^- > Va^+$  (Bedingung für Vertrauenswahl) gilt, lässt sich der Entscheidungszusammenhang gemäß Deutsch (1976: 142f.) wie folgt zusammenfassen:

„Angenommen,  $Va^-$  ist stärker als  $Va^+$ , so wird eine Vertrauenswahl getroffen, wenn

$$Va^+ \cdot S.W.^+ > Va^- \cdot S.W.^- + K$$

ist.“

$K$  ist dabei eine Individualkonstante, ein Sicherheitsgrad, den ein Individuum benötigt, um eine Wahl auf Vertrauensbasis zu treffen.  $S.W.^+$  und  $S.W.^-$  bezeichnen jeweils die geschätzte Eintrittswahrscheinlichkeit eines positiven bzw. negativen Ereignis.

Der durch die Ungleichung formalisierte Zusammenhang lässt sich wie folgt illustrieren: Die Frage, auf steigende Kurse zu setzen wird dann zur Vertrauensfrage, wenn der Verlust bei sinkenden Kursen höher ist als der Gewinn bei steigenden Kursen ( $Va^- > Va^+$ ). Diese Vertrauenswahl wird getroffen, wenn die geschätzte Wahrscheinlichkeit eines Kursanstiegs  $S.W.^+$  so hoch ist, dass die Höhe des Erwartungsnutzen bei Kursanstieg ( $Va^+ \cdot S.W.^+$ ) die Höhe des Risikos bei Kursverlust unter Beachtung eines zusätzlichen Sicherheitsgrads übersteigt.

Deutsch (1976: 143) identifiziert folgende Faktoren, die auf die subjektiv geschätzte Eintrittswahrscheinlichkeit von positiv bzw. negativ motivierten Ereignissen Einfluss nehmen:

- eigene vergangene Erlebnisse in ähnlichen Situationen
- frühere Erlebnisse anderer
- die Meinung anderer, die man respektiert
- die eigenen Ansichten über Vor- und Nachteile der Situation, in der man steht
- Möglichkeiten der Einflussnahme auf den Nichteintritt von  $Va^-$

Die Operationalisierung der genannten Einflussgrößen ist nicht trivial und erfordert Modellierungsmöglichkeiten, die deutlich über reine Kosten-Nutzen-Abwägungen hinausgehen. Die zwei Operationalisierungsprobleme sind die Ermittlung von Situationsähnlichkeiten (bzw. die Modellierung der Situation überhaupt) und die Rekalkulation von Eintrittswahrscheinlichkeiten auf Basis vergangener Erfahrungen.

Für das erste Problem existiert mit dem Modell der Frame-Selektion (MFS) inzwischen ein elaborierter handlungstheoretischer Ansatz, der die Geltung von alternativen Situationsdefinitionen u.a. in Abhängigkeit von der Existenz passender Situationsobjekte modellieren kann (Kroneberg 2005).

Für das zweite Problem, die Rekalkulation von Eintrittswahrscheinlichkeiten, bieten sich Lerntheorien an, wie sie Kroneberg (2011: 158) für den Prozess der Enkodierung, also dem Aufbau, Abbau und Verstärkung von Verhaltensprogrammen und Problemlösungen, benennt: etwa dem lerntheoretische Modell von Macy und Flache (2002).

### 3.1.2 Coleman: Vertrauen als Entscheidung unter Risiko

Der probabilistische Vertrauensbegriff ist neben Morton Deutsch auch auf James Coleman zurückzuführen, der sich im Rahmen seiner „Grundlagen der Sozialtheorie“ (Coleman 1991) mit Vertrauen beschäftigt.

Er charakterisiert Vertrauen als „[...] Untergruppe der Situationen, die ein gewisses Risiko beinhalten“ (Coleman 1991: 115). Das Vorliegen von Risiko ist dabei eine notwendige Bedingung dafür, dass überhaupt eine Abwägung zwischen Vertrauen und Misstrauen stattfindet.

Weiter unterscheidet Coleman (1991: 121) zwischen mindestens zwei Parteien, die an einer Vertrauensbeziehung beteiligt sind: Dem *Treugeber* und dem *Treuhänder*. Der Treugeber ist dabei diejenige Person, die dem Treuhänder Treue entgegenbringt, ihm vertraut. Coleman unterstellt beiden Parteien zielgerichtetes Handeln und geht davon aus, dass beide „ihre wie auch immer geschalteten Interessen realisieren wollen“ (ebd.).

Colemans Vertrauenskonzept, übersetzt in die Sprache der rationalen Wahlhandlung, etwa der SEU-Theorie (vgl. Esser 1993), zeigt, dass die Vergabe von Vertrauen (aus Sichtweise des Treugebers) sowie die Rechtfertigung von Vertrauen (aus Sichtweise des Treuhänders) eine mögliche Handlungsalternative von Vielen innerhalb eines nutzenmaximierenden Handlungskalküls ist: „[...] [D]er potentielle Treugeber [wird] mit genau den Überlegungen konfrontiert [...], die ein rationaler Akteur anstellt, wenn er entscheidet, ob er eine Wette abschließen soll“ (ebd.: 125).

Die Konzeptualisierung von Vertrauen als rationale Entscheidung unter Risiko umfasst folgende Aspekte (Coleman 1991: 123f.):

1. Der Treugeber verbessert durch die Vergabe von Vertrauen seine eigene Situation, sofern der Treuhänder vertrauenswürdig ist.
2. Der Treugeber verschlechtert durch die Vergabe von Vertrauen seine eigene Situation, sofern der Treuhänder *nicht* vertrauenswürdig ist.
3. Dem Treugeber wird durch die Vergabe von Vertrauen eine zuvor nicht mögliche Handlung ermöglicht.

Das Risiko des Treugebers besteht also darin, dass er mit einem je nach Entscheidungssituation unterschiedlichen Verlust konfrontiert wird, falls der Treuhänder nicht vertrauenswürdig ist. Im Gegensatz zu Deutsch werden die Konzepte von Vertrauen und Risiko nicht getrennt voneinander behandelt, sondern er definiert Vertrauen über Risiko.

### **Formalisierung der Vertrauensvergabe**

In der folgenden Formulierung von Coleman (1991: 125) sind bereits alle Grundelemente zur Formalisierung des Vertrauensbegriffs enthalten:

„Wenn die Chance zu gewinnen relativ zur Chance zu verlieren größer ist als das Ausmaß des Verlustes (falls er [der Treugeber] verliert) relativ zum Ausmaß des Gewinns (falls er gewinnt), kann er mit dem Eingehen der Wette einen Gewinn erwarten. Und wenn er rational handelt, sollte er sie abschließen.“

Basierend auf diesen Grundüberlegungen leitet Coleman Variablen ab, von denen die Vergabe von Vertrauen abhängt:

- $p$  – Gewinnchance (die Wahrscheinlichkeit der Vertrauenswürdigkeit des Treuhänders)
- $L$  – möglicher Verlust (falls der Treuhänder nicht vertrauenswürdig ist und das Vertrauen missbraucht)
- $G$  – möglicher Gewinn (falls Treuhänder vertrauenswürdig ist)

Formal lässt sich die Vergabe von Vertrauen als quasi-binäre Entscheidung formalisieren:

$$\text{Vertrauensvergabe}(p, L, G) = \begin{cases} ja & \frac{p}{1-p} > \frac{L}{G} \\ unentschieden & \frac{p}{1-p} = \frac{L}{G} \\ nein & \frac{p}{1-p} < \frac{L}{G} \end{cases} \quad (3.1)$$

Coleman (1991: 126) zufolge wird die Einschätzung des Treugebers über die Vertrauenswürdigkeit  $p$  des Treuhänders ins Verhältnis gesetzt zur Gegenwahrscheinlichkeit  $(1 - p)$ , also der geschätzten Vertrauensunwürdigkeit. Ist dieses Verhältnis größer als das Verhältnis von Verlust  $L$  zu Gewinn  $G$ , vertraut der Treugeber folgerichtig. Ist das Verhältnis kleiner, vergibt der Treugeber kein Vertrauen, er misstraut dem Treuhänder.

Der Fall  $\frac{p}{1-p} = \frac{L}{G}$  ist bei Coleman zwar definiert, jedoch bleibt unklar, welche Konsequenzen dieser Fall für die Vertrauensvergabe hat. Was bedeutet es für die Vertrauenswahl, unentschieden zu sein, wenn ein strikt rationales Handlungskalkül immer ein Entweder-Oder, also eine binäre Entscheidung zwischen zwei Handlungsalternativen verlangt (vgl. kritische Würdigung des probabilistischen Vertrauensmodell / Abschnitt 3.1.3)?

Nach der allgemeinen Formulierung des Zusammenhangs stellt sich die Frage, wie die konkreten Variablenbelegungen ermittelt werden. Coleman argumentiert, dass sich Gewinn  $G$  und Verlust  $L$  in den meisten Fällen leicht bestimmen<sup>1</sup> lassen, während man über  $p$ , also die geschätzte Vertrauenswürdigkeit des Treuhänders, am wenigsten weiß (Coleman 1991: 130).

Coleman benennt keinen Mechanismus, etwa eine soziologische Lerntheorie, der angibt, wie sich die Werte von  $p$  sozialisatorisch verändern. Er geht davon aus, dass wir *unbekannten* Personen mit einer Standardeinschätzung  $p^*$  begegnen, welche auf „vergangenen Erfahrungen basiert“ (132). Bei besonders engen Freunden, so Coleman weiter, sei das Verhältnis von Gewinn und Verlust besonders hoch. In dem Fall sind hohe  $p$ -Werte für die geschätzte Vertrauenswürdigkeit des Freundes erforderlich, um Vertrauen zu vergeben. Im Vergleich zur Standardeinschätzung  $p^*$  gilt für die Einschätzung eines Freundes  $p_{\text{Freund}}$ :

$$\frac{p_{\text{Freund}}}{1 - p_{\text{Freund}}} \gg \frac{p^*}{1 - p^*}$$

<sup>1</sup>Allerdings hängt dies auch mit der Wahl der Beispiele zusammen, etwa den Gewinn und Verlust für eine Bank bei Vergabe eines Kredits, oder dem gegenseitigen Aushelfen von Farmern bei der Heuernte. Typisch für die Beispiel ist, dass die Konsequenzen für Vertrauensbestätigung und Vertrauensbruch unmittelbar klar sind (vgl. Coleman 1991: 122). Für komplexe soziotechnische Systeme erscheint dies allerdings zu voraussetzungsreich, da diese gerade dadurch charakterisiert sind, dass sie auch hinsichtlich ihres Gefährdungspotenzials nicht vollständig durchschaubar sind.

### Typen von Vertrauensbeziehungen

Vertrauensbeziehungen sind im Kontext des probabilistischen Vertrauensbegriffs von Bedeutung. Neben der typischen Zweierkonstellationen zwischen Treugeber und Treuhänder, die bereits besprochen wurde, stellt Coleman (1991: 227) insgesamt vier Typen von Vertrauensbeziehungen heraus:

1. Vertrauensbeziehungen zwischen zwei Akteuren: ein (potentieller) Treugeber, ein Treuhänder
2. Gegenseitiges Vertrauen (beide sind wechselseitig Treuhänder und Treugeber)
3. Zusätzliches intermediäres Glied: Treuhänder für einen Akteur, Treugeber für einen anderen
4. Drittparteien-Vertrauen

Der zweite Vertrauensbeziehungstyp, *gegenseitiges Vertrauen*, setzt wiederkehrende Begegnungen und wechselseitige Abhängigkeiten voraus. Gegenseitiges Vertrauen, das Coleman explizit als dynamischen Prozess beschreibt (231), ähnelt strukturell dem spieltheoretischen Vertrauensbegriff (vgl. Abschnitt 3.2), der reziproke Interaktionserwartungen konstitutiv berücksichtigt.

Ein dritter Vertrauensbeziehungstyp ergibt sich durch Berücksichtigung eines weiteren Akteurs, einem *Intermediär*. Coleman unterscheidet drei Typen von intermediären Vertrauensakteuren: Berater, Bürgen und Unternehmer. Zum Zwecke einer passenden Einschätzung der Vertrauenswürdigkeit eines potenziellen Treuhänders vertraut ein Treugeber der Einschätzung eines *Beraters*. Der Treugeber übernimmt lediglich die *Einschätzung* des Beraters, das Risiko für einen Vertrauensbruch des Treuhänders trägt in vollem Umfang der Treugeber (und nicht der Berater). Im Gegensatz zum Berater existiert im Falle des *Bürgen* eine direkte Vertrauensbeziehung zwischen Bürge und Treuhänder. Der Treugeber vertraut dem Bürgen, der Bürge vertraut dem Treuhänder. Das Risiko für einen Vertrauensbruch des Treuhänders trägt der Bürge (und nicht der Treugeber). Als dritten intermediären Vertrauentyp benennt Coleman den *Unternehmer*. Ein Unternehmer vermittelt zwischen mehreren potenziellen Treugebern und Treuhändern. Als persönlich haftender Akteur knüpft er Vertrauensnetzwerke zwischen Treugebern und Treuhändern und kommt im Falle eines Vertrauensbruchs für die anfallenden Verluste auf (vgl. Coleman 1991: 232ff.).

Als vierten Typ von Vertrauensbeziehungen, der hauptsächlich in ökonomischen Zusammenhängen vorkommt, benennt Coleman das *Drittparteien-Vertrauen* (239ff.). Eine dritte Partei ist dabei eine vertrauensvolle Instanz, mit der Treugeber und Treuhänder bereits zahlreiche erfolgreiche Transaktionen abgewickelt haben. Obwohl sich Treugeber und Treuhänder zuvor nicht begegnet sind, kann der Treuhänder mit einer ausreichend hohen Schuldverschreibung der dritten, vertrauensvollen Partei das Vertrauen des Treugebers gewinnen.

### Ähnlichkeiten und Unterschiede zu anderen Konzepten

Wie Coleman (1991: 126) selbst bemerkt, fehlt seiner Vertrauensdefinition eine Einschränkung, die bei Deutsch vorkommt: Deutsch setzt voraus, dass der mögliche Verlust bei Vertrauensbruch größer sein muss als der mögliche Gewinn bei Vertrauensvergabe, um von einer Vertrauenswahl zu sprechen. Bei Coleman genügt das Vorliegen eines Risikos  $L$  und die Wahrscheinlichkeit  $1 - p$ , dass das Risiko tatsächlich realisiert wird, um von einer Vertrauensentscheidung zu sprechen. Zwei weitere Unterschiede zu dem Vertrauenskonzept von Deutsch beziehen sich auf die Individualkonstante (bei Deutsch) und die Modellierung mittels Gegenwahrscheinlichkeiten bei Coleman. Bei Deutsch ist die Vertrauenswahl, im Gegensatz zum Ansatz bei Coleman, keine rein objektive Entscheidung, weil eine Individualkonstante, die einen zusätzlichen Grad an Sicherheit für eine Vertrauenswahl erforderlich macht, modelliert ist. Ein weiterer Unterschied ist die Verwendung von Wahrscheinlichkeit  $p$  und Gegenwahrscheinlichkeit  $1 - p$  zur Einschätzung der Vertrauenswürdigkeit des Treuhänders. Bei Deutsch beziehen sich die modellierten Wahrscheinlichkeiten allgemein auf Ereignisse und nicht wie bei Coleman auf zugeschriebene Eigenschaften des Treuhänders.

Colemans Formalisierung von Vertrauen als Entscheidung unter Risiko findet in ähnlicher Form auch Berücksichtigung bei der von Hartmut Esser (2000b: 87f.) gewählten Formalisierung. Die Überlegungen von Esser finden hier Erwähnung, weil diese zwar strukturell den Überlegungen von Coleman ähneln, sich an einem wichtigen inhaltlichen Punkt jedoch von ebendiesen unterscheiden.

Es stehen zwei Handlungsalternativen „vertrauen“ (V) und „misstrauen“ (M) zur Verfügung. Bei Auswahl der Handlungsalternative V steht ein Gewinn von  $R$  in Aussicht, sofern der Treuhänder das Vertrauen bestätigt. Bei Wahl der Handlungsalternative M fällt eine Bestrafung in Höhe von  $P$  an. Esser definiert für beide Handlungsalternativen den Erwartungsnutzen, indem die jeweiligen Gewinne und Verluste mit der Einschätzung  $p$  der Vertrauenswürdigkeit bzw. der entsprechenden Gegenwahrscheinlichkeit multipliziert werden:

$$EU(V) = p \cdot R \quad (3.2)$$

$$EU(M) = (1 - p) \cdot P \quad (3.3)$$

Aus den Erwartungsnutzenwerten lässt sich leicht herleiten, unter welcher Bedingung Vertrauen vergeben wird. Der Erwartungsnutzen muss für die Handlungsalternative V größer sein als für die Handlungsalternative M:

$$EU(V) > EU(M) \Leftrightarrow \quad (3.4)$$

$$p \cdot R > (1 - p) \cdot P \Leftrightarrow \quad (3.5)$$

$$\frac{p}{1 - p} > \frac{P}{R} \quad (3.6)$$

Die Bedingung (3.6) zur Vertrauensvergabe ist zwar mit der von Coleman formulierten Bedingung 3.1 auf Seite 49 auf den ersten Blick identisch, unterscheidet sich von dieser allerdings an einem wichtigen inhaltlichen Punkt: bei der Interpretation des Verlusts. Bei Coleman gibt der Verlust  $L$  die Kosten für den Treugeber an, die entstehen, wenn der Treuhänder das Vertrauen missbraucht und der Treugeber vertraut, während bei Esser die Bestrafung  $P$  die Kosten angibt, die entstehen, wenn überhaupt keine Vertrauenswahl zu Stande kommt – der Treugeber misstraut, der Treuhänder hat daraufhin weder die Möglichkeit das Vertrauen zu bestätigen, noch es zu missbrauchen.

Die Gefahr des Vertrauensmissbrauchs bzw. der einseitigen Ausnutzung ist bei Essers Modellierungsansatz nicht vorgesehen.

### 3.1.3 Kritische Würdigung der probabilistischen Vertrauensbegriffe

Wie zu Anfang des Kapitels ausgeführt, richtet sich das primäre Interesse des probabilistischen Vertrauensbegriffs auf Vertrauens*handlungen* bzw. auf die Antizipation der Konsequenzen von Vertrauenshandlungen.

Verglichen mit den weit ausgearbeiteten Konstrukten aus Kapitel 2 erscheint der probabilistische Vertrauensbegriff unterkomplex. Die insgesamt zehn Elemente des Vertrauens (vgl. Tabelle 2.1 auf Seite 33) werden hier in eine einzige Variable codiert. In diesem abschließenden Abschnitt zum probabilistischen Vertrauensbegriff wird die Kritik einiger Autoren dargestellt und um eigene Kritik, insbesondere in Hinblick auf autonome Technik, ergänzt.

#### Vertrauenshandlung vs. Handlungsgrund

Eine umfassende Beschäftigung mit unterschiedlichen Vertrauensbegriffen hat Bernd Lahno (2002) vorgelegt. Folgt man seiner Einordnung (ebd.: 105), so nehmen Coleman und Deutsch eine Rational Choice-Perspektive auf Vertrauen ein.

Bei der Analyse des Coleman'schen Vertrauensbegriffs sei es wichtig, „[...] zwischen einer Vertrauenshandlung oder Vertrauensentscheidung auf der einen Seite und ‚Vertrauen‘ als einen Handlungsgrund auf der anderen Seite begrifflich zu unterscheiden“ (Lahno 2002: 116). Coleman wie auch Deutsch fokussieren, folgt man Lahno, ausschließlich auf die Vertrauenshandlung, also auf den konkreten Vollzug einer zuvor nicht möglichen Handlung, die sich auf Vertrauen *gründet*. Dieser Fokus widerspricht

dem Alltagsverständnis von Vertrauen, das insbesondere den Prozess der Zuschreibung von Vertrauen beinhaltet (118). Coleman und Deutsch können mit ihren Modellierungsansätzen, so Lahno weiter (ebd.), nicht überzeugend darlegen, dass die „[...] Durchführung einer Vertrauenshandlung mit der Zuschreibung von Vertrauen im Alltagssinne zusammenfällt“. Lahno kritisiert, dass Coleman und Deutsch Vertrauen nicht als allgemeine Disposition modellieren, sondern als eine „kontextabhängige Variable“, die sich an der Situation orientiert, in der sich der Akteur befindet.

Zumindest die letzte Problematisierung von Lahno, Vertrauen als kontextabhängige Variable zu modellieren, relativiert sich, da die geschätzte Vertrauenswürdigkeit  $p$  eines unbekanntes Gegenübers nicht qua Situation und erst recht nicht ex nihilo entsteht, sondern auf einer Standardeinschätzung  $p^*$ , „welche auf vergangenen Erfahrungen basiert“ (Coleman 1991: 132). Ebenso finden sich bei Deutsch (1976: 143) situationsunabhängige Faktoren, die Vertrauen beeinflussen z. B.: „[...] vergangene Erlebnisse in ähnlichen Situationen; die früheren Erlebnisse anderer; die Meinungen anderer [...]“.

Wie Lahno (2002: 148) später selbst bemerkt, kommt es bei der entscheidungstheoretischen Analyse von Vertrauen auf die Modellierung an. Die von Lahno kritisierten fehlenden Handlungsgründe können mittels empirischer Untersuchungen oder Brückenhypothesen im Rahmen eines entscheidungstheoretischen Modells in die Modellierung der Situation integriert werden. Dabei wäre es entweder möglich, die spezifischen Bedingungen der Situation, z. B. vorangegangene Vertrauenshandlungen, mit zu berücksichtigen, oder sogar die Frage nach der Entwicklung von  $p$  zum Explanandum zu erheben, um dieses in einer weiteren entscheidungstheoretischen Analyse zu erklären.

### **Vertrauen als bivalentes Entscheidungsproblem**

Die probabilistische Modellierung von Vertrauen fokussiert auf die Vertrauenshandlung mit der Folge, dass es streng genommen nur zwei mögliche Handlungsalternativen gibt: die Vertrauenshandlung und die Misstrauenshandlung.

Lahno argumentiert, dass ein Modell, das die Vertrauenshandlung in den Mittelpunkt rückt, keine Möglichkeiten gradueller Abstufungen (vgl. Lahno 2002: 117; 119) zulasse. Feine Differenzierungen zwischen blinder Vertrauenshandlung und völliger Misstrauenshandlung seien aber gerade notwendig, um z. B. die graduelle Entwicklung von Vertrauen in Freundschaften näher zu analysieren.

Neben dem eher inhaltlichen Argument von Lahno lässt sich seine Kritik zu einem methodologischen Problem weiterentwickeln: die Schwierigkeit der Reduktion komplexer Situationen auf bivalente Entscheidungsprobleme. Bivalente Entscheidungsprobleme sind solche, bei denen *entweder* die eine *oder* die andere Handlungsalternative gilt bzw. ausgewählt werden muss. Ein Sowohl-als-auch, Kron und Winter (2006) nennen das „Vagheit“, bleibt ausgeschlossen. Einen möglicher Ausweg, so argumentiert Kron (2005), könnte die Einführung von Fuzzy-Logik in die Soziologie bringen, die den

Umgang mit Vagheit und einem Sowohl-als-auch bereits auf formaler Ebene berücksichtigt.

In Bezug auf die Interaktion mit autonomer Technik gibt es in der Regel keine einfache Entscheidung zwischen zwei Handlungsalternativen. Vielmehr manifestiert sich Vertrauen in der Art und Weise, wie mit der Technik umgegangen wird (vgl. Abschnitt 4 zu Vertrauen in der Automationsforschung).

### **Vertrauensparadox und ethische Einwände**

Bart Nooteboom (2002) führt in seiner umfassenden Monographie, die bereits Gegenstand von Abschnitt 2.2.3 war, zwei weitere Argumente an, die insbesondere für den Fall der vollständigen Vertrauenswürdigkeit  $p = 1$  des Treuhänders gelten:

Erstens gilt die paradoxe Beobachtung, dass für den Fall  $p = 1$  keinerlei Risiko mehr besteht, obwohl Vertrauensprobleme sowohl bei Deutsch als auch bei Coleman explizit als Entscheidungen unter Risiko definiert sind – eine Vertrauensentscheidung ist dann nicht mehr möglich:

„[...] [T]rust is highest when the probability is unity, so that there is no risk left. However, another intuition is that trust entails a wager, an acceptance of risk, so that if there is no risk, we can no longer speak of trust.“ (Nooteboom 2002: 39)

Der zweite Kritikpunkt von Nooteboom enthält ein ethisches Argument: Wenn  $p = 1$  gilt, ist das Handeln unseres Treuhänders vollkommen sicher, das heißt dieser hat keinerlei Freiheiten mehr hinsichtlich seiner Entscheidungen (vgl. Nooteboom 2002: 40). Sein Handeln wäre vorherbestimmt und im Widerspruch zu einem, zumindest im westlichen Kulturkreis etablierten, individualistischen Menschenbild.

Insbesondere der erste Kritikpunkt von Nooteboom weist auf die Schwierigkeiten hin, Vertrauenswürdigkeit nonchalant als simple Wahrscheinlichkeitsvariable zu codieren. Die Analyse von Vertrauenshandlungen, so mein Fazit der Kritik Nootebooms, erfordert es, über monokausale Erklärungen hinauszugehen und nach komplexeren Erklärungsgründen für Handlungen, die von Vertrauen beeinflusst sind<sup>2</sup>, zu suchen.

### **Vertrauen ist subjektabhängig**

Preisendörfer (1995) hat das Vertrauensmodell von Coleman (und damit implizit auch das von Esser) scharf kritisiert. Er argumentiert (ebd.: 268), dass in der Variablen  $p$ , der geschätzten Vertrauenswürdigkeit, der Kern des Vertrauensproblem codiert

<sup>2</sup>Der Begriff *Vertrauenshandlung* wird hier bewusst vermieden, da eine Vertrauenshandlung erfordert, diese überhaupt als solche identifizieren zu können. Der Autor dieser Arbeit vertritt hingegen die Auffassung, dass Vertrauen die Handlungen von Akteuren und von autonomer Technik beeinflusst, diese aber nicht präzise in bivalenter Weise in Vertrauens- und Misstrauenshandlungen unterscheidbar macht.

sei: die Einholung möglichst vieler Informationen, die die Vertrauenswürdigkeit des Gegenüber abschätzbar machen. Ein weiterer Kritikpunkt Preisendörfers besteht darin, dass Gewinn  $G$  und Verlust  $V$  nicht ex ante angebbar sind und sich im Verlauf einer Vertrauensbeziehung stetig anpassen. Zuletzt kritisiert er, dass  $p$  stark personenabhängig und lebensgeschichtlich geprägt sei. Die Kritik der mangelnden Personenabhängigkeit relativiert sich zumindest für Esser jedoch, da dieser von *subjektivem* Erwartungsnutzen ausgeht und damit  $p$  subjektabhängig modelliert werden kann.

### Implikationen

Trotz der eingangs erwähnten Nähe probabilistischer Modelle zu Theorien der subjektiven Erwartungsnutzenmaximierung, die im Rahmen des eigenen theoretischen Modells von Bedeutung sind, ist der inhaltliche Ertrag probabilistischer Modelle für die vorliegende Arbeit gering. Allenfalls die Idee, auch Vertrauensbeziehungen in den Blick zu nehmen, könnte im Rahmen von Mensch-Technik-Interaktion ein zu berücksichtigender Aspekt sein, der seinen Niederschlag punktuell im Konzept der wechselseitigen Modellzuschreibungen (vgl. Abschnitt 8.2) findet.

Im folgenden Abschnitt zum spieltheoretischen Vertrauensbegriff gehen punktuell auch probabilistische Überlegungen in die Modellierung von Reputation ein, die Ähnlichkeiten zu den hier beschriebenen Konzepten aufweisen.

## 3.2 Spieltheoretischer Vertrauensbegriff

„Spieltheoretische Modellierungen und empirische Anwendungen in der Soziologie“, so der Titel eines Überblicksartikels von Raub und Buskens (2006), erfreuen sich zunehmender Beliebtheit. Immer dann, wenn es um prozessförmiges Entscheiden in strategischen Interdependenzsituationen geht, wird gerne auf spieltheoretische Modelle zurückgegriffen.

Hartmut Esser begreift die Spieltheorie als Modellierungsmöglichkeit sozialer Situationen und die „Lösung“ von Dilemmata-Situationen als möglichen Transformations- bzw. Aggregationsmechanismus: „Die Spieltheorie kann insofern als eine Ansammlung von formalen Transformationsregeln (und von Brückenhypothesen!) für bestimmte Typen sozialer Situationen angesehen werden“ (Esser 2000a: 23).

In diesem Abschnitt wird es hauptsächlich um das sogenannte *Vertrauensspiel* und seine Varianten gehen. Vor dem Hintergrund der Frage nach Vertrauen in autonome Technik ist auch hier zu prüfen, inwieweit Konzepte und Ergebnisse der spieltheoretischen Vertrauensforschung übertragbar sind auf Untersuchungen zu Vertrauen in autonome Technik. Insbesondere die Überlegung, Vertrauen im Rahmen von prozessförmigen experimentellen Settings zu untersuchen, wie sie für den Empirie-Teil der vorliegenden Arbeit umgesetzt wurde, stammt aus der spieltheoretischen Vertrauensforschung.

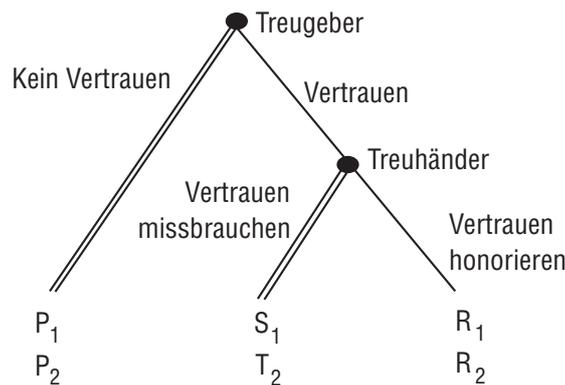
### 3.2.1 Das Vertrauensspiel

Spieltheoretisch lässt sich Vertrauen mit Hilfe des sogenannten Vertrauensspiels (in verschiedenen Varianten) modellieren (vgl. u.a. Deutsch 1958; Dasgupta 1988; Kreps 1990; Buskens 1998).

In der Grundvariante des Spiels legt der Treugeber mit einem Zug eine Entscheidung vor, er vertraut oder misstraut. Misstraut er, ist das Spiel bereits zu Ende, da der Treuhänder keine Möglichkeit mehr hat, das Vertrauen zu bestätigen oder zu missbrauchen. Hat der Treugeber im ersten Zug hingegen vertraut, kann der Treuhänder auf den Vertrauensvorschuss entweder mit Bestätigung des Vertrauens oder mit Missbrauch des Vertrauens antworten.

In der spieltheoretischen Modellierung entspricht Vertrauen der kooperativen (C) und Misstrauen der defektiven Strategie (D).

Spieltheoretisch wird Vertrauen als sequentielles Spiel modelliert, dessen extensive Form in Abbildung 3.1 dargestellt ist.



**Abbildung 3.1:** Vertrauensspiel in extensiver Form  
Quelle: Raub und Buskens (2006)

Für die Auszahlungswerte gilt:

$$S_1 < P_1 < R_1 \wedge P_2 < R_2 < T_2$$

In ausformulierter Form ist Abbildung 3.1 wie folgt zu lesen: Das Vertrauensspiel ist als sequentielles Spiel angelegt, d. h. der *Treugeber* entscheidet im ersten Zug, ob er die kooperative Strategie „Vertrauen“ oder die defektive Strategie „Kein Vertrauen“ wählt. Bei Wahl von D endet das Spiel sofort, da der Treuhänder keine Chance mehr hat, das Gegenteil zu beweisen. In diesem Fall erhält Spieler 1 die Auszahlung  $P_1$  (Punishment), Spieler 2 die Auszahlung  $P_2$ . Wählt der Treugeber im ersten Zug C, kann der Treugeber entweder „Vertrauen honorieren“ (C) oder „Vertrauen missbrauchen“ (D). Im Falle der Strategiekombination CD (Treugeber vertraut, Treuhänder missbraucht das Vertrauen) erhält Spieler 1 die Auszahlung  $S_1$  (Sucker) und Spieler 2

die Auszahlung  $T_2$  (Temptation). Im Falle der Strategiekombination CC, Treugeber vertraut, Treuhänder honoriert das Vertrauen, erhält Spieler 1 die Auszahlung  $R_1$  (Reward) und Spieler 2 die Auszahlung  $R_2$ .

Die Auszahlungsmatrix bzw. die Normalform des Vertrauensspiels findet sich in Tabelle 3.1.

<i>Treugeber / Treuhänder</i>	Vertrauen honorieren	Vertrauen missbrauchen
Vertrauen	$(R_1, R_2)$	$(S_1, T_2)$
kein Vertrauen	$(P_1, P_2)$	$(P_1, P_2)$

**Tabelle 3.1:** Auszahlungsmatrix der Grundversion des Vertrauensspiels  
Quelle: Raub und Buskens (2006: 564)

Die Normalform einer spieltheoretischen Konstellation verschleiert jedoch wichtige Details der Spielstruktur. So merken etwa Raub und Buskens (2006: 564) kritisch an: „Wir bemerken, dass die Normalform bestimmte Details des Spiels ‚verbirgt‘, die in spieltheoretischen Analysen wichtig werden können, insbesondere die Reihenfolge, in der die Akteure ziehen.“

Im Falle der hier betrachteten einfachen Version des Vertrauensspiels ist die Strategiekombination DC laut der extensiven Form (Abbildung 3.1) gar nicht möglich, sie ist allerdings in der tabellarischen Normalform mit aufgeführt.

### Spieltheoretische Analyse: Wechselseitiges Vertrauen ist unmöglich

In der hier vorgestellten Grundvariante handelt es sich zunächst um ein nicht-kooperatives, nicht-iteriertes Spiel<sup>3</sup>. Unterstellt man den Akteuren individuelle Rationalität, wird Spieler 1 immer die defektive Strategie wählen: Angenommen Spieler 1 (bezeichnet mit  $A$ ) würde kooperieren, also vertrauen, so wäre es für Spieler 2 (bezeichnet mit  $B$ ) sinnvoll, zu defektieren, um Auszahlung  $T_2$  statt  $R_2$  ( $T_2 > R_2$ ) zu realisieren. Da  $A$  die Opportunitäten von  $B$  antizipiert, ist es für  $A$  rational, auch zu defektieren, da  $A$  auf diese Weise Auszahlung  $P_1$  statt  $S_1$  realisieren kann ( $P_1 > S_1$ ). Unter der Bedingung, dass es sich um ein einmaliges, nicht-kooperatives Spiel handelt, ist das einzige und gleichzeitig ineffiziente Gleichgewicht des Spiels die Strategiekombination DD, beidseitiges Misstrauen (vgl. Kreps 1990: 101).

An die spieltheoretische Analyse des Vertrauensspiels schließt sich die Frage an, warum entgegen der Vorhersage empirisch kein dauerhaftes Misstrauen nachzuweisen ist, sondern bestimmte Bedingungen doch dazu führen, dass wechselseitiges Vertrauen möglich ist. Morton Deutsch (1958: 279) hat u. a. mittels Laborexperimenten herausgefunden, dass gegenseitiges Vertrauen auch bei individualistischer oder kompetitiver Einstellung entstehen kann, wenn die situativen Bedingungen die Entwicklung von Vertrauen unterstützen (für Details s. Abschnitt 3.2.2).

<sup>3</sup>Für Grundbegriffe der Spieltheorie vgl. Diekmann (2009).

Spätere Autoren, etwa Partha Dasgupta (1988) und David Kreps (1990) heben die Bedeutsamkeit von Reputation und unterschiedlichen Spielertypen für die Entstehung wechselseitigen Vertrauens besonders hervor und analysieren spieltheoretisch die Auswirkungen von Iteration auf die „Lösung“ des Vertrauensspiels (vgl. Abschnitt 3.2.3). Schließlich kombiniert Vincent Buskens (1995; 1998) experimentelle Untersuchungen zu Vertrauen mit Iteration und Probabilistik zu einem elaborierten Vertrauensmodell, dem *Iterated Heterogeneous Trust Game* (vgl. Abschnitt 3.2.4).

### 3.2.2 Experimentelle Untersuchung von Vertrauen

Die spieltheoretische Analyse des basalen Vertrauensspiels im vorherigen Abschnitt hat gezeigt, dass Vertrauen, zumindest aus analytischer Perspektive, eigentlich nicht möglich ist. In den frühen Arbeiten von Morton Deutsch (1958) finden sich die ersten Ansätze einer experimentell-spieltheoretischen Analyse von Vertrauen. Deutsch gilt sogar als einer der ersten Vertreter der heute so prominenten experimentellen Spieltheorie (vgl. Ostrom 2005: 211) – einem Forschungsfeld, auf dem schon mehrere Ökonomie-Nobelpreise vergeben wurden.

Den Ausgangspunkt für seine Experimente bildet ein Vertrauensbegriff, der positive und negative Konsequenzen des Handelns gegeneinander abwägt (vgl. Deutsch 1958: 266). Im Gegensatz zu den späteren Arbeiten von Deutsch (1976: 139f.), die bereits Gegenstand von Abschnitt 3.1.1 waren, ist der hier verwendete Vertrauensbegriff nicht so weitreichend formalisiert.

Die experimentelle Situation ist wie folgt aufgebaut: Spieler I kann zwischen den Handlungsalternativen X und Y wählen, Spieler II zwischen A und B (vgl. Tabelle 3.2).

<i>Spieler I / Spieler II</i>	A	B
X	(+9, +9)	(−10, +10)
Y	(+0, −10)	(−9, −9)

**Tabelle 3.2:** Auszahlungen im experimentellen Vertrauensspiel  
Deutsch (1958: 269)

Strukturell ähnelt die Auszahlungsmatrix der klassischen Gefangenendilemmakonstellation. Genau wie im Gefangenendilemma gilt hier

$$T(+10) > R(+9) > P(-9) > S(-10)$$

Deutsch führte zwei Experimente durch: ein *one-trial game* (ein Spiel mit nur einem Durchlauf) und ein *ten-trial game* (ein Spiel mit zehn Durchläufen). Innerhalb der Spiele wurden die Bedingungen variiert: *No-communication* bedeutet, dass die Entscheidung geheim getroffen wird und dass die Spieler vor den Zügen nicht kommunizieren dürfen. Bei *Communication* wird die Entscheidung zwar auch geheim getroffen,

allerdings dürfen die Spieler vor ihren Zügen kommunizieren. *Reversibility* beinhaltet die Möglichkeit, dass nach einem Zug die Entscheidungen korrigiert werden können. Im Falle von *Non-simultaneous* werden die Züge hintereinander ausgeführt.

Die praktische Umsetzung (vgl. Deutsch 1958: 269–271) dieses experimentellen Settings sah zunächst vor, dass die Spielteilnehmer zu Spielbeginn unterschiedlich instruiert wurden („motivational orientation“). Dabei wurde zwischen drei motivationalen Orientierungen unterschieden: kooperativ, individualistisch und kompetitiv. Bei der kooperativen Orientierung sollte das eigene Wohlergehen und das Wohlergehen des Gegenübers die oberste Priorität haben. Bei der individualistischen Orientierung ist ausschließlich das Eigeninteresse von Bedeutung, der andere Spieler ist unwichtig. Bei der kompetitiven Orientierung galt es, für sich selbst das Beste herauszuholen und zusätzlich besser als der andere zu sein.

Deutsch (1958: 272) fand heraus, dass „[...] the behaviour resulting from an individualistic orientation is very much influenced by situational determinants.“ Bei individualistischer Orientierung hing die Wahl der Handlungsalternative also stark von den Experimentalbedingungen (Kommunikation, keine Kommunikation, simultane Züge, Möglichkeit des Rückgängigmachens) ab. Deutsch sah in seinen Experimenten die Bedeutsamkeit motivationaler Orientierungen für die Wahl von Handlungsalternativen bestätigt: „Taken together, the results clearly support the hypotheses concerning the effects of motivational orientation“ (ebd.).

Die experimentell-spieltheoretischen Arbeiten von Deutsch haben gezeigt, dass sich Vertrauen in Form von Laborexperimenten untersuchen lässt und auf Basis der Experimente inhaltlich gehaltvolle Ergebnisse erzielt werden können.

Inhaltlich fand Deutsch (1958: 278–279) u. a. heraus, dass gegenseitiges Vertrauen vor allem dann entsteht, wenn Menschen dem Wohl des anderen gegenüber positiv eingestellt sind. Selbst bei individualistischer oder kompetitiver Einstellung kann Vertrauen entstehen, wenn die situativen Bedingungen die Entwicklung von Vertrauen unterstützen (z. B. sequentielle Entscheidungen, Präsenz einer dritten Person, Sanktionsmöglichkeiten etc.). Die von Deutsch bereits in Grundzügen formulierten situativen Bedingungen zur Entstehung von Vertrauen wurden in späteren Studien (Dasgupta 1988; Buskens 1998; Gambetta 1988a) aufgegriffen und systematisch untersucht. Diese Studien sind Gegenstand der folgenden Abschnitte.

### 3.2.3 Iterierte bzw. reputationsbasierte Vertrauensspiele

Der von Partha Dasgupta (1988) näher beschriebene iterierte Vertrauensbegriff definiert Vertrauen als korrekt eingeschätzte Erwartung an die Handlungen anderer, die Auswirkungen auf die eigene Handlungswahl haben. Das Handlungsproblem besteht darin, dass die eigene Handlungswahl erfolgen muss, *bevor* die Handlungen anderer

beobachtbar sind. Es handelt sich also um eine typische spieltheoretische Konstellation, bei der die situativen Opportunitäten wechselseitig antizipiert werden:

„[...] correct expectations about the *actions* of other people that have a bearing on one's own choice of action when that action must be chosen before one can *monitor* the actions of those others.“ (Dasgupta 1988: 51)

Als Kernproblem des Vertrauens benennt Dasgupta, dass wir bei Begegnungen mit einer Person „[...] wish to know the sort of person we are dealing with before we deal with him“ (ebd.: 54). Die Entstehung von Reputation im Kontext sich wiederholender Begegnungen und die Auswirkungen von Reputation auf die Handlungswahl des Treugebers sind die zentralen Fragestellungen, mit denen sich Dasgupta beschäftigt. Um diesem Aspekt näher zu kommen, analysiert der Autor zunächst eine Situation, die ohne Iteration auskommt. Dabei wird angenommen, dass die betrachtete Population so groß ist, dass keine Informationen aus vorherigen Interaktionen vorliegen (ebd.: 57). In dem Szenario besteht die Annahme, dass zwei unterschiedliche Typen aufeinandertreffen können. *Typ 1* handelt aus ehrlicherem Antrieb als *Typ 2*. Es ist möglich ehrlich (X) oder unehrlich (Y) zu handeln.

Gegenüber der zuvor beschriebenen Konstellation bei Deutsch unterscheidet sich der Ansatz von Dasgupta dadurch, dass unterschiedliche Typen mit je unterschiedlichen Auszahlungswerten berücksichtigt werden und dass Reputationsmechanismen in das Modell integriert werden.

Operationalisiert ist das Zusammentreffen der Typen mittels spezifischer Auszahlungswerte (vgl. Tabelle 3.3–3.5), die nicht weiter empirisch begründet werden. In Abhängigkeit eines bestimmten Schwellenwertes<sup>4</sup>  $p$  der geschätzten Ehrlichkeit bzw. Reputation des Treuhänders und des betrachteten Typs lässt sich eine „vernünftige“ Handlungswahl (ehrlich handeln vs. unehrlich handeln) ausrechnen.

Exemplarisch berechnet Dasgupta (1988: 58) dies für Spieler vom Typ 1. Der Rechenansatz besteht in der Berechnung der Grenzwahrscheinlichkeit, also der Wahrscheinlichkeit, ab der Spielertyp 1 indifferent zwischen Handlungsalternative X und Y ist:

$$\underbrace{30p + 5(1 - p)}_{S1:X; S2: X \text{ oder } Y} = \underbrace{5p + 10(1 - p)}_{S1:Y; S2: X \text{ oder } Y} \Leftrightarrow p = \frac{1}{6} \quad (3.7)$$

Wenn ein Spieler vom Typ 1 die Ehrlichkeit des Gegenübers mit  $p \geq \frac{1}{6}$  bewertet, gibt es für diesen also gute Gründe, sich für X (ehrlich handeln) zu entscheiden. Für Typ 2 kann man analog berechnen, dass dieser X wählen sollte, wenn er die Ehrlichkeit des

<sup>4</sup>Wenn der Treuhänder mit Wahrscheinlichkeit  $p = 1$  Handlungsalternative X wählt, ist es für den Treugeber ohne weitere Überlegung sinnvoll, ebenfalls X zu wählen. In dem Fall besteht gar kein Vertrauensproblem.

Gegenübers mit  $p \geq \frac{1}{4}$  einschätzt. Spieler vom Typ 1 sind also eher geneigt, ehrlich zu handeln.

In Bezug auf die in Abschnitt 3.1 vorgestellten probabilistischen Ansätze, bei denen die geschätzte Vertrauenswürdigkeit des Treuhänders nicht weiter inhaltlich ausgeführt wurde, findet sich bei Dasgupta mit dieser Berechnung, so lässt sich anmerken, eine nachvollziehbare Berechnungsgrundlage, die aus der Spielstruktur folgt.

<i>Typ 1 / Typ 2</i>	X	Y
X	( <b>30,20</b> )	(5, 5)
Y	(5, 5)	(10, 10)

**Tabelle 3.3:** Auszahlungsmatrix für das Zusammentreffen von Typ 1 / Typ 2 - Spielern

<i>Typ 1 / Typ 1</i>	X	Y
X	( <b>30,30</b> )	(5, 5)
Y	(5, 5)	(10, 10)

**Tabelle 3.4:** Auszahlungsmatrix für das Zusammentreffen von Typ 1 / Typ 1 - Spielern

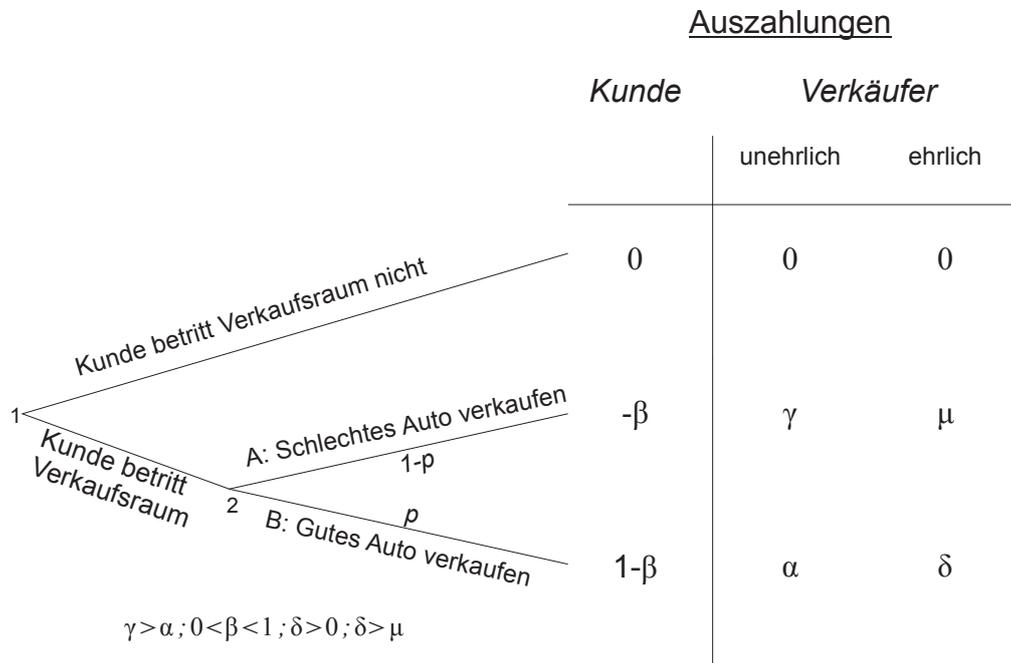
<i>Typ 2 / Typ 2</i>	X	Y
X	( <b>20,20</b> )	(5, 5)
Y	(5, 5)	(10, 10)

**Tabelle 3.5:** Auszahlungsmatrix für das Zusammentreffen von Typ 2 / Typ 2 - Spielern

Dasgupta bemerkt selbst, dass bei der betrachteten Situation sehr künstliche Annahmen getroffen wurden, da bei Vertrauen zwischen Individuen sich wiederholende Zusammentreffen und Erinnerungen an bereits stattgefundene Treffen eine große Rolle spielen: „For trust to be developed between individuals they must have repeated encounters, and they must have some memory of previous encounters“ (Dasgupta 1988: 59). Diese Überlegung führt ihn zum Kern seines Ansatzes, einem *reputationsbasierten iterierten* Vertrauensmodell.

**Das Autoverkäufer-Beispiel: Iteration und Vertrauen** Wiederkehrende Begegnungen und Erinnerungen an den Ausgang vorheriger Begegnungen sind Grundvoraussetzungen für ein reputationsbasiertes Vertrauensmodell (Dasgupta 1988: 59).

Dasgupta wählt als erklärendes Beispiel den Autoverkauf. Die beiden Spieler sind Kunde (Treugeber) und Verkäufer (Trehänder). Im ersten Schritt kann sich der Kunde entscheiden, ob er den Verkaufsraum betritt oder nicht. Er trifft diese Entscheidung ausschließlich auf der Grundlage seines individuellen Nettonutzens, der sich aus der spezifischen, vollständig bekannten Auszahlungs- und Spielstruktur (vgl. Abbildung 3.2) ergibt. Betritt der Kunde den Verkaufsraum nicht, erhalten beide Spieler die Auszahlung 0. Betritt der Kunde den Verkaufsraum, verkauft der Verkäufer (Schritt 2) ihm entweder ein gutes Auto oder ein schlechtes Auto.



**Abbildung 3.2:** Spielbaum des iterierten Vertrauensspiels. In Schritt 2 wird mit Wahrscheinlichkeit  $p$  ein zuverlässiges Auto verkauft, mit Wahrscheinlichkeit  $1 - p$  ein unzuverlässiges.  
Quelle: Dasgupta (1988: 62) (Übersetzung: R. F.)

Den jeweiligen Spielausgängen sind spezifische Nutzenwerte (Payoffs) zugeordnet: Entscheidet sich der Kunde für das Betreten (und damit für den Kauf eines Autos), erhält er eine Auszahlung von  $-\beta$ , wenn er ein schlechtes Auto erhält, und  $1 - \beta$ , wenn er ein gutes Auto erhält, wobei  $0 < \beta < 1$  gilt. Die Auszahlungswerte für den Verkäufer hängen vom Verkäufertyp ab. Handelt es sich um einen unehrlichen Verkäufer, so erhält dieser eine höhere Auszahlung, wenn er ein schlechtes Auto verkauft ( $\gamma > \alpha$ ). Handelt es sich hingegen um einen ehrlichen Verkäufer, so erhält er eine höhere Auszahlung beim Verkauf eines guten Autos ( $\delta > \mu$ ).

Die **Reputation**  $p$  eines Verkäufers ist entscheidend dafür, ob der Kunde den Verkaufsraum betritt oder nicht. Dabei kann  $p$  die durchschnittliche Ehrlichkeit der Population der Autoverkäufer (vgl. Dasgupta 1988: 63f.) darstellen, oder einen verkäuferspezifischen Wert, der sich aus der bisherigen Ehrlichkeit des Verkäufers in zuvor stattgefundenen Transaktionen ergeben hat (vgl. ebd.: 66ff.).

Aus der Spielstruktur, den Auszahlungswerten (Abbildung 3.2) und der geschätzten Wahrscheinlichkeit  $p$  auf einen ehrlichen Verkäufer zu treffen, lässt sich der Nettutzen (NN) für die Handlungsalternative „Verkaufraum betreten“ berechnen (vgl. Dasgupta 1988: 63):

$$NN(\text{„Verkaufsraum betreten“}) = \underbrace{p(1 - \beta)}_{\text{ehrlich}} - \underbrace{(1 - p)\beta}_{\text{unehrlich}} = p - p\beta - (\beta - p\beta) = p - \beta$$

Wenn also  $p$  den Wert  $\beta$  überschreitet, ist es für den Käufer rational, den Verkaufsraum zu betreten und damit eine Vertrauenshandlung zu vollziehen.

Der entscheidende Unterschied zur rein probabilistischen (und nicht prozessförmigen) Modellierung von Vertrauen ist die Berücksichtigung und Antizipation eines möglichen Vertrauensmissbrauchs. Der probabilistische Vertrauensbegriff (siehe vor allem die Ausführungen zu der Modellierung bei Esser auf Seite 51) klammert einen Vertrauensmissbrauch in der Modellierung aus.

**Tatsächliche vs. zugeschriebene Vertrauenswürdigkeit** Ob eine Transaktion zustande kommt, hängt, wie im vorherigen Abschnitt gezeigt wurde, wesentlich von der geschätzten Vertrauenswürdigkeit  $p$  des Verkäufers, m. a. W. von seiner Reputation ab. Entscheidend für die Einschätzung eines bis dato unbekanntes Verkäufers sind die *Erfahrungen* mit anderen Autoverkäufern. Dasgupta (1988: 64f.) argumentiert, dass aus Bewertungen vergangener Interaktionen mit einzelnen Populationsmitgliedern induktiv Schlüsse für die gesamte Population gezogen werden:

„In dealing with someone you learn something about him, but also about others in his society. You learn something about population statistics.“

Wenn einem Käufer bereits fünf gute Autos verkauft worden sind, so geht dieser davon aus, dass Autoverkäufer *im Allgemeinen* ziemlich ehrliche Personen sind. Zur wahrscheinlichkeitstheoretischen Fundierung der Aussagen sind folgende zwei Wahrscheinlichkeiten zu betrachten:

- $p$  – tatsächlicher Anteil ehrlicher Verkäufer in der Population<sup>5</sup>
- $p'(t)$  – subjektiv geschätzter Anteil ehrlicher Verkäufer nach  $t$  Begegnungen

Es werden solange neue Transaktionen eingegangen, wie  $p'(t) > \beta$  (s. o.) gilt. Führt nun eine zufällige Begegnung mit einem unehrlichen Verkäufer zum Zeitpunkt  $t + 1$  dazu, dass für die aktualisierte Wahrscheinlichkeit<sup>6</sup> gilt:  $p'(t) > p'(t + 1) < \beta$ , wird der Käufer ab sofort keinerlei Transaktionen mehr mit einem Autoverkäufer

<sup>5</sup>Das hier verwendete  $p$  ist keinesfalls zu verwechseln mit der geschätzten Vertrauenswürdigkeit  $p$  bei den probabilistischen Vertrauenskonzepten, bei denen  $p$  eine Individualekonstante und keine Populationskonstante ist.

<sup>6</sup>Die Aktualisierung der geschätzten Wahrscheinlichkeiten kann z. B. mit Hilfe der Satzes von Bayes für bedingte Wahrscheinlichkeiten erfolgen. Die konkrete Anwendung erfordert jedoch zusätzliche Informationen, die in einer spezifischen Anwendung als Modellannahmen hinzugefügt werden müssen.

eingehen, unabhängig davon, ob tatsächlich  $p < \beta$  gilt. Solange für die aktualisierte Wahrscheinlichkeit stets gilt  $p' > \beta$ , gilt auf lange Sicht  $\lim_{t \rightarrow \infty} p'(t) = p$ : Die geschätzte und tatsächliche Vertrauenswürdigkeit der Population der Autoverkäufer nähern sich einander an.

Sofern also  $p < \beta$  gilt, wird auf lange Sicht die Transaktionskette in jedem Fall beendet werden. Gilt hingegen  $p > \beta$ , hängt es vom Zufall ab, ob die Transaktionen unendlich weitergehen. Insbesondere die ersten Begegnungen können schnell zur Beendigung der Transaktionen führen. Wichtige Eigenschaften von Vertrauen, z. B. dass sich dieses bei Nutzung mehrt und anfällig für Zerstörung ist, lassen sich folglich mittels Reputationsmechanismen formal nachvollziehen.

Aus den Betrachtungen von Dasgupta lassen sich einige Erkenntnisse destillieren: Vertrauen hängt von den Erwartungen an das Handeln des Gegenübers ab und kann sinnvollerweise nur in einem iterativen Zusammenhang mit wiederkehrenden Begegnungen betrachtet werden. Treugeber gründen ihr Vertrauen auf vorherige Erfahrungen, die sich entweder auf die Reputation einer Population (z. B. die der Autoverkäufer) beziehen oder auf vorangegangene Erfahrungen mit einem bestimmten Interaktionspartner und dem daraus resultierenden individuellen Reputationswert. Die zugeschriebene bzw. geschätzte Vertrauenswürdigkeit ist keinesfalls fest, sondern wird nach jeder abgeschlossenen Interaktion, die sich auf Vertrauen gründete, aktualisiert. David Kreps (1990), der seine Arbeiten offenbar parallel zu Partha Dasgupta verfasst hat, beschreibt ebenso ein basales Vertrauensspiel, das von der Auszahlungsstruktur weitgehend dem Autoverkäuferbeispiel von Dasgupta (1988) gleicht. Im Unterschied zu Dasgupta berücksichtigt Kreps allerdings keine unterschiedlichen Verkäufertypen (eher ehrlich, eher unehrlich), sondern befasst sich aus einer allgemein analytischen Perspektive mit Auswegen aus dem Vertrauensdilemma, das strukturell einem Gefangenendilemma entspricht. Als mögliche Lösung, so argumentiert Kreps, komme Iteration in Frage (ebd.: 101). Iteration ermöglicht es zwar prinzipiell Vertrauens-Belohnungs-Arrangement („trust-honor arrangement“) durchzusetzen und als Gleichgewichte zu etablieren (ebd.: 103), allerdings stellen sich diese Gleichgewichte nicht automatisch ein. Zur Realisierung von Gleichgewichten müssten folgende drei Probleme gelöst werden (vgl. Kreps 1990: 103–105): Erstens können verschiedene Handelspartner und die damit verbundenen unterschiedlichen Auszahlungsstrukturen zu multiplen Gleichgewichten führen<sup>7</sup>. Zweitens führt das Argument der Rückwärtsinduktion dazu, dass kein Vertrauen zustande kommen kann, da im Falle von endlichen Spielen der Treuhänder in der letzten Runde in jedem Fall betrügt und die Antizipation rückwirkend auf die Überlegungen des Treugebers wirkt. Drittens weist Kreps (1990: 104f.) auf den Umstand hin, dass der Missbrauch von Vertrauen nicht in jedem

<sup>7</sup>Diesen Einwand greift Dasgupta (1988) zumindest prinzipiell mit der Berücksichtigung von eher ehrlichen und eher unehrlichen Verkäufertypen auf. Buskens (1995; 1998) modelliert mit dem iterierten heterogenen Vertrauensspiel (vgl. Abschnitt 3.2.4) explizit beliebig viele Verkäufertypen.

Fall direkt beobachtbar ist<sup>8</sup> und damit nicht sicher festgestellt werden kann, welche Strategie der Treuhänder gewählt hat.

Trotz Iteration, so lässt sich folgern, ist das Vertrauensdilemma nicht zwangsläufig aufzulösen.

### 3.2.4 Buskens: Synthese spieltheoretischer Vertrauensmodelle

Das im Folgenden dargestellte iterierte heterogene Vertrauensspiel (*Iterated Heterogeneous Trust Game*, IHTG, Buskens 1995; 1998) lässt sich als Synthese des basalen Vertrauensspiels (vgl. Abschnitt 3.2.1) und der Überlegungen zu einem reputationsbasierten, spieltheoretischen Vertrauensbegriff (vgl. Abschnitt 3.2.3) beschreiben.

Im Gegensatz zu den Ausführungen zur Iteration bei Dasgupta (1988) werden Iteration und Reputation bei Buskens präziser und systematischer berücksichtigt. Wie gleich gezeigt wird, werden die zentralen Eigenschaften seines Vertrauensbegriffs unmittelbar in der abstrakten mathematischen Modellierung aufgegriffen.

Eine wesentliche Erweiterung von Buskens im Vergleich zu den bisher vorgestellten Ansätzen besteht in der zusätzlichen Berücksichtigung von Informationszirkulationsnetzwerken zwischen Treugeber-Typen. Die verbundenen Treugeber können sich je nach Intensität ihrer jeweiligen Beziehung über den Missbrauch von Vertrauen eines Treuhänders austauschen und auf diese Weise den Ruf eines Treuhänders in ihren eigenen Abwägungen berücksichtigen.

Zunächst geht Buskens (1995: 3) vom basalen Vertrauensspiel mit festen Payoffs je Strategiekombination zu einer variablen Variante über. Dabei variieren die Anreize für einen Vertrauensmissbrauch qua Naturzustand in der Größenordnung  $\theta$ . Dem eigentlichen Zug der Spieler ist ein Zug der Natur vorgelagert, der zufällig einen Wert für den Parameter  $\theta \in [0, \infty[$  ermittelt. Als Auszahlungsstruktur für das Spiel  $\Gamma_\theta$  ergibt sich:

<i>Treugeber / Treuhänder</i>	C	D
C	$(R_1, R_2)$	$(S_1, R_2 + \theta)$
D	$(-)$	$(P_1, P_2)$

**Tabelle 3.6:** Auszahlungsmatrix für das heterogene Vertrauensspiel  $\Gamma_\theta$ , bei dem der Anreiz für einseitigen Vertrauensmissbrauch vom Zug der Natur abhängt und je Spiel variabel ist. Die Strategiekombination (D,C) ist nicht möglich, da das Spiel sequentiell ist und im Falle der Wahl von D durch Spieler 1 der Spieler 2 immer D wählt. Abgeleitet aus den Ausführungen bei Buskens (1995: 3)

<sup>8</sup>Buskens (1998) löst dieses Problem, indem er die Bedeutsamkeit globaler Netzwerkstrukturen für die Bewertung komplexer Produkte bzw. Transaktionen hervorhebt und schlüssig zeigt (vgl. Abschnitt 3.2.4), dass Vertrauen auch in solchen Fällen entstehen kann, bei denen die Qualitätsbewertung einer Transaktionen durch einen Spieler alleine nicht bewerkstelligt werden kann.

Das vollständige IHTG  $\Gamma$  wird beschrieben durch ein Sechs-Tupel  $\Gamma(\Gamma_\theta, F, \boldsymbol{\delta}, \boldsymbol{\pi}, \mathbf{A}, w)$ , wobei die fettgedruckten Zeichen Vektoren bzw. Matrizen sind:

- $\Gamma_\theta$  – Vertrauensspiel mit spezifischer Auszahlungsstruktur gemäß Tabelle 3.6. Der Anreiz für den einseitigen Missbrauch durch Spieler 2 variiert und folgt einer Zufallsverteilung  $F_a(\theta)$ , die vom Parameter  $a$  abhängt, der als „average incentive for opportunistic behaviour“ (Buskens 1995: 3) interpretiert werden kann.
- $F$  – bezeichnet die Zufallsverteilung, die für die variierende Spielstruktur  $\Gamma_\theta$  bzw. für den Zug der Natur notwendig ist.
- $\boldsymbol{\delta}$  – beinhaltet einen Vektor von Überlebensraten  $\delta_i$ , die angeben, mit welcher Wahrscheinlichkeit  $1 - \delta_i$  ein Treugeber vom Typ  $i$  die aktuelle Runde überlebt.
- $\boldsymbol{\pi}$  – beinhaltet einen Vektor, der angibt, mit welcher Wahrscheinlichkeit  $\pi_j$  ein neuer Treugeber vom Typ  $j$  ausgewählt wird.  $\boldsymbol{\pi}$  segmentiert die Menge aller Treugeber-Typen vollständig, d. h. es gilt  $\sum_{\pi_j \in \boldsymbol{\pi}} \pi_j = 1$
- $\mathbf{A}$  – beinhaltet für alle Kombinationen vom Spielertypen  $\alpha_{ij}$  Informationszirkulationsbeziehungen, die die Wahrscheinlichkeit angeben, mit der Spielertyp  $i$  die Information über die Defektion des Treuhänders an Spielertyp  $j$  weitergibt. Diese Information kann Spielertyp  $j$  bei Interaktion mit dem selben Treuhänder berücksichtigen und beim Eingehen einer Transaktion entsprechend vorsichtig sein. Die Matrix  $\mathbf{A}$  trägt damit dem Umstand Rechnung, dass Informationen über unehrliche Treuhänder in der Regel nicht global verfügbar sind, sondern sich im Kontext von Beziehungsnetzwerken je nach Intensität der Informationsaustauschbeziehung unterschiedlich schnell verbreiten<sup>9</sup>.
- $w \in [0, 1]$  – Exponentielle Minderung der Payoffs mit fortschreitender Zeit um den Faktor  $w^t$ . Damit wird gewährleistet, dass sich initiales Vertrauen absolut betrachtet mehr lohnt als eine spätere Vergabe von Vertrauen.

Neben dem methodologischen Nutzen der *Formalisierung* von Vertrauensinteraktionen unter Berücksichtigung von Netzwerkstrukturen (Buskens 1995) hat Vicent Buskens (1998) auch den Einfluss lokaler und globaler Netzwerkindikatoren auf den Grad des Vertrauens untersucht:

Er bedient sich bei seiner Analyse zweier unterschiedlicher Methoden: Er nutzt spieltheoretisch-analytische Methoden zur Ermittlung von Gleichgewichten bzw. Lösungen für das Vertrauensproblem in Abhängigkeit von spezifischen Netzwerkstrukturen. Als zweite Methode bedient er sich der Computersimulation, wobei Veränderun-

<sup>9</sup>Die unvollständige, nicht globale Informationsdiffusion erinnert stark an das Paradigma der agentenbasierten Modellierung und Simulation (vgl. Abschnitt 7.3).

gen des Vertrauensgrades in Abhängigkeit von unterschiedlichen Netzwerkstrukturen simuliert werden (ebd.: 268). Der Nachteil der Computersimulation besteht zwar darin, dass man nie ganz sicher sein kann, wirklich alle möglichen relevanten Netzwerkstrukturen betrachtet zu haben. Sie hat im Vergleich zur analytischen Methode jedoch den Vorteil, dass auch sehr komplexe Netzwerkstrukturen untersucht werden können. Zur Validierung der Simulationsmodelle schlägt Buskens vor zu überprüfen, ob die Computersimulation bereits bekannte analytische Resultate reproduzieren kann (ebd.: 286).

Als zentralen Einflussfaktor auf den Grad des Vertrauens betrachtet Buskens (1998: 269f.) den Vorgang der Informationsdiffusion, also der Verbreitung von Informationen innerhalb des Netzwerks. Die Informationsdiffusion hängt ihrerseits von Netzwerkeigenschaften ab (ebd.): Erstens steigt der erreichbare Grad des Vertrauens mit der Dichte des Netzwerkes, zweitens können zentrale Akteure Informationen schneller verbreiten (Degree-Zentralität) und drittens können sogenannte „weak ties“, also einzelne Brücken zwischen in sich stark zusammenhängenden Communities, einen großen Einfluss auf den Vertrauensgrad ausüben.

Ein besonders interessantes Resultat ist der Zusammenhang von situativem Kontext und dem daraus resultierenden unterschiedlichen Einfluss globaler und lokaler Netzwerkstrukturen. So hat Buskens (1998: 285) herausgefunden, dass globale Netzwerkindikatoren insbesondere bei problematischen bzw. komplexen Situationen ausschlaggebender für die Vergabe von Vertrauen sind als lokale Indikatoren: „[...] global network measures have a larger influence on the level of trust a buyer can have in a seller than local network measures if the dyadic trust situation is more problematic“. Buskens (1998: 285f.) bezieht sich bei seinen Betrachtungen auch hier auf ein Käufer-Verkäufer-Beispiel: Bei günstigen und mittelpreisigen Produkten, bei denen eine Überprüfung der Qualität ex ante zwar schwierig, aber nicht unmöglich ist, beeinflussen im Wesentlichen die unmittelbaren (also lokale) Beziehungen des Käufers zu anderen Verkäufern die Kaufentscheidung. Globale Strukturmerkmale sind von Bedeutung, wenn Produkte hochpreisig sind und sich wegen ihrer Komplexität einer Qualitätsüberprüfung weitgehend entziehen.

Als weiteres Ergebnis hat Buskens (1998: 286) herausgefunden, dass nur vier Netzwerkmaßzahlen<sup>10</sup> genügen, um 90% der mit Vertrauen in Verbindung stehenden Effekte zu erklären: Outdegree, Dichte, individuelle Outdegree-Zentralität und die Outdegree-Varianz. Bei homogenen Netzwerken genügen, folgt man Buskens (ebd.), in der Regel die Betrachtung von Dichte und Outdegree.

Zusammenfassend kann konstatiert werden, dass Buskens (1995; 1998) mit seiner spieltheoretischen Betrachtung die Bedeutsamkeit von Iteration und Interaktion sowie den Einfluss von lokalen und globalen Informationsflüssen innerhalb eines Netzwerks

<sup>10</sup>Für Definitionen der Begriffe Outdegree, Dichte, Degree-Zentralität und Degree-Varianz sei hier auf einschlägige Einführungen zur sozialwissenschaftlichen Netzwerkanalyse verwiesen: Weyer (2011); Jansen (2006); Knoke und Yang (2008).

von Treugebern für die Herausbildung von Vertrauen hervorhebt. Diese Überlegungen zu einer interaktiven und prozessförmigen Sichtweise auf Vertrauen sind in Abschnitt 3.3 als forschungsleitende Annahme (FLA 5) zusammengefasst.

### 3.2.5 Kritische Würdigung

Eine Gemeinsamkeit des spieltheoretischen und probabilistischen Vertrauensbegriff besteht darin, dass Vertrauens*beziehungen* in den Vordergrund gerückt werden. Dabei wird Vertrauen nicht wie in Kapitel 2 als Disposition für daran anschließende Handlungen aufgefasst, sondern die rationale Abwägung, eine Vertrauens- oder Missbrauchshandlung zu vollziehen, wird zum Untersuchungsgegenstand erhoben.

Ein wichtiger Unterschied zum probabilistischen Vertrauensbegriff besteht in der Annahme, dass Vertrauenssituationen durch Interdependenz und reziproke Antizipation von Situationsopportunitäten charakterisiert sind. Das bedeutet, dass ein Treugeber bei der Frage, ob es rational ist, zu vertrauen, die Opportunitäten des Treuhänders, ebendieses Vertrauen zu missbrauchen stets in den rationalen Abwägungsprozess mit einbezieht. Gambetta (1988a: 218f.) argumentiert, dass sich eine spieltheoretische Betrachtung von Vertrauen geradezu aufdrängt: „For trust to be relevant, there must be the possibility of exit, betrayal, defection“ – typische spieltheoretische Elemente einer strategischen Interdependenzsituation.

Ein weiterer Unterschied besteht in der prozessförmigen Anlage der spieltheoretischen Sichtweise auf Vertrauen: Den ersten Zug vollzieht der Treugeber. Entscheidet dieser sich dafür, kein Vertrauen zu geben, so endet das Spiel, da dem Treuhänder damit die Möglichkeit genommen wird, das Vertrauen entweder zu bestätigen oder zu missbrauchen. Vergibt der Treugeber Vertrauen, so kann sich im zweiten Zug der Treuhänder entscheiden, das Vertrauen zu missbrauchen oder zu bestätigen. Eine solche Sichtweise scheint, sofern explizite Vertrauenshandlungen erforderlich sind, realistischer zu sein, als die Entscheidung ausschließlich, wie beim probabilistischen Vertrauensbegriff, beim Treuhänder zu verorten.

Den Ausgangspunkt zu spieltheoretischen Überlegungen zu Vertrauen bildet das in Abschnitt 3.2.1 vorgestellte basale Vertrauensspiel. Die spieltheoretische Analyse des basalen Vertrauensspiels hat gezeigt, dass eine Vertrauenshandlung bei einmaliger Begegnung von Treuhänder und Treugeber eigentlich nicht möglich ist. Experimentelle Ansätze, bei denen Probanden mit Situationen konfrontiert wurden, die laut analytischer Spieltheorie eine Vertrauenshandlung unmöglich erscheinen lassen, haben nachgewiesen, dass unter bestimmten Voraussetzungen (situative Bedingungen, individuelle Orientierungen, bestimmte Typ-Konstellationen) dennoch Vertrauen zu Stande kommen kann (Abschnitt 3.2.2). Die im weiteren Verlauf dargestellten iterativen und reputationsbasierten Modelle (Abschnitt 3.2.3, 3.2.4) konnten sogar analytisch zeigen, dass Vertrauen, selbst bei egoistisch handelnden Akteuren, entstehen kann, sofern unterschiedliche Typen in wiederkehrenden Begegnungen aufeinander

treffen. Beim sogenannten iterativen heterogenen Vertrauensspiel (Abschnitt 3.2.4) wurden zur Verbreitung von Reputationsinformationen Netzwerke verwendet, die dem Umstand Rechnung tragen, dass sich Informationen besonders schnell in dichten Netzwerken von einander bekannten Treugebern verbreiten.

Aus der Spielstruktur des basalen Vertrauensspiels und den Überlegungen zu Iteration und Reputation lassen sich zwei wichtige *inhaltliche* Schlussfolgerungen ziehen (ähnlich auch bei Gambetta 1988a: 234):

1. Vertrauen ist anfällig für Zerstörung.
2. Vertrauen mehrt sich bei Nutzung.

Die Anfälligkeit von Vertrauen für Zerstörung ergibt sich daraus, dass spieltheoretische Modelle voraussetzen, dass der Treuhänder nur zum Zug kommt, wenn der Treugeber einen Vertrauensvorschuss leistet. Misstraut der Treugeber bereits im ersten Zug, hat der Treuhänder keine Chance mehr, diesen vom Gegenteil zu überzeugen. Bei wiederkehrenden Begegnungen verfügt der Treugeber über keine neuen Informationen über den Treuhänder, da er ihm unmittelbar misstraut hat. Durch das anfängliche Misstrauen hat somit der Treugeber das Vertrauen bzw. die Chance, überhaupt Vertrauen aufzubauen, bereits mit dem ersten Zug verspielt. Dieses Argument lässt sich umkehren und auf die zweite inhaltliche Erkenntnis, dass sich Vertrauen bei Nutzung mehrt, anwenden: Sofern der Treugeber unmittelbar im ersten Zug vertraut und das Vertrauen bestätigt wird, verfügt er über die Information, dass der Treuhänder prinzipiell vertrauenswürdig ist. Es ist dann rational auch in weiteren Zügen zu vertrauen und den Eindruck der Vertrauenswürdigkeit des Treuhänders zu bestärken. Wird das Vertrauen im weiteren Interaktionsverlauf vom Treuhänder stets bestätigt, so mehrt sich das Vertrauen des Treugebers in den Treuhänder bei der Entscheidung für die Vertrauenswahl des Treugebers im jeweils ersten Zug. Gambetta (1988a: 234) fasst diesen Zusammenhang in der paradoxal klingenden Feststellung (zumindest wenn man nicht den spieltheoretischen Vertrauensbegriff kennt) zusammen, dass „[...]trust is depleted through *not* being used“

### **Spieltheoretische Modellierung von Vertrauen in autonome Technik?**

Abschließend sollen einige Überlegungen angestellt werden, inwiefern sich ein spieltheoretischer Vertrauensbegriff für die Analysen von Vertrauen in autonome Technik eignet. Alle spieltheoretischen Modellierungen von Vertrauen basierten auf Varianten des basalen Vertrauensspiels. Für autonome Technik ist demnach zu prüfen, inwieweit es sinnvoll ist, Vertrauen in autonome Technik als Variante des Vertrauensspiels zu modellieren.

Vorteilhaft ist die gute Operationalisierbarkeit. Im Gegensatz zu den inhaltlich reichhaltigen Konstrukten aus Kapitel 2 kommen spieltheoretische Vertrauensansätze mit einer relativ sparsamen Modellierung aus. Dabei ist es möglich, auf einen breiten

Fundus mathematisch bewiesener Theorien zurückzugreifen und zu berechnen, unter welchen Bedingungen Vertrauen vergeben wird. Weiterhin charakterisieren spieltheoretische Modellierungen zu Vertrauen den interaktiven Charakter von Vertrauenskonstellationen. Hinsichtlich autonomer Technik, die ein zunehmendes Maß an Eigenaktivität und Interaktivität aufweist (s. u. a. Abschnitt 5.3), ist die Berücksichtigung von Interaktion von Vorteil. Eine weitere Eigenschaft des spieltheoretischen Vertrauensbegriffs, die reziproke Antizipation der situativen Opportunitäten des jeweils Anderen, findet zumindest in Form wechselseitiger Erwartungen von Mensch und autonomer Technik (vgl. Kapitel 8) seinen Niederschlag.

Allerdings gibt es vor dem Hintergrund des Untersuchungsgegenstands der vorliegenden Arbeit auch erhebliche Einwände gegen die Verwendung eines *ausschließlich* auf Spieltheorie fußenden Vertrauensbegriffs: Spieltheoretische Modelle unterstellen stets eine *strategische* Interdependenzsituation, die insbesondere dem Treuhänder unterstellt, dass dieser Vertrauen bewusst missbraucht. Diese Annahme erscheint für autonome Technik unrealistisch, da sich diese nicht strategisch gegen ihre Interaktionspartner richtet (vgl. Kapitel 8.1). Weiterhin wird bei der spieltheoretischen Modellierung unterstellt, dass eine explizite Entscheidung zwischen den Handlungsalternativen „vertrauen“ und „misstrauen“ gegeben ist. Bei der Interaktion mit autonomer Technik, die in der Regel auf die Lösung gemeinsamer Aufgaben gerichtet ist, wird eine solche explizite Unterscheidung nicht eingefordert. Ein weiterer Kritikpunkt, der auch gegen den probabilistischen Vertrauensbegriff vorgetragen wurde, ist die auch hier zugrundeliegende bivalente Denkweise. Es existiert nur ein *Entweder-oder*, ein Vertrauen oder Misstrauen. Zuletzt trägt der Rückgriff auf Spieltheorie nicht unmittelbar zu einer soziologischen Tiefenerklärung (vgl. Abschnitt 6) bei. Durch die Reduktion des Vertrauensbegriffs auf eine rationale Entscheidung auf Grundlage von Auszahlungsmatrizen beraubt man sich der Möglichkeit, zu einem Verständnis von Gründen für Vertrauen (oder Misstrauen) zu kommen.

Im abschließenden Abschnitt zu Vertrauen in der Sozialtheorie werden die Erkenntnisse zum probabilistischen und spieltheoretischen Vertrauensbegriff zu einigen forschungsleitenden Annahmen verdichtet, die insbesondere für das Experimentaldesign im empirischen Teil der Arbeit von Belang sind.

### 3.3 Implikationen

Wie zu Anfang des Kapitels zu „Vertrauen in der Sozialtheorie“ benannt, verband sich mit der Beschäftigung mit probabilistischen und spieltheoretischen Vertrauensmodellen die Erwartung, ggf. zu einer im Vergleich zu Kapitel 2 sparsameren und formaleren Modellierung von Vertrauen zu kommen, die dennoch dem Untersuchungsgegenstand der vorliegenden Arbeit angemessen ist. Zumindest für den probabilistischen Vertrauensbegriff hat sich diese Erwartung nicht bestätigt. Die Modellierung von Vertrauen

ist dort zu unterkomplex angelegt, zumindest dann, wenn *Ursachen* für Vertrauen und Misstrauen ergründet werden sollen.

Einige Erkenntnisse aus den spieltheoretischen Modellierungen lassen sich hingegen für die Untersuchung von Vertrauen in autonome Technik nutzen.

Zunächst haben die frühen Arbeiten von Morton Deutsch (1958) gezeigt (vgl. Abschnitt 3.1.1), dass sich Vertrauen in Form von Laborexperimenten untersuchen lässt. Wie ein Experimentaldesign für die Untersuchung von Vertrauen in autonome Technik genau aufgebaut sein muss, lässt sich erst vor dem Hintergrund einer soziologischen Methodologie zur Behandlung autonomer Technik, die Gegenstand von Kapitel 8 ist, beantworten. Diese Überlegung ist in der forschungsleitenden Annahme (FLA 4) zusammengefasst.

#### **FLA 4 (Laborexperimente)**

Vertrauen lässt sich mittels Laborexperimenten untersuchen.

Weiterhin haben die spieltheoretischen Überlegungen gezeigt, dass sich Vertrauen in einem interaktiven Prozess entwickelt. Ermöglicht wird eine solche Sichtweise durch die Durchführung iterierter Spiele, bei denen auch Reputationsmechanismen zum Tragen kommen. Wird als Charakteristikum autonomer Technik unterstellt, dass diese sich insbesondere durch Interaktivität auszeichnet (vgl. dazu insbesondere Abschnitte 5.2 und 5.3), so lässt sich daraus die forschungsleitende Annahme (FLA 5) formulieren, die diese Interaktivität bei der Entwicklung von Vertrauen berücksichtigt.

#### **FLA 5 (Interaktiver Prozess)**

Vertrauen entwickelt sich in einem interaktiven Prozess.

Eine letzte forschungsleitende Annahme lässt sich aus der reziproken Antizipation von Opportunitäten des Spielpartners herleiten. Wie weiter oben argumentiert, ist zwar die Unterstellung einer *strategischen* Interdependenzsituation für autonome Technik unangemessen, dennoch existieren zwischen den Interaktionspartnern (Mensch, autonome Technik) wechselseitige Erwartungen, wie in Abschnitt 8.1 noch ausführlich dargelegt werden wird. Diese wechselseitigen Erwartungen und Handlungsmuster, so ist in (FLA 6) ausgedrückt, stehen im Zusammenhang mit Vertrauen.

#### **FLA 6 (Erwartungen)**

Vertrauen beinhaltet die Unterstellung von Erwartungen und Handlungsmustern an den Interaktionspartner.

Wie sich dieser Zusammenhang darstellt, muss im Zuge der empirischen Überlegungen je Fall geklärt und ausgewertet werden.

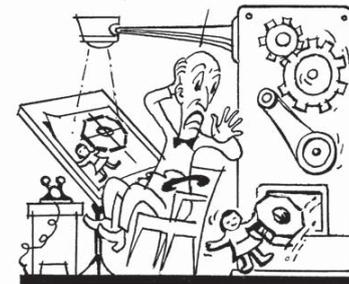
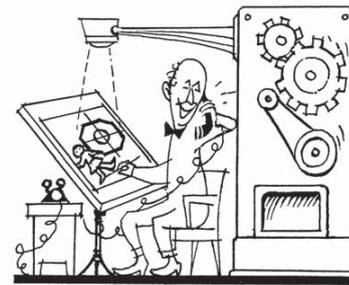
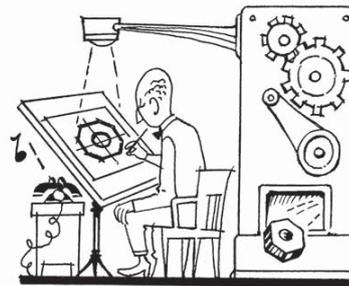


## 4 Vertrauen in der Automationsforschung

Die vorangegangenen Kapitel 2 und 3 haben sich mit Vertrauen als ein allgemeines Konzept ohne direkten Bezug zu einem spezifischen Anwendungsfeld beschäftigt. In den kritischen Betrachtungen zu den unterschiedlichen Vertrauenskonzepten wurde zwar immer wieder auf den Gegenstand *autonome Technik* hingewiesen und überprüft, ob und wie sich bestimmte Konzepte eignen. Allerdings erfolgten diese Betrachtungen vornehmlich auf konzeptioneller Ebene und mit Hinblick auf die in Kapitel 8 vorgeschlagene Methodologie einer genuin soziologischen bzw. handlungstheoretischen Betrachtung der Interaktion von Mensch und autonomer Technik.

Die Funktion dieses Kapitels besteht vornehmlich darin, einige ausgewählte *empirische* Ergebnisse zu Vertrauen aus der Automationsforschung aufzuarbeiten und systematisch zusammenzufassen. Weiterhin sollen dabei mögliche Lücken identifiziert werden, die mit dem eigenen Untersuchungsansatz weitgehend ausgeräumt werden sollen.

Im Kern befasst sich dieses Kapitel mit vier unterschiedlichen Studien zu Vertrauen. Ein besonderes Augenmerk soll dabei neben den inhaltlichen Ergebnissen auch auf das Experimentaldesign gerichtet werden, um daraus Anregungen für die eigenen Untersuchungen zu erhalten. Wie in den vorherigen Kapiteln gezeigt und in Form von (FLA 4) zusammengefasst, eignen sich experimentelle Settings, um Vertrauen zu untersuchen. In diesem Kapitel soll im Unterschied zu den spieltheoretischen Methoden mit relativ artifiziellen Modellannahmen (vgl. Abschnitt 3.2) insbesondere herausgestellt werden, wie Vertrauen anhand realitätsnäherer Laborexperimente untersuchbar ist.



**Abbildung 4.1:** Higher orders of automaticity: Automationslevel  $A_6$  bis  $A_9$  bei Amber und Amber (1962: 8)

### **Einschub: Autonome Technik vs. automatisierte Systeme**

Zwischen Automation und autonomer Technik existierenden weitreichende Schnittmengen (die der Grund sind, warum Vertrauen in *Automation* in diesem Kapitel überhaupt behandelt wird). Autonome Technik und ihre Charakteristika, die aus einer genuin soziologischen Perspektive in Kapitel 5 besprochen werden, sind als Untergruppe automatisierter Systeme aufzufassen. Zur Klassifizierung von Typen der Automation existieren verschiedene Taxonomien.

Einer der ältesten Systematisierungsansätze findet sich in den Arbeiten von Bright (1958), der sich mit Automatisierung aus Managementperspektive beschäftigt hat, sowie die „Anatomy of automation“ von Amber und Amber (1962).

Bright (1958) teilt eine komplexe Aufgabe in mehrere Teilschritte auf, denen dann einer von insgesamt 18 Mechanisierungsleveln zugeordnet wird, beginnend mit „Manual by hand“ (0) über „Measures characteristic of work“ (9) bis „Anticipates action required and adjusts to provide it“ (17). Amber und Amber (1962: 2f.) verfolgen einen gänzlich anderen Ansatz und klassifizieren die Automationsgrade anhand von *menschlichen* Attributen. Die einfachste Art der Automatisierung beginnt mit „energy“ (mechanische Energie,  $A_1$ ) und endet mit „higher orders of automaticity“, die „learning“ (lernen,  $A_6$ ), „reasoning“ (Folgern,  $A_7$ ), „creativity“ (kreativ sein,  $A_8$ ) und schließlich „commands others“ (Menschen und Maschinen Befehle erteilen,  $A_9$ ) umfassen (vgl. Abbildung 4.1).

Neuere Ansätze wie die „Levels of automation“ von Parasuraman u. a. (2000: 288) richten den Fokus auf Entscheidungsprozesse und definieren zehn Automationslevel, die getrennt für vier funktionale Dimensionen, die Stationen eines typischen Automationsprozesses abbilden, angegeben werden können. Es wird unterschieden (ebd.: 287) zwischen Informationsbeschaffung („information acquisition“), Informationsanalyse („information analysis“), Entscheidungs- und Handlungsselektion („decision and action selection“) sowie Handlungsvollzug („action implementation“). Die Automationslevel starten bei Level 1 mit „The computer offers no assistance: human must take all decisions and actions“ und enden bei Level 10 mit „The computer decides everything, acts autonomously, ignoring human“. (ebd.)

Mit der „human-automation collaboration taxonomy“ (HACT) haben Cummings und Bruni (2009) ein Klassifizierungsschema vorgelegt, das den Schwerpunkt auf die Zusammenarbeit von Mensch und Automation legt. Dort werden drei typischen Rollen (ebd.: 440) identifiziert: „moderator“, „generator“ und „decider“. Die Aufgabenverteilung zwischen Mensch und Technik kann sich für diese Rollen auf unterschiedlichen Leveln (ebd.: 441f.) bewegen, von  $-2$  (vollständig automatisiert) bis  $+2$  (vollständig in menschlicher Hand). Es können dann Tupel angegeben werden, die Systeme entlang dieser drei Rollen charakterisieren.

Hinsichtlich der Einordnung des Gegenstandsbereiches der vorliegenden Arbeit muss zunächst festgestellt werden, dass die Automationsgrade, so elaboriert die unter-

schiedlichen Taxonomien auch sein mögen, stets als *objektive* Eigenschaft des Systems bzw. des Interaktionsprozesses konzipiert sind. Demnach werden die Grade der Automation von demjenigen festgelegt, der sich mit der Architektur des automatisierten Systems befasst. Diese Vorstellung widerspricht einer strikt akteurszentrierten und attributionstheoretischen Perspektive, wie sie zwecks Anschlussfähigkeit an soziologische Ansätze in der vorliegenden Arbeit eingenommen wird (vgl. Kapitel 8). Sofern dennoch der Versuch unternommen werden soll, *autonome Technik* einzuordnen, dann dürfte sich diese im Schema von Parasuraman u. a. (2000) auf Ebene der Informationsanalyse, Handlungsselektion und -ausführung oberhalb von Level 5 und bei Cummings und Bruni (2009) zwischen Level -2 und 0 für die Rollen des Entscheiders und Generators bewegen.

## 4.1 Vertrauen und Funktionsallokation

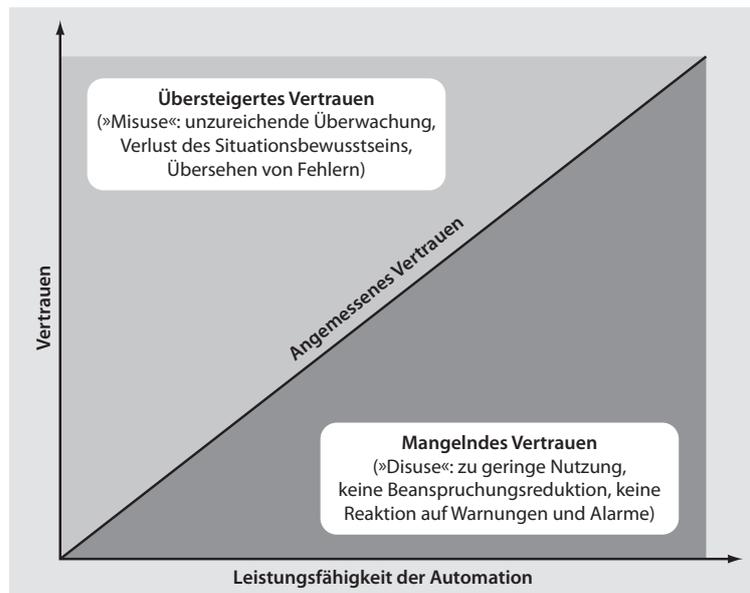
Einer der zentralen Untersuchungsgegenstände der Automationsforschung ist die sogenannte *Funktionsallokation*. Die Funktionsallokation befasst sich mit der Frage, welche Funktionen von der Automation und welche Funktionen von einem menschlichen Operateur übernommen werden. Der Begriff „Allokation“ geht insbesondere auf die Arbeiten von Jordan (1963) und Kantowitz und Sorkin (1987) zurück und ist im englischen Sprachgebrauch unter dem Label „allocation of functions“ zu finden. Die Frage nach der Allokationsstrategie ist immer eine Frage danach, welche (Teil-)aufgaben eines komplexen Systems an den menschlichen Operateur und welche an die Automation delegiert werden. In der Studie von Lee und Moray (1994) haben die Operateure beispielsweise die Möglichkeit, selbst über die Funktionsallokation und damit über ihre Allokationsstrategie zu entscheiden. Eine Übereinstimmung der Vorstellung des menschlichen Operateurs über den aktuellen Allokationszustand, also die momentan eingestellte Aufgabenverteilung zwischen Mensch und Automation, mit der tatsächlichen Verteilung ist ein wichtiger Faktor für funktionierende Mensch-Maschine-Interaktion. Sollten die Vorstellung von der tatsächlichen Allokation differieren, herrscht „mode confusion“ (Weyer 1997: 247). Vor allem in Gefahrensituationen, in denen schnell reagiert werden muss, können Unklarheiten über den Automationsmodus fatale Folgen haben.

Studien zu Vertrauen in Automation deuten auf einen engen Zusammenhang zwischen Allokationsverhalten (wann wird die Automation genutzt) und Vertrauen hin (vgl. Muir und Moray 1996: 454). Lee und Moray (1994: 177) identifizieren Vertrauen als wichtige Einflussgröße, um das Allokationsverhalten vorauszusagen. Weiter weisen Bisantz und Seong (2001) auf einen Zusammenhang von Vertrauen, Allokationsverhalten und die Angabe von Gründen im Falle von Fehlfunktionen hin. Allokationsverhalten kann gemäß der Definition in Abschnitt 2.1.5 als auf Vertrauen bezogenes Verhalten („trust-related behaviour“) identifiziert werden.

Die im Abschnitt 4.3 behandelten Studien bzw. Experimente sind so aufgebaut, dass diese den Zusammenhang von Allokationsverhalten und Vertrauen systematisch untersuchen und unterschiedliche Aspekte des Vertrauensproblems in der Automationsforschung (vgl. nächster Abschnitt) in den Blick nehmen.

## 4.2 Das Vertrauensproblem in der Automationsforschung

Eine kompakte Zusammenfassung zur Verortung des Vertrauensproblems in der Automationsforschung findet sich bei Dietrich Manzey (2012). Manzey geht davon aus, dass eine Automation eine angebbare Leistungsfähigkeit<sup>1</sup> besitzt. Ein *angemessener Vertrauenslevel* (vgl. auch Lee und See 2004: 55) orientiert sich an den wesentlichen Leistungsmerkmalen der Automation wie z. B. Zuverlässigkeit, Verständlichkeit und Nützlichkeit (Manzey 2012: 339). Lee und See (2004) bezeichnen diesen Zusammenhang als korrekte Kalibrierung des Vertrauenslevels: „[...] correspondence between a person’s trust in the automation and the automation’s capabilities“ (ebd.: 55). Der Zusammenhang von Leistungsfähigkeit und Vertrauenslevel ist schematisch in Abbildung 4.2 dargestellt.



**Abbildung 4.2:** Das Problem des Vertrauens in der Automationsforschung  
Quelle: Manzey (2012: 339)

Aus der Definition des angemessenen Vertrauenslevels ergeben sich unmittelbar zwei Teilprobleme mit je unterschiedlichen Implikationen: Ist der Vertrauenslevel höher als

<sup>1</sup>In der englischsprachigen Automationsliteratur wird Leistungsfähigkeit (automation capability) auch mit „trustworthiness“ (Vertrauenswürdigkeit) gleichgesetzt (vgl. Lee und See 2004: 55). Da allerdings Verwechslungsgefahr zu den spieltheoretischen und probabilistischen Konzepten der Vertrauenswürdigkeit aus Kapitel 3 besteht, erscheint der von Manzey gewählte Begriff – Leistungsfähigkeit – passender.

die Fähigkeiten der Automation, so liegt *übersteigertes Vertrauen* vor. Ist hingegen der Vertrauenslevel niedriger als die Leistungsfähigkeit der Automation, kann von *mangelndem Vertrauen* gesprochen werden.

### Übersteigertes Vertrauen

Übersteigertes Vertrauen subsumiert Manzey (2012: 339) unter dem Label „Misuse“, das u. a. „unzureichende Überwachung“, „Verlust des Situationsbewusstseins“ und das „Übersehen von Fehlern“ bedeuten kann. Folgt man dieser Charakterisierung, sieht Manzey übersteigertes Vertrauen vor allem als ein Problem des Operateurs an, im Gegensatz etwa zu Inagaki und Itoh (2013: 2), die dafür auch System-Design-Fehler verantwortlich sehen.

Die Beschäftigung mit übersteigertem Vertrauen, im Automationsdiskurs meist unter dem Label „Complacency-Effekt“ subsumiert, geht auf die Arbeiten von Moray und Inagaki (2000) zurück.

In jüngster Zeit haben sich vor allem Inagaki und Itoh (2013) mit dem Problem des übersteigerten Vertrauens (engl. *overtrust*) in Bezug auf moderne Fahrerassistenzsysteme für das Auto beschäftigt. Folgt man den Autoren, so bezeichnet übersteigertes Vertrauen „[...] an incorrect situation diagnostic decision claiming that the object is trustworthy when it is actually not“ (ebd.: 2). Dabei liegt ebenso wie der Übersetzung „Leistungsfähigkeit der Automation“ (vgl. Manzey 2012: 339) die Idee zugrunde, dass sich eindeutig erkennen lässt, ob ein Objekt vertrauenswürdig ist bzw. ob eine Automation eine definierte Leistung erbringt.

Inagaki und Itoh (2013: 2) schlagen vor statt der im englischen Sprachraum gebräuchlichen Formulierung „automation complacency“ den Begriff „overreliance“ oder „overtrust“ zu verwenden. Die Automation nicht zu überwachen, so die Autoren, wird häufig mit Complacency (Selbstzufriedenheit) gleichgesetzt, was aber nicht zwangsläufig der Fall sein muss, da z. B. auch parallel laufende, hochprioräre Aufgaben ein Grund für die mangelnde Überwachung sein können. Übersteigertes Vertrauen ist kein Problem, das ausschließlich auf den Operateur zurückgeführt wird, sondern kann auch durch eine fehlerhaft gestaltete Benutzerschnittstelle entstehen. In diesem Fall könnte man dem Problem durch Verbesserungen der Mensch-Maschine-Schnittstelle begegnen (ebd.: 5).

Inagaki und Itoh (2013: 2f.) entwickeln ein Framework, um übersteigertes Vertrauen präzise beschreiben zu können. Dabei unterscheiden sie insgesamt drei Klassifizierungskriterien für übersteigertes Vertrauen:

- Vertrauensdimension, z. B. Leistungsindikatoren oder ein gewisses Prozessverständnis
- Zielobjekt, auf das sich übersteigertes Vertrauen bezieht, z. B. Hardware oder Software
- Möglichkeiten der Beobachtung

Zumindest die ersten beiden Kriterien erinnern an die grundsätzlichen Überlegungen zu Facetten des Vertrauens aus Abschnitt 2.2 – auch dort wird zwischen unterschiedlichen Vertrauensobjekten und Vertrauensdimensionen unterschieden. Die Überlegungen von Inagaki und Itoh beziehen sich allerdings speziell auf Vertrauen in Automation und setzen ein festes Vertrauenssubjekt, den Bediener der Automation, voraus. *Möglichkeiten der Beobachtung* kommen in keinem der in den Einführungskapiteln vorgestellten Vertrauenskonzepte vor. Inagaki und Itoh (2013: 3) unterscheiden in Bezug auf fortgeschrittene Fahrerassistenzsysteme (*advanced driver assistance systems*, ADAS) zwischen Systemen, die beim normalen Fahren unterstützen, und solchen, die in Notfallsituationen eingreifen. Systeme in Notfallsituationen können kaum vom Fahrer beobachtet werden, da diese äußerst selten aktiv werden. Systeme, die selten aktiv werden, verhindern auf Seiten des Fahrers die Herausbildung eines entsprechenden mentalen Modells dieser Systeme (ebd.: 3).

### **Mangelndes Vertrauen**

Der Gegenpart zu übersteigertem Vertrauen ist mangelndes Vertrauen und umfasst nach Manzey (2012: 339) die Aspekte der zu geringen Automationsnutzung, fehlenden Beanspruchungsreduktion und der fehlenden Reaktion auf Warnungen und Alarmer. Mangelndes Vertrauen, so zumindest die Sichtweise der Automationsforschung, manifestiert sich in einem Allokationsverhalten, das der ursprünglich angedachten Aufgabe der Automation, nämlich Arbeitserleichterung für den Operateur, nicht gerecht wird. Auf die Nutzung der Automation wird im Falle des mangelnden Vertrauens leichtfertig verzichtet – zu Gunsten einer schlechteren Performance des Gesamtsystems. In der Studie von Dzindolet u. a. (2003), die noch ausführlich in Abschnitt 4.3.3 vorgestellt wird, wurde nachgewiesen, dass Operateure trotz offensichtlich schlechterer Leistung eher den eigenen Entscheidungen als den Entscheidungen eines automatischen Entscheidungshelfers vertrauen: „Even among participants provided with feedback that their aid’s performance was far superior to their own, the majority (81 %) chose to rely on their own decisions rather than on the decisions of the automated aid“ (ebd.: 703).

Manzey (2012: 340) sieht mangelndes Vertrauen vor allem bei solchen Systemen als Problem an, bei denen die Operateure selbst über die Nutzung und Nicht-Nutzung entscheiden können. Meist bricht das Vertrauen, so Manzey weiter, auf Grund einzelner Automationsfehler zusammen, so dass nachfolgend z. T. vollständig auf eine Automationsnutzung verzichtet wird. Manzey betont, dass insbesondere bei automatischen Alarmsystemen, die den Operateur bei der Erkennung kritischer Systemzustände unterstützen sollen, mangelndes Vertrauen ein erhebliches Problem darstellt. Dabei werden zwei Aspekte unterschieden: Erstens sollte das Alarmsystem alle kritischen Zustände auch tatsächlich erkennen, und zweitens sollte beim Auslösen eines Alarms auch tatsächlich ein kritischer Systemzustand vorliegen (vgl. ebd.). Bei diesen

beiden Anforderungen handelt es sich um z. T. konfligierende Ziele. Eine möglichst vollständige Erkennung kritischer Systemzustände wird mit einer höheren Sensitivität der Alarmsysteme erkaufte, die wiederum dazu führen kann, dass zunehmend falsche Alarme ausgelöst werden, die ihrerseits zu einem abnehmenden Vertrauen in die Automation führen, und damit zu einer zu geringen Automationsnutzung beitragen.

Parasuraman und Riley (1997) haben sich in ihrem Aufsatz „Humans and Automation: Use, Misuse, Disuse, Abuse“ ausführlich mit dem Problem des mangelnden Vertrauens und der damit einhergehenden zu geringen Automationsnutzung befasst. Folgt man den Autoren, so ist der Hauptgrund für mangelnde Automationsnutzung die Auslösung falscher Alarme (ebd.: 230). Ausgehend von der Diagnose, dass mangelndes Vertrauen bei Alarmsystemen auf Grund des Problems falscher Alarme weit verbreitet ist, skizzieren die Autoren mögliche Strategien im Umgang mit falschen Alarmen (ebd.: 244). Ein möglicher Lösungsansatz könnte dabei der Rückgriff auf sogenannte „Likelihood-Alarme“ sein, die zusätzlich angeben, für wie wahrscheinlich die Automation das Vorliegen eines tatsächlich kritischen Zustands einschätzt. Systematisch erforscht wurde die Angabe derartiger Klassifizierungswahrscheinlichkeiten in der Studie von Bisantz und Seong (2001), die ausführlich in Abschnitt 4.3.2 behandelt wird.

Ein weiteres Problem, das Parasuraman und Riley (1997: 244f.) ansprechen, bezieht sich auf den Umstand, dass kritische Bedingungen für die Auslösung eines Alarms äußerst selten vorliegen. Gepaart mit der Anforderung möglichst keinen tatsächlich kritischen Systemzustand zu verpassen, ergibt sich trotz hoher Sensitivität des Erkennungssystems eine Vielzahl von Fehlalarmen. Der fatale Schluss, den Operateure aus einer Vielzahl von Fehlalarmen ziehen, ist in der Automationsforschung als „Cry-Wolf-Effekt“ bezeichnet (vgl. Breznitz 1983) und bezieht sich darauf, dass Alarmsignale fortan ignoriert werden und bei einem tatsächlichen Alarm keine Aktionen seitens des Operateurs eingeleitet werden (vgl. auch Manzey 2012: 340).

### 4.3 Automationsstudien zu Vertrauen

In diesem Abschnitt werden insgesamt vier Fallstudien vorgestellt, die sich mit Vertrauen aus Perspektive der Automationsforschung beschäftigen. Die Auswahl lässt sich dadurch begründen, dass diese Studien oft zitiert werden und zudem theoretisch und methodisch innovative Ansätze enthalten, von denen sich auch die vorliegende Arbeit hat inspirieren lassen. Die Studien beschäftigen sich mit dem Zusammenhang von Vertrauen und Allokationsverhalten im Kontext des Vertrauensproblems, das im vorherigen Abschnitt 4.2 beschrieben wurde.

Die erste Studie, ein Klassiker der Automationsforschung von Lee und Moray (1994), untersucht die Steuerung einer Pasteurierungsanlage mit unterschiedlichen Subsys-

temen und fragt nach dem Zusammenhang von Allokationsverhalten und Vertrauen. Nach Angabe der Autoren handelt es sich um die erste Laborstudie zur Untersuchung des Zusammenhangs von Vertrauen und Automationsnutzung in der Leitstandsarbeit (vgl. ebd.: 154).

Die zweite Studie von Bisantz und Seong (2001) führt laborexperimentelle Untersuchungen durch, bei denen Probanden am Computer feindliche Flugzeuge klassifizieren sollen. Die Autoren versuchen mit den Experimenten einen Zusammenhang zwischen Gründen für Fehlfunktionen, Allokationsverhalten und Vertrauen herzustellen.

Die Studie von Dzindolet u. a. (2003) nimmt insbesondere Vertrauensentwicklungen in den Blick und schließt damit an die bereits zuvor ausgeführte Annahme (FLA 5) in Kapitel 3 an. Bei dem dort durchgeführten Laborexperiment besteht die Aufgabe der Probanden darin, zusammen mit einem automatischen System zu entscheiden, ob sich auf bestimmten Fotos versteckte Soldaten befinden. In den Blick genommen werden dabei anfängliche Vertrauenslevel, der Einfluss von Feedback-Mechanismen und der Einfluss von Gründen für Fehlfunktionen – dies verweist auf Informationsvertrauen und intentionales Vertrauen in der Nomenklatur von Nooteboom aus Abschnitt 2.2. Die zuletzt vorgestellte Studie von Grabowski und Sanborn (2003) ist methodisch besonders interessant, da sie Aspekte des Vertrauens an einem *Realsystem* untersucht. Gegenstand der Untersuchung ist ein kürzlich neu eingeführtes automatisiertes System an Bord eines Ölfrachters.

### 4.3.1 Steuerung einer Pasteurierungsanlage

Die Studie von Lee und Moray aus den 1990er Jahren befasst sich mit „[...] the role of trust in mediating human-machine relationships in supervisory control situation“ (1994: 154). Den Autoren geht es in ihrer Studie um den Zusammenhang von Vertrauen, Selbstvertrauen und Allokationsstrategie. Die Operateure haben die Möglichkeit, selbst über die Funktionsallokation und damit über ihre Allokationsstrategie zu entscheiden.

Im Rahmen von Laborexperimenten bestand die Aufgabe von Probanden darin, eine simulierte Pasteurierungsanlage für Orangensaft zu steuern (s. Abbildung 4.3).

#### Versuchsaufbau und Experimentaldesign

Die Steuerungsaufgabe der Probanden besteht darin, eine aus mehreren Subsystemen bestehende Pasteurierungsanlage zu steuern. Für die jeweiligen Subsysteme kann von den Probanden eine Allokationsstrategie gewählt werden: jeweils manuell oder automatisch. Insgesamt führten Lee und Moray (1994: 161) 28 Versuchsläufe à zwei Stunden an zwei Tagen mit insgesamt zwölf Probanden durch.

Die Probanden werden dabei vor einen Steuerungsbildschirm gesetzt, über den sie verschiedene Informationen zum Pasteurierungsprozess erhalten. Für die jeweiligen



Am zweiten Tag, der Experimentphase, konnten sie ebenfalls nach Belieben steuern. Nach einigen Versuchsläufen traten (künstlich induzierte) Fehler auf. Je nach Gruppe (s. o.) geschah dies zunächst im manuellen, dann im automatischen Modus, oder umgekehrt. Die Idee von Lee u. Moray bestand vermutlich<sup>2</sup> darin, herauszufinden, ob es abhängig von Fehlertypen unterschiedliche Einflüsse auf die Vertrauenseinschätzung („trust“) und das Selbstvertrauen<sup>3</sup> („self-confidence“), also die Einschätzung der eigenen Fähigkeiten, gibt. Die Vertrauensmessung, die nicht näher beschrieben ist, rekurriert sehr wahrscheinlich auf die Skala von Muir (1987), die zumindest in einer früheren Veröffentlichung der beiden Autoren (vgl. Lee und Moray 1992) als Grundlage genannt wird.

### Ergebnisse

Trotz einiger problematischer Aspekte hinsichtlich des Experimentaldesigns, auf die ich im anschließenden Abschnitt eingehen werde, haben Lee und Moray interessante Ergebnisse erzielt:

Zentrales Ergebnis ist der von Lee und Moray (1994: 176) modellierte Zusammenhang zwischen Automationsnutzung, Vertrauen und Selbstvertrauen (4.1).

$$\%Auto(t) = \underbrace{\phi \cdot \%Auto(t-1)}_{\textcircled{1}} + \underbrace{A_1 \cdot (T - SC)(t)}_{\textcircled{2}} + \underbrace{A_2 \cdot IB}_{\textcircled{3}} + \underbrace{a(t)}_{\textcircled{4}} \quad (4.1)$$

Die von den Autoren gewählte Formulierung des gefundenen Kausalzusammenhangs als Differentialgleichung ist nichts anderes als eine Zusammenfassung der qualitativen Ergebnisse. Verbalisiert bedeutet die Differentialgleichung, dass die prozentuale Automationsnutzung  $\%Auto(t)$  (gemessen an der Gesamtdauer des jeweiligen Experimentdurchlaufs) zum Zeitpunkt  $t$  abhängt von ...

- ① ... der Automationsnutzung zum Zeitpunkt  $t - 1$  gewichtet mit der Abhängigkeit  $\phi$ , die den Einfluss vergangener Automationsnutzungen auf die aktuelle Allokationsentscheidung modelliert,
- ② ... der Differenz zwischen Vertrauen  $T$  und Selbstvertrauen  $SC$  und dem Einfluss  $A_1$  dieser Differenz,
- ③ ... dem individuellen Bias ( $IB$ ) in Richtung manueller Steuerung und dem Einfluss dieses Bias  $A_2$ ,
- ④ und einer Störgröße  $a(t)$ .

<sup>2</sup>„Vermutlich“, da die Autoren den Einfluss der Reihenfolge der Fehlertypen, trotz ausführlicher Beschreibung und Ankündigung nicht ausgewertet haben.

<sup>3</sup>Obwohl *Selbstvertrauen* ( $SC$ ) als wichtiges Konstrukt in der Studie verwendet wird, sind die zugehörigen Items nirgendwo erklärt. Self-confidence sei „anticipated performance during manual control“ (Lee und Moray 1994: 154). Vermutlich wurden die Probanden im Modus der manuellen Steuerung gefragt, wie sie ihre eigene Performanz einschätzen.

Das Kernergebnis der von Lee und Moray (1994: 177) gefundenen Zusammenhänge ist der Einfluss der Differenz von Vertrauen und Selbstvertrauen ② auf das Allokationsverhalten: „The results provide convincing evidence that operators’ allocation policy depends on the difference between trust and self-confidence“ (ebd.). Ein weiteres Ergebnis ist der Einfluss der vergangenen Automationsnutzung ① auf das aktuelle Allokationsverhalten: „[...] the use of automatic controllers depends [...] on the past use of the automatic controllers“ (ebd.). Schließlich fanden die Autoren heraus, dass die Probanden sich hinsichtlich ihres individuellen Bias in Richtung manueller Steuerung ③ unterschieden: „The data show that [...] the use of automatic controllers [...] also depends upon individual differences between operators“ (ebd.).

### **Kritische Würdigung / Lessons learned**

Das Experimentaldesign von Lee und Moray (1994) ist insofern inspirierend, als es eine Möglichkeit eröffnet, Vertrauen in Automation in einer kontrollierten Laborumgebung zu untersuchen. Damit bestätigt sich noch einmal die Annahme von (FLA 4), dass sich Vertrauen im Rahmen von Laborexperimenten untersuchen lässt.

Einige methodische Probleme wie die fehlende Beschreibung der verwendeten Vertrauens- und Selbstvertrauenskonstrukte sind bereits angeklungen. Weitere Probleme ergeben sich aus der geringen Probandenzahl: Z. T. werten die Autoren Einzelfälle (ebd.: 168f.) aus und generalisieren diese zu einem formal-mathematischen Zusammenhang (ebd.: 176). An der Generalisierung ist grundsätzlich nichts zu bemängeln, allerdings versäumen es die Autoren darauf hinzuweisen, dass der formulierte quantitative Zusammenhang zwar plausibel aus den qualitativen Beobachtungen folgt, dass dieser aber noch in größeren Settings untersucht werden müsste. Auch wurde keine systematische Auswertung hinsichtlich der Reihenfolge der Fehlertypen vorgenommen. Es findet sich weder ein Verweis, dass diese einen Einfluss hat, noch dass sie keinen hat. Hinzu kommt, dass die gebildeten Gruppen mit nur sechs Probanden ohnehin keine haltbaren Folgerungen auf die Reihenfolge von Fehlern zulassen würden.

Aus soziologischer Perspektive ist zu konstatieren, dass die Automationsstudie von bivalentem Denken geprägt ist: *Entweder* übernimmt die Automation die *vollständige* Kontrolle über ein Teilsystem, *oder* der Operateur übernimmt die *vollständige* Kontrolle. Die Kritik, die in extenso auch schon für den probabilistischen Vertrauensbegriff (vgl. Abschnitt 3.1.3) ausgeführt wurde, trifft auch auf diese Studie zu. Im Gegensatz zu den spieltheoretischen und auch probabilistischen Modellen wird Vertrauen nicht als Vertrauensentscheidung modelliert, sondern als Einflussgröße, die auf Vertrauen bezogenes Verhalten (vgl. Abschnitt 2.1.5), in dem Fall das Allokationsverhalten, beeinflusst.

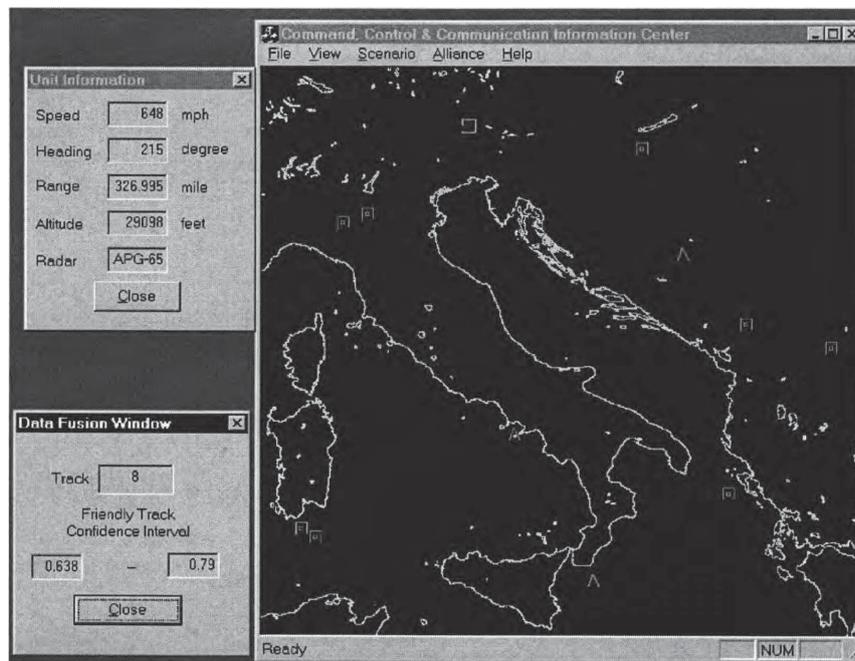
Echtes „Zusammenhandeln“ im Sinne von verteilter Handlungsträgerschaft (vgl. Kapitel 5.3) ist in dem Versuchsaufbau nicht vorgesehen.

### Zentrale Ergebnisse von Lee und Moray (1994)

- Die Differenz von Vertrauen und Selbstvertrauen beeinflusst das Allokationsverhalten des Operateurs. (177)
- Die Automationsnutzung hängt von vorheriger Nutzung und dem Auftreten von Fehlern ab. (177)
- Operateure haben einen Bias zu manueller Kontrolle. (177)
- Insbesondere beim ersten Kontakt mit der Automation hängt die Automationsnutzung von der individuellen Einstellung der Operateure ab. (177)

### 4.3.2 Klassifizierung feindlicher Flugzeuge

Die Laborexperimente von Bisantz und Seong (2001) untersuchen den Einfluss von Erklärungen und Gründen für Fehler auf das Vertrauen in automatisierte Systeme, genauer: „automated decision-aids“. Erklärungen als Element des Vertrauens sind mit dem Typ des Informationsvertrauens (vgl. Tabelle 2.1 aus Abschnitt 2.2.3) zu identifizieren. Die Aufgabe der Probanden besteht darin, Funkkontakte auf einer Radarkarte als feindlich oder friedlich zu klassifizieren. Die Probanden haben die Möglichkeit, unterschiedliche Informationsfester kurz einzublenden, die ihnen Auskunft über den zu klassifizierenden Kontakt geben (vgl. Abbildung 4.4).



**Abbildung 4.4:** Klassifizierung feindlicher Flugzeuge  
Quelle: Bisantz und Seong (2001: 89)

Hinsichtlich des Experimentaldesigns lehnt sich die Studie an Lee und Moray (1994) an, allerdings mit einer etwas anderen Fragestellung (Einfluss von Fehlergründen) und einer anderen Aufgabe für die Probanden (Klassifizierung von Funkkontakten vs. Steuerung einer Pasteurierungsanlage).

### Experimentaldesign

Das Versuchsszenario besteht aus einer Radarkarte, auf der Funkkontakte eingezeichnet sind, von denen nicht bekannt ist, ob diese feindlich oder friedlich sind. Die Probanden können diese Kontakte via Maus und Tastatur klassifizieren. Sie haben dabei die Möglichkeit, Informationsfenster als Hilfe in Anspruch zu nehmen, die sich nach fünf Sekunden wieder automatisch ausblenden. Es stehen zwei Hilfsfenster zur Verfügung, von denen das eine ausschließlich Informationen anzeigt und das andere als Entscheidungshelfer zu Rate gezogen werden kann:

- *Information Window*: Enthält Informationen über den Kontakt (z. B. seine Geschwindigkeit, Richtung, Funkfrequenz, ...)
- *Data Fusion Window / Decision Aid*: Zeigt ein Konfidenzintervall für die Wahrscheinlichkeit an, mit der es sich bei der betrachteten Einheit um eine friedliche handelt.

Zu Anfang des Experiments wurden zunächst drei Versuchsgruppen gebildet, die hinsichtlich der Fehlerursachen des „Decision Aid“-Systems unterschiedlich instruiert wurden (Bisantz und Seong 2001: 89): die Sabotagegruppe, die Nichtintentionalitätsgruppe und die Kontrollgruppe. Der *Sabotagegruppe* wurde anfänglich erklärt, dass der Entscheidungshelfer im Falle einer Fehlfunktion (z. B. die Anzeige eines fehlerhaften Konfidenzintervalls) sabotiert wurde. Die *Nichtintentionalitätsgruppe* wurde instruiert, dass ein auftretender Fehler immer auf ein Hardwareproblem zurückgeführt werden kann und keine bewusste Sabotage im Spiel ist (deswegen nicht intentional). Die dritte Gruppe, die *Kontrollgruppe*, wurde nicht über Fehlerursachen aufgeklärt. Es wurden sechs Durchläufe à 20 Minuten durchgeführt, in denen zwischen 37 und 50 Funkkontakte identifiziert wurden (ebd.: 89). In den ersten zwei Durchläufen lieferte der Entscheidungshelfer die korrekten Wahrscheinlichkeiten, im dritten Durchlauf wurden bewusst Fehler eingebaut, in den letzten drei Durchläufen arbeitete der Entscheidungshelfer wieder korrekt.

Nach Auftreten der Fehler, die in den ersten zehn Minuten des dritten Durchlaufs auftraten, wurden die Probanden (mit Ausnahme der Kontrollgruppe) über das Auftreten des Fehlers und seine Behebung informiert. Das Auftreten des Fehlers wurde laut Bisantz und Seong (2001: 90) den Probanden mitgeteilt, „[...] because the intent of the experiment was not to assess fault detection, but rather operator responses to a known fault“.

Ein weiteres wichtiges Detail beim Experimentaldesign besteht darin, dass die Probanden nur 1,5mal so viele Hilfsfenster öffnen konnten wie Funkkontakte in einem Durchlauf zu klassifizieren waren. Damit sollte untersucht werden, ob sich das Verhältnis der Nutzung des Informationsfensters („Information Window“) zum Entscheidungshelfer („Decision Aid“) nach Auftreten von Fehlern zu Gunsten des Informationsfensters verschiebt (ebd.: 90).

In Anschluss an den ersten, dritten und sechsten Durchlauf mussten die Probanden einen Fragebogen ausfüllen, der sich an die Vertrauensskala von Jian u. a. (2000) anlehnte und unterschiedliche Fragen hinsichtlich der subjektiven Vertrauenseinschätzung der Probanden abfragte. Zusätzlich erfolgte eine Datenaufzeichnung im Hintergrund, um herauszufinden, welche Fenster vor der Klassifizierung der Funkkontakte geöffnet waren.

### Zusammenfassung der Ergebnisse

Die Methodik der Experimente von Bisantz und Seong (2001) ist interessant und in Grundzügen in das Experimentaldesign der in Abschnitt 10 und 11 beschriebenen eigenen Experimente eingegangen. Besonders innovativ erscheint die Idee, durch Verknappung der Möglichkeit, auf die unterschiedlichen Informationsfenster zuzugreifen, Rückschlüsse darauf ziehen zu wollen, inwieweit das Auftreten von Fehlern die Allokationsstrategie (und damit das Vertrauen) beeinflusst.

Für das eigene Experiment SIMHYBS (s. Kapitel 11) wurde eine ganz ähnliche Vorgehensweise gewählt. Dort werden Probanden nach getätigten Versuchsläufen mit einem Vertrauensfragebogen konfrontiert.

Anhand der Untersuchung von Bisantz und Seong (2001) wird deutlich, ähnlich wie schon bei McKnight und Chervany (2001: 48f.) bemerkt und in der Vertrauensskala von Jian u. a. (2000) aufgegriffen, dass Vertrauen ein multidimensionales Konstrukt ist, wie mit (FLA 1) bereits formuliert.

Inhaltlich haben Bisantz und Seong herausgefunden, dass sich die Reaktionen der Probanden auf Fehler des Entscheidungshelfers je nach anfänglicher Instruktion bzgl. des Fehlerursachen messbar unterscheiden: „[...] differences in fault causation may impact operators' trust in, and use, of information systems [...]“ (97). Die Autoren konstatieren, dass sowohl in der Sabotagegruppe als auch in der Nichtintentionalitätsgruppe die gemessenen Vertrauenswerte über die Zeit sinken. Die niedrigsten positiven Vertrauenswerte treten in der Nichtintentionalitätsgruppe auf, in der als Fehlerursache Hardwareprobleme genannt wurden (ebd.: 95). Dieses Ergebnis deckt sich mit den frühen experimentellen Vertrauensstudien von Morton Deutsch (1958) (vgl. Abschnitt 3.2.2), bei denen herausgefunden wurde, dass situative Bedingungen wie die anfängliche Instruktion der Probanden einen wesentlichen Einfluss auf Vertrauen haben.

### 4.3.3 Erkennung versteckter Soldaten

Bei den von Dzindolet u. a. (2003) durchgeführten Studien werden drei verschiedene Laborexperimente, bei denen Probanden versteckte Soldaten ggf. unter Zuhilfenahme einer Automation erkennen sollen, durchgeführt. Die Experimente haben unterschiedliche Erklärungsansprüche: Im ersten Experiment geht es um die Messung des Vertrauenslevels *vor* der Interaktion mit der Automation. Im zweiten Fall untersuchen die Autoren den Einfluss von Feedback-Mechanismen auf Vertrauen. Im dritten Szenario beleuchten die Autoren den Einfluss von Erklärungen hinsichtlich Automationsfehlern auf das Vertrauen in die Automation, ähnlich wie in der zuvor behandelten Studie von Bisantz und Seong (2001).



**Abbildung 4.5:** Erkennung versteckter Soldaten  
Quelle: Dzindolet u. a. (2003: 701)

Bei allen drei Experimenten wurde den Probanden erklärt, dass ihnen pro Durchlauf insgesamt 200 Folien gezeigt würden. Die Folien (vgl. Abbildung 4.5) wurden jeweils kurz eingeblendet und nach der Einblendung mussten die Probanden entscheiden, ob sich auf der gezeigten Folie ein versteckter Soldat (*camouflaged soldier*) verbarg oder nicht. Nach ihrer Entscheidung wurde zusätzlich die Einschätzung eines automatischen Entscheidungshelfers („Contrast Detector“) angezeigt. Die Probanden wurden vorab instruiert, dass der Entscheidungshelfer nicht perfekt funktioniert: „All participants were told that their automated decision aid was not perfect“ (ebd.: 700). Bei jedem der drei Experimente gab es eine kurze Eingewöhnungsphase mit vier Probeläufen, auf denen die versteckten Soldaten einfach erkennbar waren und der Entscheidungshelfer keine Fehler machte.

Einige Überlegungen aus dem Experimentablauf, z. B. die Frage nach der Übernahme der Entscheidung des Entscheidungshelfers und die Instruktion, dass dieser noch nicht perfekt funktioniert, haben auch Eingang in das eigene Experiment MensaAssist (vgl. Kapitel 10) gefunden.

### Experiment 1: Vertrauenslevel vor Interaktion mit Automation

Im ersten Laborexperiment wurden ausschließlich vier Probeläufe durchgeführt und im Anschluss nach der erwarteten eigenen Performance und der des Entscheidungshelfers für die nächsten 200 Durchläufe gefragt.

Die Probanden, die nur sehr unspezifisch über Fehlerraten des Entscheidungshelfers aufgeklärt wurden, waren der Automation gegenüber positiv eingestellt: „Taken together, it appears that people generally have positive expectations of unfamiliar automated decision aids“ (ebd.: 702). Interessanterweise schätzten sie ihre zukünftige eigene Performance hinsichtlich der korrekten Identifikation versteckter Soldaten schlechter ein als die der Automation: „People expected automated aids to outperform human partners“ (ebd.).

### Experiment 2: Einfluss von Feedback-Mechanismen

Bei diesem Experiment, bei dem der Einfluss von Feedback-Mechanismen untersucht wurde, wurde ein  $2 \times 2 \times 3$ -Faktordesign<sup>4</sup> verwendet:

- *Aid's Performance Level* (2): Die Automation macht entweder doppelt so viele Fehler wie der Proband (inferior), oder halb so viele Fehler (superior).
- *Provision of Aid's Decision* (2): Die Entscheidung der Automation wird entweder angezeigt (Decisions Present) oder nicht angezeigt (Decisions Absent)
- *Type of Feedback* (3): Entweder wird gar kein Feedback gegeben (No Feedback), ein kumulatives Feedback (Cumulative Feedback), bei dem die Performance der Automation der Performance des Probanden ganz am Ende des Experiments gegenübergestellt wird, oder es wird ein kontinuierliches Feedback (Continuous Feedback) gegeben, bei dem eine Balkengrafik während der kompletten Versuchsreihe sichtbar ist und nach jeweils fünf Durchläufen aktualisiert wird.

Die Probanden absolvieren zunächst gemäß der zugehörigen Experimentalbedingung 200 Durchläufe. Nach Beendigung der Durchläufe können sie sich entscheiden, ob zehn zufällig ausgewählte eigene Entscheidungen oder zehn Entscheidungen der Automation als Bewertungsgrundlage für den Erhalt einer Belohnung bei möglichst vielen korrekten Entscheidungen zu Grunde gelegt werden soll (707). Mit dieser „one-time reliance decision“ (710) wurde untersucht, ob und unter welchen Bedingungen sich die Probanden auf die Automation verlassen. Dabei wurde ein starker „bias toward self-reliance“ (707) gefunden, also eine Tendenz eher der eigenen Entscheidung als der der Automation zu vertrauen. Zudem zeigte sich, dass Feedback-Mechanismen besonders stark wirken, wenn keine andere Information zur Verfügung stand, also insbesondere

<sup>4</sup>Für eine umfangreiche Übersicht zur Verwendung und systematischen Auswertung dreifaktorieller Designs s. Keppel und Wickens (2004: 465ff.).

unter der „Decisions Absent“-Bedingung. Wurde im Falle von Feedback eine besonders gute Automationsperformance angezeigt, verließen sich nur 13% der Probanden auf ihre eigene Entscheidung. Umgekehrt ist es bei „inferior“-Performance. In dem Fall wählen 93% ihre eigenen Entscheidungen als Bewertungsgrundlage. M. a. W. : Ein messbarer Einfluss von Feedback ist bei folgenden vier Experimentalbedingungen festzustellen:

$$\{superior\} \times \{Decisions Absent\} \times \{continuous, cumulative\}$$

Beachtenswert ist, dass, unabhängig von der Gesamtpformance des Entscheidungshelfers, die Probanden bei unmittelbarer Sichtbarkeit der Entscheidungen der Automation (Decisions Present) einen starken „self-reliance bias“ haben und folglich eher ihren eigenen Entscheidungen vertrauen, da offensichtliche Fehler, unabhängig von ihrer Häufigkeit, offenbar zu einen irreparablen Vertrauensverlust führen. Auf der anderen Seite widerspricht dieses Ergebnis einer üblichen Ansicht in der Automationsforschung, dass sich Transparenz positiv auf Vertrauen auswirkt (Manzey und Bahner 2005: 95).

In realen Systemen ist ein direktes Feedback über die Performance der Automation in der Regel nicht vorhanden. Für das dritte Experiment wurde zur Steigerung der externen Validität der direkte Feedback-Mechanismus abgeschafft. Stattdessen wurden den Probanden erklärt, aus welchen Gründen die Automation möglicherweise Fehler machen könnte. Gemessen werden soll dabei der Einfluss von Erklärungen und Gründen auf Vertrauen (ebd.: 710).

### Experiment 3: Einfluss von Erklärungen

Beim dritten Szenario wurden demnach realitätsnähere Experimentalbedingungen festgelegt, insbesondere gab es kein kumulatives oder kontinuierliches Feedback mehr. Die Entscheidung der Automation wurde unmittelbar nach Anzeige des Bildes angegeben und der Proband konnte dann entscheiden, ob er dem Vorschlag der Automation folgt oder eine andere Entscheidung trifft. Im Gegensatz zu den vorherigen Szenarien wurde also hier keine *einmalige* Vertrauensentscheidung getroffen, sondern kontinuierlich gemessen, ob die Probanden der Automationsentscheidung folgen und ihr damit vertrauen.

Durch die Einführung einer neuen Experimentalbedingung, *provision of a rationale for the aid's error*, wurde der Einfluss von Erklärungen auf das Vertrauen in die Automation untersucht. Die Probanden bekamen vorab entweder eine Erklärung, warum die Automation Fehler macht (information), oder nicht (Kontrollgruppe, control). Zusätzlich wurde der mentale Workload durch Hinzufügen einer weiteren Aufgabe (auf



fürte dazu, dass Probanden dem Entscheidungshelfer deutlich mehr vertrauten als solche, die keine Erklärung bekamen und zwar *unabhängig* davon, ob die Performance des Entscheidungshelfers besser oder schlechter war als die der Probanden (714).

#### Zentrale Ergebnisse von Dzindolet u. a. (2003)

- Es bestehen in der Regel hohe Ausgangserwartungen an die Automation, verknüpft mit der Erwartung, dass die Automation keine Fehler macht. (702)
- Undurchschaubarkeit verringert die Vertrauenswürdigkeit in die Automation. (704)
- *Offensichtliche* und *unmittelbar wahrnehmbare* Automationsfehler führen zu einem Vertrauensverlust und dienen als Begründung für übersteigertes Selbstvertrauen. (709)
- Begründete Automationsfehler wirken sich weit weniger stark auf Vertrauen aus als unbegründete. (714)

#### 4.3.4 Langzeitstudie an einem Realsystem

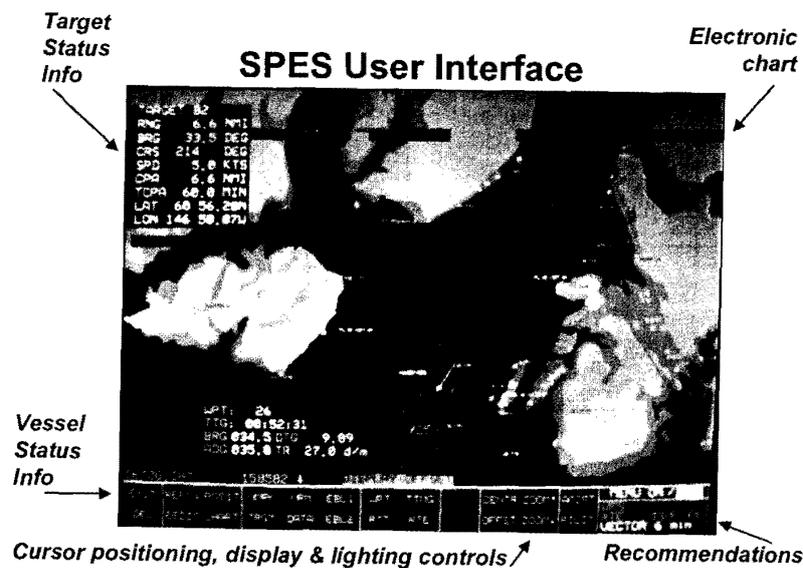
Bei den bisher vorgestellten Studien handelt es sich ausnahmslos um Laborstudien. Laborstudien bieten den Vorteil kontrollierter Experimentalbedingungen, sind i. d. R. risikolos und mit überschaubarem Aufwand durchführbar. Als letztes Beispiel wird in diesem Abschnitt eine Automationsstudie von Grabowski und Sanborn (2003) vorgestellt, die unmittelbar an einem Realsystem, nämlich an einem Ölfrachter, durchgeführt wurde. Die Durchführung von Studien an Realsystemen, ähnlich den Workplace Studies<sup>5</sup> von Suchman (1987), bringt einige Vorteile, aber auch Probleme mit sich, die hier näher ausgeführt werden. Im Gegensatz zu den Workplace Studies, die sich hauptsächlich auf Praktiken innerhalb von Räumen fokussieren, versuchen die Autoren der Ölfrachter-Studie trotz widriger Experimentalbedingungen an einem Realsystem zu messbaren Ergebnissen zu kommen.

Gegenstand der über zwei Jahre durchgeführten Langzeitstudie auf dem Ölfrachter „Exxon Valdez“ ist die Untersuchung der Auswirkungen eingebetteter intelligenter Technik („embedded intelligent technology, *EIT*“) auf Vertrauen und die Erfüllung der nautischen Aufgaben trackkeeping (Kurshalten), maneuvering (Steuern) und collision avoidance (Kollisionsvermeidung). Die Autoren gehen davon aus, dass Vertrauen nicht direkt messbar ist, sondern nur menschlicher Verlass auf technische Systeme: „In this research, we were able to measure several constructs that influence human reliance on technology, although we were not able to measure trust directly“ (Grabowski und Sanborn 2003: 640). Diese Annahme deckt sich mit der zuvor formulierten (FLA 2),

<sup>5</sup>Für einen guten Überblick über Workplace Studies in der Technikforschung vgl. Knoblauch und Heath (1999).

also einem engen Zusammenhang von Vertrauen und Verhalten, das auf Vertrauen basiert.

Die in der Studie untersuchte EIT heißt SPES (Shipboard Piloting Expert System, vgl. Abbildung 4.6) und ist eingebettet in das reguläre Schiffsnavigationssystem. Die Aufgabe des SPES besteht darin, aus gesammelten Echtzeitdaten ein Lagebild zu generieren, dieses grafisch ansprechend darzustellen und im Falle von kritischen Systemzuständen Hinweise und Lösungsvorschläge zu geben (Grabowski und Sanborn 2003: 643).



**Abbildung 4.6:** Grafische Anzeige des Shipboard Piloting Expert System (SPES), ein in das Schiffsnavigationssystem eingebettetes intelligentes System zur Unterstützung von Navigationsaufgaben  
Quelle: Grabowski und Sanborn (2003: 643)

Ein Versuchsdurchlauf erstreckte sich über eine vierzehntägige Fahrt zwischen Valdez in Alaska und den Ölterminals an der Westküste in Kalifornien. Grabowski und Sanborn (2003: 648) wählten für ihre Untersuchung ein  $2 \times 3 \times 3$ -Design:

$$\underbrace{\text{technology treatment}}_{\{EIT, host\}} \times \underbrace{\text{subject type}}_{\{ship's master, mate on watch, pilot\}} \times \underbrace{\text{level of voyage stress}}_{\{low, medium, high\}}$$

Die einzelnen Design-Elemente haben folgende Bedeutung:

- *technology treatment*: Es wurde entweder ohne Zuhilfenahme von SPES gefahren (*host*), oder unter Zuhilfenahme von SPES, der intelligenten eingebetteten Technik (*EIT*).

- *subject type*: Auf einem Schiff fallen sehr unterschiedliche Tätigkeiten an. Der *ship's master* ist der Kapitän auf einem Schiff, der *mate on watch* ist der erste Offizier und steht in der Hierarchie direkt unter dem Kapitän, der *pilot* ist zuständig für die Steuerung im operativen Sinne, z. B. in Hafenanlagen oder kritischen Gewässern.
- *voyage stress*: Je nach äußeren Bedingungen (wie Sicht, Fahrzeugverkehr, Wind, Strömungsstärke, räumliche Begrenztheit bei Hafendurchfahrten) befindet sich die Crew in unterschiedlichen Stresszuständen (ebd.: 647). Voyage stress stellt eine Beschreibungsmöglichkeit eines Systemzustands in einer sicherheitskritischen Umgebung dar.

Grabowski und Sanborn (2003) konnten im Rahmen ihrer Langzeitstudie folgende Ergebnisse erzielen:

*Neuer Technologie wird anfänglich misstraut*: „New technology, when first introduced, tends to be looked at suspiciously and perhaps mistrusted“ (ebd.: 664). Dieses Ergebnis steht allerdings teilweise im Widerspruch zu den Ergebnissen von Dzindolet u. a. (2003), die herausgefunden habe, dass hohe Ausgangserwartungen an eine Automation bestehen und von dieser zunächst erwartet wird, dass sie bessere Entscheidungen trifft als menschliche Akteure (Dzindolet u. a. 2003: 702). Nach einiger Zeit waren die an der Studie teilnehmenden Crewmitglieder aber durchaus bereit, das SPES punktuell zu nutzen (da es offenbar sehr zuverlässig funktionierte). Interessanterweise steigerte die Nutzung allerdings nur in Situationen geringer Belastung die Performance<sup>6</sup> der Operateure, woraus die Autoren folgern, dass „[...] operator reliance and benefit were not calibrated“ (Grabowski und Sanborn 2003: 664). Diese Feststellung entspricht der Folgerung der Autoren Dzindolet u. a. (2003: 615) nach einem „appropriate level of trust“. Entscheidend ist die Frage, wie dieses erreicht werden kann. Während Dzindolet et al. fordern, dass die Operateure entsprechend trainiert werden müssen, gehen Grabowski und Sanborn (2003: 664) davon aus, dass eine Kalibrierung und damit eine spürbare Verbesserung der Operator-Performance bereits bei der Systemkonstruktion berücksichtigt werden muss, damit das System überhaupt sinnvoll eingesetzt werden kann.

Weiter beschreiben die Autoren, dass Unterschiede zwischen den untersuchten Akteurstypen (*subject type*) hinsichtlich Vertrautheit, Nutzungsroutiniertheit und Arbeitspensum (workload) existieren. Erklärbar sei dies über die unterschiedlichen Erfahrungswerte, über die z. B. ein Kapitän im Vergleich zum ersten Offizier und zum Piloten verfüge (ebd.).

---

<sup>6</sup>Die Performance der Operateure wurde anhand verschiedener nautischer Kenngrößen wie „track-keeping“ (das Abfahren eines vorgegebenen Kurses) oder „threat avoidance“ (Vermeidung von Gefahrensituationen) gemessen (vgl. Grabowski und Sanborn 2003: 645).

Ein weiteres wichtiges Ergebnis bzw. eine Schlussfolgerung von Grabowski und Sanborn (2003) besteht darin, dass ein Verständnis über *Gründe* für Automationsnutzung und Nicht-Nutzung auf Seiten der Operateure einen wesentlichen Beitrag zur Verbesserung der Sicherheit sicherheitskritischer Systeme leisten kann: „Understanding how and why operators utilize and rely on technology in safety-critical systems is thus an important step in enhancing safety and performance in safety-critical systems“ (Grabowski und Sanborn 2003: 665).

### Zusammenfassung und kritische Würdigung

Die Besonderheit der Arbeit von Grabowski und Sanborn (2003) besteht darin, dass es sich nicht, wie in der Automationsforschung weit verbreitet, um eine Laborstudie handelt, sondern dass direkt am Realsystem (in dem Fall an einem großen Ölfrachter) geforscht wird. Methodisch ist dieser Ansatz ungewöhnlich, kommt er doch mit wenigen Probanden aus, benötigt viel Zeit (zwei Jahre) und schafft es doch, prüfbare Hypothesen zu formulieren und schließlich zu empirisch gehaltenen Aussagen zum Zusammenhang von Automationsnutzung und Operator-Performance zu gelangen. Dabei wird mit dem Realsystem experimentiert, als handele es sich um ein Laborexperiment mit einem vollständigen  $2 \times 3 \times 3$ -Design, bei dem unabhängige Variablen systematisch verstellt werden, um eine abhängige Variable, nämlich die Performance verschiedener Operateure, zu messen.

Kritisch an der Studie ist zu bemerken, dass die Schwierigkeiten einer direkten Vertrauensmessung zwar thematisiert werden, („we were not able to measure trust directly“, ebd.: 646), die Berücksichtigung von Vertrauen im verwendeten Forschungsmodell (ebd.: 647) jedoch unterkomplex bleibt bzw. gar nicht systematisch aufgegriffen wird. Es wird lediglich ein Einfluss von Vertrauen auf die wahrgenommene Zuverlässigkeit, die ihrerseits gleichgesetzt wird mit der Automationsnutzung, konstatiert. Dieser Einfluss wird in der Studie allerdings nicht näher untersucht.

#### Zentrale Ergebnisse von Grabowski und Sanborn (2003)

- Vertrauen beeinflusst das Allokationsverhalten. (640)
- Neuer Technologie wird anfänglich misstraut. (664)
- Ein Verständnis über Praktiken und Gründe zur Automationsnutzung ist ein wichtiger Faktor zur Verbesserung der Sicherheit in sicherheitskritischen Systemen. (665)

## 4.4 Implikationen

Vor dem Hintergrund der zahlreichen Einzelergebnisse ist zu klären, welche Implikationen sich aus den behandelten Studien für die vorliegende Arbeit ergeben.

An mehreren Stellen im Text fanden sich bereits Vorwärtsverweise auf die beiden selbst durchgeführten Experimente, die sich mit Vertrauen in autonome Technik beschäftigen. Abgesehen von der vierten Studie, die an einem Realsystem, auf einem Ölfrachter, durchgeführt wurde, handelte es sich bei den anderen drei Studien um Laborexperimente.

Alle Studien waren so angelegt, dass Vertrauen und Automationsnutzung in einer *interaktiven Perspektive* untersucht wurden. Dies deckt sich mit der schon zuvor formulierten (FLA 5).

Allen vorgestellten Studien gemeinsam ist ihr Fokus auf die *Optimierung* einer wie auch immer gearteten Systemperformance. Menschliche Operateure spielen insofern eine Rolle, als sie durch ihren angemessenen Vertrauenslevel dazu beitragen, dass die Systemperformance möglichst gesteigert wird oder zumindest auf hohem Niveau gehalten wird.

Eine weitere Gemeinsamkeit der Studien ist ferner die Betonung, dass *Begründungen und Erklärungen* für Fehlfunktionen einen Einfluss auf das Vertrauen in ebendiese Systeme haben. Vor dem Hintergrund der in Abschnitt 8 vorgestellten Methodologie zu einer soziologischen Analyse der Interaktion von Mensch und autonomer Technik sind Gründe und Erwartungen zweifelsfrei von entscheidender Bedeutung. Allerdings, so lässt sich argumentieren, findet in der Automationsforschung eine Überbetonung der Bedeutung von Fehlfunktionen statt, die dazu noch äußerst voraussetzungsreich ist. Der Begriff der Fehlfunktion bedingt, dass diese auch als solche erkannt wird – dass also zweifelsfrei klar ist, ob eine Fehlfunktion vorliegt oder nicht. Eine angemessenere „Optimierungsgröße“ im Gegensatz zu Fehlfunktionen und Systemperformance, so mein Gegenargument, ist eine stärkere Berücksichtigung der gemeinsamen Aufgabe. Durch Befragungen lässt sich herausfinden, ob die Vorstellungen des menschlichen Akteurs über die Art der Unterstützung mit der tatsächlich erfahrenen Unterstützung übereinstimmt.

Die Begriffe „Gründe“ (bzw. Begründungen) und „Erklärungen“, wie sie in der Automationsforschung verwendet werden, unterscheiden sich konnotativ von den gleichnamigen soziologischen Konzepten (vgl. Abschnitt 6.2). In der Automationsforschung beziehen sich Gründe eher auf eine Art Vorabvorstellung bzw. Disposition, die, vermittelt über einen unklaren Mechanismus, Einfluss auf das Allokationsverhalten und das Vertrauen nehmen. Soziologische Makro-Mikro-Makro-Modelle, wie z. B. das Esser'sche Modell soziologischer Erklärung, berücksichtigen Gründe bzw. Ziele hingegen konstitutiv bei der Handlungswahl.

Die Bedeutsamkeit von Gründen bzw. das Verständnis für die Gründe (trotz unterschiedlicher Nuancierung in der Automationsforschung und der soziologischen Handlungstheorie) einer bestimmten Handlungswahl des Interaktionspartners hat einen positiven Einfluss auf Vertrauen. Diese Erkenntnisse lassen sich als (FLA 7) verdichten.

**FLA 7 (Gründe)**

Ein Verständnis über Gründe von Automations-Aktionen wirkt sich positiv auf das Vertrauen aus.

Neben der Bedeutsamkeit von Gründen fanden sich in den Studien Erkenntnisse dafür, dass Ausgangserwartungen bzw. Ausgangsvertrauen einen Einfluss auf die Interaktion mit der Automation haben. Allerdings sind die Ergebnisse dazu widersprüchlich. So argumentieren Dzindolet u. a. (2003: 702), dass an eine unbekannte Automation hohe Ausgangserwartungen bestehen, während Grabowski und Sanborn (2003: 664) feststellen, dass neuer Technik anfänglich misstraut wird.

Daher wird dieser Zusammenhang in (FLA 8) offen formuliert und nur allgemein vom Einfluss des initialen Vertrauens auf spätere Interaktionen gesprochen.

**FLA 8 (Ausgangsvertrauen)**

Das anfängliche Vertrauen in Automation beeinflusst die späteren Interaktionen.

Eine Berücksichtigung eines initialen Vertrauensbias findet sich auch in den sehr allgemeinen Überlegungen zum probabilistischen Vertrauensbegriff bei Coleman (vgl. Abschnitt 3.1.2) in Form einer Standardeinschätzung, mit der unbekanntem Interaktionspartnern begegnet wird.

Der Übersicht halber werden die Einzelergebnisse der vier behandelten Studien nochmals zusammenhängend auf Seite 97 aufgelistet.

**Lessons learned**

- Undurchschaubarkeit verringert die Vertrauenswürdigkeit in die Automation. (Dzindolet u. a. 2003)
- Offensichtliche und unmittelbar wahrnehmbare Automationsfehler führen zu einem Vertrauensverlust und dienen als Begründung für übersteigertes Selbstvertrauen. (Dzindolet u. a. 2003)
- Begründete Automationsfehler wirken sich weit weniger stark auf Vertrauen aus als unbegründete. (Dzindolet u. a. 2003)
- Neuer Technologie wird anfänglich misstraut. (Grabowski und Sanborn 2003)
- Insbesondere beim ersten Kontakt mit der Automation hängt die Automationsnutzung von der individuellen Einstellung der Operateure ab. (Lee und Moray 1994)
- Es bestehen in der Regel hohe Ausgangserwartungen an die Automation, verknüpft mit der Erwartung, dass die Automation keine Fehler macht. (Dzindolet u. a. 2003)
- Vertrauen beeinflusst das Allokationsverhalten. (Muir und Moray 1996)
- Die Differenz von Vertrauen und Selbstvertrauen beeinflusst das Allokationsverhalten des Operateurs. (Lee und Moray 1994)
- Die Automationsnutzung hängt von vorheriger Nutzung und dem Auftreten von Fehlern ab. (Lee und Moray 1994)
- Die Aufklärung über mögliche *Ursachen* für Automationsfehler hat maßgebliche Auswirkungen auf Vertrauen. (Bisantz und Seong 2001)
- Ein Verständnis über Praktiken und Gründe zur Automationsnutzung ist ein wichtiger Faktor zur Verbesserung der Sicherheit in sicherheitskritischen Systemen. (Grabowski und Sanborn 2003)



## **Teil II**

# **Verbindende Konzepte zwischen Computersimulation, Handlungstheorie und autonomer Technik**



## 5 Autonome Technik als Gegenstand der Soziologie

Während sich die vorherigen Kapitel vornehmlich mit unterschiedlichen *Vertrauens*konzepten und ihrer empirischen Erforschung beschäftigten, befassen sich dieses und auch die folgenden Kapitel schwerpunktmäßig mit dem Gegenstand *autonome Technik* aus Sichtweise der Soziologie.

Punktuell werden mit Verweisen auf die vorherigen Kapitel Hinweise gegeben, welche Ansätze besonders geeignet erscheinen, um an unterschiedliche Vertrauenskonzepte anzuknüpfen. Eine explizite Verbindung der Überlegungen zu autonomer Technik in der Soziologie wird im Anschluss an die integrierende Methodologie (Kapitel 8), insbesondere im Empirieteil ab Kapitel 11 erfolgen.

Die Ansicht, Technik bzw. autonome Technik als probaten Gegenstandsbereich der Soziologie aufzufassen, ist eine noch relativ junge Erkenntnis, deren Ursprung u. a. in der von Bruno Latour postulierten symmetrischen Ontologie im Rahmen seiner Akteur-Netzwerk-Theorie zu suchen ist (vgl. Abschnitt 5.1). Der darauffolgende Abschnitt 5.2 stellt einige Fallbeispiele und eigene Forschungsarbeiten zu autonomer Technik vor, die unterstreichen, dass autonome Technik ein gesellschaftlich und damit soziologisch relevanter Untersuchungsgegenstand ist. Es wird sich zeigen, dass autonome Technik *handlungswirksam* und im Rahmen einer soziologischen Analyse zu berücksichtigen ist. Abschließend stellt Abschnitt 5.3 mit dem *Graduellen Modell verteilter Handlungsträgerschaft* (GMVH) ein Theorieangebot vor, das auf Arbeiten von Werner Rammert und Ingo Schulz-Schaeffer basiert und vorschlägt, das Mit-Handeln von Technik attributionstheoretisch aufzufassen. Es wird sich am Ende des Kapitel allerdings zeigen, dass das GMVH das Mit-Handeln von autonomer Technik auf einer *deskriptiven* Ebene zwar recht genau beschreiben kann, zentrale Fragen der Operationalisierung und Formalisierung des Zusammenhandelns allerdings unbeantwortet lässt.

### 5.1 Die Provokation der Akteur-Netzwerk-Theorie<sup>1</sup>

Der folgende Abschnitt beschränkt sich bewusst auf die Actor Network Theory (ANT) Latour'scher und Callon'scher Lesart (Callon 1986; Callon und Latour 1992; Latour 1996a;b; 1998b). Die theoretische Debatte um die ANT hat sich seit einer breiteren

---

<sup>1</sup>Der folgende Abschnitt orientiert sich z. T. an Fink und Weyer (2011: 93f.).

Rezeption in verschiedene Richtungen weiterentwickelt, so sprechen etwa Gad und Bruun Jensen (2010), die einen umfangreichen Überblick über den Stand der Forschung zur ANT vorgelegt haben, von „Post-ANT“. Die ANT habe sich weiterentwickelt zur Post-ANT, die „stretches ANT’s empirical-ethnographic interest to include everything. There are no a priori limits as to where the empirical can be found or to what kind of settings will enable insights about a given theme“ (ebd.: 74).

In den beiden folgenden Abschnitten werden zwei zentrale Konzepte der ANT näher erläutert: die *generalisierte Symmetrie* und der *Hybrid-Akteur*. Aus der generalisierten Symmetrie ergibt sich das Prinzip der symmetrischen Ontologie bzw. des symmetrischen Vokabulars. Beide Konzepte bilden einen wichtigen Orientierungspunkt für den eigenen theoretischen Ansatz, der in Abschnitt 8.1 ausführlich dargelegt wird.

### 5.1.1 Generalisierte Symmetrie

Das Konzept der generalisierten Symmetrie spielt insbesondere in frühen Arbeiten von Callon und Latour eine wichtige Rolle (vgl. Callon 1986; Callon und Latour 1992).

Unter generalisierter Symmetrie ist laut Callon (1986: 196) „the commitment to explain conflicting viewpoints in the same terms“ zu verstehen. Mit den konfligierenden Betrachtungsweisen nimmt Callon vor allem den Gegensatz zwischen Natur und Gesellschaft in den Blick. Callon diagnostiziert eine Überbetonung des Sozialen bei der (technik-)soziologischen Erklärung von Wissenschaft und Technik (ebd.: 197). Er wendet sich damit gegen das sozialkonstruktivistische Erklärungsprogramm von Pinch und Bijker (1984), die in ihren Arbeiten eine *soziale* Konstruktion von Wissenschaft und Technik in der Vordergrund stellen. Durch eine Asymmetrie von Natur und Gesellschaft, so das Argument von Callon (1986: 197), beraube man sich der Möglichkeit Gesellschaft und ihre Bestandteile in offener Weise zu diskutieren. Callon (1986: 200) fordert alle „a priori distinctions between natural and social events“ zu Gunsten einer symmetrischen Betrachtungsweise aufzugeben.

Als einen möglichen Weg zur Überwindung der Asymmetrien empfiehlt Callon (1986: 200) den Beobachtern von Natur und Gesellschaft ein einheitliches Wortrepertoire zur Beschreibung beider Aspekte zu verwenden. In einem späteren Aufsatz sind Callon und Latour (1992: 353) noch deutlicher und schlagen vor „to extend our principle of symmetry to vocabulary and to decide that whatever term is used for humans, we will use it for nunhumans as well“. Die Erfindung einer neuen gemeinsamen Metasprache könne dabei helfen, die blickeinengenden Asymmetrien zu überwinden (ebd.: 154).

Latour und Callon sehen das generalisierte Symmetrieprinzip als einen dritten Weg zwischen Realismus und Konstruktivismus an, der Natur und Gesellschaft als Resultat wechselseitiger Vernetzungsaktivitäten auffasst (ebd.: 348).

### Symmetrisches Vokabular und begriffliche Umbenennungen

Eine logische Konsequenz, die sich aus dem Postulat der generalisierten Symmetrie ergibt, ist die Umbenennung von genuin soziologischen Begriffen, wie sie im Rahmen der ANT vorgenommen wird. Wie Callon (1986: 227) bemerkt, ist die konkrete Auswahl passender Begriffe im Prinzip komplett beliebig, die einzige Voraussetzung ist, „that [the word repertoire] must relate both to nature and society“.

Callon und Latour (1992: 347) erfinden neue hybride Begriffe: z. B. Aktant statt Akteur, Akteur-Netzwerk statt soziale Beziehung, Übersetzung statt Interaktion oder Delegation statt soziale Rolle.

Die wohl bekannteste Umbenennung ist die von Akteur zu *Aktant*. Latour (2000: 372) begründet die Verwendung des aus der Semiotik abgeleiteten Begriffs damit, dass der Akteursbegriff im Englischen und Deutschen begrifflich vorbesetzt sei und stattdessen Aktant verwendet wird, „um nicht-menschliche Wesen in die Definition einzubeziehen“. Fast zehn Jahre später gesteht Latour (1999: 15) allerdings ein, dass die neu verwendeten Begriffe z. T. unglücklich gewählt sind und zu einigen Missverständnissen geführt haben: So spricht die ANT zwar von Aktanten, allerdings von Akteur-Netzwerken statt Aktanten-Netzwerken. Ebenso spricht Latour (2000: 218) in dem Buch „Die Hoffnung der Pandora“ nicht etwa vom „Hybrid-Aktant“, sondern vom „Hybrid-Akteur“, der Gegenstand des nächsten Abschnitts ist.

#### 5.1.2 Der Hybrid-Akteur

Ausgangspunkt für die argumentative Zuspitzung hin zum Begriff des Hybrid-Akteurs ist die beispielhafte Behandlung der Frage, wer für den Akt des Tötens mit einer Schusswaffe verantwortlich ist (vgl. Latour 2000: 216).

Typischerweise gibt es auf diese Frage zwei Antworten, die jeweils davon abhängen, an wen man die Frage richtet. Latour (2000: 214f.) karikiert diese wie folgt: Vertreter einer Organisation, die sich für die Einschränkung des freien Verkaufs von Waffen einsetzen, werden materialistisch argumentieren: Durch ihren physikalischen Aufbau ist der Waffe das Ziel des Tötens bereits eingeschrieben. Im Gegensatz dazu würden Waffen-Lobbyisten argumentieren, dass Menschen töten und Waffen lediglich „neutrale Boten eines Willens [sind], der zur Handlung nichts hinzufügt“ (ebd.: 215). Latour kritisiert beide Positionen, da sie mal das Soziale, mal das Materielle überbetonen. Eine Waffe in der Hand eines nicht tötungswilligen guten Bürgers macht aus diesem nicht automatisch einen Mörder. Hat jedoch jemand die Absicht zu töten, eröffnet eine Waffe zumindest neue Möglichkeiten und kann den Prozess des Tötens massiv beschleunigen und ist folglich nicht so neutral, wie von der Waffen-Lobby behauptet (ebd.).

Die „Lösung“ von Latour (2000: 216) für dieses Problem besteht darin, durch das Zusammentreffen von Bürger (Agent 1) und Waffe (Agent 2), einen neuen, zusammen-

gesetzten, dritten Agenten erzählerisch entstehen zu lassen<sup>2</sup>. Ausgehend von diesem neuen Agenten können die ursprünglichen Ziele der Agenten in neue Ziele „übersetzt“ bzw. transformiert werden. Der Begriff der Übersetzung ist nicht im Sinne einer Übersetzung von einer Fremdsprache in eine andere zu verstehen. Vielmehr bezieht sich der Begriff darauf, dass Akteure durch ihre Zusammenkunft auch ihre Ziele wechselseitig verändern und verschieben (vgl. Latour 2000: 381). Unter Rekurs auf den Übersetzungsbegriff (engl.: *translation*) bei Callon (1986: 203ff.) beschreiben Peucker und Voss (2006: 14) den Begriff als Aushandlungsprozess von Rollen und Eigenschaften im Verlaufe der Netzwerkkonstruktion, bei dem „die spezifischen Aktionspotentiale der Entitäten aufeinander ausgerichtet“ werden.

In Bezug auf das Waffenbeispiel führt Latour aus, dass aus dem ursprünglichen Ziel nur jemanden verletzen zu wollen, *zusammen* mit der Waffe das neue Ziel jemanden tatsächlich töten zu wollen entstehen kann (ebd.: 217). Durch die Zusammenkunft von Bürger und Waffe entsteht „Jemand anderes“, ein „Waffen-Bürger“ oder eine „Bürger-Waffe“. Die beiden in den Übersetzungsprozess involvierten Elemente verändern sich durch die neue Verknüpfung wechselseitig: „Mit der Waffe in der Hand bist du jemand anderes, und auch die Waffe ist in deiner Hand nicht mehr dieselbe“ (ebd.: 218). Dieser so genannte „*Hybrid-Akteur*“, der zum Adressat zugeschriebener Handlungen mit neuen, vorher nicht dagewesenen, Zielen wird, kann somit in das Zentrum der soziologischen Analyse rücken.

### 5.1.3 Fazit ANT

Die ANT bzw. Post-ANT, so Gad und Bruun Jensen (2010) will sich weder als Methode noch als Theorie verstanden wissen (ebd.: 56), sondern möchte dem Forschungsprozess lediglich eine postplurale Einstellung und explizit nichthumanistische Disposition vermitteln (ebd.: 73f.).

Trotz zahlreicher Weiterentwicklungen der ANT konstatieren Gad und Bruun Jensen (2010: 73f.), dass zentrale Postulate der klassischen ANT, etwa die generalisierte Symmetrie (vgl. Abschnitt 5.1.1), nach wie vor Gültigkeit besitzen, wenngleich Ingo Schulz-Schaeffer (2011: 292) dies relativiert und zeigt, dass neuere sozialtheoretische Überlegungen von Latour (2007) an eben dieser zentralen Grundannahme rütteln<sup>3</sup>. Teile der vorliegenden Arbeit, insbesondere das Modell soziologischer Erklärung hybrider Systeme (HMSE, vgl. dazu ausführlich Abschnitt 8.1), sind von der ANT

<sup>2</sup>Wobei sich Latour um die Frage windet, von wem das Zusammenfinden ausgeht, zunächst heißt es: „Agent 1 [Bürger] greift auf Agent 2 [Waffe] zurück“, im nächsten Satz heißt es, dass der Bürger von der Waffe rekrutiert würde (Latour 2000: 216).

<sup>3</sup>Schulz-Schaeffer verweist dabei insbesondere auf die neue Monografie „Eine neue Soziologie für eine neue Gesellschaft“ von Latour (2007: 123) und die Kombination der dort entwickelten „assoziationalistische[n] Sozialtheorie“ mit der bereits in früheren Schriften von Callon und Latour (1992: 359) entwickelten Idee der „Partizipation [...] von Nicht-Menschen, insbesondere von Maschinen und Artefakten“ (ebd.), ohne die kein soziales Leben denkbar sei (zit. nach Schulz-Schaeffer 2011: 349f.).

inspiriert. Der folgende Abschnitt unternimmt den Versuch, die Verdienste der ANT hinsichtlich einer Soziologie autonomer Technik zu würdigen, und zwar in einer Weise, die nicht ontologisch eine Symmetrie von Mensch und Technik postuliert, sondern eine empirische Offenheit und Neugier hinsichtlich der Grenzziehung zwischen Mensch und autonomer Technik zulässt.

Latour und Callon haben mit ihren z. T. polemisch-provokativen Beispielen und Forderungen<sup>4</sup> sowie den dahinterstehenden ontologischen Postulaten, z. B. dem Symmetrieprinzip, zu einer äußerst kritischen Rezeption der ANT beigetragen (vgl. etwa Collins und Yearley 1992: 310ff.; Schulz-Schaeffer 2000: 132f.; Rammert und Schulz-Schaeffer 2002a: 48)<sup>5</sup>.

Latour und Callon nehmen mit ihren Beispielen eher konventionelle Technik in den Blick, etwa Muscheln (Callon 1986), Hotelschlüssel (Latour 1991: 104f.) und Pistolen / „Waffenbürger“ (Latour 2000: 218). Die Beispiele bei Callon und Latour haben in der Regel keinen direkten empirischen Nutzen, sondern dienen dem Zweck, zentrale Konzepte der ANT zu illustrieren und „das Bewusstsein für die Artefakte zu schärfen“.

Die von Latour (1996c) ausgegebene Lösung: „Follow the actors themselves, is the slogan of our sociology; indeed, but it is not said how to follow them“ charakterisiert den Anspruch der ANT insofern, als sie sich weder als Theorie, noch als Beschäftigung mit Netzwerken oder Akteuren, erst recht nicht mit Akteur-Netzwerken verstanden wissen will: „I will start by saying that there are four things that do not work with actor-network theory; the word actor, the word network, the word theory and the hyphen! Four nails in the coffin“ (Latour 1999: 15). Führt man Latours Gedanken gang weiter, so ist es ein zweckloses Unterfangen, nach dem konkreten theoretisch-methodologischen Nutzen der ANT zu suchen. Nach meiner Lesart verordnet die ANT lediglich eine *Forschungsdisposition*, also eine Haltung gegenüber einem sozialwissenschaftlich zu erforschenden Thema, die weitgehend auf begriffliche Vorfestlegungen verzichtet und ex ante Natur, Technik und Gesellschaft als gleichberechtigte Elemente begreift.

Im Kontext dieser Arbeit fasse ich die ANT, ähnlich wie schon Gad und Bruun Jensen (2010), als eine ebensolche dem Forschungsprozess zugrundeliegende Disposition auf, die eine (möglichst) vorurteilsfreie Beschäftigung mit (autonomer) Technik zulässt.

---

<sup>4</sup>So schreibt Latour etwa, dass man die Nicht-Menschen in den Sozialwissenschaften ebenso mit offenen Armen begrüßen solle wie einst die Arbeiter des 19. Jahrhunderts: „I have sought to show researchers in the social sciences that sociology is not the science of human beings alone – that it can welcome crowds of nunhumans with open arms, just as it welcomed the working masses in the nineteenth century“ (Latour 1998a: viii).

<sup>5</sup>Callon und Latour 1992: 356 wehren sich gegen die ontologische Interpretation der generalisierten Symmetrie und damit auch gegen die entgegengebrachte Kritik: „We do not deny differences; we refuse to consider them a priori [...]“. Wenig später heißt es noch deutlicher die empirische Offenheit der ANT betonend (ebd.: 360f.): „Our empirical program does not claim either that humans and artifacts are exactly the same or that they are radically different. We leave this question entirely open.“

Obwohl sich Latour selbst nie ernsthaft mit autonomer Technik auseinandergesetzt hat<sup>6</sup>, ist es im Falle autonomer Technik umso mehr erforderlich, einen Mittelweg zu finden zwischen dem radikalen Technikdeterminismus, der prägend für die Technikfolgenabschätzung der 1970er Jahre war, und dem Sozialkonstruktivismus der 1980er Jahre (Pinch und Bijker 1984; Bijker u. a. 1987) auf der anderen Seite<sup>7</sup>.

Wird Technik zum aktiven Mitspieler, kann für einen ganzheitlich *soziologischen* Erklärungsansatz weder der Primat des Sozialen noch der Primat des Technischen ex ante vorausgesetzt werden. Avancierte Technik, so die Implikation für die Forschungshaltung, die sich für die vorliegende Arbeit daraus ableiten lässt, ist nie bloß als passives Objekt zu berücksichtigen, sondern stets im Kontext seiner soziotechnischen Einbettung in ein System verteilter Handlungsträgerschaft. Werner Rammert und Ingo Schulz-Schaeffer haben sich zu diesem Zweck eines methodologischen Tricks bedient und das Zusammenspiel von Mensch und autonomer Technik aus der Perspektive eines Beobachters als Zuschreibungsphänomen konzeptualisiert (vgl. dazu ausführlich Abschnitt 5.3). Die zunehmende Eigenaktivität von Technik wird im Rahmen der vorliegenden Arbeit in Rechnung gestellt, indem die Frage nach der Zuschreibung von Handlungsträgerschaft empirisch offen gestellt wird, ohne etablierte Analysekatgorien wie Mensch und Nicht-Mensch aufzugeben und der Gefahr aufzusitzen „im begrifflichen Niemandsland der Akteur-Netzwerk-Theorie zu stranden“ (Rammert und Schulz-Schaeffer 2002b: 48).

Insbesondere in den Abschnitten 8.1 und 8.2, in denen eigene theoretische Überlegungen zu einem Modell soziologischer Erklärung hybrider Systeme angestellt werden und das Konzept der wechselseitigen Modellzuschreibung vorgestellt wird, wird deutlich werden, dass sich durch die Aufrechterhaltung der Analysekatgorien Mensch und Nicht-Mensch empirisch gehaltvolle und soziologisch interessante Aussagen ableiten lassen. Aus der ANT stammt zwar die Anregung, die Perspektive von Mensch- und Nicht-Mensch gleichermaßen zu berücksichtigen. Die forschungsmethodologischen Implikationen sind allerdings gänzlich andere: Während eine empirische Anwendung der ANT in der Regel darin besteht, Geschichten zu erzählen („Story Telling“) oder wie Latour (1998a: ix) diese quasi-fiktionale Erzählform auch nennt „scientifiction“, versucht die vorliegende Arbeit das Zusammenspiel von Mensch und autonomer Technik mit Hilfe einer symmetrisch angewendeten intentionalistischen Modellierungsspra-

---

<sup>6</sup> Abgesehen von einer „Fallstudie“ zu Aramis (Latour 1998a), einem führerlosen Bahnsystem, bei dem sich einzelne, frei fahrende Wagenteile an Kreuzungen mit Bahnstationen rekonfigurieren können. Ein wissenschaftlicher Wert, im Sinne von Erkenntniszugewinn, ist in dieser Studie schwer auszumachen, da Latour weitgehend auf der deskriptiven Ebene verhaftet bleibt und die Bedeutung von Miniaturmodellen und Versammlungen um ebendiese Modelle hervorhebt. Der Untersuchungsgegenstand von Latour ist in dem Fall zwar autonome Technik, sein Interesse richtet sich allerdings nicht auf das spezifisch „autonome“ dieses Systems, sondern kommt über eine (relativ banalisierende) Betrachtung von Artefakten nicht hinaus.

<sup>7</sup> Für eine ausführliche Auseinandersetzung mit den Forschungsdispositionen des Technik- und Sozialdeterminismus insbesondere in Hinblick auf die Technikfolgenabschätzung bzw. auf das Technology Assessment (TA) s. Grunewald (2007).

che, dem HMSE, darzustellen und dabei dennoch dem Anspruch einer soziologischen Tiefenerklärung zu genügen. In (FLA 9) sind diese Überlegungen zusammengefasst.

### FLA 9 (Eigenaktivität / Gleichberechtigte Partner)

Autonome Technik mit zunehmender Eigenaktivität wird aufgefasst als gleichberechtigter Partner in einer Mensch-Technik-Konstellation, ohne dabei die Analysekategorien Mensch und Technik aufzugeben.

## 5.2 Praxisbeispiele und Handlungswirksamkeit autonomer Technik

Als *handlungswirksam* soll autonome Technik bezeichnet werden, die durch ihr selbsttätiges Verhalten Einfluss auf die Wahl von Handlungsalternativen eines menschlichen Akteurs nimmt und den Interaktionsprozess zwischen Mensch und Technik aktiv mitgestaltet.

Diese begriffliche Annäherung steckt gleichzeitig den Rahmen für die soziologische Relevanz autonomer Technik ab: Relevant ist nur solche Technik, die handlungswirksam ist, also an irgendeiner Schnittstelle mit menschlichen Akteuren (allgemein: mit der Gesellschaft) in Berührung kommt. Dabei ist es unerheblich, ob es sich um physikalische Berührungspunkte (etwas bei Robotern oder Fahrerassistenzsystemen), virtuelle Berührungspunkte (z. B. bei autonom agierenden Börsencomputern) oder um über Netzwerke vermittelte Berührungspunkte (z. B. bei Wikipedia-Bots<sup>8</sup>) handelt.

Das Ziel dieses Abschnitts besteht darin, anhand von drei ausgewählten Beispielen zu erläutern, in welcher Hinsicht autonome Technik als handlungswirksam bezeichnet werden kann. Dabei werden schlaglichtartig sogenannte *autonomous trading agents*, also autonom agierende Börsencomputer, unterschiedliche Fahrerassistenzsysteme mit zunehmender Eigenaktivität und ein von der Firma Google entwickeltes selbstfahrendes Auto vorgestellt. Es schließen sich etwas ausführlichere Erläuterungen zu Wikipedia-Bots an, die an Forschungsarbeiten von Fink und Liboschik (2012) anschließen und einige neue Entwicklungen hinsichtlich besonders avancierter Bots, die sich durch ein hohes Maß an Eigenaktivität und Multifunktionalität auszeichnen, herausstellen.

### 5.2.1 Autonomous Trading Agent

Bei dem Begriff Börsenhandel hat man bisher typischerweise ein Bild hektisch agierender Börsenmakler vor Augen, die sich auf dem Handelsparkett Orders (Käufe und

<sup>8</sup>Bots sind (teil-)autonom agierende Softwareagenten, die bestimmte Aufgaben z. B. innerhalb der Wikipedia durchführen.

Verkäufe von Aktien) mündlich zurufen. Seit dem „Flash Crash“ vom 6. Mai 2010, dem plötzlichen Abrutschen des Dow-Jones-Index um über 1 000 Punkte, ist dieses Bild auch außerhalb der ökonomischen Fachöffentlichkeit überholt. Der Begriff High-Frequency-Trading (HFT) bzw. der algorithmische Börsenhandel macht seitdem die Runde und ist Gegenstand der öffentlichen Diskussion geworden (vgl. Kuls und Mohr 2010; Spiegel Online 2010).

Dabei handeln leistungsfähige Computer ohne weiteres menschliches Zutun und tätigen auf Grundlage komplexer Entscheidungsroutrinen Käufe und Verkäufe von Aktien. Dabei kommt es auf jede gewonnene Millisekunde (inzwischen sogar Mikrosekunde) an (vgl. Sietmann 2011: 150f.), was z. B. dazu geführt hat, dass an großen Handelsplätzen Banken ihre Server direkt in den Gebäuden der elektronischen Börse untergestellt haben. Z. T. werden sogar Glasfaserverbindungen, die in der Regel entlang von Autobahnen oder Kanälen verlegt sind, neu gebaut und aufwendig als möglichst kurze, Luftlinien ähnliche, Verbindung umgesetzt.

Am weltweiten Aktienhandel hat HFT inzwischen einen wesentlichen Anteil: Im Jahr 2011 betrug der Anteil des HFT in den USA 55 %, gefolgt von Europa mit 35 % und Japan mit 28 % (vgl. Siedenbiedel 2013).

Nach letztem Kenntnisstand (ebd.) zum sogenannten „2010 Flash Crash“ kann keine eindeutige Schuld des Hochfrequenzhandels an dem Kursrutsch konstatiert werden. Offenbar gab es einen selbstverstärkenden Effekt, der daraus resultierte, dass ein sehr hohes Order-Volumen die Überschreitung von Schwellenwerten bewirkte. Die High Frequency Trader reagierten auf die Überschreitung und tätigten daraufhin massenhaft automatische Verkäufe (vgl. Siedenbiedel 2013), die zu weiteren Überschreitungen führten. Siedenbiedel beschreibt in seinem Artikel, dass die ökonomische Forschung uneinig darüber ist, ob HFT eine insgesamt stabilisierende Wirkung auf den Markt hat oder ob HFT die Marktstabilität gefährdet. Die verwendeten HFT-Algorithmen basieren fast ausschließlich auf komplexen *technischen* Analysen von Kursentwicklungen und Marktreaktionen. Algorithmen, die auf externe Informationen (z. B. bestimmte Schlagwörter in Nachrichten) reagieren, sind noch die Ausnahme, befinden sich laut Siedenbiedel jedoch bereits in der Entwicklung und würden zukünftig zu weiteren großen Veränderungen führen und den „Hochfrequenzhandel aus seiner Selbstreferenzialität [befreien]“.

Seit jüngster Zeit ist der Hochfrequenzhandel in Deutschland reguliert<sup>9</sup>. Besonders perfide Handelspraktiken wurden zwar verboten, jedoch sind keine Mindesthaltefristen für Käufe vorgesehen.

Nach der relativ kurzen Beschreibung der Funktionsweise von HFT sollte deutlich geworden sein, dass autonom agierende Börsencomputer in höchstem Maße handlungswirksam sind. Gehandelt wird nicht mit Spielgeld oder virtuellem Geld, sondern

<sup>9</sup>Die Drucksache 17/11631 vom 26.11.2012: „Gesetz zur Vermeidung von Gefahren und Missbräuchen im Hochfrequenzhandel (Hochfrequenzhandelsgesetz)“ wurde am 28.02.2013 vom Bundestag beschlossen.

HFT hat im Verlustfalle Auswirkungen auf die Kunden einer Bank, die algorithmischen Handel verwendet und auf HFT vertraut. Auswirkungen haben die Aktivitäten von HFT auch auf andere menschliche und nicht-menschliche Händler, da durch Käufe und Verkäufe Aktienkurse verändert werden. Andere HFT können auf die Orders blitzschnell reagieren, während menschliche Händler mitunter gar nicht mitbekommen, wenn binnen einer Sekunde mehrere HFTs zu Beträgen von Millionen Dollar miteinander gehandelt haben. Vertrauen spielt in diesem Zusammenhang eine wesentliche Rolle, da menschliche Vertreter und Händler einer Bank die monetären Verluste und Gewinne zu verantworten haben. Sie müssen darauf vertrauen, dass die HFTs die von der Bank vorgegebenen Ziele möglichst optimal umsetzen, ohne auf Grund der Geschwindigkeit die Möglichkeit zu haben die Aktivitäten in Echtzeit kontrollieren oder überwachen zu können.

Mit Bezug auf die Vertrauensdefinitionen aus Kapitel 2 spielen im Kontext von HFT insbesondere Kompetenzvertrauen (die HFTs setzen die mathematischen Analysetechniken korrekt um), intentionales Vertrauen (die HFTs agieren im Sinne der Ziele der Bank bzw. Anleger) sowie Verhaltensvertrauen (auf die von HFTs durchgeführten Käufe und Verkäufe) eine Rolle. Die Entscheidung für den Einsatz von HFTs kann zudem als bewusste Vertrauenswahl interpretiert werden, als Entscheidung unter Risiko (Abschnitt 3.1.2): Sofern der Treuhänder (in dem Fall ein HFT) im Sinne der Bank agiert, wird die Situation der Bank durch den Einsatz von HFT verbessert. Im Falle eines Programmierfehlers oder eines unvorhersehbaren selbstverstärkenden Effekts kann die Bank durch den Einsatz von HFT auch ihre eigene Situation verschlechtern. In jedem Fall werden durch die enorme Geschwindigkeit von HFT Handlungen möglich, in dem Fall massenhafte Käufe und Verkäufe in kürzester Zeit, die ohne Vergabe von Vertrauen, ergo ohne die Nutzung von HFT, nicht erreicht worden wären.

### 5.2.2 (Teil-)autonomes Fahren

In den letzten Jahren ist eine enorm voranschreitende Automatisierung im Automobil-Sektor zu beobachten. So beschreibt etwa Charette (2009), dass ein aktuelles Oberklasse-Fahrzeug ca. 100 Millionen Zeilen Quellcode zum Betrieb der elektronischen Systeme ausführt, etwa zehnmal so viel wie das neue Großraum-Flugzeug von Boeing, der 787 Dreamliner<sup>10</sup>. Bei der Fahrzeug-Automatisierung geht es längst nicht mehr um einfache Komfort- und Warnfunktionen, sondern Fahrerassistenzsysteme (FAS) greifen *aktiv* in das Fahrgeschehen ein. Die aktuell verbauten FAS beziehen bei ihren Entscheidungen (bisher) in der Regel den menschlichen Fahrer mit ein und geben ihm die Möglichkeit, die Entscheidungen des Assistenzsystems zu übersteuern.

<sup>10</sup>Diese Feststellung klingt spektakulärer als sie tatsächlich ist. Speziell in der Avionik werden technische Systeme mit formal-mathematischen Methoden verifiziert. Vor dem Hintergrund ist die Anzahl der Codezeilen im Vergleich zur Fahrzeug- und Unterhaltungselektronik geringer, da eine formale Verifikation aufwendig ist.

Bei der adaptiven Geschwindigkeitsregulierung *Adaptive Cruise Control* (ACC) versucht das Assistenzsystem beispielsweise, eine eingestellte Wunschgeschwindigkeit zu halten und bremst nur im Falle eines langsamer vorausfahrenden Fahrzeugs ab. Trotz des aktiven Eingreifens von ACC in das Fahrgeschehen hat der Fahrer zu jedem Zeitpunkt die Möglichkeit, durch eine selbst initiierte Brems- oder Beschleunigungsaktion einzugreifen.

Der folgende Abschnitt wird zunächst in Rekurs auf einen aktuellen Forschungsbericht der Firma Bosch (vgl. Robert Bosch GmbH 2013) beschreiben, welche fortgeschrittenen Fahrerassistenzsysteme bereits heute eingesetzt werden oder kurz vor der Marktreife stehen und wie Anwender den Einsatz dieser Assistenzsysteme bewerten. Zwar werden im genannten Bericht Anwender befragt, kritische Fragen, etwa nach einem möglichen Kontrollverlust durch zunehmende Fahrzeug-Automatisierung (vgl. Weyer u. a. 2014) bleiben allerdings weitgehend aus.

Glaukt man den Technik-Optimisten aus dem Silicon Valley (vgl. Vanderbilt 2012) werden die Straßen in wenigen Jahrzehnten von vollständig autonomen Fahrzeugen bevölkert sein, bei denen sich die Fahrer dann voll und ganz auf ihr „Digital Life“, etwa die Präsenz und Aktivität auf Social Networking Sites (SNS) wie Facebook, konzentrieren können (ebd.: 124). Die Firma Google hat einen funktionsfähigen Prototyp, das „Google Car“, entwickelt, das bereits über 300 000 Kilometer autonom auf den Straßen Kaliforniens gefahren ist (vgl. Lindner 2012).

Abschließend wird argumentiert, dass bereits jetzt davon gesprochen werden kann, dass Fahrerassistenzsysteme handlungs- bzw. fahrwirksam sind und es in zunehmendem Maße sein werden.

### **Fahrerassistenzsysteme aus Nutzersicht**

Ein kürzlich erschienener Bericht der Robert Bosch GmbH (2013) ist in zweierlei Hinsicht interessant: Erstens gilt Bosch als wichtiger Hersteller für Fahrerassistenzsysteme und ist selbst aktiv mit der Entwicklung ebendieser Systeme betraut. Die dortige Beschreibung eingesetzter und künftiger Assistenzsysteme gibt somit adäquat Auskunft über den aktuellen Entwicklungsstand. Zweitens beschreibt Bosch in dem Bericht die Ergebnisse einer quantitativen Studie, bei der vor allem nach den Einschätzungen und Bedürfnissen von Endanwendern gefragt wurde.

Wohl wissend, dass Hersteller von Fahrerassistenzsystemen mit derartigen Studien ein gewisses Eigeninteresse verfolgen, scheint die Studie methodisch solide angelegt und gibt einen guten Überblick über den aktuellen technischen Stand von Fahrerassistenzsystemen und darüber, in welche technologische Richtung sich Fahrerassistenzsysteme weiterentwickeln könnten.

Ausgangspunkt für die Autoren der Studie ist die Feststellung, dass 90 % aller Unfälle auf menschliches Fehlverhalten zurückzuführen sind, bei dem der Fahrer „too slowly or incorrectly“ (ebd.: 3) agiert hat. Die (kaufmännisch logische) Antwort von Bosch

auf diesen Befund ist das Versprechen, durch weitere Automatisierung des Fahrens zu einer Erhöhung der Sicherheit beizutragen.

Bosch hat in dem Forschungsbericht die Nutzer von Fahrerassistenzsystemen als Zünglein an der Waage identifiziert und möchte herausfinden, welche FAS besonders geschätzt und gewünscht werden und für welche Assistenzsysteme eine Bereitschaft besteht, Geld auszugeben. Ambivalent ist insbesondere der Wunsch der Fahrer nach Assistenzsystemen, die aktiver und stärker in das Fahren eingreifen (ebd.: 12) und dennoch übersteuerbar und deaktivierbar sind (ebd.: 15).

In der empirischen Studie wurden 2 261 Online-Fragebögen quantitativ und 120 Tiefeninterviews qualitativ ausgewertet. Befragt wurden dabei nur solche Personen, die sich in den letzten drei Jahren ein neues Auto (allerdings kein Oberklasse-Fahrzeug) gekauft haben<sup>11</sup>.

Inhaltlich ging es in der Befragung um Informationen über die alltägliche Fahrerfahrung und den damit verbundenen Aufmerksamkeitslevel, um Kenntnis moderner Assistenzsysteme sowie schließlich um Einschätzungen und Einstellungen bzgl. eines vollständig autonomen Fahrens. Im folgenden Abschnitt werden einige wenige ausgewählte Ergebnisse nebst Fahrerassistenzsystemen vorgestellt, die sich hinsichtlich ihrer Handlungswirksamkeit besonders auszeichnen.

### **Die Ambivalenz zwischen Autonomie und Kontrolle**

Die Robert Bosch GmbH (2013: 7) stellt zunächst fest, dass Fahrer glauben, stets ihr Fahrzeug unter Kontrolle zu haben, obwohl 90 % der Unfälle auf menschliches Fehlverhalten zurückzuführen seien. Zwei Drittel der Probanden geben an, dass sie in den letzten zwölf Monaten mindestens einmal eine Notbremsung wegen eines plötzlichen Hindernisses durchführen mussten (ebd.: 9). Ein hohes Stresslevel wird vor allem von Fahrern kleinerer Fahrzeuge bekundet: Insbesondere werden von diesen Fahrern plötzlich notwendige Bremsmanöver beim Durchqueren von Baustellen als besonders stressbehaftet empfunden (ebd.: 8).

Im Bericht wird mehrfach hervorgehoben, dass die Befragten eine enorm gute Kenntnis über FAS besitzen, wenngleich ihr eigenes Auto nicht unbedingt mit diesen ausgestattet ist. So kennen etwas 90 % der Befragten einen Spurhalteassistenten, nur 12 % haben einen solchen im eigenen Auto eingebaut. Selbst sehr neuartige und eher selten verwendete Assistenzsysteme sind bekannt: 79 % der Befragten kennen eine automatische Fahrerbenommenheitserkennung (*Driver Drowsiness Detection*), die den Wachheitszustand anhand des Lidschlages des Fahrers überprüft und bei zunehmender Müdigkeit Lenkradvibrationen auslöst. Diese Automatik befindet sich in nur 4 %

---

<sup>11</sup>Eine derartige Auswahl verfälscht einige Fragen z. B. nach der Bekanntheit von Fahrerassistenzsystemen erheblich, da davon ausgegangen werden kann, dass sich gerade solche Personen intensiv mit auf dem Markt verfügbaren FAS beschäftigt haben und diese entsprechend kennen.

der Fahrzeuge der Befragten und dennoch ist die Automatik einer Mehrzahl der Fahrer bekannt.

In der durchgeführten Studie unterbleiben fast ausnahmslos FAS-kritische Fragen, allerdings deutet sich an, dass die Fahrer, ähnlich wie schon bei Weyer u. a. (2014) thematisiert, einen eventuellen Kontrollverlust durch die zunehmende Automatisierung fürchten: So stimmen 44 % der Befragten der Aussage zu oder voll zu, dass Fahrerassistenzsysteme dazu führen, dass man ggf. vergisst, wie man richtig Auto fährt oder korrekt in schwierigen Situationen reagiert (ebd.: 12).

Bei der Frage nach dem wichtigsten Assistenzsystem landen ausnahmslos solche Systeme auf den ersten Plätzen, die helfen, Situationen mit einem hohen Gefährdungspotenzial zu vermeiden: 12,3 % halten eine Notfallbremse zum Schutz von Fußgängern, 11,9 % eine Notfallbremse für Geschwindigkeiten über 50 km/h und 10,8 % eine Totwinkelerkennung für das wichtigste Assistenzsystem.

Neben bereits existierenden Assistenzsystemen, die Fahrer bei der Bewältigung der Fahraufgaben auf verschiedene Weisen unterstützen (s.u.) wird von besonders technikaffinen Entwicklern bereits die Vision des vollständig autonomen Fahrens beschrieben (vgl. Vanderbilt 2012). So interessiert sich auch die vorliegende Studie für die Einstellung der Fahrer zu vollständig autonomen Fahren: Nur 14 % der Befragten glauben, dass autonomes Fahren technisch nicht möglich sein wird, 48 % glauben, dass Autos zukünftig autonom fahren werden, gleichzeitig halten es über 50 % für wichtig, selbst verantwortlich zu sein und das automatische System jederzeit ausschalten zu können (ebd.: 15). Ebenso wie bei den Fragen zu stärker eingreifenden Assistenzsystemen zeigt sich, dass die Befragten jederzeit die Kontrolle behalten wollen, obwohl dies sowohl im Widerspruch zur Idee des *vollständig* autonomen Fahrens als auch im Widerspruch zu aktiv eingreifenden Assistenzsystemen (z. B. der Notfallbremse) steht, die vor allem einen zusätzlichen Sicherheitsaspekt durch schnelle Reaktionszeiten gewährleisten sollen. Inagaki (2008) schlägt zur Auflösung dieses Dilemmas eine Art der adaptiven Automation vor, die in Echtzeit den mentalen und physischen Zustand des Fahrers erfasst (ebd.: 257) und im Gefahrenfall eingreift, denn „[...] humans may not always be powerful enough to cope with any given situation“ (ebd.: 260).

Eine wesentliche Herausforderung, so meine Prognose, wird für die Automobilhersteller darin bestehen, das ambivalente Verhältnis zwischen Autonomie und Kontrolle so auszutarieren, dass eine hohe Akzeptanz bei den Kunden gewährleistet bleibt.

Aktiv eingreifende Assistenzsysteme tangieren in besonderem Maße unterschiedliche Aspekte des Vertrauens. Fehlfunktionen der Assistenzsysteme können fatale Folgen haben. Deswegen bedingt der Einsatz von autonomer Technik mit interventionistischem Charakter eine Entscheidung unter Risiko (im Sinne des probabilistischen Vertrauensbegriff, vgl. Abschnitt 3.1). Von dispositioneller Bedeutung (vgl.

Abschnitt 2.2.3) sind insbesondere Verhaltensvertrauen, Kompetenzvertrauen, intentionales Vertrauen sowie Informationsvertrauen. Das Verhaltensvertrauen bezieht sich darauf, dass ein Eingriff des Assistenzsystems eine tatsächliche Wirkung entfaltet. Kompetenzvertrauen beinhaltet die Zuschreibung, dass das Assistenzsystem einen notwendigen Eingriff korrekt durchführt, z. B. auf eine bestimmte Geschwindigkeit abbremst. Intentionales Vertrauen rekurriert auf bestimmte zugeschriebene Gründe für einen Eingriff. Schließlich bezieht sich Informationsvertrauen darauf, dass die Gründe für einen Eingriff, z. B. zunehmende Müdigkeit des Fahrers, korrekt angezeigt werden.

### Übersicht ausgewählter Assistenzsysteme

Der folgende Abschnitt gibt einen Eindruck, welches Maß an Unterstützung (und damit partieller Handlungswirksamkeit) bereits mit existierenden Assistenzsystemen möglich ist (vgl. Robert Bosch GmbH 2013: 16):

**Verkehrszeichenerkennung** (*Road Sign Recognition with speed control*): Externe Kameras nehmen die Umgebung des Fahrzeugs auf und erkennen vornehmlich durch rote Kreise gekennzeichnete zulässige Höchstgeschwindigkeiten. Diese werden im Display des Fahrzeugcockpits eingeblendet. Im Falle einer Geschwindigkeitsüberschreitung wird darauf hingewiesen und, falls gewünscht, automatisch abgebremst.

**Spurhalteassistent** (*Lane Keeping Support*): Mit moderaten Steueraktionen wird das Fahrzeug automatisch in der Spur gehalten. Die Spurerkennung erfolgt auf Basis von Bilderkennungsverfahren. Das System kann jederzeit abgeschaltet werden und soll partielle Unaufmerksamkeit (z. B. durch Müdigkeit des Fahrers) kompensieren.

**Baustellenassistent** (*Construction Zone Assist*): Für enge Baustellen warnt das FAS vor kritischen Situationen, in denen sich beispielsweise Fahrzeuge einander annähern.

**Totwinkel-Erkennung** (*Blind Spot Detection*): Sofern sich ein anderes Fahrzeug im toten Winkel befindet, leuchtet im Seitenspiegel eine Warnlampe.

**Ausweichsteuerungsunterstützung** (*Evasive Steering Support*): Falls die unmittelbare Gefahr eines Auffahrunfalls besteht, überprüft das Fahrzeug, ob eine der benachbarten Spuren frei ist und steuert das Fahrzeug auf diese.

**Querverkehr-Alarm** (*Rear Cross Traffic Alert*): Auf Parkplätzen überprüft das System beim Ausparken, ob von links oder rechts Autos kommen und bremst im Gefahrenfall sofort ab, um eine Kollision zu vermeiden.

**Adaptive Leuchthöhenverstellung** (*High Beam Control*): Bei entgegenkommenden Fahrzeugen wird die Leuchthöhe automatisch heruntergestellt, um ein Blenden entgegenkommender Fahrzeuge zu verhindern.

**Notfallbremse bei niedrigen Geschwindigkeiten** (*Automatic Emergency Braking at low speeds*): Dieser Typ Notfallbremse wird vor allem in Innenstädten aktiv, um Auffahrunfälle zu vermeiden (bei plötzlichem Bremsen des vorausfahrenden Fahrzeugs).

**Notfallbremse bei höheren Geschwindigkeiten** (*Automatic Emergency Braking at higher speeds*): Diese Notfallbremse wird bei Geschwindigkeiten über 50 km/h aktiv und funktioniert mehrstufig: Zunächst wird vor einer möglichen Kollision gewarnt, bei Nicht-Reaktion wird automatisch der notwendige Bremsdruck aufgebaut, bleibt eine Bremsreaktion des Fahrers immer noch aus, erfolgt eine vollautomatische Notbremse.

**Notfallbremse für Fußgänger** (*Automatic Emergency Braking of pedestrians*): Plötzlich auftauchende Fußgänger erkennt das System, warnt und bremst falls notwendig automatisch um den Aufprall zu verhindern oder zumindest abzumildern.

**Einparkassistent** (*Parking Maneuver Assist*): Dieses Assistenzsystem startet auf Knopfdruck eine Einparkaktion in eine freie Parklücke oder in eine Garage.

**Adaptive Geschwindigkeitsregulierung** (*ACC*): Über die Fahrzeugkonsole kann der Fahrer eine Wunschgeschwindigkeit einstellen. Sollte es wegen eines vorausfahrenden Fahrzeugs nicht möglich sein, die Wunschgeschwindigkeit zu halten, orientiert sich das Assistenzsystem unter Berücksichtigung eines konfigurierbaren Sicherheitsabstands an der Geschwindigkeit des vorausfahrenden Fahrzeugs.

**Adaptive Geschwindigkeitsregulierung innerstädtisch** (*ACC Stop&Go*): Im Gegensatz zum Standard-ACC-System (s.o.) ermöglicht es ACC Stop&Go auch in Zeiten hoher Verkehrsdichte bis zum Fahrzeugstillstand abzubremsen. Als Orientierung dient ebenfalls das vorausfahrende Fahrzeug.

**Stauassistent** (*Traffic Jam Assist at up to 60 km/h*): Das System führt Brems- und Beschleunigungsaktionen bis zur Geschwindigkeit von 60 km/h durch. Der Fahrer kann jederzeit die Aktionen des Stauassistenten übersteuern.

**Kreuzungsassistent** (*Intersection Assist*): Beim Linksabbiegen überprüft das System, ob beim Abbiegen eine Kollision mit dem Gegenverkehr stattfindet und bremst im Gefahrenfall abrupt ab.

Nach der Vorstellung einiger moderner Fahrassistenzsysteme ist besonders deutlich geworden, dass diese zunehmend in das Fahrverhalten eingreifen.

Inagaki und Itoh (2013) sprechen in diesem Fall von sogenannten Advanced Driver Assistance Systems (ADAS). Bei einer typischen Fahraufgabe (ebd.: 1) sind folgende vier ständig wiederholende Tätigkeiten erforderlich: *perception* (a), *cognition* (b), *action selection* (c) und *action implementation* (d). Weiter wird argumentiert, dass für die Tätigkeiten (a)–(c) ein menschenzentrierter („human-centered“) Automationsansatz genügt, bei dem die letztliche Entscheidung über den Handlungsvollzug (*action implementation*) beim menschlichen Fahrer liegt. Moderne Fahrerassistenzsysteme, und damit schließt sich der Bogen zu den oben vorgestellten Systemen, zielen zunehmend darauf hin, auch Schritt (d) zu automatisieren, deswegen sind „careful investigations [...] needed regarding to what extent the system may be given authority for deciding and acting autonomously without asking the human driver’s approval or consent [...]“ (ebd.: 2).

Im Falle der Bremsassistenten erfolgt im Gefahrenfalle eine Bremsung, im Falle von ACC eine Geschwindigkeitsanpassung, bei der Ausweichsteuerungsunterstützung sogar ein aktiver Eingriff in die Lenkung des Fahrzeugs. Außerdem ist zu erkennen, dass sich die Assistenzsysteme nicht verstecken und im Hintergrund arbeiten (wie etwa bei ESP<sup>12</sup>), sondern aktiv mit dem Nutzer interagieren und auf Eingaben von diesem warten. Einer besonders bedienerfreundlichen Mensch-Maschine-Schnittstelle kommt folglich eine wichtige Bedeutung zu. Letztlich hat sich gezeigt, dass selbst die Architektur modernster Assistenzsysteme in der Regel eine Deaktivierung oder Übersteuerung durch den menschlichen Fahrer vorsieht. Vor dem Hintergrund wird die Frage virulent, ob dem Fahrer durch derartige Assistenzsysteme und die Notwendigkeit ihrer ständigen Überwachung (und evtl. Eingriffs) nicht zusätzliche Belastungen zugemutet werden, die das Fahren insgesamt nicht erleichtern, sondern zukünftig eher erschweren. Weyer (1997: 244) hat diesen Typ von Tätigkeit für die Cockpit-Automation (im Flugzeug) als „Gewährleistungsarbeit“ bezeichnet. Gemäß den Ergebnissen der Robert Bosch GmbH (2013: 11) geben immerhin 60 % der Befragten an, dass Fahrerassistenzsysteme den Fahrkomfort steigern und das Fahren insgesamt entspannter machen. Ob diese Zustimmungsraten auf dem hohen Niveau bleiben, wenn sich die Anzahl der zu überwachenden Systeme stetig erhöht, bleibt abzuwarten.

Ein möglicher technologischer Entwicklungspfad im Bereich der Fahrerassistenzsysteme ist die Vision des vollständig autonomen Fahrens: Der menschliche Fahrer gibt lediglich den Zielort ein und alle zur Erreichung des Zielortes notwendigen Aktionen

---

<sup>12</sup>Die Abkürzung ESP steht für Elektronisches Stabilitätsprogramm.

erledigt das Fahrzeug selbsttätig. Um diese, zugegebenermaßen recht technikaffine, aber dennoch nicht unrealistische<sup>13</sup> Vision geht es im nächsten Abschnitt.

### Vision des autonomen Fahrens und des fehlerbehafteten Menschen

In der technikaffinen Diskussion um das Automobil wird längst nicht mehr nur leise über die Vision des vollständig autonomen Fahrens nachgedacht. Das US-Technologie-Magazin Wired, das sich selbst als „[...] the first word on how ideas and innovation are changing the world“<sup>14</sup> versteht, hat einen Artikel von Tom Vanderbilt (2012) veröffentlicht, in dem ausgehend von dem angeblich unvollkommenen und fehlerbehafteten Menschen eine technisch schon längst mögliche Ablösung durch autonome Fahrzeuge propagiert wird: „Self-driving cars may seem like science fiction, but most of the technology already exists. [...] It never panics. It never gets angry. It never even blinks. In short, it is better than human in just about every way“ (ebd.: 90f.). Der Autor ist sogar der Ansicht, dass nach fast hundert Jahren, in denen Autofahren weitgehend unverändert geblieben ist, in den letzten fünf Jahren eine nahezu vollständige Transformation stattgefunden hat: „In short, after almost a hundred years in which driving has remained essentially unchanged, it has been completely transformed in just the past half decade“ (ebd.: 88).

Abgesehen von dem relativ naiven Technikoptimismus, der zwischen den Zeilen mitschwingt, bietet der Artikel interessante Einblicke in den Entwicklungsstand des autonomen Fahrens. Die Entwicklung autonomer Fahrzeuge wird massiv von der Firma Google im Rahmen ihres Projekts *Google Car* gefördert. Das Google-Auto, ein umgebauter Toyota Prius, hat nach Angaben des Chefsingenieurs Chris Urmson bereits im August 2012 insgesamt 500 000 selbsttätig gefahrene Kilometer absolviert<sup>15</sup>. Google stellt damit den klassischen ingenieurwissenschaftlichen Ansatz, der auf realweltliche Probleme die jeweils passende technisch-analytische Antwort sucht, gewissermaßen auf den Kopf: Gemäß der tief verwurzelten Google-Philosophie „[t]he data can make better rules“ vertritt Google die Auffassung, das selbstfahrende Auto „[...] needs to learn how people *really* drive“ (Vanderbilt 2012: 92, Herv. R. F.). Autonomes Fahren wird neben den technisch zu lösenden Aspekten folglich vor allem zu einem Data-Mining-Problem, das darin besteht, die „vernünftigen“ Regeln für gutes Autofahren herauszufinden. Dabei liegt die Idee zugrunde, dass sich menschliche Autofahrer nicht strikt nach Regeln verhalten, sondern bestimmte angepasste Verhaltensweisen

<sup>13</sup>Die Vision erscheint durchaus realistisch, da Unternehmen wie Bosch bereits jetzt über einen derartigen Entwicklungspfad nachdenken und mit der vorgestellten Studie beispielsweise ausloten möchten, welcher Automationsgrad (bis hin zum vollständig autonomen Fahren) ohne größeren Protest der Autofahrer realistischerweise umzusetzen ist (vgl. Robert Bosch GmbH 2013).

<sup>14</sup>[http://www.wired.com/about/press\\_about/](http://www.wired.com/about/press_about/), Abruf: 31. Januar 2014.

<sup>15</sup><http://googleblog.blogspot.hu/2012/08/the-self-driving-car-logs-more-miles-on.html>, Abruf: 31. Januar 2014.

und Problemlösungsstrategien herausgebildet haben, die sich der Algorithmus des Google-Autos zu Nutze machen kann.

Interessant ist dabei die Frage, nach welchem Ideal zukünftiges Autofahren ausgerichtet sein wird. Orientiert sich Google bei der Entwicklung an menschlichen Fahrern, bedeutet dies eine strikt symmetrische Sichtweise: Algorithmen sollen sich genauso verhalten wie Menschen und in ihrer Gesamtwirkung von einem menschlichen Fahrer nicht unterscheidbar sein. Denkt man diesen datengetriebenen Ansatz konsequent zu Ende und geht davon aus, dass zu irgendeinem Zeitpunkt nur noch autonome Fahrzeuge die Straßen bevölkern, werden diese ganz ähnlich fahren wie die heutigen menschlichen Fahrer. Ob es allerdings sinnvoll ist, ein aus ausschließlich autonomen Fahrzeugen bestehendes System Straßenverkehr am Ideal des heutigen menschlichen Fahrens zu orientieren, erscheint fraglich, zumal Vanderbilt (2012: 90) die Qualitäten menschlicher Fahrer für defizitär hält.

Erkenntnistheoretisch wird die Frage interessant sein, und dieser Gedanke soll hier nur kurz angedeutet werden, ob auch Algorithmen von anderen Algorithmen lernen können und damit die Gesamtperformanz des zukünftigen Straßenverkehrssystems weiter gesteigert werden kann, oder ob das System schließlich in seiner Selbstreferenzialität „gefangen“ bleibt. Unterstellt man, dass Algorithmen echt lernfähig sind und Adaptivität herausbilden können, so wäre die konsequente Folgerung, dass auch Algorithmen (genauer: autonome Agenten mit jeweils eigenem Erfahrungsschatz und eigenem Gedächtnis) voneinander lernen können. Vanderbilt (2012: 90) beschreibt die Funktionsweise des autonomen Google-Auto bereits heute wie die eines wissbegierigen und lernfähigen Autofahrers, der ständig die Umgebung beobachtet, diese interpretiert und darauf Schlussfolgerungen ableitet.

Der futuristische Text kann ähnlich eingeordnet werden wie Marc Weisers Vision des eingebetteten allgegenwärtigen Computers (Ubiquitous Computing), der damals versprach, dass „[m]achines that fit the human environment instead of forcing humans to enter theirs will make using a computer as refreshing as taking a walk in the woods.“ (Weiser 1991: 11). Für das autonome Fahren heißt es ähnlich verheißungsvoll bei Vanderbilt (2012: 91): „It is only now, well more than a century after the invention of the so-called automobile – an independently powered car – that we are finally developing a *true automobile*, a car that can drive itself“ (Herv. R. F.). Die jüngeren Entwicklungen um Smartphones und Tablet-PCs haben Marc Weiser zwar weitgehend in seiner *technischen* Voraussicht recht gegeben, allerdings führte die zunehmende Durchdringung des Alltags mit Computern nicht zu der erhofften Entlastung, sondern wird zunehmend als Belastung empfunden (Carr 2010). Schließlich bleibt abzuwarten ob und in welcher Form das prophezeite „true automobile“ tatsächlich Wirklichkeit werden wird.

### 5.2.3 Wikipedia Bots – Nicht-menschliche Mitglieder der Wikipedia-Gemeinschaft<sup>16</sup>

Aktivitäten autonomer Softwareagenten finden ihren Vollzug nicht ausschließlich im virtuellen Raum, sondern sind innerhalb hybrider Akteurskonstellationen mit Aktivitäten menschlicher Akteure gekoppelt und damit Gegenstand der Soziologie. Dabei wird ihnen nicht nur die Rolle passiver technischer Artefakte zuteil, sondern sie sind bereits als aktive Mitspieler einer hybriden Gemeinschaft, im hier betrachteten Fall der Wikipedia-Gemeinschaft, konzipiert.

#### Was sind Bots?

Unter Bots versteht man nicht, wie die Begriffsnähe zu „Robot“ und Darstellungen von Bots, etwa in Filmen wie „Krieg der Sterne“ andeuten könnten, humanoide Roboter, sondern vielmehr Software, die im Dienste ihrer Programmierer insbesondere repetitive Tätigkeiten (teil-)autonom erledigt. Im Falle der Wikipedia nehmen Bots zumeist kleinere Editierungen vor, geben Hinweise auf formale Mängel und Inkonsistenzen in Artikeln. Die Wikipedia-Gemeinde charakterisiert Bots als „[...] Computerprogramme oder Skripte, die ihren Betreibern stupide, langweilige und häufig auftretende Aufgaben abnehmen“<sup>17</sup>.

Bots tauchen in der Wikipedia, ebenso wie menschliche Nutzer, als angemeldete Benutzer auf. Sie sind allerdings durch ihren Namen und durch eine spezielle Kennzeichnung als Bots erkennbar. In der Regel lassen sich Bots über ihre Selbstbeschreibungseite (vgl. Abbildung 5.1) eindeutig einem Programmierer zuordnen. Neben dieser Zuordnung finden sich auf der Selbstbeschreibungseite in der Regel Angaben zur Funktionsweise des Bots und in manchen Fällen auch der Quelltext. Einige Bots sehen über ihre Selbstbeschreibungseite eine Sperrfunktion vor, die bei Bot-Fehlfunktion (z. B. massenhaft falschen Editierungen) vom menschlichen Nutzer genutzt werden kann.

**Namensräume** (engl.: namespaces) sind ein zentrales Konzept der Wikipedia, um unterschiedliche Bereiche, in denen editiert wird, voneinander abzugrenzen. Jedem Namensraum ist eine eindeutige Nummer zugeordnet (für Details s. Tabelle A.1 im Anhang). Die Haupteditieraktivität spielt sich im Namensraum 0, dem Artikel-Namensraum, ab. Passive Nutzer der Wikipedia (Leser) kommen hauptsächlich mit diesem Namensraum in Berührung. Diskussionen zu Artikeln sind einem anderen Namensraum zugeordnet als die Artikel selber. Die Berücksichtigung der Namensraumaktivität als Merkmal gibt Aufschluss darüber, in welchen Bereichen editiert wird bzw. um welche Art der Editierung es sich handelt. Editierungen im Artikel-

<sup>16</sup>Der folgende Abschnitt ist angelehnt an Fink und Liboschik (2010) und Fink und Liboschik (2012) und erweitert um aktuelle Entwicklungen.

<sup>17</sup>Quelle: <http://de.wikipedia.org/wiki/Wikipedia:Bots> (Abruf: 31. Januar 2014).

## Benutzer:PortalBot

Hallo Leute!

Ich bin ein reiner Pywikipediabot. Gesteuert werde ich vom Benutzer:Augiasstallputzer.

+	Dieser Benutzer ist ein <b>Bot</b> .
<b>Wichtige Hinweise für andere Benutzer</b>	
Wer möchte, dass ein Bot etwas für ihn erledigt, der soll <a href="#">hier</a> anfragen.	
<b>Wenn der Bot läuft, dann bin ich grundsätzlich unter Benutzer Diskussion:Augiasstallputzer zu erreichen.</b>	
Wenn nach einer Verschiebung das Redirect gelöscht werden soll, dann fixiere ich gelegentlich auch Links auf Benutzerseiten, denn es ist für einen Benutzer in aller Regel besser, wenn ein Bot einen Link ändert, als dass dieser Link plötzlich tot ist und der Benutzer nicht weiß, ob und welchen Ersatz es gibt. Wer das trotzdem nicht möchte, der kann auf seiner Benutzerseite, am Besten ganz oben, den Kommentar: <code>&lt;!--NOBOTLINKFIX--&gt;</code> einfügen. Mein Bot überspringt diese Seite dann.	

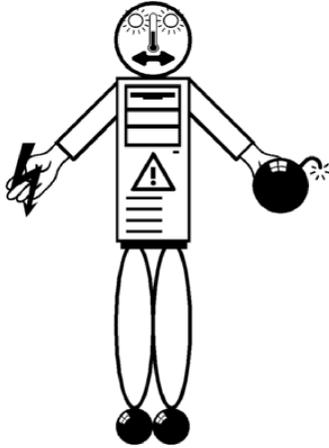


Abbildung 5.1: Selbstbeschreibungseite eines Bots

Namensraum sind von besonderer Bedeutung, da mit dortigen Änderungen der weit größte Teil der Wikipedia-Nutzerschaft, die passiven Leser, in Berührung kommen. In anderen Namensräumen spielen sich z. T. interne Verwaltungs- und Koordinationsprozesse ab, die für das Wikipedia-Projekt zwar wichtig sind, aber für die passiven Nutzer eher „hinter den Kulissen“ stattfinden.

### Was macht Bots als Forschungsobjekt interessant?

Ein erster quantitativer Überblick zeigt, dass pro Woche durchschnittlich etwa 60 Bots aktiv sind, die etwa 15 % des gesamten Editiervolumens<sup>18</sup> in der Wikipedia ausmachen. Ihre Bedeutung ist somit auf ähnlich hohem Niveau wie das Editiervolumen der unregistrierten Nutzer, das bei etwa 10 % liegt.

Im Falle der Wikipedia werden Aufgaben wie z. B. Tippfehlerkorrekturen oder Kategorisierungen, die durchaus zentral organisiert werden könnten, durch Bots dezentral gelöst. Teilweise werden Bots auf den Computern der Programmierer und damit unabhängig von der Wikipedia-Infrastruktur ausgeführt. In vielen Fällen erfolgt auch die

<sup>18</sup>Die 15.2 % beziehen sich auf einen Zeitraum aus 2010 ( $Z_1$ : 01.02.2010–30.04.2010). In einer neueren Stichprobe ( $Z_2$ : 01.11.2012–31.01.2013) liegt der Anteil etwas höher, nämlich bei 18.6 %, sofern alle Editierungen (auch außerhalb des Artikel-Namensraums) berücksichtigt werden und bei 16.1 % wenn nur Editierungen an Artikeln berücksichtigt werden. Für Details und einen Vergleich zwischen den beiden Zeiträumen vgl. Tabelle 5.2 auf Seite 123.

Erstellung eines Bots nach dem Wiki-Prinzip. Eine Idee für einen Bot wird gemeinsam diskutiert und nach gründlicher Diskussion dann softwaretechnisch umgesetzt<sup>19</sup>.

In einer näheren Analyse der Aktivitäten der Bots, die u.a. Tätigkeitsfelder benennt, in denen Bots besonders aktiv sind, stellt sich heraus, dass deren Editierungen eher „lektorativen“ Charakter haben, was zunächst vielleicht wenig überraschend ist. Die zur Zeit aktive Bot-Generation ist noch weit davon entfernt, autonom längere, inhaltlich gehaltvolle Textbeiträge zu verfassen. Die Handlungswirksamkeit von Bots wird allerdings dann interessant, wenn man sie im Kontext hybrider Interaktionsketten, also beim Zusammenwirken von Bots und menschlichen Nutzern, betrachtet. Ein Beispiel wird dabei anschaulich zeigen, wie Bots Initiatoren derartiger Ketten werden und damit handlungswirksam sind. Ein weiteres Beispiel wird zeigen, dass Bots z. T. bereits hybrid konzipiert, d. h. auf die Zusammenarbeit von menschlichen und nicht-menschlichen Akteuren ausgelegt sind.

### Tätigkeitsfelder der Bots

Eine hier nicht näher dargestellte quantitative Analyse<sup>20</sup> der Bot-Aktivitäten zeigt, dass diese in sehr unterschiedlichen Bereichen der Wikipedia aktiv sind. Um die Aktivitäten der Bots weiter zu präzisieren zeigt Tabelle 5.1 das Ergebnis einer qualitativen Analyse der 112 im Zeitraum  $Z_1$  (Februar bis April 2010) aktiven Bots. Eine erste Durchsicht der Bot-Editierungen und einschlägiger Informationen aus der Wikipedia<sup>21</sup> führte zu einer Liste potenzieller Bot-Tätigkeitsfelder (1. Tabellenspalte von Tabelle 5.1). Betrachtet wurden dann alle 112 Bots, die aktiv waren, also mindestens eine Editierung vorgenommen haben. Um etwaige Veränderungen herauszustellen, wurde die Tabelle aus der früheren Veröffentlichung um einen weiteren Zeitraum  $Z_2$  (November 2012 bis Januar 2013) ergänzt. Neben den Selbstbeschreibungsseiten der Bots, die allerdings hinsichtlich der durchgeführten Tätigkeiten z. T. übertrieben viele angebliche Tätigkeiten auflisten, wurden vor allem die Versionshistorien<sup>22</sup> herangezogen, die sämtliche Änderungs-Details enthalten. Auf Basis dieser Informationen lassen sich die Bots einem oder mehreren der in der Tabelle aufgeführten Tätigkeitsfelder zuordnen. Die Information in der zweiten Tabellenspalte gibt an, wie viele Bots in dem jeweiligen Tätigkeitsfeld aktiv waren. Dabei wurde bewusst nicht auf die formalen Namensräume als Differenzierungsmerkmal zurückgegriffen, da die Angabe,

<sup>19</sup>Für ein Beispiel einer Diskussion über Bot-Architektur und -Funktionsweise s. [http://de.wikipedia.org/wiki/Vorlage\\_Diskussion:Infobox\\_Gemeinde\\_in\\_Frankreich/Archiv/2010#Botlauf\\_M.C3.A4rz\\_2010](http://de.wikipedia.org/wiki/Vorlage_Diskussion:Infobox_Gemeinde_in_Frankreich/Archiv/2010#Botlauf_M.C3.A4rz_2010) (Abruf: 31. Januar 2014).

<sup>20</sup>Für Details vgl. Fink und Liboschik (2010: 18).

<sup>21</sup>Liste der Bots, Link: [https://de.wikipedia.org/w/index.php?title=Wikipedia:Bots/Liste\\_der\\_Bots&oldid=112996843](https://de.wikipedia.org/w/index.php?title=Wikipedia:Bots/Liste_der_Bots&oldid=112996843) (Abruf: 31. Januar 2014).

<sup>22</sup>Zu jedem Wikipedia-Artikel existiert mit der Versionshistorie eine feingranulare Auflistung von allen jemals vorgenommenen Änderungen. So lässt sich im Nachhinein präzise rekonstruieren, wer, wann an welchem Artikel welche Editierung vorgenommen hat.

Tätigkeitsfeld	# $Z_1$	# $Z_2$
Interwiki-Links setzen/ändern/entfernen	86	76
Administrative Aufgaben	10	15
Links korrigieren	9	5
Kategorien verschieben/umbenennen/löschen	6	4
Vorlagen aktualisieren (z.T. mit externen Daten)	5	2
Rechtschreib- und Typographiefehler korrigieren	4	1
Bilder auf korrekte Auszeichnung überprüfen	3	3
Listen pflegen	2	5
Vandalismus erkennen	1	2
Redaktionelles	0	5
<i>Artikel verfassen</i>	0	0

**Tabelle 5.1:** Tätigkeitsfelder der Bots  
( $Z_1$ : 01.02.2010–30.04.2010 ;  $Z_2$ : 01.11.2012–31.01.2013)

dass ein Bot beispielsweise ausschließlich im Artikel-Namensraum editiert hat, keine Auskunft darüber gibt, *welche Tätigkeit* er dort genau durchgeführt hat.

Da sich die Zeiträume  $Z_1$  und  $Z_2$  bezüglich der Verteilung auf die unterschiedlichen Tätigkeitsfelder sehr ähnlich sind, beziehen sich die folgenden Ausführungen ausschließlich auf  $Z_1$ :

Auffällig an der Tabelle ist zunächst, dass ein Großteil, nämlich 86 von 112 Bots, sogenannte *Interwiki-Bots* sind. Interwiki-Bots kümmern sich um die Verknüpfung von thematisch deckungsgleichen Artikeln in unterschiedlichen Sprachversionen der Wikipedia. Der Grund für die Vielzahl von Interwiki-Bots liegt zum einen darin, dass ein standardisiertes Framework zur Erstellung von Interwiki-Bots existiert, womit die Hürde zur Programmierung eines solchen Bots relativ niedrig ist. Zum anderen sind angemeldete Interwiki-Bots häufig keine nativen Bots, d.h. Helfer von vornehmlich in der deutschsprachigen Wikipedia aktiven Programmierern, sondern entstammen den zahlreichen fremdsprachlichen Versionen.

Weitere Tätigkeitsfelder sind *administrative Aufgaben*, die sich mit internen Wikipedia-Abläufen befassen, z. B. Archivierung, Artikelumbenennungen und Organisation und Zusammenfassung umfangreicher Diskussionen. Wie in Tabelle 5.1 genannt, korrigieren neun Bots fehlerhafte *Links*. Dies ist z. B. notwendig, wenn sich interne Artikel- oder Bildernamen ändern oder wenn externe Web-Ressourcen nicht mehr verfügbar sind. Wikipedia verfügt über ein umfangreiches *Kategoriensystem*, das von aktuell sechs Bots unterstützend gepflegt wird. *Vorlagen*, die in Artikeln verwendet werden, vereinheitlichen Artikel (z. B. für tabellarische Angaben zu Städten). Werden Vorlagen allerdings verändert, müssen zahlreiche Artikel auf die neuen Vorlagenformate angepasst werden – eine typische Tätigkeit für einen der fünf dort aktiven Bots. Immer wiederkehrende *Rechtschreib- und Typographiefehler* werden von vier Bots korrigiert. Weitere drei Bots kümmern sich um die *korrekte Auszeichnung*

von Bildern und überprüfen beispielsweise, ob Lizenzinformationen korrekt vermerkt sind. Listen, die z. B. auf Themenportalen zu finden sind, müssen regelmäßig gepflegt werden, woran zwei Bots mitwirken. Die Erkennung und Meldung von *Vandalismus*, der z. B. durch massenhafte oder sinnlose Editierungen böswilliger Nutzer entsteht, wurde im Betrachtungszeitraum durch einen Bot durchgeführt.

Wie der Tabelle zu entnehmen ist, stellen Bots, die sich um Interwiki-Verlinkungen kümmern, quantitativ die größte Tätigkeitsgruppe dar. Auf den ersten Blick scheinen Interwiki-Bots keine besonders intelligente Arbeit zu verrichten, sie verlinken lediglich thematisch identische Artikel aus unterschiedlichen Sprachversionen. Betrachtet man die Arbeit der Bots allerdings genauer, erscheint dieser erste Eindruck wesentlich differenzierter, insbesondere in Hinblick auf das Vermögen der Bots, Folgehandlungen anderer Nutzer zu initiieren. Auf diesen Aspekt wird in Abschnitt 5.2.3 weiter eingegangen.

Die Bots der verschiedenen Kategorien sind hinsichtlich ihres Verhaltens unterschiedlich weit fortgeschritten. Besonders avancierte Bots nehmen sogar Editierungen auf Grundlage selbst beschaffter externer Daten vor (PatDi BOT)<sup>23</sup>. Deutlich geworden ist aber auch, dass Bots (bisher) Artikel nicht eigenständig verfassen.

Insgesamt lässt sich konstatieren, dass die einzelnen von Bots durchgeführten Aktivitäten der Selbstcharakterisierung der Wikipedia sehr nahe kommen, in der es heißt, dass sie „ihren Betreibern stupide, langweilige und häufig auftretende Aufgaben abnehmen“. Dennoch sind bereits erste Tendenzen einer intelligenteren Bot-Generation erkennbar wie sich an einigen der nachfolgenden Beispiele zeigen wird.

Bevor auf die Funktionsweise einiger besonders avancierter Bots eingegangen wird, werden zunächst grobe quantitative Kenngrößen angegeben, um die Aktivität von Bots besser einordnen zu können.

### **Bot-Aktivität im Vergleich: 2010 vs. 2012/13**

Wie oben bereits genannt wurde neben der ursprünglichen Stichprobe im Zeitrum  $Z_1$ , die in Fink und Liboschik (2012) verwendet wurde, eine zusätzliche Stichprobe  $Z_2$ <sup>24</sup> gezogen, um zu überprüfen, ob es Verschiebungen gab und ob die auf  $Z_1$  bezogenen Folgerungen immer noch Gültigkeit besitzen. Hinsichtlich der Tätigkeitsfelder konnten keine wesentlichen Verschiebungen festgestellt werden (vgl. Tabelle 5.1 auf Seite 121).

Wesentliche Kennzahl für einen quantitativen Vergleich ist die Anzahl der Editierungen, da sie den Anteil der Bots an der Gesamteditierungsaktivität, also der Gesamtzahl aller Editierungen, messen kann. Das Resultat dieses Vergleichs findet sich in Tabelle 5.2. Dabei werden insgesamt drei Kenngrößen miteinander verglichen: Die

<sup>23</sup>Koordinaten für französischen Gemeinden wurden externen, im Web verfügbaren, Tabellen entnommen.

<sup>24</sup> $Z_1$ : 01.02.2010–30.04.2010 ;  $Z_2$ : 01.11.2012–31.01.2013.

absolute Anzahl der von Bots durchgeführten Editierungen (# Bot-Editierungen), die Gesamtanzahl aller Editierungen (# Editierungen) und der relative Anteil von Bot-Editierungen an allen Editierungen (% Bot-Editierungen). Vergleichend dargestellt sind die Zeiträume  $Z_1$  und  $Z_2$  differenziert nach allen Editierungen und Editierungen, die nur im Artikel-Namensraum erfolgt sind. Editierungen außerhalb des Artikel-Namensraums beziehen sich meist auf administrative Aufgaben und Diskussionen.

	$Z_1$ : 01.02.2010–30.04.2010		$Z_2$ : 01.11.2012–31.01.2013	
	alle Edits	nur Artikel	alle Edits	nur Artikel
# Bot-Editierungen	507 659	360 998	615 229	361 672
# Editierungen	3 350 524	2 351 860	3 304 642	2 243 431
% Bot-Editierungen	15.15 %	15.35 %	18.62 %	16.12 %

**Tabelle 5.2:** Anteil Bot-Editierungen im Vergleich zweier Zeiträume der Jahre 2010 und 2012/13

Auffällig ist bei dem Vergleich der Zeiträume  $Z_1$  und  $Z_2$ , dass sich der Anteil der Bot-Editierungen im Artikel-Namensraum nur marginal erhöht hat (15.35 %  $\rightarrow$  16.12 %), während die Bot-Editierungen in Bezug auf alle Namensräume (inkl. Artikeln) stärker zugenommen haben (15.15 %  $\rightarrow$  18.62 %) bzw. von 507 659 Editierungen in  $Z_1$  auf 615 229 Editierungen in  $Z_2$  gestiegen sind, was einer absolute Zunahme von gut 100 000 Editierungen entspricht. Stellt man in Rechnung, dass bei der Betrachtung aller Namensräume der Artikel-Namensraum dominiert, kann man folgern, dass die Bot-Aktivitäten vor allem außerhalb des Artikel-Namensraums, also bei administrativen Tätigkeiten, stark zugenommen haben.

### Bots und Hybridität

Die Tätigkeiten von Bots haben auf den ersten Blick keine größere inhaltliche Bedeutung, ihre Editierungen haben im Wesentlichen „lektorativen“ Charakter. In den folgenden Beispielen, in denen die betrachteten Bots weniger durch Quantität denn durch Qualität ihrer Editierungen auffallen, wird sich zeigen, dass Bots hochgradig handlungswirksam sein können. Zur Verdeutlichung ihrer Handlungswirksamkeit genügt es nicht, die Aktivitäten der Bots isoliert zu betrachten. Vielmehr wird ihre Wirksamkeit deutlich, wenn die Einzelaktivitäten im Kontext ihrer zugehörigen Interaktionsketten betrachtet werden.

#### Beispiel I: Bots als Initiatoren hybrider Interaktionsketten

In Abbildung 5.2 ist die Versionshistorie für eine Grafikdatei, die in der Wikipedia eingestellt ist, dargestellt. Für jeden Wikipedia-Artikel und jede Datei (z. B. eine Grafik) existiert eine solche Historie, in der sämtliche Bearbeitungsschritte chronologisch aufgeführt sind. Die Protokollierung umfasst neben der genauen Änderungszeit auch

den Benutzernamen desjenigen, der die Änderung durchgeführt hat, die Größe der Änderung, gemessen in Bytes und einen optionalen Kommentar. Mit einer Vergleichsfunktion ist es auch möglich, Änderungen zwischen zwei Artikel- oder Dateiversionen anzuzeigen.

In dem hier diskutierten Beispiel geht es um die Klärung einer Lizenzfrage bzgl. eines eingescannten 10.000 Dong-Geldscheins (vietnamesische Währung). In Tabelle 5.3 sind die Änderungen der Nutzer ausformuliert, so dass deutlich wird, wie die hier betrachtete Interaktion zustande kommt.

15:41, 21. Jun. 2009	Sa-se (Diskussion   Beiträge)	K (153 Bytes)	(Informationsmängel beseitigt)	(entfernen)	[automatisch gesichtet]
00:20, 18. Nov. 2007	Augiasstallputzer (Diskussion   Beiträge)	K (156 Bytes)	(entfernen)	gesichtet von Bondom]	
01:45, 26. Jul. 2007	Forrester-Bot (Diskussion   Beiträge)	K (138 Bytes)	(ersetze Vorlage: Bild-UN durch Vorlage: Bild-frei)	(entfernen)	
10:47, 23. Jul. 2007	ChristianBier (Diskussion   Beiträge)	(136 Bytes)	(geldscheine sind gemeinfrei)	(entfernen)	
21:12, 16. Jul. 2007	ChristianBier (Diskussion   Beiträge)	(155 Bytes)	(entfernen)		
23:31, 5. Jul. 2007	BuschBot (Diskussion   Beiträge)	K (148 Bytes)	(Bot: Füge Vorlage:Dateiüberprüfung ein. Bei Fragen: WP:DÜP.)	(entfernen)	
16:46, 3. Feb. 2007	Ar-ras (Diskussion   Beiträge)	K (125 Bytes)	(entfernen)		
20:50, 3. Aug. 2006	Sebbot (Diskussion   Beiträge)	K (177 Bytes)	(clean up)	(entfernen)	
20:38, 19. Dez. 2005	Crux (Diskussion   Beiträge)	K (129 Bytes)	{{Bild-PD}} (Scannen stellt keine Schöpfungshöhe dar)	(entfernen)	
14:07, 13. Dez. 2005	Senegur (Diskussion   Beiträge)	(131 Bytes)	(entfernen)		
13:59, 13. Dez. 2005	Senegur (Diskussion   Beiträge)	(45 Bytes)	(Eigenes Digitalfoto 10000 Đông Rückseite)		

**Abbildung 5.2:** Versionshistorie eines Bildes: Die Einzelaktivitäten von Bots und menschlichen Nutzern sind aufeinander bezogen.

Entscheidend für die Interaktion zwischen Bots und menschlichen Nutzern ist die Editierung des Benutzers *Ar-ras*, der die Lizenzinformation zu der Grafik entfernt, da er diese nicht für gemeinfrei hält. *Ar-ras* unterbreitet allerdings keinen Gegenvorschlag. Auf die fehlende Lizenzinformation wird der Bot *BuschBot* aufmerksam und kennzeichnet die Grafik als zu überprüfend. Dieser Aufforderung kommt der Benutzer *ChristianBier* wenige Tage später nach und präzisiert das Problem, das der Bot bezüglich der Grafik festgestellt hat. Es handelt sich offenbar um ein Lizenzproblem, das der Benutzer *ChristianBier* nicht direkt klären kann. Einige Tage später stellt er fest, dass Geldscheine gemeinfrei sind und kennzeichnet die betroffene Datei entsprechend. Der Bot *Forrester-Bot* ersetzt wenige Tage später eine vom Benutzer *ChristianBier* verwendete englischsprachige Vorlage zur Grafikbeschreibung durch eine deutschsprachige Version für gemeinfreie Bilder.

Insgesamt führte die Interaktion zwischen verschiedenen Bots und den menschlichen Nutzern schließlich dazu, dass ein Lizenzproblem, nämlich die Frage, unter welcher Lizenz eingescannte Geldscheine stehen, geklärt wurde. An diesem Beispiel lässt sich folglich verteiltes Problemlösen bzw. verteiltes Handeln in Aktion beobachten. Beachtenswert ist, dass ohne die Kennzeichnung durch den Bot *BuschBot*, die auf ein Problem mit der Grafik hindeutete, der Benutzer *ChristianBier* höchstwahrscheinlich nicht darauf aufmerksam geworden wäre und die Lizenzfrage ungeklärt geblieben

Bearbeitungszeitpunkt	Benutzer	Bot?	Beschreibung der Bearbeitung
13:59, 13. Dez 2005	Senegur	✗	Benutzer stellt erste Version eines eingescannten Geldscheins ein.
14:07, 13. Dez 2005	Senegur	✗	Die Lizenz des Geldscheins wird auf „GFDL“ (Copyleft-Lizenz mit Pflicht zur Nennung des Autors) eingestellt.
20:38, 19. Dez. 2005	Crux	✗	Lizenzinformation wird auf 'Public Domain' (gemeinfrei) geändert, da Benutzer der Auffassung ist, dass Einscannen nicht dazu berechtigt eigene Rechte an der Grafik anzumelden, wie z. B. die Nennung von Benutzer Senegur als Schöpfer.
20:50, 3. Aug. 2006	Sebbot	✓	Anpassen der Bildbeschreibung an neues Format.
16:45, 3. Feb. 2007	Ar-ras	✗	Entfernen der Lizenzinformation, da Benutzer nicht die Meinung teilt, dass es sich um eine gemeinfreie Grafik handelt.
23:31, 5. Jul. 2007	BuschBot	✓	Benutzer stellt fest, dass wichtige Informationen zu der Grafik (u. a. die Lizenz) fehlen und kennzeichnet die Grafik zur Überprüfung.
21:12, 16. Jul. 2007	ChristianBier	✗	Benutzer überprüft die Grafik und stellt fest, dass das Problem eine fehlende Lizenzinformation ist.
10:47, 23. Jul. 2007	ChristianBier	✗	Nach einigen Tagen findet der Benutzer heraus, dass Geldscheine gemeinfrei sind und kennzeichnet die Grafik entsprechend.
01:45, 26. Jul. 2007	Forrester-Bot	✓	Einfache Vorlagenersetzung für gemeinfreie Bilder in der deutschsprachigen Wikipedia.

**Tabelle 5.3:** Kommentierte Versionshistorie, die die Interaktion zwischen Bots und menschlichen Nutzern illustriert.

wäre. Die folgende Editierung durch den Bot *Forrester-Bot* wäre zudem nicht ohne die vorherige Editierung des Benutzers *ChristianBier* erfolgt.

## Beispiel II: Hybride Konzeption eines Bots

Das vorherige Beispiel hat gezeigt, dass Bots Initiatoren hybrider Interaktionsketten werden können. Das folgende Beispiel geht noch einen Schritt weiter und verdeutlicht, dass Bots z. T. bereits *hybrid konzipiert* sind.

Der sogenannte Interwiki Link Checker (IWLC)<sup>25</sup> ist dabei ein Bot (genauer: ein aus mehreren Bots bestehendes Tool), der auf die „Mitarbeit“ menschlicher Nutzer angewiesen ist. Die Verlinkung von Artikeln unterschiedlicher Sprachversionen ist, wie bereits erläutert, ein wichtiges Anliegen der Wikipedia. Insbesondere für Sprachen, die auf Grund ihrer geringen Verbreitung nicht auf eine derart große (potenzielle) Autorenschaft wie beispielsweise die englischsprachige Wikipedia zurückgreifen können, ist die Methode der Interwiki-Verlinkung ein probates Mittel, um zu einem Artikel in der eigenen Sprache weiterführende Informationen zu erhalten.

Die Oberfläche des IWLC ist in Abbildung 5.3 dargestellt. Unter Zuhilfenahme von Bots und des Betreibers des IWLC werden Listen von Artikeln generiert, die jeweils in zwei unterschiedlichen Sprachversionen ein vermeintlich identisches Thema behandeln.

The screenshot shows the IWLC interface for the article 'Susie'. It displays two versions of the article side-by-side. The left version is in Japanese, and the right version is in Norwegian. Below the articles is a table with statistics and a summary of the comparison.

Last Change	Statistics	At least one of the pages is ...	Both pages deal with the same subject	Change databases	
same	not the same	4 ... a redirect	1 Yes	3 Don't know	(ja) (no)
for bot-operator	6 ... a disambiguation		2 No	7 Article(s) deleted	Cancel

**Abbildung 5.3:** Der Interwiki Link Checker (IWLC): Beispiel für die hybride Konzeption eines Bots. Menschliche Nutzer werden zum „Mit-Handeln“ ermutigt.

<sup>25</sup><http://toolserver.org/~flacus/IWLC/> (das Tool ist leider nicht mehr online verfügbar, Stand: 31. Januar 2014).

Im hier präsentierten Beispiel werden einem menschlichen Nutzer zwei Artikel mit dem Titel *Susie* auf einer Webseite nebeneinander dargestellt. Der Interwiki Link Checker ruft die thematisch potenziell identischen Artikel selbstständig ab und präsentiert sie einem menschlichen Nutzer, der beide Artikelsprachen (hier: Japanisch und Norwegisch) beherrscht. Dieser kann leicht beurteilen, ob die Artikel ein identisches Thema behandeln. Mit einem Mausklick auf die vom IWLC bereitgestellten Buttons beantwortet er diese Frage mit 'Yes', 'No', 'Don't know' oder 'Article(s) deleted'. Für einen Bot wäre es hier nur schwer entscheidbar, ob die Artikel das gleiche Thema behandeln. Aus dem Grund unterstützt der IWLC menschliche Nutzer bei der Klassifizierungsaufgabe durch das Abrufen der Artikel und eine möglichst übersichtliche Gegenüberstellung.

Nachdem die Artikel von sprachkundigen Nutzern entsprechend klassifiziert wurden, kümmert sich täglich ein Bot um das Einpflegen der Verlinkung. Dabei verlässt sich der Bot nicht auf die Entscheidung eines einzelnen Nutzers, sondern verlinkt zwei Artikel erst dann miteinander, wenn mindestens fünf Benutzer diese als identisch klassifiziert haben.

### Neuere Bots, Bot-Typen und Inter-Bot-Kooperation

Im Vergleich zu den in Fink und Liboschik (2012) untersuchten Bots (die dortigen Daten stammen aus 2010) fällt auf, dass Bots zunehmend multifunktional werden und untereinander kooperieren (Inter-Bot-Kooperation) sowie Aufträge von menschlichen Nutzern entgegennehmen. In diesem Abschnitt werden daher einige dieser Bots vorgestellt. Dabei wird ausführlicher auf **DrTrigonBot** eingegangen, der besonders avancierte Funktionen beinhaltet (z. B. Kategorisierung von Bildern auf Grundlage von Bilderkennungsverfahren), und cursorisch auf drei weitere Bots, die im wesentlichen auf Grundlage ihrer Namensraumaktivität charakterisiert werden: **GiftBot**, **SpBot** und **SebBot**.

Tabelle 5.1 zeigt eine Übersicht über einige quantitative Kenndaten der hier beschriebenen Bots. Die Spalte „# Edits“ bezeichnet die absolute Anzahl Editierungen des jeweiligen Bots. Die Spalte „Aktivität“ listet die Namensräume auf, in denen der Bot aktiv war. Dabei sind die Namensräume mit der größten Aktivität je Bot (>20%) gesondert gekennzeichnet. In Klammern steht der genaue Anteil an Editierungen im entsprechenden Namensraum.

#### GiftBot

Anhand der quantitativen Kenngrößen in Tabelle 5.4 ist abzulesen, dass **GiftBot** vielfältige Betätigungsfelder hat, da dieser in sehr unterschiedlichen Namensräumen aktiv ist. 40.3% der Aktivitäten im betrachteten Dreimonatszeitraum  $Z_2$  spielen sich im Artikel-Namensraum ab. Diese hohe Zahl ist allerdings verfälscht, da die Edits im Artikel-Namensraum aus einer im betrachteten Zeitraum durchgeführten Spezialauf-

	# Edits	Aktivität <sup>1</sup>	Kurzbeschreibung
GiftBot	32 278	<b>0</b> (40.3%), <b>4</b> (25.6%), 3, 100, 2, 1, 5, 101, 10, 11, 13, 13	Mehrere Aufgaben, auch hier: Hinweise auf Inter-Bot-Kooperation, scheint eher typischer Verwaltungs-Bot zu sein, der hilft, interne Prozesse zu organisieren.
SpBot	22 710	<b>4</b> (55.2%), <b>3</b> (26.0%), 1, 5, 101, 100, 2, 11, 13, 9, 12	Befasst sich hauptsächlich mit Kategorien, scheint sehr intelligent zu sein, KI-Auszeichnung.
Sebbot	22 710	<b>0</b> (68.6%), 14, 4, 2, 10, 6	Aktualisiert Einwohnerzahlen durch gezielten Datenbankabrufe.
DrTrigonBot	10 438	<b>2</b> (25.4%), <b>100</b> (24.7%), Com <sup>2</sup> . <b>6</b> (24.6%) 4, 10, 3, Com.6	Multifunktionseller Bot, insbesondere spannend, dass er mit anderen Bots interagiert bzw. sie sich auf ihn und er sich auf sie verlässt.

<sup>1</sup> Die fett gedruckten Namensräume kennzeichnen Namensräume, in denen der jeweilige Bot hauptsächlich tätig ist. Von hauptsächlichlicher Tätigkeit wird ausgegangen, wenn der Bot mindestens 20 % seiner Gesamteditierungen in dem gekennzeichneten Namensraum durchführt. Die Reihenfolge der Namensräume gibt die absteigend sortierte Editieraktivität wieder.

<sup>2</sup> Das Präfix *Com* kennzeichnet, dass der Bot auf Wikimedia Commons, dem zentralen Bilderarchiv für alle Wikipedia-Projekte, aktiv ist.

**Tabelle 5.4:** Aktivitäten ausgewählter Bots im Zeitraum  $Z_2$ : 01.11.2012–31.01.2013, aufgeschlüsselt nach Anzahl Editierungen (# Edits) und Aktivität je Namensraum.

gabe resultieren: der Markierung einiger Artikel für das ggf. demnächst eingeführte Artikel-Feedback-Tool<sup>26</sup>.

25.6 % der Editierungen spielen sich im Namensraum 4, dem Metabereich der Wikipedia, ab. Ein Blick auf die Benutzerseite<sup>27</sup> des Bots bestätigt, dass hier die Hauptaufgaben von GiftBot zu verorten sind, die sich vor allem auf die Vandalismusbekämpfung, also die Verhinderung destruktiver Artikeländerungen, und Artikelsichtungen beziehen. Auf der Benutzerseite wird zudem unterschieden zwischen ständigen, regelmäßigen, sporadischen und einmaligen Aufgaben. Bei den ständigen Aufgaben lauscht der Bot in einem speziellen Wikipedia-Chat-Kanal auf Anweisungen, die nahezu in Echtzeit prozessiert werden können. Insgesamt sind vier ständige, sechzehn regelmäßige, vier sporadische und fünfzehn einmalige Aufgaben dokumentiert. Eine Durchsicht der dort genannten Tätigkeitsfelder deutet darauf hin, dass dieser Bot der üblichen Bot-Definition in der Wikipedia weitgehend gerecht wird, die Bots als Helfer zur Ab-

<sup>26</sup>s. <http://de.wikipedia.org/wiki/Wikipedia:Artikel-Feedback> (Abruf: 31. Januar 2014).

<sup>27</sup><http://de.wikipedia.org/w/index.php?title=Benutzer:GiftBot&oldid=114377794> (Abruf: 31. Januar 2014).

nahme „stupide[r], langweilige[r] und häufig auftretende[r]“ Aufgaben (vgl. S. 118) charakterisiert.

Im Wikipedia-Namensraum 4 aktualisiert der Bot im Zuge der Vandalismusbekämpfung z. B. auf Bearbeitungslisten für Seitensperrungen die Sperrgründe, die in entsprechenden Vorlagen von menschlichen Nutzern angegeben werden, oder die Anzahl ungesichteter<sup>28</sup> Artikel.

Aus der Perspektive der Handlungswirksamkeit autonomer Technik ist nicht abzusehen, dass **GiftBot** etwas *bewirkt*, allerdings erscheinen seine Änderungen für sich genommen als wenig komplex und seine Architektur nicht hybrid, sondern als Ansammlung einzelner, klar definierter Teilaufgaben.

### **SpBot**

Betrachtet man die Editierungen von **SpBot** fällt auf, dass ein Großteil der Editierungen (55.2%) im Wikipedia-Namensraum stattfindet – typisch für einen Bot, der hilft, interne Verwaltungsprozesse zu organisieren. 26% der Editierungen erfolgen im Benutzerdiskussionsnamensraum (4)<sup>29</sup>, die dadurch zu erklären sind, dass der Verwaltungs-Bot u. a. bei der Vandalismus-Bekämpfung hilft und z. B. gesperrte Nutzer über ihre Diskussionsseite über eine Sperrung und ihren Grund informiert. Bei näherer Betrachtung der Benutzerseite von **SpBot**<sup>30</sup> wird deutlich, dass der Bot eine Vielzahl weiterer, kleiner Aufgaben erfüllt: Z. B. die Archivierung von Diskussionen, Verwaltung einer Datenbank zu persönlichen Bekanntschaften oder die Aktualisierung von Bot-Statistiken.

Für die Frage nach der Handlungswirksamkeit gilt für diesen Bot die gleiche Aussage wie für **GiftBot**: Die Funktion von **SpBot** ist am ehesten als Werkzeug zu verorten, das verschiedene, klar umrissene und relativ einfache Aufgaben durchführt.

**Sebbot** Die Aktivitäten von **Sebbot** erstrecken sich mit 68.6% hauptsächlich auf den Artikel-Namensraum. Bei näherer Betrachtung der Aktivitäten und der Bot-Beschreibungsseite<sup>31</sup> wird deutlich, dass eine wesentliche Aufgabe des Bots in der Umbenennung und Verschiebung von Kategorien besteht. Kategorien werden zwar in einem eigenen Namensraum (14) verwaltet, die Zuordnung einer bestimmten Seite zu einer Kategorie erfolgt hingegen im Artikel-Namensraum, womit sich die Vielzahl der Editierungen in ebendiesem erklären lässt. Außerhalb des Kategoriensystems befasst

<sup>28</sup> Artikeländerungen, die in der Wikipedia von unangemeldeten Nutzern vollzogen werden, erfordern eine „Sichtung“ bzw. Freigabe durch einen fortgeschrittenen angemeldeten Nutzer. Damit soll verhindert werden, dass offensichtlich falsche Artikeländerungen (z. B. wegen fehlender Quellen) oder Vandalismus unmittelbar für Leser der Wikipedia sichtbar sind.

<sup>29</sup> In der Wikipedia ist der Benutzerdiskussionsnamensraum (Namensraum 4) der übliche Weg, um einem Benutzer eine Nachricht zu hinterlassen. Private Kommunikation ist nicht vorgesehen, jegliche Kommunikation ist öffentlich einsehbar und wird auf den Benutzerdiskussionsseiten geführt.

<sup>30</sup> <http://de.wikipedia.org/w/index.php?title=Benutzer:SpBot&oldid=102487151> (Abruf: 31. Januar 2014).

<sup>31</sup> <http://de.wikipedia.org/wiki/Benutzer:Sebbot> (Abruf: 31. Januar 2014).

sich **Sebbot** mit Verwaltungsaufgaben, etwa der Aktualisierung von Löschanträgen und der Aktualisierung von Qualitätssicherungsanträgen.

**Sebbot** übernimmt mit seinen umfangreichen Tätigkeiten im Umfeld der Kategorien zwar wichtige Aufgaben, der Bot wird allerdings nur auf Anweisung menschlicher Nutzer tätig, die nach langer Diskussion z. B. beschlossen haben, eine Kategorie umzubenennen oder mit einer anderen Kategorie zusammenzufügen. Im Sinne instrumenteller Technik arbeite der Bot lediglich eine klar umrissene Aufgabe ab. Von einem autonom agierenden Software-Agenten kann im vorliegenden Fall nicht gesprochen werden, da **Sebbot** ausschließlich auf Anweisung aktiv wird und eher im Sinne eines Werkzeugs denn als (teil-)autonomer Mithelfer agiert.

### **DrTrigonBot**

In Tabelle 5.2 sind einige quantitative Kenngrößen zu **DrTrigonBot** aufgeführt. Es wird deutlich, dass es sich um einen multifunktionalen Verwaltungsbots handelt, der vor allem im Hintergrund (also nicht im Artikel-Namensraum 0) aktiv ist.

Beachtenswert ist dieser Bot nicht wegen der Anzahl seiner Editierungen, die mit 10 438 nicht zu den Spitzenreitern mit über 80 000 Editierungen zählt, sondern wegen seiner avancierten Funktionen. Stellvertretend für die Vielzahl von Aufgaben, die **DrTrigonBot** erfüllt, werden zwei besonders interessante Funktionen näher beschrieben. Die Erste umfasst die Grobkategorisierung von Bildern auf Grundlage von Bilderkennungsverfahren. Dabei können Bilder auf Basis bestimmter Merkmale z. B. als Porträt-Aufnahmen von Personen vorkategorisiert werden. Der Aktivitätsanteil von 24.6 % im Namensraum Com.6, dem Datei- und Bildnamensraum von Commons<sup>32</sup>, zeigt, dass diese Kategorisierung eine der Hauptaufgaben des Bots darstellt.

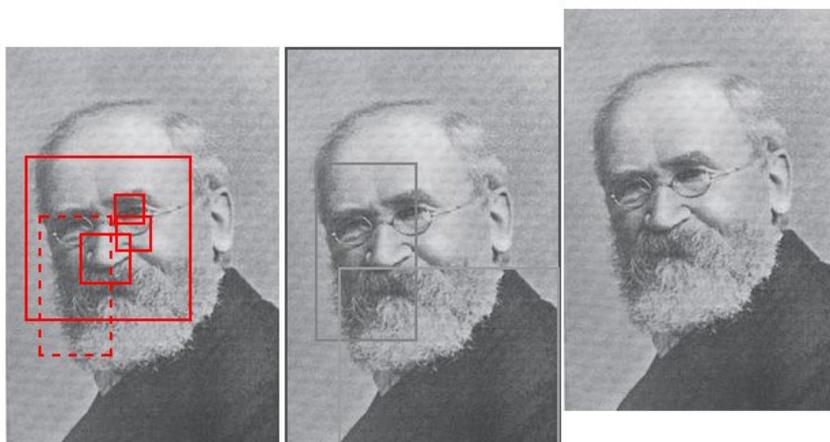
Die zweite vorgestellte Funktion beinhaltet die Inter-Bot-Kooperation, bei der **DrTrigonBot** gezielt anderen Bots zuarbeitet und damit Folgeaktionen von Bots und menschlichen Nutzern initiiert.

Die typische Arbeitsweise der **Bilderkategorisierung** ist in Abbildung 5.4 dargestellt: Im ersten Schritt durchsucht der Bot das Dateiarchiv von Commons nach nicht kategorisierten Bildern. Hat er ein solches Bild gefunden, werden darauf spezielle Bilderkennungsalgorithmen angewendet, die z. B. typische Merkmale einer Porträt-aufnahme feststellen können (Abbildung 5.4a). Im vorliegenden Fall wurden Augen, Nase und Gesichtsproportionen erkannt, die zu einer Kategorisierung des Fotos in die Kategorie *Portraits* und *Unidentified people* führen.

Nach der Einordnung weist der Bot zusätzlich auf die automatische Kategorisierung hin (Abbildung 5.4b) und ermuntert andere (menschliche) Nutzer, den Automatismus zu überprüfen und das Foto einer eindeutigen Personen-Kategorie zuzuordnen.

<sup>32</sup>Wikimedia *Commons* (<http://commons.wikimedia.org>) ist das Medienarchiv aller Wikipedia-Versionen. Dort werden Fotos hochgeladen, die in unterschiedlichen Wikipedia-Sprachversionen eingebunden werden können.

File:A.M. Fairbairn.jpg



**automatic categorization** Portraits, Categorized by DrTrigonBot, Unidentified people

**Info:** DrTrigonBot (Diskussion) has evaluated this file and left hidden information. View it! For a deeper understanding of how the bot works, please consult the [documentation](#).

(a) Bot erkennt Bild als Portrait einer Person



(b) Hinweis auf Überprüfung der automatischen Kategorisierung

• (Aktuell | Vorherige) 07:48, 8. Feb. 2013 UAwiki (Diskussion | Beiträge) .. (2.421 Byte) (-141) .. (*add Category:Scholars from Scotland*) (rückgängig machen)

• (Aktuell | Vorherige) 20:15, 26. Jan. 2013 DrTrigonBot (Diskussion | Beiträge) K...; (2.562 Byte) (+2.156) .. (*bot automatic categorization; adding Category:Categorized by DrTrigonBot, Category:Portraits, Category:Unidentified people*) (rückgängig machen)

(c) Folgeinteraktion eines menschlichen Nutzers

**Abbildung 5.4:** DrTrigonBot kategorisiert ein Bild auf Grundlage bestimmter automatisch erkannter Bildmerkmale und initiiert damit Folgeinteraktionen menschlicher Nutzer. Quellen s. Anhang A.2

Der am 26. Januar 2013 platzierte Hinweis von DrTrigonBot zeigt sehr schnell eine Wirkung, so dass der Benutzer *UAwiki* am 8. Februar 2013 reagiert, den Hinweis von DrTrigonBot zwecks Überprüfung löscht und das Foto der eindeutigen Personen-Kategorie *Scholars from Scotland* zuordnet. Die Bilderkennungsfunktion von DrTrigonBot befindet sich noch in der Entwicklung. Bereits jetzt kann der Bot Menschen (Porträts, Gruppenaufnahmen, Gesichter), Grafiken, Schachbrettmuster und Barcodes erkennen. In der Entwicklung befindet sich u. a. eine Erkennung von Gebäuden, Pflanzen, Wappen, Karten, Zügen, Autos und Diagrammen<sup>33</sup>.

Anhand der quantitativen Kenngrößen aus Tabelle 5.2 lässt sich angeben, wie viele erfolgreiche automatische Kategorisierungen DrTrigonBot vollzogen hat: Im Namensraum Com.6, dem Bildernamensraum, hat der Bot 24.6% seiner 10 438 Editierungen durchgeführt und damit 2 568 Bilder auf Basis von bilderkennenden

<sup>33</sup><http://en.wikipedia.org/w/index.php?title=User:DrTrigonBot/doc&oldid=534966927#Categorization> (Abruf: 31. Januar 2014).

Verfahren innerhalb von drei Monaten erfolgreich kategorisiert. Stellt man die Schwierigkeit dieser Aufgabe in Rechnung ist die Zahl durchaus beeindruckend. Es bleibt abzuwarten, wie sich diese Funktion von DrTrigonBot weiterentwickelt, zumal für die Zukunft maschinelle Lernverfahren geplant sind, die die Erkennungsquoten zusätzlich steigern sollen<sup>34</sup>.

Als zweite beachtenswerte Funktion bzw. Merkmal der Softwarearchitektur wird die **Inter-Bot-Kooperation** vorgestellt.



**Abbildung 5.5:** Zusammenhandeln von Xqbot, einem menschlichen Uploader (*The weaver*), DrTrigonBot, der auf die Antwort aufmerksam macht (an anderer Stelle, hier nicht dargestellt) und einem menschlichen Problemlöser (*Nolispanmo*).  
Quelle: Anhang A.2

Auf der Benutzerseite von DrTrigonBot<sup>35</sup> heißt es zur Inter-Bot-Kooperation: „Unterstützung von Kollege BLUbot bzw. Xqbot: Assistiere regelmässig meinem Kollegen (Inter-Bot-Kooperation) und liefere Informationen, wer auf automatische Nachrichten direkt auf der Benutzer-Diskussion geantwortet hat“.

Der Bot Xqbot benachrichtigt Nutzer, die beim Upload neuer Bilder vergessen haben, diese mit einer Lizenz zu versehen über ihre Benutzerdiskussionsseite und bittet sie, diese umgehend nachzutragen. Der Xqbot eröffnet mit dem Hinweis „Du kannst aber auch gern hier antworten, damit dir individuell geholfen wird.“ die Möglichkeit auf den vom Bot hinterlassenen Hinweis direkt zu antworten. Da der Bot, der auf das Problem hingewiesen hat, natürliche Sprache (noch) nicht verstehen kann, ist eine Reaktion eines anderen menschlichen Nutzers gefragt, der sich näher mit Lizenzen auskennt.

<sup>34</sup>Quelle: [https://en.wikipedia.org/w/index.php?title=User:DrTrigonBot/doc&oldid=564401372#Machine\\_learning](https://en.wikipedia.org/w/index.php?title=User:DrTrigonBot/doc&oldid=564401372#Machine_learning) (Abruf: 31. Januar 2014).

<sup>35</sup><http://de.wikipedia.org/w/index.php?title=Benutzer:DrTrigonBot&oldid=113617689> (Abruf: 31. Januar 2014).

Nur wie soll ein kompetenter menschlicher Nutzer von dem Problem erfahren, dass *Xqbot* aufgeworfen hat und zu dem es offenbar eine Rückfrage gibt? Durch *DrTrigonBot*: Dieser durchsucht gezielt Benutzerdiskussionsseiten, auf denen zuvor *Xqbot* aktiv war und überprüft, ob auf den Hinweis von *Xqbot* eine Antwort eines anderen Nutzers erfolgt ist. Ist dies der Fall, wird davon ausgegangen, dass der Uploader des Bildes eine Rückfrage zu dem Lizenzproblem hat. *DrTrigonBot* trägt daraufhin einen direkten Link zu der Reaktion des menschlichen Uploaders auf die Seite *Benutzer:Xqbot/Antworten*<sup>36</sup> ein, damit ein anderer menschlicher Nutzer sich der Rückfrage zu der Bot-Aktivität annehmen kann. In Abbildung 5.5 ist ein typischer Ablauf zur Klärung eines Lizenzproblems dargestellt: *Xqbot* macht auf die fehlende Lizenz aufmerksam und der ursprüngliche Uploader des Bildes, *The weaver*, antwortet auf den Hinweis des Bots mit „Ich habe das Foto selbst erstellt und mir die Veröffentlichung genehmigt.“ *DrTrigonBot* bemerkt die Antwort, informiert an anderer Stelle darüber und leitet damit die Aktion eines menschlichen Lizenzfachmanns, *Nolispanmo*, ein. Dieser antwortet: „ne das ist ein Bildschirmfoto von einer google Suche mit Foto und daran kannst du dir keine Rechte 'genehmigen'. So schwer ist die Beachtung der Bildrechte anderer doch nicht“.

An dem Beispiel wird deutlich, dass Bots selbsttätig Aufgaben durchführen, auf Probleme hinweisen und damit Folgeaktionen menschlicher Nutzer bewirken, die ihrerseits wiederum Bot-Aktivitäten anstoßen usw. Es wäre möglich, noch unzählige weitere Beispiele für Inter-Bot-Kooperation und das Zusammenhandeln von Bots und menschlichen Nutzern anzuführen. Interessant ist die Beobachtung, dass die Lösung von Lizenzproblemen nicht etwa durch eine zentrale Verwaltungseinheit gelöst wird, sondern vielmehr als Zusammenspiel dezentral agierender Einheiten funktioniert. Das gemeinsam erklärte (und in dem Fall auch erreichte) Ziel ist dabei die Lösung eines Problems.

### Plädoyer für eine hybride Sicht auf Bots

Anhand des ersten Beispiels (vgl. Abbildung 5.2 auf Seite 124) konnte gezeigt werden, dass zwar die isoliert betrachtete Aktivität eines Bots eine relativ geringe Bedeutung hat, diese jedoch Folgehandlungen von menschlichen Nutzern initiiert, die ihrerseits weitere Folgehandlungen hervorrufen, entweder von Bots oder wiederum von menschlichen Nutzern. Im Falle dieser aufeinander folgenden Handlungsstränge kann man von hybriden Interaktionsketten sprechen. Hybrid deswegen, weil sich die Gesamt-Handlung, zum Beispiel die Lösung eines Copyright-Problems, in einem „Handlungsstrom“ (Rammert und Schulz-Schaeffer 2002a) vollzieht, in dem die einzelnen Aktivitäten auf menschliche und nicht-menschliche Nutzer verteilt und die Handlungen aufeinander bezogen sind.

<sup>36</sup><http://de.wikipedia.org/wiki/Benutzer:Xqbot/Antworten> (Abruf: 31. Januar 2014).

Im zweiten Beispiel hat sich gezeigt, dass der Interwiki Link Checker stellvertretend für einen Bot-Typus steht, der bereits hybrid konzipiert ist und auf die Mitarbeit von menschlichen Nutzern angewiesen ist bzw. sie zur Mitarbeit ermutigt.

Da der Interwiki Link Checker inzwischen eingestellt ist, wurde mit **DrTrigonBot** ein weiterer Bot ausführlich vorgestellt, der sich durch eine Vielzahl von Funktionen, hohe Komplexität (automatische Kategorisierung auf Grundlage von bilderkennenden Verfahren) und eine hybride Architektur auszeichnet. Im Gegensatz zum ersten Beispiel, in dem Bots nur indirekt Initiatoren hybrider Interaktionsketten werden, fördert **DrTrigonBot** aktiv Interaktion, indem dieser menschliche Nutzer „anspricht“ und auffordert, seine Arbeit zu kontrollieren bzw. zu verfeinern.

Die anderen kurz vorgestellten Bots, namentlich **GiftBot**, **SpBot**, **Sebbot** wurden hier beschrieben, da sie relativ viele Editierungen durchgeführt haben (jeweils über 20 000) und weil die jeweiligen Benutzer-Seiten eine Vielzahl von Funktionen versprochen. Bei näherer Betrachtung bestätigte sich zwar die Funktionsvielfalt, die Komplexität der einzelnen Aufgaben stellte sich allerdings als nicht besonders hoch heraus. Vielmehr handelt es sich bei den Aufgaben dieser Bots um klar definierte, ohne Feedback-Schleifen und ohne eine explizit hybride Architektur. In diesem Fall kann im Sinne der Abgrenzung aus Kapitel 4 *nicht* von autonomer Technik gesprochen werden.

### **Zwischenfazit: Handlungswirksamkeit**

In diesem Abschnitt wurden insgesamt drei unterschiedliche Fälle von Handlungswirksamkeit autonomer Technik vorgestellt: automatische Börsencomputer, Fahrerassistenzsysteme und Wikipedia-Bots. Es ist deutlich geworden, dass ein rein instrumenteller Technikbegriff, der Technik als Werkzeug zur Erfüllung eines bestimmten Zwecks auffasst, nicht ausreicht, um autonome Technik angemessen beschreiben zu können (vgl. dazu auch Abschnitt 5.3.1). Autonome Technik greift aktiv in gesellschaftliche Interaktionszusammenhänge ein und ist innerhalb soziotechnischer System ein Mitspieler neben weiteren menschlichen und nicht-menschlichen. Wie bereits in (FLA 9) formuliert sind Mensch und autonome Technik als gleichberechtigte Partner zu betrachten.

Im Falle der selbsttätig Käufe und Verkäufe tätigen Autonomous Trading Agents hat sich gezeigt, dass diese durch ihre Aktivitäten mit immensen Geldbeträgen handeln und das Kauf- und Verkaufsverhalten menschlicher und weiterer nicht-menschlicher Börsenhändler steuern bzw. beeinflussen. Zur Zeit basieren die Handelsagenten auf komplexen mathematischen Prognosemodellen, die sich ausschließlich an der vergangenen Kursentwicklung orientieren. Wie allerdings in Abschnitt 5.2.1 erläutert, wird aktuell eine neue Algorithmengeneration erprobt, die aktiv Nachrichten durchforstet, daraus Schlussfolgerungen generiert und auf deren Grundlage Kaufentscheidungen trifft.

Neben Autonomous Trading Agents wurde auf den technischen Entwicklungsstand zum (teil-)autonomen Fahren eingegangen. Dabei ist deutlich geworden, dass aktuell vor allem Fahrerassistenzsysteme, die aktiver in das Fahrverhalten eingreifen und den menschlichen Fahrer sogar übersteuern können, diskutiert werden. Die Vision des vollständig autonomen Fahrens scheint greifbar nahe und technisch bereits heute prinzipiell machbar. Ob allerdings die Autofahrer mitspielen und auf das individuell von ihnen gesteuerte Fahrzeug verzichten bleibt fraglich. Realistisch erscheint eine relativ lange Übergangszeit zum (teil-)autonomen Fahren, in der Fahrerassistenzsysteme zwar aktiver werden, jedoch nach wie vor mit dem Handeln menschlicher Fahrer in einem hybriden System verteilter Handlungsträgerschaft (vgl. Abschnitt 5.3) verstrickt sein werden.

Als letztes Beispiel wurden Wikipedia-Bots, kleine (teil-)autonom agierende Computerprogramme in der Online-Enzyklopädie Wikipedia, vorgestellt. Dabei hat sich gezeigt, dass diese einen nennenswerten Anteil an der Gesamtzahl der Editierungen umfassen, Folgeaktionen menschlicher Nutzer und anderer Bots initiieren, z. T. aktiv mit anderen Bots kooperieren und zunehmend intelligenter werden. Bei näherer Betrachtung der thematischen Bereiche (Namensräume), in denen Bots aktiv sind, hat sich ergeben, dass diese vor allem repetitive Änderungen in Artikeln vornehmen und Verwaltungsaufgaben hinter den Kulissen der Artikel automatisieren und effizienter gestalten. Es wurde argumentiert, dass nicht in jedem Fall von autonomer Technik ausgegangen werden kann, dass aber ebenso Tendenzen erkennbar sind, dass Bots zunehmend eigenaktiver und „intelligenter“ (im Sinne komplexerer Algorithmen) werden.

Insgesamt lässt sich aus der Vorstellung dieser drei Fälle autonomer Technik ableiten, dass sowohl qualitativ als auch quantitativ deutlich wurde, dass Technik „mit-handelt“ und ihre Wirkung im Kontext soziotechnischer Interaktionszusammenhänge entfaltet. Aus soziologischer Perspektive schließt sich die Frage an, wie mit dieser empirisch evidenten Eigenaktivität von Technik umgegangen werden kann. Eine vorläufige Antwort darauf wird der anschließende Abschnitt 5.3 geben.

## **Implikationen**

In allen drei Fällen spielt Vertrauen eine wichtige Rolle. Gemäß der Ausführungen von James Coleman (vgl. Abschnitt 3.1.2) umfasst Vertrauen stets auch die Abgabe von Kontrolle. In den Beispielen müssen die jeweiligen menschlichen Mitspieler partiell Kontrolle an autonome Technik abgeben, wohl wissend, dass sie dadurch verletzlich werden und eventuelle Fehler verantworten müssen: Börsencomputer verschleudern Geld, Wikipedia-Bots führen unsinnige Editierungen durch, Fahrerassistenzsystem verursachen Unfälle.

Unter Rekurs auf die Handlungswirksamkeit, um die es in diesem Abschnitt hauptsächlich ging, kann festgestellt werden, dass autonome Technik *aktiv* in Interaktionszusammenhänge eingreift und diese (mit-)gestaltet. Sie kann dementsprechend als handlungswirksam betrachtet werden, da die Eingriffe ganz reale Auswirkungen haben: Autonome Börsencomputer handeln mit „echtem“ Geld, Wikipedia-Bots führen Editierungen durch, die für andere Nutzer sichtbar sind und an die weiter angeschlossen wird, und Assistenzsysteme verändern schließlich aktiv das Fahrverhalten eines Fahrzeugs und haben Auswirkungen auf den umgebenden Straßenverkehr. Diese Beobachtungen lassen sich schließlich zu (FLA 10) zusammenfassen.

**FLA 10 (Handlungswirksamkeit)**

Autonome Technik wird in Mensch-Maschine-Zusammenhängen handlungswirksam und ist im Rahmen einer soziologischen Analyse zu berücksichtigen.

### 5.3 Verteilte Handlungsträgerschaft und hybride Systeme

In den vorangegangenen Abschnitten wurde gezeigt, in welcher Weise (teil-)autonome Technik handlungswirksam wird und aktiv in Interaktionszusammenhänge menschlicher und nicht-menschlicher Akteure eingreift. Auf Seiten der Soziologie herrscht ein gewisses Theoriedefizit, um diesem empirischen Sachverhalt angemessen Rechnung zu tragen, da Kategorien wie „Handlung“, „Akteur“ oder „Intentionen“ als Eigenschaften aufgefasst werden, die ausschließlich menschlichen Wesen vorbehalten sind. Das radikale Gegenprogramm zu dieser Position hat Bruno Latour mit seiner Actor Network Theory (vgl. Abschnitt 5.1) formuliert, in der eine generalisierte Symmetrie von Natur, Technik und Gesellschaft postuliert wird. Abgesehen davon, dass die ANT bei ihrer Selbstanwendbarkeit schließlich in einem unendlichen Regress endet (vgl. Schulz-Schaeffer 2000: 132), bringt sie das forschungspragmatische Problem mit sich, dass durch die Aufhebung jedweder Unterscheidungen auch analytische Schärfe verloren geht.

Eine mögliche Antwort auf das Theoriedefizit der Soziologie und das Anwendbarkeitsdefizit der ANT für empirische Forschungsfragestellungen liefern Werner Rammert und Ingo Schulz-Schaeffer (2002a) mit ihrem, zumindest im technikoziologischen Kontext, breit rezipierten Aufsatz „Technik und Handeln. Wenn soziales Handeln sich auf menschliches Verhalten und technische Abläufe verteilt“. Darin schlagen sie ein graduelles Modell verteilter Handlungsträgerschaft als soziologische Antwort auf die zunehmende Eigenaktivität avancierter Technik vor, ohne damit im „[...] begrifflichen Niemandsland der Akteur-Netzwerk-Theorie zu stranden“ (ebd.: 48).

Die Autoren (ebd.: 11f.) argumentieren direkt zu Beginn ihrer Abhandlung, dass nach neueren Erkenntnissen der Psychologie Menschen weit weniger autonom handeln als gemeinhin angenommen. Moderne Technik hingegen gehe inzwischen weit über ihre bloße Instrumentalfunktion hinaus: „Menschen handeln, Maschinen funktionieren nur“ (ebd.), dieses soziologische Mantra sei durch avancierte Technik, die einen zunehmenden Grad an Eigenaktivität aufweist, in seiner Allgemeinheit nicht mehr haltbar.

Zunächst beantworten Rammert und Schulz-Schaeffer (2002b: 13) die Frage, welchen Typ Technik sie adressieren: Es geht vor allem um *verteilte* Aktivitäten, die sich durch Interaktivität und Dezentralität auszeichnen, keiner zentralen Steuerung bedürfen und nicht technisch determiniert sind. Die zunehmende Verteiltheit, folgt man den Autoren weiter, führt dazu, dass das Augenmerk vor allem auf das Zusammenspiel („gemeinsames Geflecht“) von Aktivitäten gerichtet werden müsse. Die Handlungsverteilung beschreiben die Autoren anhand eines Beispiels aus der Fahrzeugautomation:

„Die Verfügungsmacht über den Handlungsablauf ist nicht mehr allein auf den menschlichen Lenker konzentriert, sondern sie ist auf Mikrochips, Programme und Fahrer verteilt.“ (ebd.: 16)

Eine ähnliche Verteilung kann man auch beim Fliegen moderner Flugzeuge feststellen. Spätestens seit Einführung der Fly-by-wire-Technik<sup>37</sup> ist deutlich geworden, dass der Pilot das Flugzeug nicht (mehr) alleine fliegt. Vielmehr ist die Gesamttätigkeit „Fliegen“ ein komplexes soziotechnisches Geflecht, bestehend u. a. aus der Flugzeughülle, einer Vielzahl von Komponenten und Assistenzsystemen, dem Bord-Personal, der Flugsicherung, dem Auto-Piloten, der Fluggesellschaft usw. (ebd.: 42). Handlungsverteilung ist auch im Falle von Wikipedia-Bots (vgl. Abschnitt 5.2.3) zu beobachten: Die verteilte Lösung eines Problems wird dort im Zuge von Bot-Mensch-Interaktionen dezentral gelöst – die Gesamthandlung, z. B. die Lösung eines Lizenz-Problems, verteilt sich auf verschiedene menschliche und nicht-menschliche Akteure.

Werner Rammert und Ingo Schulz-Schaeffer beschäftigen sich mit der Frage, wie der neuartige Charakter avancierter Technik angemessen im Rahmen soziologischer Theorieentwicklung berücksichtigt werden kann. Zur Vermeidung metaphysischer Debatten um den Akteurstatus von Technik bedienen sich Rammert u. Schulz-Schaeffer eines methodologischen Tricks, indem sie Handeln als Zuschreibungsphänomen konzeptualisieren (ebd.: 20) und die Aufmerksamkeit auf die Art und Weise und die Bedingungen der Attributierungen konzentrieren. Dabei handelt es sich im Luhmann'schen Sinne um eine Beobachtung zweiter Ordnung (Luhmann 1991: 149), bei der die „[...] Realitätsgarantie [...] dann nur noch im faktischen Stattfinden des Beobachtens [liegt]. Das Beobachten zweiter Ordnung ist immer auch eine Beobachtung erster Ordnung, aber es konzentriert sich dabei auf ein Beobachten von Beobachtern, was als Selbstbeobachtung oder als Fremdbeobachtung durchgeführt werden kann.“ Wenn es nur noch um das Beobachten von Beobachtungen bzw. Zuschreibungen geht, schließt sich unmittelbar die Frage an, wie die Wirksamkeit von Handlungsträgerschaft überhaupt feststellbar werden kann. In Anlehnung an das Thomas-Theorem<sup>38</sup> (vgl. Thomas und Thomas 1928: 572) definieren Rammert und Schulz-Schaeffer (2002a: 31) die Wirksamkeit von Handlungsträgerschaft wie folgt:

„[...] gesellschaftlich wirksam wird die Handlungsträgerschaft von Technik in diesen Fällen allein dadurch, dass sich menschliche Akteure auf der Grundlage der von ihnen bewusst oder unbewusst vorgenommenen Handlungszuschreibungen im Umgang mit Technik anders verhalten als es sonst der Fall wäre.“

<sup>37</sup>Fly-by-wire bedeutet, dass Steuerungsanweisungen des Piloten nicht mehr mechanisch an z. B. die Höhenruder übermittelt werden, sondern die Impulse werden elektronisch vermittelt und letztlich führen Aktuatoren die physikalischen Bewegungen durch.

<sup>38</sup>Das Thomas-Theorem lautet: „If men define situations as real, they are real in their consequences.“ Thomas und Thomas (1928) beschäftigten sich in ihrer Abhandlung mit geisteskranken Kindern. Dabei kamen sie zu der Erkenntnis, dass sich jeder seine je eigene Realität konstruiert und diese real (und damit wirksam) wird, wenn sie hinsichtlich ihrer Konsequenzen real ist. Die Realitätskonstruktionen geisteskranker Menschen sind nicht weniger real als die davon differierenden subjektiven Realitätskonstruktionen nicht geisteskranker Menschen (ebd. 572). Hartmut Esser (1999: 64) hebt die Bedeutsamkeit des Thomas-Theorems für die gesamten Sozialwissenschaften hervor: „Das Thomas-Theorem soll verstehen helfen, warum die Subjektivität der Menschen für die Erklärung der objektiven sozialen Prozesse so wichtig ist.“

Diese Definition von gesellschaftlicher Wirksamkeit des Mit-Handelns von Technik bezieht sich darauf, dass Zuschreibungen von Handlungsträgerschaft an Technik Rückwirkungen auf den Zuschreibenden haben. Diese Rückwirkungen, so meine weiterführende Interpretation, sind besonders stark ausgeprägt bei avancierter Technik, also solcher Technik, der zunehmend Handlungsträgerschaft zugeschrieben wird.

Neben der Wirksamkeit beschäftigen sich Rammert und Schulz-Schaeffer auch mit unterschiedlichen Qualitäten des Handelns, mit einem „gradualisierten Handlungsbegriff“ (vgl. ebd.: 20f.). Dass Technik instrumentelles (Mit-)Handeln zugeschrieben wird, ist relativ unstrittig. Doch wie steht es um die Zuschreibung von Intentionen an Technik? Der Begriff Intention, also das bewusste Verfolgen von Zielen und die Ausrichtung des eigenen Handelns an ebendiesen Zielen, wird in seiner humanistischen Tradition i. d. R. ausschließlich für Menschen reserviert. Die Autoren beobachten (ebd.: 36) allerdings, dass technische Systeme wegen ihrer Komplexität nicht mehr mechanistisch beschrieben (oder zumindest verstanden) werden können, sondern nach ihren Intentionen beschrieben werden. Rammert und Schulz-Schaeffer (2002a: 36) verweisen darauf, dass diese Zuschreibung von Intentionen auch bei menschlicher Interaktion vorzufinden ist: Durch die Beobachtung des Verhaltens des Gegenüber schließen wir auf bestimmte innere Zustände und schreiben angenommene Intentionen zu.

### 5.3.1 Einschub: Technik und Intentionen

Da das Argument der zugeschriebenen Intentionalität als wesentliches Kennzeichen avancierter Technik zentral ist und in der eigenen Theorieentwicklung eine große Rolle spielt, soll es an dieser Stelle vertieft werden. Besonders relevant scheint dabei die Frage der Abgrenzung: Wo (ungefähr) ist die Grenze von autonomer und konventioneller Technik zu verorten?

Unter Rekurs auf Automationstaxonomien (vgl. Kapitel 4), die Levels of Automation von Parasuraman u. a. (2000) oder die Human Automation Collaboration Taxonomy von Cummings und Bruni (2009) würde man versuchen, *objektiv* anhand systemimmanenter Handlungsqualitäten Grade der Mithandelns zu bestimmen. In dem Fall handelt es sich nicht um eine Zuschreibung, sondern um eine Festlegung.

Zugespitzt könnte man fragen, woran der Unterschied zwischen einer Bohrmaschine und einem Navigationssystem für das Auto festgemacht werden kann. Techniksoziologisch gesprochen – ganz im Sinne eines instrumentellen Technikbegriffs – erweitert eine Bohrmaschine lediglich die zur Verfügung stehenden Handlungsoptionen. Sie wird als Werkzeug bzw. als Ressource des eigenen Handelns genutzt. Die Bohrmaschine verfolgt dabei keine eigenen Ziele, sondern bietet je nach Anwendungsfall neue Möglichkeiten. Sofern einem Techniknutzer die *Wirkungsweise* einer Bohrmaschine bekannt ist, kann er diese nicht nur zum Bohren von Löchern einsetzen, sondern sie beispielsweise auch als Ventilator verwenden.

Im Gegensatz zur Bohrmaschine dürfte den allermeisten Anwendern die interne Funktionsweise eines modernen Navigationssystems nicht exakt bekannt sein. Vielmehr wird mit der Nutzung des Navigationssystems die Erwartung verknüpft, dass uns dieses unter Berücksichtigung einstellbarer Randbedingungen (z. B. schnellste vs. kürzeste Route) zu einem definierten Ziel bringt. Die genaue Kenntnis über die Funktionsweise eines Navigationssystems würde praktisch jeden Nutzer überfordern und ist im Kontext des zgedachten Anwendungsfeldes auch nicht erforderlich.

Führt man den Gedanken weiter, dass bei einem Navigationssystem nur bekannt ist *wofür* es konzipiert ist, nicht aber wie es funktioniert, so geht die Komplexität und Eigenaktivität des Gerätes mit einer Einschränkung des Verwendungskontextes einher. Obwohl das Navigationssystem die technischen Voraussetzungen dafür mitbringt, kann dieses nicht zum automatischen Hinzufügen von Positionskordinaten in Fotos genutzt werden: Es verfügt über einen Speicherkartenslot (für die Navigationskarten) sowie ein GPS-Positionssignal und kann dennoch nicht zweckentfremdet werden bzw. andere als die ursprünglich zgedachten Ziele verfolgen.

Sofern eine betrachtete Technik nur intentional beschrieben oder verstanden werden kann, existiert ein enger Zusammenhang zwischen Intentionalität und Vertrauen – wie schon mit (FLA 3) formuliert und in Form intentionalen Vertrauens (vgl. Abschnitt 2.2) beschrieben. Eine Beurteilung, ob die avancierte Technik tatsächlich im Sinne der eigenen menschlichen Ziele (mit-)gehandelt hat, lässt sich erst im Nachhinein und im Kontext einer längeren Interaktionshistorie beurteilen. Im Fall der Wikipedia-Bots konnte dieser Abgleich über die Versionshistorie erfolgen und nachvollzogen werden, ob am Ende ein Problem gemeinsam gelöst wurde. Zunächst ist es erforderlich, der Technik mit einem Vertrauensvorschuss zu begegnen, damit diese die Möglichkeit hat, sich im Sinne der zgedachten Ziele zu verhalten und zu beweisen.

Um beim Beispiel mit der Bohrmaschine und dem Navigationssystem zu bleiben: Ob eine Bohrmaschine *funktioniert*, ist unmittelbar klar, sobald sich der Bohrer dreht. Wollen wir allerdings beurteilen, ob ein Navigationssystem seine Aufgabe gut verrichtet hat, können wir dies erst am Ende der Fahrt beurteilen, wenn wir die geplante Route mit bekannten Routen, die als besser oder schlechter bekannt sind, vergleichen. Verfügt das Navigationssystem über die Möglichkeit Staus zu umfahren, kann auch die Erfüllung dieses Ziels erst am Ende der Fahrt beurteilt werden. Ein intentionales Verständnis von Technik, wie im Falle des Navigationssystems unterstellt, bedeutet, dass ein Nutzer weiß, *was* die Technik für ihn erledigt, allerdings nicht *wie*.

Abschließend soll dieser Einschub erneut eine Brücke zurück zur attributionstheoretischen Perspektive von Rammert und Schulz-Schaeffer schlagen. Sofern man Intentionalität nicht als systemimmanente Eigenschaft von Technik auffasst, sondern als Zuschreibung, so bedeutet dies, dass Technik nur im Auge des Betrachters avanciert oder komplex sein kann. Besonders deutlich werden diese unterschiedlichen Sichtweisen von Entwicklern und Nutzern einer (avancierten) Technik. Ingenieuren von Navigationssystemen sind die internen Algorithmen, z. B. der Dijkstra-Algorithmus zur Auffindung kürzester Pfade, in ihrer Funktionsweise bekannt, während Nutzer des Navigationssystems nur im Nachhinein beobachten können, dass das Navigationssystem erfolgreich kürzeste Wege berechnet.

Worin besteht nun die Konsequenz in der Anerkennung eines beobachterabhängigen Avanciertheits- bzw. Komplexitätskonzepts von Technik? Pessimistisch könnte konstatiert werden, dass wir über einen Sachverhalt (smarte/avancierte/autonome Technik) reden möchten, dessen Charakter erst durch Zuschreibungen konstituiert wird und damit als Untersuchungsgegenstand beobachterabhängig schwimmt. Wie bei allen attributionstheoretischen bzw. konstruktivistischen Ansätzen gibt es auch hier einen Ausweg: die Suche nach Konsens und akteurübergreifenden Zuschreibungen. Sofern eine bestimmte Gruppe (z. B. Nutzer eines Navigationssystems) in Mehrheit die Arbeitsweise einer bestimmten Technik intentional beschreibt, so kann dieses technische Gerät als avanciert bezeichnet werden.

Die Idee, Technik als Mitspieler zu konzipieren und ihr Intentionen zuzuschreiben, findet sich in Teilen auch schon in attributionstheoretischen Ansätzen von Hans Geser. Geser (1989: 233) konstatiert, dass menschliche Akteure gegenüber technischen Gegenständen „Erwartungshaltungen ausbilden und Verhaltensweisen praktizieren, die in der interpersonellen Sphäre erworben worden sind.“

Folgt man Geser, so gibt es keinen Grund, die Interaktion beispielsweise mit Computern prinzipiell anders zu behandeln als die zwischenmenschliche Interaktion. Durch bestimmte, dem menschlichen Gehirn „weitläufig analoge Fähigkeiten zur Rezeption, Speicherung, Rekombination und Transformationen von Informationsbeständen“ tritt der Computer eher als Interaktionspartner denn als Werkzeug auf und ist je nach Programmierung sogar lernfähig (Geser 1989: 230f.).

In der Informatik und in der Forschung zu (Multi-)Agenten-Systemen ist es mittlerweile üblich, Software-Agenten mit BDI-Architekturen auszustatten (vgl. Rao und Georgeff 1995), sie also mit der Fähigkeit zu versehen, *beliefs*, *desires* und *intentions* zu prozessieren. Auf diese Weise können sie Verhaltensweisen generieren, die Geser (1989: 242) zufolge vom menschlichen Interaktionspartner in „soziomorphen Sinndeutungsmuster[n]“ interpretiert werden und demnach den Unterschied zur zwischenmenschlichen Interaktion verschwinden lassen.

### 5.3.2 Ein gradualisiertes Modell verteilter Handlungsträgerschaft

Auf Basis der vorangegangenen Überlegungen kommen Rammert und Schulz-Schaeffer (2002a: 41ff.) zum theoretischen Kern, dem gradualisierten Modell verteilter Handlungsträgerschaft (GMVH).

Gemäß der Begrifflichkeit der Autoren umfasst verteiltes Handeln in hybriden Zusammenhängen drei Aspekte:

1. einen Handlungsstrom (statt einzelner klar abgrenzbarer Handlungen).
2. Handlungen in medias res (statt Handlungen ex nihilo).
3. Verteiltheit in zwei Dimensionen (auf Aktivitäten, auf menschliche und nicht-menschliche Akteure).

Das Ausgehen von einem *Handlungsstrom* greift die Latour'sche Provokation auf, der etablierte Subjekt-Objekt-Asymmetrien mit seiner ANT in Frage gestellt hat (s. Abschnitt 5.1). In hybriden Systemen, so der Ansatz von Werner Rammert und Ingo Schulz-Schaeffer, ist eine eindeutige Zuordnung von Handlungen auf zugehörige Handlungsträger nicht mehr möglich und muss fallspezifisch als offene Frage gestellt werden, bei der es möglich wird, das Vorliegen oder Fehlen einer Subjekt-Objekt-Asymmetrie empirisch festzustellen (ebd.: 41).

Handlungen *in medias res* gehen auf ein Konzept von Steve Fuller (1994) zurück. Handlungen (Fuller spricht von „Agency“) in medias res finden im Kontext bereits existierender, relativ stabiler, sozialer Strukturen statt, in denen die Handlungsspielräume etwa durch Gesetze eingeschränkt werden (ebd.: 746). Die soziale Ordnung bzw. die Realität, so Fuller weiter, wird nicht in Frage gestellt, sondern als gegeben vorausgesetzt (ebd.: 747). Agency in medias res konzipiert Fuller gegenläufig zum Konzept der absoluten Agency ex nihilo, bei der die Agenten gemäß ihrer initialen Zielvorstellungen als feste Einheiten agieren und soziale Strukturen die „ongoing joint constructions of many such agents“ (ebd.: 747) sind. Fuller (1994: 747f.) argumentiert, dass der methodologische Individualismus (und damit implizit auch alle Rational Choice-Ansätze) von James Coleman (1991) ebenfalls von Agency ex nihilo ausgehe und die Wirkmächtigkeit existierender sozialer Strukturen und Gesetze unterschätze. Fuller (1994: 748) spannt mit seinem Konzept der Agency in medias res einen Bogen zwischen strukturalistischen Konzepten, die einzelne Agenten für unwichtig erklären und sie auf ihre Quidditas<sup>39</sup> „ultimately to a mere repetition of something that happened at some other time and place“ (ebd.: 748) reduzieren und individualistischen Konzepten, die das Individuum in seiner Bedeutsamkeit maßlos

<sup>39</sup> *Quidditas* ist ein Begriff aus der Philosophie und bezeichnet typische (Wesens-)Eigenschaften einer Entität, die vom konkreten Individuum abstrahieren, sondern sich auf Spezifika gleichartiger Entitäten beziehen. Im Gegensatz dazu bezeichnet die *Haecceitas* spezifische Eigenarten und Wesenszüge eines Individuums (Cross 2010).

überhöhen. Fuller geht mit der Agency in medias res einen dritten Weg, der sich weder für die strukturalistische noch für die individualistische Denkrichtung entscheidet, sondern die Möglichkeit eröffnet zwischen Situationen, in denen mal „high structure“ oder mal „high agency“ dominiert, variabel zu changieren (ebd.: 749).

Nach dem kurzen Einschub zu Handlungen in medias res folgt der dritte Aspekt des verteilten Handelns in hybriden Zusammenhängen: die *Verteiltheit*. Nach Rammert und Schulz-Schaeffer (2002a: 42) erfolgt die Verteilung des Handelns innerhalb zweier Dimensionen: In der ersten Dimension ist Handeln verteilt auf viele Aktivitäten und Instanzen und „emergiert aus dem Gesamtzusammenhang der Interaktivitäten“ (ebd.: 42). In der zweiten Dimension wird davon ausgegangen, dass Handeln auf verschiedene menschliche und nicht-menschliche Instanzen verteilt ist und unterschiedliche Grade der Handlungsträgerschaft aufweisen kann. Verteiltes Handeln umfasst demnach eine analytische Dekomposition des Gesamthandlung in auf Instanzen verteilbare Aktivitäten. Diese Aktivitäten sind in der zweiten Dimension auf menschliche und nicht-menschliche Handlungsträger verteilt und bewirken in ihrem Wirkzusammenhang eine verteilte Handlung.

Mit den drei Elementen des verteilten Handelns in hybriden Konstellationen ist der erste Aspekt des GMVH erschöpfend erfasst – die unterschiedlichen Facetten der Handlungs*verteilung*. Neben der Erkenntnis, dass ein wesentliches Kennzeichen hybrider Handlungszusammenhänge die Verteilung der Gesamthandlung auf menschliche und nicht-menschliche Akteure ist, rücken Rammert und Schulz-Schaeffer (2002a: 44ff.) von einem absoluten Handlungsbegriff ab und schlagen unterschiedliche Grade der Handlungsträgerschaft vor. Absolute Handlungsträgerschaft würde lediglich eine binäre Entscheidung ermöglichen, ob eine Instanz Handlungsträgerschaft aufweist oder nicht. Die Autoren (ebd.: 44) sehen die Einführung eines gradualisierten Handlungsbegriffs im wesentlichen als Vorbedingung ihres Konzepts der Handlungsverteilung: „Wenn wir ein Konzept verteilten Handelns entwickeln, [...] so benötigen wir dazu zunächst einen Handlungsbegriff, der voraussetzungsärmer und voraussetzungsreichere Aspekte dessen, was es heißen kann zu handeln, differenziert.“

Insgesamt werden folgende Ebenen bzw. Grade der Handlungsträgerschaft unterschieden:

1. **Kausalitätsebene:** Handlungsträgerschaft auf dieser Ebene ist gekennzeichnet durch die Bewirkung von Veränderungen (ebd.: 44f.). Sofern es einem Bankkunden nur um den Erhalt eines bestimmten Geldbetrags gehe, sei es auf der „Vollzugsebene“ unerheblich, ob diese Aufgabe von einem Geldautomaten oder von einem Schalterbeamten durchgeführt werde. Charakteristisch für die Kausalitätsebene ist ferner die relativ mechanistische Abarbeitung vorgegebener Handlungsprogramme. Die Abgrenzung zur nächsthöheren Ebene ist dabei das Unvermögen, auf Störungen flexibel reagieren zu können (vgl. ebd.: 45): Der

Schalterbeamte kann ohne Probleme auch anders handeln und z. B. den Wunsch des Kunden nach einer anderen Stückelung der Geldscheine erfüllen, während der Bankautomat sein festes Programm abarbeitet und die vorgesehenen Scheine herausgibt<sup>40</sup>.

2. **Kontingenzebene:** Auf dieser Ebene geht es um das „Auch-anders-handeln-Können“. Ein modernerer Terminus würde Adaptivität oder auch Flexibilität als Kennzeichen dieser Ebene benennen. Zwei Aspekte sind kennzeichnend für Handlungsträgerschaft auf der Kontingenzebene: Erstens das flexible Eingehen auf neuartige Gegebenheiten und veränderte Umweltbedingungen. Bezogen auf einen kontingent agierenden Geldautomaten würde dies z. B. bedeuten, dass dieser je nach Bankkunde eine unterschiedliche Geldstückelung anbietet. Zweites umfasst Kontingenz, die Möglichkeit, das eigene Verhalten so zu verändern, dass es für andere Akteure nicht exakt vorhersagbar ist. Ein weiteres typisches Merkmal für Handeln auf der Kontingenzebene ist die Möglichkeit, aus vorherigen Interaktionen zu lernen und das eigene Handeln auf das Gelernte abzustimmen.
3. **Intentionalitätsebene:** Rammert und Schulz-Schaeffer (2002a: 46) argumentieren, dass der Begriff der Intentionalität, sofern man ihn im humanwissenschaftlichen Sinne versteht und Bewusstseinsfähigkeit und Sinnhaftigkeit explizit einschließt, ausschließlich Menschen vorbehalten ist. Allerdings, so die Autoren weiter, weisen komplexe Systeme die Eigenart auf, nicht mehr nach ihrer Wirkungsweise, sondern nach ihren Zielen beschrieben und programmier-technisch definiert zu werden. Technik steuert interne Abläufe über Konzepte, die Ähnlichkeiten mit den Semantiken menschlicher Intentionalitätsbegriffe haben<sup>41</sup>.

Die intentionale Deutung menschlichen Verhaltens beruhe wie die Deutung technischen Verhaltens auf Blackbox-Beobachtungen<sup>42</sup>, also einem Schließen auf innere Beweggründe durch äußerlich beobachtbares Verhalten (ebd.: 47). Diese Feststellung wird zugespitzt auf die These, dass man auch auf der Intentionalitätsebene den Fokus auf die gesellschaftlichen Praktiken der Verwendung

<sup>40</sup>Es soll nicht unerwähnt bleiben, dass inzwischen einige Bankautomaten auch eine solche flexible Stückelung anbieten. Dies zeigt einmal mehr, dass Grade der Handlungsträgerschaft immer vom jeweils betrachteten empirischen Fall abhängen. Ein Bankautomat (oder auch jeder andere Automat) agiert nicht zwangsläufig nur auf der Kausalitätsebene. Der Grad der Handlungsträgerschaft hängt vom technischen System selbst ab und von den zugeschriebenen Graden der Handlungsträgerschaft, die je nach menschlichem Nutzer, der mit dem Geldautomat interagiert, höchst unterschiedlich ausfallen können.

<sup>41</sup>Rammert u. Schulz-Schaeffer rekurren dabei vor allem auf das BDI-Programmiermodell (vgl. Bratman 1987; Rao und Georgeff 1995), das Algorithmen mit Hilfe intentionalistischen Vokabulars nicht nur beschreibt, sondern auch definiert – nach **B**eliefs, **D**esires und **I**ntentions.

<sup>42</sup>Meine Weiterauslegung der Blackbox-These lautet wie folgt: Komplexe technische Systeme sind von ihrer Funktionsweise her nicht mehr vollständig durchschaubar. Aus dem Grund beschränken sich Interaktionspartner auf die Beobachtung des Eingabe-Ausgabe-Verhaltens der Technik und auf dieser Basis wird auf die Ziele bzw. die intendierten Zielzustände geschlossen.

intentionalen Vokabulars richten und diese nicht von vornherein ausschließlich menschlichen Akteuren vorbehalten solle:

„[...] wir plädieren dafür, das Augenmerk auf die empirisch beobachtbaren gesellschaftlichen Praktiken der Verwendung intentionaler Begriffe bei der Steuerung und Interpretation menschlichen wie technischen Verhaltens zu richten.“ (Rammert und Schulz-Schaeffer 2002a: 47)

Die Autoren plädieren somit für ein pragmatisches Verständnis von Intentionalität, das einen menschlich-intentionalen Primat nicht ex ante voraussetzt, sondern diesen zur empirisch offenen Untersuchungsfrage macht.

Die zentralen Bestandteile des GMVH lassen sich wie folgt zusammenfassen:

#### **Das gradualisierte Modell verteilter Handlungsträgerschaft**

Handeln vollzieht sich in einem Handlungsstrom im Kontext relativ gefestigter gesellschaftlicher Strukturen und ist verteilt auf unterschiedliche Aktivitäten, die ihrerseits auf menschliche und nicht-menschliche Akteure verteilt sind. Dabei lassen sich unterschiedliche Grade bzw. Ebenen der Handlungsträgerschaft unterscheiden: Kausalitätsebene, Kontingenzebene und Intentionalitätsebene. Entscheidend sowohl für die Verteilung wie auch für die Grade der Handlungsträgerschaft sind die beobachtbaren gesellschaftlichen Praktiken und Zuschreibungen.

Wenn Handlungsträgerschaft autonomer Technik durch Zuschreibungen konstituiert wird und Vertrauen gemäß (FLA 6) die Unterstellung gewisser Erwartungen und Handlungsmuster beinhaltet, dann sind Zuschreibungsprozesse auch für die Untersuchung von Vertrauen in autonome Technik von Bedeutung. Mit (FLA 3) wurden Vertrauen allgemein bereits als gerichtete Zuschreibung identifiziert, was sich mit der hier vorgenommenen Schlussfolgerung deckt.

#### **Handlungsträgerschaft von Technik?**

In diesem Abschnitt wird der Frage nachgegangen, welche Konsequenzen sich aus dem GMVH für die Mithandlungsträgerschaft von Technik ergeben.

Bleibt man im Analyserahmen des GMVH, lässt sich die Frage der Handlungsträgerschaft von Technik beantworten, indem man die gesellschaftlich vorgenommenen Zuschreibungen<sup>43</sup> in den Blick nimmt. Gehen die Mitglieder der Gesellschaft mit avan-

<sup>43</sup>Ingo Schulz-Schaeffer beschäftigte sich in seiner Habilitationsschrift ausführlich mit „Zugeschriebenen Handlungen“ (vgl. Schulz-Schaeffer 2007).

cierter Technik so um, als wäre diese ein Handlungsträger? Um das Modell empirisch anzuwenden sind zumindest folgende Schritte erforderlich<sup>44</sup>:

1. Identifizierung der zu untersuchenden Gesamthandlung. Insbesondere in komplexen soziotechnischen Konstellationen ist dieses Abgrenzungsproblem nicht immer leicht lösbar.
2. Identifizierung der an der Gesamthandlung beteiligten Handlungsträger.
3. Dekomposition des Gesamthandlung in Aktivitäten.
4. Verteilung der Aktivitäten auf menschliche und nicht-menschliche Akteure.
5. Zuordnung von Graden der Handlungsträgerschaft für die identifizierten Aktivitäten.

Die Frage nach der Handlungsträgerschaft von Technik ist folglich nicht ontologisch beantwortbar, sondern muss sich an den Attributierungen und den damit verbundenen gesellschaftlichen Praktiken orientieren. In welchem Maße Technik (mit-)handelt, ist mit dem begrifflichen Instrumentarium der drei Handlungsträgerschaftsebenen qualifizierbar. Ob das identifizierte Mit-Handeln schließlich als menschenähnliches Handeln identifiziert wird, hängt davon ab, auf welcher Ebene menschliches Handeln verortet wird. Um dabei begriffliche Präjudizierungen zu vermeiden, erscheint es konsequent – im Sinne einer attributionstheoretischen Perspektive – auf die Angabe einer harten Grenze entweder ganz zu verzichten und stattdessen nur die vorgefundene Handlungspraxis zu beschreiben, oder aber diese Grenze im jeweiligen empirischen Fall von den beteiligten Handlungsträgern selbst bestimmen zu lassen.

Mit den Worten der Autoren wird die Frage nach der Mithandlungsträgerschaft von Technik wie folgt beantwortet:

„Beantwortbar wird die Frage nach dem Handeln von Technik dadurch, dass in diesem hybriden Strom von Aktivitäten die empirische Verteilung auf die menschlichen und nicht-menschlichen Instanzen wie auch die Zurechnung von Handlungsqualitäten zum Gegenstand der Analyse gemacht werden kann.“ (Rammert und Schulz-Schaeffer 2002a: 43)

### 5.3.3 Schlussüberlegungen und kritische Würdigung

Rammert u. Schulz-Schaeffer bezeichnen ihren eigenen Ansatz interessanterweise als eine Art dritten Weg zwischen Technikgeneseforschung und Technikfolgenabschätzung (vgl. ebd.: 22). Mit dem GMVH sei es nun möglich, die Wechselwirkungen von

<sup>44</sup>Bei den genannten Schritten (insbesondere bei 3 und 4) besteht die Aufgabe des Sozialwissenschaftlers vor allem darin, die wirksamen Zuschreibungen zu untersuchen. Allerdings kann man auch bei den anderen Schritten die Frage stellen (mit der ich mich in Abschnitt 5.3.3 kritisch auseinandersetzen werde) an welcher Stelle eine gewisse analytische Vorstrukturierung erforderlich ist, um das Modell nicht der völligen Beliebigkeit preiszugeben.

soziotechnischen Konstellationen zu untersuchen. Die Technikfolgenabschätzung fragt nach den Auswirkungen von Technik auf die Gesellschaft. Die Technikgeneseforschung befasst sich mit den gesellschaftlichen Ausgangsformationen, die zur Entstehung einer bestimmten Technik beitragen. Der Fokus des GMVH liegt auf der Frage, wie unterschiedliche Technik-Gesellschafts-Konstellationen miteinander wechselwirken.

Dieser hohe Anspruch, den das GMVH für sich reklamiert, kann meines Erachtens nicht oder nur sehr bedingt eingelöst werden. Hinsichtlich der Modellierungsmöglichkeiten umfasst das Modell vordergründig ein präzises begriffliches Instrumentarium zur Beschreibung des IST-Zustandes einer soziotechnischen Konstellation bzw. eines hybriden Systems verteilter Handlungsträgerschaft. Wirkzusammenhänge zwischen unterschiedlichen Konstellationen werden nicht näher in den Blick genommen und theoretisch nicht weiter präzisiert. Es fehlt ein integrierendes Modell, das die Beobachtungen von Rammert u. Schulz-Schaeffer in einen größeren prozessförmigen Erklärungszusammenhang einordnet. Welche Auswirkungen haben bestimmte Arten des Zusammenhandelns auf eine darüber liegende Makro-Ebene? An diesem Defizit wird die eigene Methodologie (Kapitel 8) anknüpfen und die Interaktion von Mensch und autonomer Technik in ein einheitliches Makro-Mikro-Makro-Modell integrieren.

Darüber hinaus erscheint die analytische Trennung von Graden der Handlungsträgerschaft zwar inhaltlich elegant, methodologisch allerdings anspruchsvoll. Die Handlungsverteilung, bzw. genauer: die Verteilung der Aktivitäten der Gesamthandlung auf die beteiligten Handlungsträger, ist nicht unabhängig zu sehen von den Graden der Handlungsträgerschaft. Eng zusammenhängend mit dem methodologischen Problem der Aktivitäten, das im nächsten Abschnitt betrachtet wird, ist es wenig sinnvoll für „atomare“ Aktivitäten Grade der Handlungsträgerschaft festzustellen, denn dies wird nur in Bezug auf bestimmte Bündel von Aktionen sinnvoll möglich sein. Folgert man weiter, bedeutet dies, dass die Verteilung bestimmter Aktivitätskombinationen, zumindest nicht völlig unabhängig von den Graden der Handlungsträgerschaft zu betrachten ist.

Neben den genannten modellimmanenten Schwierigkeiten wäre es darüber hinaus wünschenswert gewesen, wenn sich die Autoren in benachbarten wissenschaftlichen Disziplinen umgesehen hätten, etwa der Automationsforschung, da dort ebenfalls Probleme der verteilten (und gradualisierten) Handlungsträgerschaft diskutiert werden, wenn auch unter anderen Termini. Handlungsverteilung heißt dort „allocation of functions“ und mit den „Levels of Automation“ (vgl. Parasuraman u. a. 2000) oder auch der HACT, existieren Modelle, die feingranulare Möglichkeiten zur Modellierung von Handlungsverteilung und Qualifizierung von Aktivitäten zwischen Mensch und Maschine ermöglichen. Der konzeptionelle Unterschied zwischen den beiden Ansätzen besteht darin, dass Parasuraman u.a. die Automationslevel als *objektive* Eigenschaft des betrachteten Mensch-Maschine-Settings ansehen, während Rammert und Schulz-Schaeffer eine attributionstheoretische Perspektive einnehmen und sowohl Handlungs-

verteilung als auch Grade der Handlungsträgerschaft ausschließlich als Beobachtungen zweiter Ordnung analysieren können, indem sie beschreiben, wie gesellschaftlich mit den relevanten Zuschreibungen umgegangen wird.

### **Das methodologische Problem der Aktivitäten**

Rammert und Schulz-Schaeffer betonen, dass man vor dem Hintergrund zunehmender Aktivität von Technik nicht mehr von einzelnen Handlungen ausgehen kann, sondern von einem Geflecht „vermischter Aktivitäten“ (ebd.: 13). Das GMVH sieht vor, diese Aktivitäten dann verteilt auf menschliche und nicht-menschliche Akteure zu beschreiben.

Die Auffassung von Aktivitäten als eine Art atomare Einheit des verteilten Handelns bringt das methodologische Problem der Abgrenzung zwischen Aktivität und Handlung mit sich. Dieser Aspekt lässt sich an einem Beispiel (ebd.: 42) der Autoren gut verdeutlichen: Es wird argumentiert, dass die Gesamthandlung „Fliegen“ aus einer Vielzahl von Aktivitäten besteht, die sich auf beteiligte Akteure verteilen (z. B. auf den Piloten, die Passagiere, die Fluggesellschaft, den Autopiloten). Als eine Aktivität wird beispielsweise das „Bedienen des Steuerknüppels“ (ebd.) angegeben. Allerdings lässt sich feststellen, dass mit der Einführung der Fly-by-wire-Steuerung in der zivilen Luftfahrt erstmals mit dem Airbus A320 in den 80er Jahren (vgl. Weyer 1997: 247) die Bedienung des Steuerknüppels keine atomare Aktivität mehr ist. Sie ist wiederum in ein soziotechnisches Geflecht bzw. ein System verteilter Handlungsträgerschaft bestehend aus Pilot, Steuerungscomputer und Aktuatoren eingebettet. Fly-by-wire bedeutet dabei keinesfalls, dass die physikalischen Steuerungssignale des Piloten 1:1 (in verstärkter Form) an die Aktuatoren, die z. B. die Störklappen bedienen, weitergegeben werden. Steuerungscomputer verarbeiten die Steuerungsimpulse des Piloten und überprüfen, ob die momentane Situation des Flugzeugs eine Umsetzung „sinnvoll“ zulässt und sich im Rahmen bestimmter Grenzwerte befindet (Weyer 2008b: 250). Dieses Beispiel zeigt, dass die Festlegung auf einzelne Aktivitäten problematisch sein kann und keinesfalls klar ist, ob es sich um eine Aktivität oder doch um eine verteilte Handlung handelt.

Eine weitere Schwierigkeit besteht darin, dass für eine sinnvolle Analyse verteilter Aktivitäten eine derartige Trennung und Identifizierung relevanter Aktivitäten vorgegeben sein muss. Im strikt attributionstheoretischen Sinne müsste die Identifikation von einzelnen Aktivitäten spezifisch für jeden Zuschreibenden sein. In diesem Fall würde man das Modell allerdings der völligen Beliebigkeit opfern. Andererseits scheint es, zumindest aus ingenieurwissenschaftlicher Perspektive, nicht völlig abwegig bestimmte Einzelaktivitäten angeben zu können, die dann allerdings ggf. so feingranular sind, dass eine gehaltvolle soziotechnische Analyse ausgeschlossen ist.

Zusammenfassend erscheint mir eine sinnvolle Identifikation von Einzelaktivitäten als eine der größten Herausforderungen bei der Anwendung des GMVH – eine Art

Quadratur des Kreises: Eigentlich ist keine präzise Zurechnung von Handlungen auf die beteiligten menschlichen und nicht-menschlichen Instanzen möglich, auf der anderen Seite soll ebendiese mit Hilfe atomarer Aktivitäten erfolgen. Ein möglicher Ausweg aus dem Dilemma könnte dabei eine enge Zusammenarbeit zwischen Technikern und Sozialwissenschaftlern sein, um ein passendes Aktivitätsmodell für die zu analysierende soziotechnische Konstellation zu entwickeln, das einen Mittelweg zwischen ingenieurwissenschaftlicher Feingranularität und attributionstheoretischer Beliebigkeit geht. Als weitere Lösung, bzw. als möglichen forschungspragmatischen Umgang mit diesem Problem bietet sich die Durchführung von computersimulativen Experimenten, z. B. agentenbasierten Simulationen (vgl. Abschnitt 7.3) an, bei denen die Möglichkeit besteht, Interaktionsdaten auf feingranularer Ebene aufzuzeichnen. Ein solcher Ansatz wurde auch für das Experiment SIMHYBS im Rahmen der vorliegenden Arbeit gewählt (vgl. Abschnitt 11).

### **Was bleibt vom GMVH?**

Mit dem GMVH wurde der empirisch evidente Sachverhalt des verteilten Handelns in hybriden Systemen auf eine solide theoretische Grundlage gestellt, die sich nicht nur in bloßer Polemik übt, wie es bei Latour der Fall war, sondern das (Mit-)Handeln von Technik konstruktiv aufgreift und ein begriffliches Instrumentarium zur angemessenen Beschreibung dieses Sachverhaltes anbietet. Unmittelbar profitieren von dem Modell allerdings eher Arbeiten der *qualitativen* Sozialforschung, etwa die im Sammelband von Rammert und Schubert (2006) versammelten Technografie-Studien. Für eine unmittelbare Anwendung in der quantitativen Sozialforschung fehlt dem Modell an einigen Stellen die analytische Präzision, z. B. bei der Abgrenzung von Aktivitäten und der Operationalisierung von Verteiltheit und Graden der Handlungsträgerschaft. Es wird zwar von unterschiedlichen Graden der Handlungsträgerschaft und Verteilung gesprochen, diese zentralen Konzepte werden allerdings eher auf einer lebensweltlich-deskriptiven Ebene abgehandelt. Trennschärfere Kriterien für die Grade der Handlungsträgerschaft, insbesondere aber für die Handlungsverteilung bleiben die Autoren schuldig, wenngleich in Rechnung zu stellen ist, dass sie ihr Modell strikt attributionstheoretisch konzipieren und eine inhaltliche Präzisierung dieser Idee ggf. zuwiderlaufen würde.

An den aufgelisteten Defiziten knüpft der eigene Ansatz in Kapitel 8 an. Dabei wird das Zusammenhandeln von Mensch und autonomer Technik im Kontext eines gemeinsamen Ziels, bzw. einer gemeinsam zu erfüllenden Aufgabe betrachtet. Der Rückgriff auf soziologische Makro-Mikro-Makro-Erklärungen in einer interaktiven Prozess-Perspektive ermöglicht es, Wechselwirkungen zwischen dem Zusammenhandeln von Mensch und autonomer Technik und der übergeordneten soziotechnischen Einbettung in den Blick zu nehmen. Im Gegensatz zum GMVH existiert ein einheitlicher, integrierender Bezugsrahmen, der die Aktivitäten von Mensch und autonomer Technik

auf identische Weise operationalisieren kann. Die attributionstheoretische Sichtweise ist dabei nach wie vor von Bedeutung, da insbesondere *Modellzuschreibungen* (vgl. Abschnitt 8.2) betrachtet werden. Die Betrachtung von Modellzuschreibungen und die Systematisierung von Modelldifferenzen (Abschnitt 8.2.1) kann Probleme, die sich aus wechselseitigen Zuschreibungen zwischen Mensch und Technik ergeben, analysieren.

Das methodische Problem der Abgrenzung von Aktivitäten lässt sich zwar nicht grundsätzlich ausräumen, aber mit Hilfe der entwickelten Methodologie (Abschnitt 8.1), die anschlussfähig ist an Computersimulationen und Laborexperimente, zumindest kontrollieren.

Die Kritik am GMVH lässt sich wie folgt knapp zusammenfassen:

#### **Kritik am GMVH**

- Es existiert keine prozessförmige Sichtweise sowie kein integrierender Makro-Mikro-Makro-Rahmen.
- Die Grade der Handlungsträgerschaft und die Verteilung der Aktivitäten sind nicht unabhängig voneinander.
- Es besteht ein Abgrenzungsproblem von Aktivität und Handlung.
- Zentrale Konzepte wie Handlungsträgerschaft und Grade der Handlungsträgerschaft sind nicht operationalisierbar.

## 6 Handlungstheoretische Mikro-Makro-Modelle

Am Ende des vorherigen Kapitels ist einige Kritik am graduellen Modell verteilter Handlungsträgerschaft zur Sprache gekommen (vgl. Abschnitt 5.3.3). Die eigene Methodologie, das Modell soziologischer Erklärung hybrider Systeme (Kapitel 8), kann durch ihre handlungstheoretische Fundierung und Integration in einen Makro-Mikro-Makro-Zusammenhang einige der Kritikpunkte ausräumen.

Das Esser'sche Modell soziologischer Erklärung (Esser 1996) mit seinem handlungstheoretischen Kern, der SEU-Theorie, stellt die Modellierungsgrundlage für den eigenen Ansatz dar und wird ausführlich in Abschnitt 6.2 besprochen. Das Esser'sche Erklärungsmodell basiert seinerseits im Kern auf den von James Coleman (1991) formulierten Überlegungen zu einer allgemeinen Sozialtheorie, die Wechselwirkungen zwischen der gesellschaftlichen Makroebene und der Mikroebene der Akteure in den Blick nimmt (Abschnitt 6.1).

Das vorliegende Kapitel versteht sich nicht als Repetitorium handlungstheoretischer Literatur oder als umfangreiche Einführung in verschiedene handlungstheoretische Ansätze. Vielmehr liegt der Fokus insbesondere auf erkenntnistheoretischen Überlegungen und der damit verbundenen Frage, welche Erklärungskraft von Makro-Mikro-Makro-Modellen ausgeht. Neben erkenntnistheoretischen Überlegungen wird der heuristische Modellierungswert eines formalisierbaren Erklärungsmodells – „Methodologie statt Ontologie“ (Kroneberg 2009) – betont, der auch in Abschnitt 7.2 im Hinblick auf Simulationsmodelle besprochen wird.

Handlungstheoretische Mikro-Makro-Ansätze, wie sie im Folgenden vorgestellt werden, sind anschlussfähig an die vorangegangenen Überlegungen. Im Rahmen der forschungsleitenden Annahmen ist der enge Zusammenhang von Vertrauen und Handlungen und den damit verbundenen Gründen schon mehrfach thematisiert worden. So ist in (FLA 2) der Zusammenhang von Vertrauen und Verhalten, in (FLA 6) der Zusammenhang von Vertrauen und Handlungsmustern sowie Erwartungen, in (FLA 7) die Notwendigkeit des Verständnisses von Gründen sowie schließlich in (FLA 10) die Handlungswirksamkeit autonomer Technik betont worden. Demnach scheint es inhaltlich gerechtfertigt, zunächst eine nähere Beschäftigung mit handlungstheoretischen Modellen voranzustellen, bevor der eigene Ansatz auf ebendiese rekurren kann. Die in Abschnitt 6.2 behandelte SEU-Theorie stellt für diesen Zweck einen sinnvollen Modellierungsansatz dar. Es werden Gründe und Handlungen auf Modellierungsebene

systematisch berücksichtigt. Damit ist eine methodische Anschlussfähigkeit an Computersimulationen und Experimente (vgl. Kapitel 7) gewährleistet, so dass Vertrauen gemäß (FLA 4) im Rahmen experimenteller Settings untersucht werden kann.

## 6.1 Das Coleman'sche Erklärungsmodell

Die vorliegende Arbeit schließt sich dem Anspruch der Coleman'schen Sozialtheorie weitgehend an, dass soziale Strukturphänomene über die an dem Phänomen beteiligten Einheiten (meist Akteure) tiefenerklärt werden müssen. Eng mit dieser Denkweise verbunden ist die sogenannte „Coleman'sche Badewanne“. Wenngleich Karl-Dieter Opp die Urheberschaft für ebendiese reklamiert (Opp 2009: 28), erstrecken sich die hier vorgetragenen Erklärungen auf das Mikro-Makro-Modell von James Coleman (1991: 10), da spätere Autoren meist auf sein sozialtheoretisches Grundlagenwerk und nicht auf Opp rekurren.

Das Hauptaugenmerk sozialwissenschaftlicher Forschung liegt laut Coleman (1991: 2) „in der Erklärung sozialer Phänomene, nicht in der Erklärung von Verhaltensweisen einzelner Personen“ – das ist Aufgabe der Psychologie, ließe sich ergänzen. Coleman fügt seinen Überlegungen gleich eine soziologische Stopp-Regel<sup>1</sup> hinzu: Eine Erklärung sollte nur so tief unter die Systemebene gehen, dass ein Eingriff auf dieser Ebene (z. B. eine Verhaltensänderung, Anm. R. F.) noch zu einer Änderung des Systemverhaltens beitragen kann (ebd.: 5). Es handelt sich bei dieser Stopp-Regel um eine relative Grenze, die von den zu erklärenden Phänomenen abhängt. Diese Beschränkung auf das Wesentliche kann sowohl horizontal als auch vertikal interpretiert werden: Horizontal bedeutet sie, dass nur die Facetten und Einstellungen eines Akteurs Berücksichtigung finden sollten, die für die Erklärung des sozialen Phänomens relevant sind. Die vertikale Grenze berücksichtigt, dass gerade so weit auf die Individualebene heruntergegangen wird, wie es für die Erklärung notwendig ist. Genügt zur Erklärung eine Grobunterscheidung typisierter Akteure<sup>2</sup>, so können auch abstrahierte Individuen Strukturphänomene erklären.

Coleman weist darauf hin, dass sich sein Ansatz in zwei Aspekten vom methodologischen Individualismus unterscheidet (ebd.: 6) bzw. eine Spezialisierung dessen darstellt: Erstens fasst Coleman Makro-Phänomene nicht als einfache Aggregate von Mikro-Handlungen auf, sondern spricht von Emergenz. Emergenz bedeutet, dass das Zusammenwirken von Akteuren zu unvorhersagbaren und unbeabsichtigten Effekten auf der Strukturebene führt. Als zweiten Unterschied führt er den bereits zuvor erwähnten Punkt an: Befriedigende Erklärungen erfolgen nicht ausschließlich über die Individualebene wie beim reinen methodologischen Individualismus, sondern können

<sup>1</sup>Diese Formulierung stammt von Johannes Weyer (zit. nach Kron 2004: 201).

<sup>2</sup>So z. B. erfolgt bei Kroniger und Lücke (2010), die unterschiedliche Akteurstypen zur Erklärung von spezifischen Stauphänomenen bei Fußballspielen herangezogen haben.

von Individualakteuren abstrahieren, sofern die Abstraktion nach wie vor Veränderungen des Systemverhaltens abbilden kann.

Die Abgrenzungsüberlegungen zum rein individualistischen Programm führen Coleman schließlich zum Hauptproblem seines sozialtheoretischen Ansatzes, dem Mikro-Makro-Problem (Coleman 1991: 7ff.). Gemeint ist damit die Frage, wie Handlungen von Akteuren auf der Individualebene zu Makrophänomenen auf der Strukturebene aggregiert werden. Er betont, dass eine einfache Aggregation zu unterkomplex sei und dass es komplexerer Aggregationsmechanismen bedürfe (ebd.: 6; 27). Das Argument ist plausibel, allerdings bleibt Coleman eine nähere Erläuterung dieser Aggregationsmechanismen schuldig. Zwar beschreibt er beispielartig wie komplexere Mechanismen funktionieren (ebd.: 25f.), beschäftigt sich aber nicht näher mit einer Generalisierung dieser Mechanismen in Form von Transformationsregeln. Weiter bleibt er eine Erklärung schuldig, was unter einfachen Aggregationen zu verstehen ist.

### 6.1.1 Die drei Übergänge

Zentral im Coleman'schen Erklärungsschema sind die drei Übergänge (Coleman 1991: 14f.): Der Übergang von der Makro- zur Mikroebene (Typ 1), von der Mikro- zur Mikroebene (Typ 2) und schließlich von der Mikroebene zurück auf die Makroebene (Typ 3).

Für die Übergänge vom Typ 1 und Typ 3 sind die „Spielregeln“ von besonderer Bedeutung. Die Spielregeln geben an, welche Auswirkungen das Individualhandeln auf die Makroebene hat (Mikro-Makro-Übergang), z. B. in Form von veränderten Marktpreisen. Zum anderen geben die Spielregeln an, welche Einschränkungen für das Individualhandeln gelten und welche Auswirkungen die Handlungen anderer Akteure auf das Individualhandeln haben (Makro-Mikro-Übergang). Coleman beschäftigt sich nicht näher mit einer Theorie der Situation, wie sie sich in späteren Ausarbeitungen des Modells (vgl. Esser 1991: 47, 1993: 47) oder noch später bei Kroneberg (2005) finden. So bleibt bei Coleman nicht nur die Deutung der Spielregeln unterspezifiziert, sondern auch, wer die Spielregeln überhaupt festlegt. In der Coleman'schen Sozialtheorie jedenfalls sind sie nicht das Produkt der beteiligten Akteure, wie man strikt individualistisch argumentieren müsste, sondern eher eine aufoktroyierte analytische Sichtweise des Forschers, der sich mit den Wechselwirkungen zwischen Mikro- und Makroebene in dem jeweils betrachteten sozialen System befasst und die zugrunde gelegten Spielregeln für sinnvoll erachtet.

Deutlich präziser beschreibt Coleman hingegen den Mikro-Mikro-Übergang und befasst sich ausführlich (ebd.: 16ff.) mit der Frage, welche Struktur eine Handlungstheorie der Individualebene haben sollte. Demnach soll eine Handlungstheorie von zielgerichteten Handlungen ausgehen. Coleman selbst hält die Idee der zielgerichteten Handlungen längst für common sense (ebd.: 17). Als weitere Spezialisierung schlägt er das in der Ökonomie verbreitete Konzept des nutzenmaximierenden Akteurs vor,

„wonach der Akteur diejenige Handlung auswählt, die den Nutzen maximiert“ (ebd.). Im Gegensatz zu späteren Auspezifizierung seines Makro-Mikro-Makro-Modells z. B. bei Esser (1996) geht Coleman von einem *objektiven Rationalitätsbegriff* aus, der auf situative Opportunitäten für beliebige Akteure genau eine Antwort kennt – die nutzenmaximierende Handlungsalternative.

Coleman tritt vehement für eine Theorie der zielgerichteten Handlung ein, die er selbst als „explizit teleologisch“ (ebd.: 19) bezeichnet, wenngleich er sich der wissenschaftstheoretischen Probleme teleologischer Erklärungen bewusst ist: „Es ist den in der Wissenschaft üblichen kausalen Erklärungen entgegengesetzt.“ Dem Argument der scheinbaren Unwissenschaftlichkeit teleologischer Erklärungen entgeht er geschickt dadurch, dass das zu erklärende Phänomenen auf einer anderen Ebene verortet ist: Auf der Mikroebene kommen zwar teleologische Erklärungen vor, der Übergang von der Mikroebene zurück auf die Makroebene ist hingegen nicht teleologisch, sondern kausal und „stützt sich auf unmittelbar wirksame Ursachen“ (ebd.: 20). Für Coleman bedeutet das Verstehen einer Handlung eines Akteurs, seine Beweggründe, seine Ziele herauszufinden, so dass eine Handlung im Lichte seiner Ziele rational erscheint.

Die Gründe für den Rückgriff auf eine spezialisierte Version der zielgerichteten Handlungstheorie, der nutzenmaximierenden, sind für Coleman vor allem von forschungspragmatischer Natur (ebd.: 23f.). Er misst in seinem dreischrittigen Erklärungsmodell den Makro-Mikro- und Mikro-Makro-Übergängen eine größere Bedeutung zu und möchte insbesondere einer Psychologisierung entgegenwirken, weshalb er auf eine einfache Variante des Individualhandelns zurückgreift.

### 6.1.2 Erkenntnistheoretische Überlegungen zum Makro-Mikro-Makro-Modell

In diesem Abschnitt werden einige erkenntnistheoretische Überlegungen zum Coleman'schen Makro-Mikro-Makro-Modell angestellt und erläutert, inwiefern es sich um eine wissenschaftliche Erklärung im Sinne des deduktiv-nomologischen Modells von Hempel und Oppenheim (1948) handelt. Argumentativ werden sich die Überlegungen unter Rekurs auf Kroneberg (2009) dahingehend zuspitzen, dass Makro-Mikro-Makro-Modelle eher als Methodologie denn als Ontologie aufzufassen sind. Diese Überlegung wird, wie schon in der Einleitung dieses Kapitel angedeutet, für Computersimulationsmodelle eine Rolle spielen, bei denen es nicht zwangsläufig um möglichst realitätsnahe Modelle geht, sondern durch den Prozess des Modellierens ein heuristischer Modellierungsmehrwert realisiert werden kann.

#### Einschub: Das Hempel-Oppenheim-Schema

Das Hempel-Oppenheim-Schema (H-O-Schema), das auf einen Aufsatz von Carl Gustav Hempel und Paul Oppenheim zurückgeht (Hempel und Oppenheim 1948), be-

fasst sich mit der Frage, wie eine wissenschaftliche Erklärung aufgebaut ist bzw. wie eine Erklärung aufgebaut sein *sollte*, damit diese als wissenschaftlich gelten kann. Hempel und Oppenheim (1948: 136) unterscheiden zunächst zwischen *Explanandum*, dem zu Erklärenden und *Explanans*, dem Erklärenden. Das Explanandum beschreibt das zu erklärende Phänomen, das Explanans enthält jene Elemente, die für eine Erklärung des Explanandums herangezogen werden können. Für eine Erklärung zugelassene Elemente sind (ebd.: 137):

- Randbedingungen (Circumstances):  $C_1, C_2, \dots, C_k$
- Allgemeingültige Gesetze (Laws):  $L_1, L_2, \dots, L_r$

In Tabelle 6.1 ist die Struktur und logische Schlussweise einer wissenschaftlichen Erklärung nach Hempel und Oppenheim dargestellt: Demnach besteht eine wissenschaftliche Erklärung aus allgemeinen Gesetzen  $L_j$ , „lawlike sentences [...] of universal form“ (ebd.: 153) die zusammen mit dem Vorliegen bestimmter Rand- bzw. Vorbedingungen  $C_i$  eine logische Schlussfolgerung auf das Explanandum  $E$  erlauben. Diese Art der Erklärung wird auch *deduktiv-nomologisch* genannt. Deduktiv ist sie, weil allgemeine Gesetze unter Geltung bestimmter Randbedingungen auf einen zu erklärenden empirischen Fall angewendet werden. Nomologisch ist sie, weil zur Erklärung von Phänomenen allgemeine Gesetze verwendet werden.

Hempel und Oppenheim (1948: 157ff.) spezifizieren eine eigene Modellsprache, die u. a. zulässige logische Schlussfolgerungen innerhalb des H-O-Schemas angibt sowie Bedingungen an „lawlike sentences“ spezifiziert. Die vollständige Darstellung dieses Logik-Kalküls würde für den Zweck der vorliegenden Arbeit zu weit führen, da in Anlehnung an Kroneberg (2009) „nur“ das Makro-Mikro-Makro-Modell von Coleman in der Sprache des H-O-Schemas betrachtet werden soll, nicht jedoch eine allumfassende Betrachtung sämtlicher wissenschaftlicher Erklärungen.

$C_1, C_2, \dots, C_k$	Randbedingungen	}	Explanans
$L_1, L_2, \dots, L_r$	Allgemeine Gesetze		
$\rightarrow E$		}	Explanandum
	Beschreibung des zu erklärenden empirischen Phänomens		

**Tabelle 6.1:** Struktur einer wissenschaftlichen Erklärung (Hempel und Oppenheim (1948: 138), Übersetzung R. F.), Der Pfeil „ $\rightarrow$ “ bezeichnet die logische Implikation

Für die Anwendung des H-O-Schemas muss die *Adäquatheitsbedingung* (ebd.: 137f.) gelten, die aufgeteilt ist in *logische Adäquatheit* und *empirische Adäquatheit*. Die logische Adäquatheit setzt sich aus drei Teilbedingungen zusammen:

- $R_1$ : Das Explanandum muss logische Konsequenz des Explanans sein.
- $R_2$ : Das Explanans darf ausschließlich allgemeine Gesetze enthalten.

- $R_3$ : Es muss prinzipiell möglich sein das Explanans empirisch oder durch Beobachtung zu überprüfen.

Die empirische Adäquatheit  $R_4$  fordert schließlich, dass das Explanans wahr sein muss bzw. nach allen bereitstehenden Mitteln überprüft worden sein muss.

### Rekonstruktion der rationalen Handlungserklärung

Nachdem im vorangegangenen Abschnitt das Hempel-Oppenheim-Schema in seinen Grundzügen erläutert wurde, wird im folgenden Abschnitt in Anlehnung an Kroneberg (2009: 232ff.) das Makro-Mikro-Makro-Modell von Coleman mit den Begrifflichkeiten des H-O-Schemas rekonstruiert.

Im Gegensatz zu Esser (1999: 206), der nur den Mikro-Mikro-Übergang im H-O-Schema rekonstruiert, verbindet Kroneberg mit seinen Überlegungen die Rekonstruktion *aller* Übergänge im H-O-Schema.

In Anlehnung an Esser (1999: 206) modelliert Kroneberg den *Mikro-Mikro-Übergang* wie folgt: Den Kern des Übergangs bildet eine Handlungstheorie, die die Aufgabe hat, individuelle Effekte des Handelns, IE, unter den Randbedingungen RB zu erklären. In der Sprache der Prädikatenlogik handelt es sich dabei um eine Implikation die mit dem Symbol „ $\rightarrow$ “ gekennzeichnet wird. Aus der sprachlichen Formulierung „*Aus den Randbedingungen folgen individuelle Effekte*“ lässt sich wie folgt in die Sprache der formalen Logik übersetzen:

$$RB(x) \rightarrow IE(x)$$

Kroneberg spricht bei diesem Zusammenhang von einer „allgemeinen gesetzesartigen Hypothese“ (ebd.: 233). Der Zusammenhang von individuellen Effekten und Randbedingungen genügt allerdings noch nicht, um den *speziellen* individuellen Effekt und *spezielle* Randbedingungen zu erklären. Nehmen wir als Beispiel folgende allgemeine Aussage: „Wenn ein Autofahrer den Motor ausschaltet, hält das Auto an.“ Der allgemeine gesetzesartige Zusammenhang würde folglich lauten:

$$\text{Motor\_aus}(x) \rightarrow \text{Auto\_stoppt}(x)$$

Übertragen auf den speziellen Fall eines Autos von Person  $i$  muss gefolgert werden können, dass im Falle des Ausschaltens dieses Autos zusammen mit dem allgemeinen Gesetz ( $\text{Motor\_aus} \rightarrow \text{Auto\_stoppt}$ ) tatsächlich auch das Auto von Person  $i$  stoppt. In der Sprache der Prädikatenlogik ausgedrückt:

$$\underbrace{[(\text{Motor\_aus}(x) \rightarrow \text{Auto\_stoppt}(x)) \wedge \text{Motor\_aus}(i)]}_{\text{allgemeines Gesetz}} \rightarrow \underbrace{\text{Auto\_stoppt}(i)}_{\text{individueller Effekt } i}$$

Abstrahiert von dem konkreten Beispiel und unter Verwendung der Kurzschreibweise von Kroneberg (IE für  $IE(x)$ ,  $IE_i$  für  $IE(i)$  usw.) lässt sich die Schlussweise für den Mikro-Mikro-Übergang wie folgt (vgl. ebd.) formulieren:

$$((RB \rightarrow IE) \wedge RB_i) \rightarrow IE_i$$

Wie eingangs erwähnt, modelliert Kroneberg nicht nur den Mikro-Mikro-Übergang mit Hilfe des H-O-Schemas, sondern auch die anderen Übergänge:

Das Resultat des *Makro-Mikro*-Übergangs bildet die Ausgangssituation für den Mikro-Mikro-Übergang. Folgt man Kroneberg (2009: 234), werden Kontextvariablen (KV), die die Bedingungen einer Situation beschreiben, überführt in Randbedingungen (RB), die ihrerseits den Ausgangspunkt für den Mikro-Mikro-Übergang bilden. Den Zusammenhang zwischen Kontextvariablen und Randbedingungen beschreiben die sogenannten *Brückenhypothesen*, die Situationsbedingungen in Randbedingungen der Handlungstheorie überführen. Der allgemeine gesetzesartige Zusammenhang, der durch die Brückenhypothesen beschrieben wird, umfasst also die Implikation:

$$KV \rightarrow RB$$

Ähnlich wie bei dem zuvor beschriebenen Mikro-Mikro-Übergang führt der allgemeine Zusammenhang zwischen Kontextvariablen und Randbedingungen zusammen mit den speziellen Kontextvariablen  $KV_i$  für die betrachtete Situation  $i$  zu den speziellen Randbedingungen  $RB_i$ . Zusammengenommen ergibt sich daraus folgende logische Schlussweise für den Makro-Mikro-Übergang:

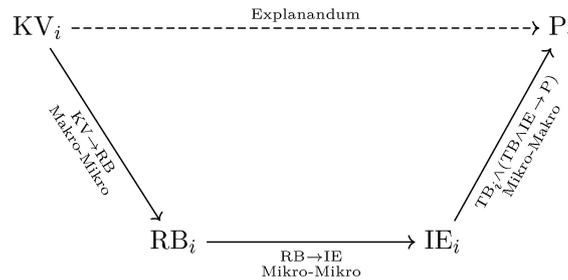
$$((KV \rightarrow RB) \wedge KV_i) \rightarrow RB_i$$

Der *Mikro-Makro*-Übergang überführt schließlich die individuellen Effekte IE bzw. die Auswirkungen des Handelns auf der Mikroebene zu einem zu erklärenden Phänomenen P. Folgt man Kroneberg (2009: 233f.), benötigt man für eine Erklärung dieses Übergangs im Sinne des H-O-Schemas Transformationsregeln und Transformationsbedingungen (TB). Die Transformationsregel gibt an, wie individuelle Effekte zusammen mit den Transformationsbedingungen zu dem zu erklärenden Phänomenen führen. In der Sprache der Logik lässt sich dieser Zusammenhang kompakt mit der Implikation

$$TB \wedge IE \rightarrow P$$

schreiben. Soll nun ausgehend von einer Situation  $i$  das Phänomen  $P_i$  über die individuellen Effekte  $IE_i$  erklärt werden, lässt sich die Transformationsregel mittels UND-Verknüpfung um die situationspezifischen Variablen ergänzen:

$$((TB \wedge IE \rightarrow P) \wedge TB_i \wedge IE_i) \rightarrow P_i$$



**Abbildung 6.1:** Das Coleman'sche Erklärungsmodell im H-O-Schema (eigene Darstellung in Anlehnung an die Überlegungen von Kroneberg 2009: 232ff.)

Fasst man die Überlegungen zu den drei Übergängen zusammen, so gelangt man zu einer Darstellung wie in Abbildung 6.1 gezeigt. Die Pfeile in der Zeichnung sind als Implikation, also als logische Schlussfolgerung zu verstehen. Da die logische Implikation transitiv<sup>3</sup> ist, lässt sich das zu erklärende Phänomen  $P_i$  aus einer Kette von Schlussfolgerungen erklären. Aus den Kontextvariablen  $KV_i$  folgt über eine soziologische Tiefenerklärung, die einen „Umweg“ über die Mikroebene der Akteure nimmt, logisch das Explanandum  $P_i$ .

Mit der *vollständigen* Rekonstruktion der Erklärung im H-O-Schema geht Kroneberg mit seinen Überlegungen über Coleman hinaus, der nur den Mikro-Mikro-Übergang als Erklärung im Sinne des H-O-Schemas rekonstruiert, während Kroneberg jeden der Übergänge in der Weise konzeptualisiert.

Aus der logischen Rekonstruktion von Kroneberg ergeben sich einige wichtige Konsequenzen für Makro-Mikro-Makro-Erklärungen und für den heuristischen Wert des Modells insgesamt. Trotz aller logischen Konsistenz kann Kroneberg (2009: 235) die strengen Adäquatheitsbedingungen des H-O-Schemas, zumindest für den Makro-Mikro- und für den Mikro-Makro-Übergang nicht vollständig erfüllen:

*Makro-Mikro / Logik der Situation:* Der logische Schluss  $KV \rightarrow RB$  verlangt, dass aus den Kontextvariablen, die die Situation beschreiben, die Randbedingungen der Handlungstheorie folgen. Kroneberg schreibt (ebd.: 236), dass dafür Brückenhypothesen notwendig seien, die „unter Umständen komplexe Aussagen- oder Gleichungssysteme“ umfassen können. Allgemein werden die Brückenhypothesen BH als Funktion modelliert, die Kontextvariablen in Randbedingungen der Handlungstheorie übersetzen:  $BT: RB = f(KV)$ . Der Versuch, die Adäquatheitsbedingungen aus dem H-O-Schema direkt auf diesen Zusammenhang anzuwenden, zeigt schnell die Limitierungen auf: Das Explanans, in dem Fall also die Brückenhypothesen, ist keinesfalls wie in Bedingung  $R_2$  im H-O-Schema gefordert ein „ausschließlich allgemeines Gesetz“. Die Brückenhypothesen verletzen  $R_2$  in zweierlei Hinsicht: Erstens können Brückenhypothesen immer nur im Lichte eines speziellen Anwendungsfalls

<sup>3</sup>Wenn  $A \rightarrow B$  und  $B \rightarrow C$  gilt, so gilt auch  $A \rightarrow C$ .

angegeben werden. Zweitens schränkt die Handlungstheorie auf der Mikroebene den Allgemeinheitsgrad der Brückenhypothesen radikal ein, da eben nur solche Brückenhypothesen möglich sind, die zu der gewählten Handlungstheorie passen.

*Mikro-Makro / Logik der Aggregation:* Ganz ähnlich lässt sich für den Übergang von der Individualebene zur Makroebene argumentieren. Zumindest die Adäquatheitsbedingung  $R_2$  ist auch hier verletzt. Die Aggregation von Individualhandlungen zu kollektiven Effekten kann immer nur im Lichte eines spezifischen zu erklärenden sozialen Phänomen betrachtet werden. Würde man verlangen, für die Transformationsregeln ausschließlich allgemeine Gesetze zuzulassen, würde dies u. a. die Existenz einer Modellierungssprache für alle möglichen sozialen Phänomene voraussetzen. Dies erscheint vor dem Hintergrund der enormen Diversität sozialer Phänomene ein zweckloses Unterfangen.

Kroneberg mildert in seinen Betrachtungen die Adäquatheitsbedingungen stark ab (ebd.: 235): „Das Explanandum muss (prinzipiell) logisch aus dem Explanans ableitbar sein. Das Explanans muss mindestens eine gesetzesartige Hypothese enthalten, empirischen Informationsgehalt besitzen und wahr (oder zumindest empirisch gut bestätigt) sein“. Weiter heißt es, dass das H-O-Schema für das Makro-Mikro-Makro-Modell lediglich als „regulative Idee“ (ebd.: 237) dient.

Kroneberg plädiert dafür, neben der erklärenden Bedeutung der Handlungstheorie im Makro-Mikro-Makro-Modell insbesondere ihre heuristische Bedeutung (vgl. auch Abschnitt 7.2) stärker in den Vordergrund zu stellen (ebd.: 238; 240). Durch die Modellierungsmöglichkeiten der Handlungstheorie auf der Mikroebene ergeben sich Heuristiken zur Konstruktion von Brückenhypothesen, die Kontextvariablen in die Variablen der Handlungstheorie überführen. Das primäre soziologische Interesse verortet Kroneberg (2009: 237) ohnehin bei der Konstruktion angemessener Brückenhypothesen und Transformationsregeln.

Die erkenntnistheoretischen Überlegungen dienen Kroneberg der Hinführung auf seine zwei Kernthesen: Erstens, das Makro-Mikro-Makro-Modell nicht als ontologisches Postulat, sondern als *Methodologie* aufzufassen. Zweitens dienen die Überlegungen zum heuristischen Wert der Handlungstheorie auf der Mikroebene dazu, das Kriterium der *Modulierbarkeit* einzuführen.

Der methodologische Nutzen des Modells im Allgemeinen und der Handlungstheorie im Speziellen wurde bereits erläutert. Abschließend soll noch auf das Kriterium der Modulierbarkeit eingegangen werden. Das Kriterium der Modulierbarkeit zielt auf die notwendige Komplexität der Handlungstheorie. Gemäß den Überlegungen von Kroneberg (2009: 240) umfasst das Kriterium der Modulierbarkeit folgende Anforderungen an eine Handlungstheorie:

- Starke Vereinfachungen müssen ebenso erlaubt sein wie detaillierte, auf empirische Adäquatheit abzielende Handlungserklärungen,
- es müssen Möglichkeiten vorhanden sein die Distanz zwischen Adäquatheit und Abstraktion systematisch zu überbrücken.

Der Aspekt der Modulierbarkeit erscheint auch für die vorliegende Arbeit von besonderer Bedeutung. In den Empirie-Abschnitten ab Kapitel 9 werden u. a. interaktive Simulationsexperimente vorgestellt, die eine gewisse Abstraktion bei der Modellierung soziotechnischer Zusammenhänge erfordern. Die Abstraktionen basieren auf den Kroneberg'schen Überlegungen zur Modulierbarkeit. Dabei wird der Detailgrad der Modellierungen gerade so gewählt, wie es für den betrachteten Anwendungsfall erforderlich erscheint. Somit stellt das Kriterium der Modulierbarkeit eine soziologische Präzisierung der modellierungspragmatischen Binsenweisheit „So kompliziert wie nötig, so einfach wie möglich“ dar.

Letztlich muss man sich bei jeder Modellierung eines sozialen oder soziotechnischen Systems die Frage stellen, ob die Abstraktionsannahmen für den zu untersuchenden Sachverhalt angemessen sind. Wird zu feingranular modelliert, kann dies zur Folge haben, dass bestimmte Effekte nicht mehr auf bestimmte Ausgangsbedingungen oder Ausgangsverhalten zurückzuführen sind. Modelliert man hingegen zu grob besteht die Möglichkeit, dass Effekte, die empirisch beobachtbar sind in der Simulation wegen der ungenauen Modellierung gar nicht erst eintreten.

## 6.2 MSE und SEU-Theorie

Die Grundlage für das in Abschnitt 8.1 entwickelte Modell ist das Modell soziologischer Erklärung (MSE) von Hartmut Esser (1993). Da das MSE inzwischen zur soziologischen Grundlagenausbildung gehört (vgl. exemplarisch Kron 2009), beschränken sich die Ausführungen in diesem Abschnitt auf die wesentlichen Elemente. Das MSE ist zu verstehen als eine Weiterentwicklung des zuvor vorgestellten Coleman'schen Makro-Mikro-Makro-Modells. Im Unterschied zu Coleman hat Esser die Handlungstheorie auf der Mikroebene näher spezifiziert, die Coleman in Form abstrakter Eigenschaften umrissen hat.

Essers handlungstheoretischer Kern gründet sich auf die im Kontext von Rational Choice-Ansätzen breit rezipierte Idee der Maximierung des Erwartungsnutzens (s. bereits Esser 1991; Coleman 1991), vgl. zusammenfassend Diekmann und Voss (2004). Im Unterschied zu klassischen RC-Ansätzen betont Esser die *subjektive* Komponente des Erwartungsnutzenbegriffs – nicht Erwartungsnutzen *EU*, sondern subjektiver Erwartungsnutzen *SEU* genannt. Der klassische Erwartungsnutzenbegriff setzt eine objektive Rationalität voraus, die Akteuren feste Ziele, Zielbewertungen und Erwartungen unterstellt. Die Anreizstrukturen einer Situation „zwingen“ praktisch jeden

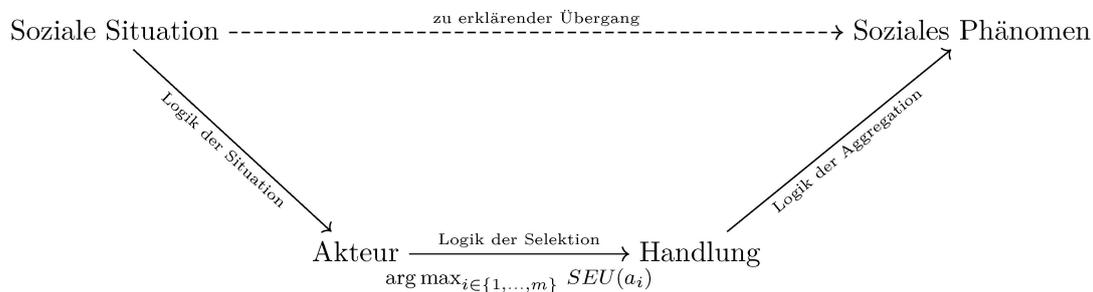
Akteur eine bestimmte Handlungsalternative auszuwählen, nämlich die, die für objektiv rational gehalten wird<sup>4</sup>.

Die fehlende Möglichkeit, unterschiedliche Zielbewertungen zu berücksichtigen, bedeutet, so lässt sich folgern, dass von einem einheitlichen Akteur ausgegangen wird und dass es folglich keine Möglichkeit gibt, unterschiedliche Akteurstypen zu modellieren.

Als implizite Folgerung eines uniformen Erwartungsnutzens, ergibt sich eine, soziologisch zumindest hinterfragbare, einheitliche Wahrnehmung von Anreizstrukturen und Situationsbedingungen. Für sehr einfache Situationen, bei denen zu erwartende Verluste bei Auswahl bestimmter Handlungsalternativen zweifelsfrei feststehen, mag diese Annahme noch berechtigt sein. In komplexeren Entscheidungssituationen hängt die Einschätzung von Gewinnen und Verlusten hingegen vom jeweiligen Akteur und seiner subjektiven Deutung situativer Opportunitäten ab. Eine Handlungstheorie, die ex ante eine solche Modellierungsmöglichkeit ausschließt, ist im Sinne des Kroneberg'schen Modulierbarkeitskriteriums unterkomplex.

### Modellbestandteile und Erklärungsschritte

Die Grundstruktur einer soziologischen Erklärung nach Esser ist in Abbildung 6.2 dargestellt.



**Abbildung 6.2:** Das Modell soziologischer Erklärung (MSE)  
In Anlehnung an: Esser (1993: 98)

Das Modell orientiert sich dabei am Coleman'schen Erklärungsmodell. Im erkenntnistheoretischen Sinne erklärungsbedürftig ist der Übergang von einer sozialen Ausgangssituation hin zu einem sozialen Phänomen. Die Erklärung erfolgt nicht unmittelbar kausal auf der Makroebene, sondern wird über den Umweg der Mikroebene *tiefenerklärt*. Dabei durchläuft die Erklärung den Coleman'schen Dreischritt, den Esser als Abfolge von *Logik der Situation*, *Logik der Selektion* und *Logik der Aggregation* bezeichnet.

<sup>4</sup>Eine solch verengte Betrachtungsweise muss, je nach zu erklärendem Phänomen, keinesfalls nachteilig oder falsch sein. Allerdings bedarf die Annahme der objektiven bzw. uniformen Rationalität über alle Akteure immer einer empirischen Begründung.

Zur Modellierung der Logik der Situation sind, so Esser (1999: 15), *Brückenhypothesen* erforderlich, die die äußeren Situationsbedingungen in Variablen der Handlungstheorie auf der Mikroebene überführen. Das handlungstheoretische Modell, das auf der Mikroebene verwendet wird, strukturiert die Art und Weise der Modellierung der Situation vor. Die Brückenhypothesen begründen, wie eine bestimmte Ausgangssituation in die spezifischen Modellbestandteile (Handlungsalternativen, Zielen, Zielbewertungen und Erwartungen) der Handlungstheorie überführt werden kann. Die Brückenhypothesen und damit die gesamte *Logik* der Situation, so meine Lesart, kann im Forschungsprozess als *Modellierungsplausibilisierung* aufgefasst werden, in der begründet wird, warum eine soziale Ausgangssituation auf diese (und nicht auf eine andere) Weise modelliert wird.

Im Esser'schen Handlungsmodell ist der zweite Erklärungsschritt, die Logik der Selektion, in Form einer zielgerichteten Handlungstheorie modelliert, die im anschließenden Abschnitt gesondert vorgestellt wird.

Der dritte Erklärungsschritt, der von der Mikroebene zurück zur Makroebene führt, die Logik der Aggregation, erfordert die Angabe bestimmter Transformationsregeln, die angeben, „unter welchen Bedingungen bestimmte individuelle Effekte bestimmte kollektive Sachverhalte erzeugen“ (Esser 1999: 16). In einigen Fällen, z. B. bei Wahlentscheidungen, liegen Transformationsregeln in explizierter Form bereits vor und müssen nur noch auf die Einzelhandlungen angewendet werden. In den allermeisten Fällen, vor allem aber bei komplexen Handlungs- und Wirkzusammenhängen heterogener Akteure, ist eine explizite Angabe von Transformationsregeln ein schwieriges Unterfangen. Esser sieht das Problem der Aggregation als „kompliziertesten der drei Schritte“ (Esser 1996: 97) in seinem Erklärungsmodell an. Esser (1993: 136f.) empfiehlt für komplexere Aggregationen die Methode der Computersimulation, wendet diese in seinem soziologischen Grundlagenwerk allerdings nicht an. Er greift zur Aggregation von Einzelhandlungen stattdessen auf spieltheoretische Modelle zurück, denen er sich in zahlreichen Abschnitten seiner Allgemeinen und Speziellen Grundlagen widmet (vgl. exemplarisch: Esser 1996: 123ff., 2000c: 55ff.; 117ff.; 165ff.). Den spieltheoretischen Aggregationen, zumindest in der Weise, wie sie Esser verwendet, ist gemeinsam, dass es sich um sehr abstrahierte und artifizielle Situationen handelt, in denen sich, meist zwei, Akteure in einer strategischen Interdependenzsituation befinden. Für das grobe Verständnis bestimmter Akteurskonstellationen mag die Spieltheorie einen Ansatzpunkt bieten, allerdings erscheint sie als Transformationsheuristik für die zugrundeliegende Handlungstheorie unterkomplex. In der vorliegenden Arbeit wird daher auf die Methode der Computersimulation (genauer der agentenbasierten Modellierung und Simulation) zurückgegriffen (vgl. ausführlich Abschnitt 7.3), da diese relativ problemlos die Einzelhandlungen einer Vielzahl heterogener Akteure zusammenführen kann.

Der Abschnitt zu den unterschiedlichen Erklärungsschritten soll mit einem Zitat von Kroneberg (2009: 235) geschlossen werden, das unterstreicht, dass das Hauptinteresse bei der empirischen Anwendung des Esser'schen Modells soziologischer Erklärung nicht bei der Logik der Selektion zu verorten ist, sondern bei den anderen beiden Erklärungsschritten: „Die Brückenhypothesen und Transformationsregeln – also die Hypothesen über *soziale* Einflüsse und Prozesse – sind vielmehr soziologisch von primärem Interesse.“

Dieser Gedanke lässt sich dahingehend ergänzen, dass zwar mit dem Makro-Mikro-Makro-Modell im Sinne von Hempel-Oppenheim soziale Phänomene erklärt werden sollen, dass aber bereits bei der Modellierung (Logik der Situation) und bei Überlegungen zur Transformation von Einzelhandlungen (Logik der Aggregation) echte soziologische Erkenntnisse generiert werden, die nicht unmittelbar mit dem zu erklärenden Phänomen in Verbindung stehen müssen. Nicht nur die Gesamterklärung kann als soziologische Erkenntnis gewertet werden, sondern ebenso der isoliert betrachtete (Forschungs-)Prozess der wohlbegründeten Situationsmodellierung und Handlungsaggregation.

### Der nomonologische Kern: SEU-Theorie

Das Ergebnis der Logik der Situation ist, wie bereits erläutert, die Überführung von Situationsdeutungen in Variablen der Handlungstheorie. Esser geht davon aus, dass sich Handlungen auf der Mikroebene durch drei Kernelemente modellieren lassen:

- Aktionen / Handlungen  $a_i$ ,
- Ziele  $o_j$  mit zugehörigen Zielintensitäten  $u(o_j)$  bzw. abkürzend  $u_j$ ,
- Erwartungen  $p_{i,j}$ , die jeder Handlungsalternative  $a_i$  eine subjektiv geschätzte Wahrscheinlichkeit zuweisen, inwiefern diese für die Erfüllung des Ziels  $o_j$  zuträglich ist.

Mathematisch lassen sich die Modellbestandteile für  $n$  Handlungsalternativen und  $m$  Ziele als Fünf-Tupel präzisieren:

$$(A, O, \mathbf{U}^{1 \times m}, \mathbf{P}^{m \times n}, s)$$

Die Modellbestandteile haben folgende Bedeutung:

- $A$  – eine Menge von Handlungsalternativen  $A = \{a_1, a_2, \dots, a_m\}$
- $O$  – eine Menge von Zielen (Outcomes)  $O = \{o_1, o_2, \dots, o_n\}$
- $\mathbf{U}$  – eine Menge von Zielbewertungen, die jedem Ziel einen Wert aus den reellen Zahlen zuordnet,  $U = \{u(o_1), u(o_2), \dots, u(o_n)\}$  mit  $u : O \rightarrow \mathbb{R}$

- $\mathbf{P}$  – eine Menge von Erwartungen, die jeder Kombination aus Handlungsalternative und Ziel eine Zielerfüllungswahrscheinlichkeit  $p_{i,j}$  zuordnet, formal  $p_{i,j}$  mit  $i \in \{1, \dots, m\}$ ,  $j \in \{1, \dots, n\}$ ,  $p_{i,j} \in [0, 1]$
- $s$  – Selektionsfunktion, die für  $SEU(a_i) := \sum_{j \in \{1, \dots, n\}} p_{i,j} \cdot u_j$  die Handlungsalternative mit dem höchsten SEU-Wert zurückgibt, also  $a_i$  mit  $\arg \max_{i \in \{1, \dots, m\}} SEU(a_i)$  berechnet.

Die präzise mathematische Formulierung<sup>5</sup> der einzelnen Modellbestandteile, die Esser in dieser Form selbst nicht vornimmt (oder sie zumindest variiert), hat den Vorteil genau angeben zu können, welche Elemente der SEU-Theorie mit subjektiven Werten belegt werden und welche Elemente für alle Akteure in der betrachteten Situation gleich bleiben.

Für die Untersuchung eines sozialen Phänomens bleibt die Menge der Ziele  $O$  und die Menge der Handlungsalternativen  $A$  für alle Akteure in der Regel<sup>6</sup> identisch. Freilich könnte man diese auch *je Akteur* variieren, allerdings würde man in diesem Fall das Modell der vollständigen Beliebigkeit preisgeben. Sollte ein bestimmtes Ziel oder eine bestimmte Handlungsalternative für einen Akteur bzw. eine Akteursgruppe nicht zur Verfügung stehen, kann man dies, wie sich leicht zeigen lässt, mit entsprechenden Belegungen für die subjektiven Modellbestandteile ausschalten. Subjektiv sind demnach je Akteur bzw. Akteursgruppe die Zielintensitäten  $\mathbf{U}$  sowie die Erwartungen  $\mathbf{P}$  zu belegen. Die Selektionsfunktion ist für jedes beliebige SEU-Modell identisch und berechnet stets die Handlungsalternative mit dem höchsten SEU-Wert.

Steht eine bestimmte Handlungsalternative  $a_x$  für einen Akteur nicht zur Verfügung, so lässt sich diese innerhalb des Modells leicht ausschließen, indem die Zielverfolgungswahrscheinlichkeit für diese Handlungsalternative für alle Ziele  $j$  auf 0 gesetzt wird, also  $p_{x,j} = 0$ . Sollte ein Ziel  $o_y$  nicht zur Verfügung stehen hat es weder negative noch positive Konsequenzen und lässt sich durch Nullsetzung der Zielbewertung neutralisieren,  $u_y = 0$ .

Zusammenfassend ist also die Modellstruktur objektiv vorgegeben und wird durch den Forscher vorstrukturiert. Die konkreten Zielintensitäten und geschätzten Eintrittswahrscheinlichkeiten sind hingegen subjektiv in Bezug auf die betrachteten Akteure bzw. Akteursgruppen zu betrachten.

### 6.3 Weiterentwicklungen

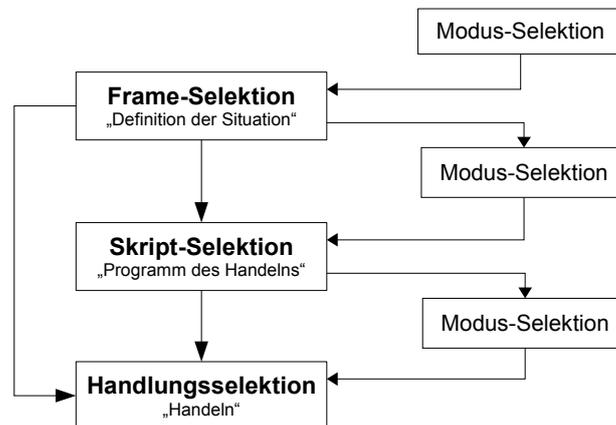
Das bisher vorgestellte Modell soziologischer Erklärung (MSE) ist ein relativ schlankes Modell, das den Modellierungsfokus auf die Handlungsselektion legt. Obwohl Kroneberg (2009) in seinem Aufsatz „Methodologie statt Ontologie“ überzeugend dargelegt

<sup>5</sup>Die hier vorgenommene Formalisierung unterscheidet sich *inhaltlich* nicht von den Formalisierungen Essers.

<sup>6</sup>Eine Systematisierung unterschiedlicher SEU-Modell differenzen findet sich in Abschnitt 8.2.1.

hat, dass die eigentliche soziologische Arbeit vor allem bei der Angabe und Begründung von Brückenhypothesen und Transformationsregeln zu verorten ist, bleibt das MSE ein relativ starres Modellierungskorsett, das alle Komplexität der Überführung von Situationsbedingungen in die Variablen der Handlungstheorie den Brückenhypothesen überlässt.

Das in diesem Abschnitt kurz angerissene *Modell der Frame-Selektion* (MFS), das in seiner letzten Version von Clemens Kroneberg (2005; 2007; 2011) vorgelegt wurde, verteilt in gewisser Weise die Komplexität auf mehrere Modellierungsschritte. Zusätzlich erhält der Aspekt der variablen Rationalität Einzug in die Modellierungs-Toolbox des MFS. Die variable Rationalität greift Erkenntnisse aus der Sozialpsychologie auf, nach denen menschliche Akteure nicht in jeder Situation (subjektiv) rational handeln, sondern z. T. auf erlernte Muster und Situationsdeutungen zurückgreifen.



**Abbildung 6.3:** Das MFS. Erweiterungen gegenüber dem MSE: Frame-, Skript- und Handlungsselektion mit jeweils vorgelagerter Modus-Selektion  
Quelle: Kroneberg (2005: 348)

Die ursprünglich von Esser (2001: 259ff.) vorgestellte, damals noch unter dem Label „Frame-Selektionstheorie“ (FST) verbreitete Theorie ist auf Grund von Inkonsistenzen und formalen Ungenauigkeiten stark in die Kritik geraten (vgl. z. B. Kron 2004; Etzrodt 2000; Rohwer 2003). Clemens Kroneberg (2005) hat die Frame-Selektionstheorie zu dem Modell der Frame-Selektion (MFS) weiterentwickelt und versucht damit, den aus seiner Sicht wesentlichen Kritikpunkten an der FST zu begegnen. Die kursorische Darstellung des MFS orientiert sich an der zuletzt vorgelegten Version von Kroneberg, der sich inzwischen auch Esser angeschlossen hat (Kroneberg 2007: 217). Den jüngsten Beitrag zur kritischen Diskussion um die Frame-Selektionstheorie lieferte Ingo Schulz-Schaeffer (2008), der die Relevanz unterschiedlicher Modi zur Situationsdefinition in Frage stellt.

Die wesentliche Erweiterung des MFS im Vergleich zum MSE besteht darin, dass die Abfolge von Situationsdefinition und Handlungsselektion weiter differenziert wird (vgl. Abbildung 6.3).

Innerhalb des MFS erfolgt die Handlungswahl in drei Selektionsschritten: Frame-Selektion, Skript-Selektion und Handlungsselektion. „Ein *Frame* ist eine mentale Repräsentation einer typischen Situation. Es enthält in typisierter Form die spezielle inhaltliche Definition der Situation, insbesondere aber das Oberziel, um das es in der betreffenden Situation geht. [...] Ein *Skript* beschreibt in ähnlicher Weise die typischen, am Code des Frames orientierten, inhaltlichen Abläufe und Verzweigungen für ganze Bündel und Sequenzen von Handlungen“ (vgl. Esser 2001: 263).

Diesen drei Selektionen vorgelagert ist jeweils eine Modusselektion (vgl. Kroneberg 2005: 350), bei der ein angemessener Modus der Informationsverarbeitung ausgewählt wird. Es wird unterschieden zwischen zwei Modi der Informationsverarbeitung (s. Esser 2001: 271ff.): dem *automatisch-spontanen* (as) Modus und dem *reflexiv-kalkulierenden* (rc) Modus.

Entscheidend für die Wahl des vorgelagerten Modus der Frame-Selektion ist der Match, „der Grad der unmittelbar erfahrenen Passung eines Frames zu einer aktuell vorliegenden Situation“ (Kroneberg 2005: 350f.). Je besser ein Frame auf eine gegebene Situation passt, um so eher wird der automatisch-spontane Modus aktiviert und einer reflexiv-kalkulierenden Suche nach einem Alternativframe vorgezogen.

Ein Skript  $S_j$  wird umso eher automatisch-spontan selektiert, je höher die mentale Verfügbarkeit des Skripts ist, je stärker Skript und Frame miteinander verknüpft sind und je genauer das bereits selektierte Frame auf die Situation passt. (vgl. Kroneberg 2005: 351). Die Selektion einer Handlung im as-Modus erfolgt umso eher, je genauer die Handlung durch das Skript vorstrukturiert ist.

Für den weiteren Verlauf der Arbeit spielt das MFS keine Rolle mehr. Es wurde hier der Vollständigkeit halber genannt, um zu zeigen, in welche Richtung sich das ursprüngliche Modell soziologischer Erklärung weiterentwickelt hat. Stattdessen wird als modellierungstheoretische Grundlage im Wesentlichen auf das MSE zurückgegriffen. Dieses erscheint für eine experimentell-computersimulative Umsetzung, die vornehmlich ihren Fokus auf die Handlungswahl und die Modellierung ebendieser setzt, angemessen. Die Frage, ob unter bestimmten Voraussetzungen automatisch-spontan oder reflexiv-kalkulierend gehandelt wird, spielt im Kontext der vorliegenden Arbeit keine Rolle und würde das Experimentalsetting unnötig kompliziert machen. Neuere Ansätze deuten jedoch darauf hin, dass auch vereinfachte Versionen des MFS mittels Computersimulationen umsetzbar sind (vgl. Adelt u. a. 2014).

## 6.4 Implikationen

Ausgehend von der „Coleman’schen Badewanne“ wurden einige erkenntnistheoretische Überlegungen zum Makro-Mikro-Makro-Modell des methodologischen Individualismus angestellt. Dabei hat sich gezeigt, dass neben dem zu erklärenden Phänomen der Erkenntnisgewinn schon mit Start der Modellierung einsetzt. Um situative Bedingungen in die Variablen einer Handlungstheorie zu übersetzen, sind abstrahierende Annahmen erforderlich, die zur Präzision zwingen. Dieser Umstand wurde als *heuristischer Mehrwert* der Modellierung zusammengefasst.

Für die vorliegende Arbeit erscheint das Modell soziologischer Erklärung mit seinem handlungstheoretischen Kern, der SEU-Theorie, in mehrerlei Hinsicht angemessen: Durch die systematische Verknüpfung von Makro-Mikro-Makro-Zusammenhängen existiert ein einheitlicher Modellierungsrahmen, mit dem die Auswirkungen von Mikro-Handlungen, die für die Frage nach dem Vertrauen in autonome Technik von Bedeutung sind, berücksichtigt werden können. Die SEU-Theorie ist zu einem Grad formalisierbar, der es ermöglicht, diese im Rahmen von Computersimulationen (vgl. nächstes Kapitel 7) insbesondere im Paradigma der agentenbasierten Modellierung und Simulation (vgl. Abschnitt 7.3) zu verwenden. In Vorgriff auf (FLA 13), in der sich noch zeigen wird, dass sich das Zusammenspiel von Mensch und autonomer Technik im Lichte jeweils gemeinsamer Ziele betrachten lässt, ist die SEU-Theorie mit ihren teleologischen Modellierungselementen anschlussfähig.

Obgleich noch keine Integration des Mithandelns von autonomer Technik in das Modell soziologischer Erklärung erfolgt ist (s. dafür Abschnitt 8), so kann schon jetzt formuliert werden, dass sich das MSE als modellierungstheoretische Leitidee für weitergehende Überlegungen eignet.

### **FLA 11 (MSE als modellierungstheoretische Leitidee)**

Das MSE ist als modellierungstheoretische Leitidee geeignet, um Vertrauen in autonome Technik in einer handlungstheoretischen Perspektive zu untersuchen.



## 7 Computersimulation und Experimente

Ein Blick in die naturwissenschaftliche Forschungspraxis, etwa der Physik oder Chemie, verdeutlicht, dass Experimente und Computersimulationen dort üblich sind und einen integralen Bestandteil des Forschungsprozesses darstellen. Ebenso findet die experimentelle Methode eine breite Anwendung in Form von psychologischen und medizinischen Experimenten.

Eine ebenso weite Verbreitung von Experimenten und Computersimulationen lässt sich für die Sozialwissenschaften nicht vorfinden. Die sozialwissenschaftliche Forschung kann zwar auf eine gewisse Tradition hinsichtlich der experimentellen Methode und der Verwendung von Computersimulation zurückblicken (vgl. Abschnitte 7.1 und 7.2), allerdings hat, zumindest in der deutschen Soziologie, die Computersimulation, speziell die agentenbasierte (vgl. Abschnitt 7.3), nach wie vor keinen festen Platz im Methodenkanon gefunden.

Die Sektion *Modellbildung und Simulation* der Deutschen Gesellschaft für Soziologie beschäftigt sich, so lässt sich zumindest an der Auflistung der letzten Tagungen ablesen, vornehmlich mit avancierten spieltheoretischen Modellen und komplexen statistischen Methoden der empirischen Sozialforschung<sup>1</sup>. Die Modellierungen orientieren sich insbesondere an Modellen sozialer Dilemma-Situationen aus der Mikro-Ökonomie. Eine genuin soziologische Perspektive auf Modellbildung und Simulation (mit Unterstützung von Informatikern) wie sie das DFG-Schwerpunktprogramm *Sozionik* (vgl. exemplarisch Malsch 1997; 1998) zum Ziel hatte, spielt im aktuellen *deutschen* Modellierungs- und Simulationsdiskurs keine wichtige Rolle mehr. Neuere Ansätze der sogenannten Analytischen Soziologie (vgl. z. B. Hedström 2008; Kron und Grund 2010; Manzo 2014) greifen hingegen neben anderen formalen Methoden auch explizit auf agentenbasierte Simulationsmodelle zurück (vgl. Hedström 2008: 112ff.). Das vorliegende Kapitel stellt einige ausgewählte Erkenntnisse und Vorgehensweisen aus der experimentellen und computersimulativen Forschungsrichtung der Sozialwissenschaften vor. Die Argumentation läuft schließlich darauf hinaus, dass insbesondere die Methode der agentenbasierten Modellierung und Simulation (Gilbert und Terna 2000; Bonabeau 2002) anschlussfähig an die soziologische Handlungstheorie ist, die im vorherigen Kapitel behandelt wurde. Im Sinne der Esser'schen Logiken (vgl. vorherigen Abschnitt 6.2) kann die Computersimulation als *mikrofundierte Logik der*

---

<sup>1</sup>Vgl. Tagungsliste der Sektion <http://www.socio.ethz.ch/modsim/tagungen> (Abruf: 31. Januar 2014). Einzige Ausnahme bildet die Veranstaltung „Evolution und Zusammenbruch sozialer Ordnung“, die auf dem DGS-Kongress im Oktober 2014 in Trier stattfinden soll.

*Aggregation* verstanden werden. Die Kombination von Computersimulation und Laborexperiment ermöglicht interaktive Computersimulationen, bei denen Probanden die Möglichkeit haben, mit den Agenten einer Simulation zu interagieren. Eine solche Art von interaktiver Computersimulation findet insbesondere beim Experiment SIMHYBS (vgl. Abschnitt 11) Verwendung.

## 7.1 Experimente in den Sozialwissenschaften

Karl-Dieter Opp (1970) hält in seinem Aufsatz „The experimental method in the social sciences“ ein Plädoyer für die verstärkte Verwendung der experimentellen Methode in den Sozialwissenschaften und sieht die Anwendung dieser als eine Art Lackmusktest für den Stand der Disziplin an: „If we define the strength of tests scientists try to apply to their propositions as an indicator of the state of a science, we may say that the more the social sciences progress the more extensively experiments will be applied.“

Opp (1970: 39f.) unterscheidet grob zwischen zwei Arten von Experimenten: kontrollierte und natürliche. Bei kontrollierten Experimenten werden die Variablen verändert, von denen ein Einfluss ausgeht, und die Effekte auf die beeinflussten Variablen gemessen. Die beeinflussenden Variablen heißen *unabhängige*, die beeinflussten Variablen sind *abhängige* Variablen. Davon zu unterscheiden sind die natürlichen Experimente, bei denen die unabhängigen Variablen unverändert bleiben und die Effekte aller anderen Variablen herausgerechnet werden.

Keppel (1973: 4) definiert in seinem experimentalwissenschaftlichen Grundlagenwerk „Design and analysis“, das inzwischen in der vierten Auflage erschienen ist (Keppel und Wickens 2004), den Begriff Experiment noch enger als Opp: Demnach kontrastiert die experimentelle Methode zwei Einflussbedingungen auf ein Untersuchungsobjekt, wobei die Einflussbedingungen sich nur hinsichtlich einer Eigenschaft unterscheiden. Das Ziel der experimentellen Methode sei die Aufdeckung von Kausalzusammenhängen (ebd.).

Opp (1970: 41ff.) befasst sich auch mit den Argumenten, die typischerweise von Gegnern der experimentellen Methode in Anschlag gebracht werden. Diese Argumente werden mit der jeweiligen Gegenposition von Opp sowie einem eigenen Kommentar dazu vorgestellt:

*Modell- und Experimentalannahmen sind artifiziell und deswegen nicht anwendbar für die Sozialwissenschaften* (Opp 1970: 41f.).

Eine experimentelle Situation, so Opp, muss in der Wahrnehmung der Probanden nicht von einer „natürlichen“ abweichen. Selbst, wenn sich das experimentelle Setting von realen Situationen unterscheidet, ist der Unterschied irrelevant, wenn die nicht berücksichtigten realen Merkmale keinen Einfluss auf die zu erklärenden Va-

riablen haben. Situationen können so gestaltet werden, dass sie nicht als artifiziell wahrgenommen werden. Weiterhin bezweifelt Opp, dass Probanden ihr Verhalten im Rahmen der Laborsituation ändern, wenn sie nicht wissen, welche Hypothesen mit der experimentellen Situation getestet werden sollen.

Es lässt sich ergänzen, dass die Auswirkungen, die eine künstliche Laborsituation mit sich bringen, vor dem Hintergrund zusätzlich beeinflussender Faktoren beurteilt werden muss. Würde beispielsweise eine unrealistische Handlungsoption geschaffen, die wesentliche Auswirkungen auf die Untersuchungsfrage hat? Weiterhin sollten nach Möglichkeit Anreize ausgeschaltet werden, die dazu führen, dass sich eine Versuchsperson anders als außerhalb des Labors verhält. Existiert beispielsweise eine persönliche Beziehung zwischen Versuchsleiter und dem Probanden, ist es, je nach Brisanz des Untersuchungsgegenstandes, möglich, dass sich der Proband aus Rücksicht auf die Beziehung anders verhält, als er es außerhalb des Labors tun würde.

Die Forderung von Opp, alle ergebnisbeeinflussenden Faktoren auch im Laborsetting zu berücksichtigen, bedingt allerdings, dass die beeinflussenden Faktoren vollständig bekannt sind. Würde man Opp strikt folgen, verböten sich von vornherein Experimentaldesigns mit explorativen Fragestellungen, bei denen Einflussgrößen erst noch gefunden werden sollen.

*Die Komplexität der sozialen Welt verhindert die Anwendung von Experimenten* (Opp 1970: 42f.).

Gerade wenn ein Verhalten erklärt werden soll, ohne dass exakte Informationen über die Randbedingungen existieren, so entgegnet Opp, ist die experimentelle Methode sehr gut geeignet, um die relevanten Einflüsse herauszufinden. Mit Verweis auf Lern- und Kognitionstheorien argumentiert er, dass diese keine immense Zahl von Variablen enthalten und die soziale Welt demnach nicht zu komplex sei. Die Frage, wie viele Variablen zur Erklärung bestimmter Phänomene notwendig sind, könne nur empirisch beantwortet werden, nicht a priori. Damit entkräftet Opp allerdings sein vorheriges Argument, demzufolge Experimentalsettings zulässig seien, sofern die beeinflussenden Variablen ausreichend berücksichtigt sind. Insofern ist Opp aber zuzustimmen, dass mittels Experimenten gezielt der Einfluss oder auch Nicht-Einfluss von Variablen gemessen werden kann.

Wären vor dem Hintergrund der Komplexität der sozialen Welt tatsächlich unendlich viele unabhängige Variablen für eine soziologische Erklärung erforderlich, dann wäre dies genauso ein Argument gegen jegliche Theoriekonstruktion. Ein ähnliches Argument findet sich bereits bei Renate Mayntz (1967: 14), die argumentiert, dass Theorie und Modell letztlich Synonyme seien. Sowohl für eine Theorie als auch für ein Modell werden Zusatzannahmen getroffen, die ihren jeweiligen Geltungsbereich einschränken. Viele Jahre später wird das Argument der Komplexität der sozialen Welt in besonderer Weise herausgefordert: Mittels agentenbasierter

Computersimulationen kann gezeigt werden, dass bereits einfache Regeln auf der Mikro-Ebene genügen, um emergente Effekte auf der Makro-Ebene zu erzeugen, die weitgehend deckungsgleich mit beobachteten sozial komplexen Strukturphänomenen sind (vgl. Epstein und Axtell (1996) / Abschnitt 7.2). Trotz der Deckungsgleichheit bedeutet dies nicht automatisch, dass die Modelle valide sind, wenn sie sich z. B. auf unzulässige Annahmen auf der Mikro-Ebene gründen. In neueren Arbeiten fordert etwa Hedström (2008: 188) „empirisch kalibrierte Modelle“, die verhindern sollen, dass sich agentenbasierte Simulationen auf unzulässige Mikro-Annahmen stützen.

*Die soziale Welt ist nicht deterministisch, deswegen können Experimente nicht angewendet werden (Opp 1970: 43f.).*

Opp erkennt das Problem an, dass Akteure in gleichen Situationen unterschiedlich handeln können, allerdings, so argumentiert er weiter, wurde dieses Anders-Handeln-Können von den Kritikern bisher nicht belegt. Die Tatsache, dass relevante Gesetze bisher nicht gefunden wurden, heiße nicht, dass es sie nicht gäbe.

Vor dem Hintergrund der in Abschnitt 6.2 vorgestellten SEU-Theorie erscheint der Einwand, sofern man eine subjektive Rationalität von Akteuren in Rechnung stellt, nicht berechtigt. Es spricht nichts dagegen, subjektive Komponenten auch bei Experimenten konstitutiv zu berücksichtigen und eben gerade nicht davon auszugehen, dass sich Akteure in der gleichen Situation immer gleich verhalten. Strukturentdeckende Verfahren aus der empirischen Sozialforschung, z. B. die Clusteranalyse, ermöglichen u. a. die Identifikation und Charakterisierung von Akteurstypen, die in einer gegebenen Situation je unterschiedlich handeln. Für computersimulative Experimente, die im nächsten Abschnitt behandelt werden, kann der Determinismus-Kritik mit Zufallsgeneratoren begegnet werden, die das Verhalten von Computeragenten nicht deterministisch werden lassen. Ein weiterer konzeptioneller Vorschlag zur *Entdeterminisierung* wurde vom Thomas Kron (2005) mit seiner Forderung nach „Fuzzy-Logik für die Soziologie“ vorgelegt, der jedoch bisher keinen wesentlichen Einzug in die Praxis sozialwissenschaftlicher Experimente gefunden hat.

*Durch Experimente können nur simple kausale Strukturen erkannt werden. (Opp 1970: 44f.)*

Opp entgegnet, dass auch komplexe relationale Zusammenhänge untersucht werden können und dass die Erklärungskraft von einem intelligenten Experimentaldesign abhängt. Es lässt sich ergänzen, dass es per se nichts Schlechtes hat, zunächst einfache kausale Strukturen mittels Experimenten aufdecken zu wollen. Um die Frage nach einem möglichst intelligenten Experimentaldesign, das auch fähig ist, komplexe Effekte aufzudecken, hat sich der große Forschungszweig des *Design of Experiments* (vgl. exemplarisch Keppel und Wickens 2004; Montgomery 2013)

herausgebildet. Die Aufdeckung komplexer Strukturen mit der experimentellen Methode ist weniger ein Machbarkeitsproblem, sondern eher ein Ressourcenproblem. Je komplexer der aufzudeckende Zusammenhang<sup>2</sup> ist, desto höhere Fallzahlen sind erforderlich, um diesen Zusammenhang statistisch signifikant nachweisen zu können. Im Fall von Experimenten mit Versuchspersonen ergeben sich dadurch starke Einschränkungen, weil die Fallzahl nicht beliebig nach oben erweiterbar ist. Abhilfe schaffen Methoden der Computersimulation, die mit hohen Fallzahlen und einer hohen Anzahl von Replikationen (dafür aber ohne Probanden) umgehen können. Ein vielversprechender Ansatz ist die von Wilensky und Stroup (2000) entwickelte Software *Hubnet* als Teil der Simulationsplattform Netlogo, mit der sogenannte *partizipative Simulationen* möglich sind. Dabei können über das Internet eine große Anzahl an Probanden gleichzeitig an einem Experiment teilnehmen.

Trotz zahlreicher Vorteile der experimentellen Methode, die Opp durch Entgegnung auf typische Kritik herausstellt (s.o.), benennt er auch einige Nachteile, denen die experimentelle Methoden unterworfen sei (vgl. Opp 1970: 46ff.). So stellt er heraus, dass einige Experimente auf Grund moralischer Erwägungen nicht durchgeführt werden können. Entweder weil wichtige Variablen (z. B. die Gesellschaft als Ganzes) nicht verändert werden können, oder weil Wissenschaftscodizes die Modifikation dieser Variablen verbieten. Zum Ende der Achtzigerjahre diagnostizieren u. a. vor dem Hintergrund des Atomkraft-Diskurses nach Tschernobyl Krohn und Weyer (1989): „In der modernen Wissenschaft existieren Tendenzen, Forschungsprozesse und die mit ihnen verbundenen Risiken über die institutionalisierten Grenzen der Wissenschaft hinauszutreiben und die *Gesellschaft mit der Durchführung von Experimenten* zu belasten“ (Hervorh. R. F.). Die Gesellschaft werde zunehmend von der Wissenschaft als Labor für gesellschaftliche Großexperimente genutzt. Wenngleich es solche Art der Experimente auch im 21. Jahrhundert geben mag, z. B. in Form unausgereifter Software, die erst beim Kunden „reift“, so liegt der Fokus der vorliegenden Arbeit auf Laborexperimenten im ursprünglichen Sinne – also solchen, die unter kontrollierten Bedingungen innerhalb definierter Grenzen stattfinden. Zwar haben „echte“ Laborexperimente nicht die Tragweite gesellschaftlicher Großexperimente, dennoch sind auch bei diesen sozialwissenschaftlich orientierten Laborexperimenten gewisse ethische Mindeststandards (z. B. eine informierte Einwilligung) zu beachten (vgl. Aronson u. a. 2011: 48f.).

Opp (1970: 49ff.) befasst sich auch mit Hinweisen zur effektiveren Nutzung von Experimenten in den Sozialwissenschaften:

---

<sup>2</sup>Mit einem „komplexen Zusammenhang“ ist ein Zusammenhang im statistischen Sinne gemeint, etwa in Form von statistischen Interaktionseffekten oder Modellen mit einer Vielzahl von Einfluss- und Abhängigkeitsgrößen. Die Erklärung komplexer Phänomene ist nicht per se ein Mengen-, sondern vielmehr ein erkenntnistheoretisches Problem.

Insbesondere sei es wichtig, neben den Experimentalvariablen auch weitere Variablen zu benennen, die einen eventuellen Einfluss auf den zu messenden Effekt haben könnten. Insbesondere werden Versuchsleitereffekte oder Kommunikation (vgl. auch Bortz und Döring 2006: 84f.) als mögliche Störgrößen genannt. Opp betont, dass diese Größe im Zuge der Experimente, wenn nicht ausgeschaltet, dann wenigstens identisch gehalten werden sollen. Als besonders fruchtbar für die Sozialwissenschaften sieht Opp (1970: 50) die Kombination von Fragebogen und Experiment. Eine solche Kombination wurde auch für die beiden empirischen Untersuchungen der vorliegenden Arbeit gewählt (vgl. Kapitel 10 und 11). Opp empfiehlt mit Hilfe von Befragungen herauszufinden, ob alle relevanten unabhängigen Variablen im Experimentaldesign berücksichtigt wurden. Eine solche Erhebung wurde insbesondere für den in Kapitel 10 behandelten Mensa-Assistenten durchgeführt, um Informationen über Aspekte zu erhalten, die bei der Essenswahl eine Rolle spielen.

### **Experimentelle Forschungspraxis in den Sozialwissenschaften**

Für aktuelle sozialwissenschaftliche Forschung bemerkt Andreas Diekmann (2009), dass in der Ökonomie mehr experimentelle Studien durchgeführt werden als in der Soziologie. Diekmann bezieht sich dabei vor allem auf das Feld der experimentellen Spieltheorie. Im vorherigen Abschnitt 3.2.4 wurde ein spieltheoretisches Vertrauensspiel vorgestellt, das sich experimenteller Methoden bedient (Raub und Buskens 2006). Das Vorgehen der experimentellen Spieltheorie sieht vor, dass theoretische Vorhersagen aus der analytischen Spieltheorie mit den Ergebnissen abgeglichen werden, die im Rahmen von Experimenten erhoben werden. Für das experimentelle Vertrauensspiel bemerkt Diekmann (2009: 61), dass „[...] Versuchspersonen bereit sind, mehr Vertrauen zu geben, als die Spieltheorie erlaubt“. Ähnliche Resultate, also Differenzen zwischen den Vorhersagen der analytischen Spieltheorie und den experimentellen Ergebnissen, ergeben sich auch für Altruismus und prosoziales Verhalten. Reinhard Selten und Elinor Ostrom (1990) erhielten für ihre Forschungen auf dem Feld der experimentellen Wirtschaftswissenschaften den Nobelpreis – Selten 1994, Ostrom 2009. In der techniksoziologischen Forschung finden experimentelle Ansätze insbesondere in der qualitativen Forschung Verwendung und werden häufig gekoppelt mit ethnografischen Methoden (vgl. exemplarisch Hahne u. a. 2006; Rammert und Schubert 2006; Braun-Thürmann 2007; Krummheuer 2010).

## 7.2 Computersimulation als Methode der Sozialwissenschaften

Im vorherigen Abschnitt sind einige Schwierigkeiten der laborexperimentellen Methode skizziert worden. Je nach Fragestellung können moralische Bedenken ihre Anwendbarkeit, gerade bei gesellschaftlichen Fragestellungen, limitieren. Bei komplexen Wirkzusammenhängen stößt die experimentelle Methode schnell an Grenzen, da mitunter hohe Fallzahlen erforderlich sind.

Das Buch „Growing artificial societies“ von Joshua Epstein und Robert Axtell aus dem Jahr 1996 kann als Initialzündung zur Verwendung computersimulativer Methoden (genauer: der agentenbasierten Modellierung und Simulation, Details s. Abschnitt 7.3) in den Sozialwissenschaften angesehen werden. Die Idee, auf Computersimulation zur Untersuchung soziologischer Fragestellungen zurückzugreifen, findet sich für den deutschen Diskurs bereits im Jahre 1967 bei Renate Mayntz, die nach längerem Forschungsaufenthalt in den USA, beeindruckt von der dort aufkommenden Mathematisierung und Modellierung sozialer Prozesse (ebd.: 5), das Potenzial dieses relativ jungen Forschungsfeldes erkannt hat. Der besondere Wert von Computersimulation und der damit untrennbar verbundenen Modellierung sozialer Systeme liegt vor allem in dem „durch die Modellkonstruktion ausgeübten Zwang zur Präzisierung“ (ebd.: 27). „Die Formalisierung“, so Mayntz weiter, „zwingt dazu, *ausdrückliche* Annahmen zu machen“ (ebd., Hervorhebung im Original). Vordergründiges Ziel der Modellierung und Simulation sei die Entdeckung sozialer Mechanismen<sup>3</sup> und weniger die „quantitative Vorhersage bestimmter Ereignisse“ (ebd.: 25).

Dieses Argument, das sich begrifflich vielleicht am besten als *heuristischer Wert von Modellen* zusammenfassen lässt, zieht sich durch ihren gesamten Aufsatz. Sie attestiert dem damaligen Stand der soziologischen Theorie-Debatte mangelnde Explizitheit. Dabei fasst sie die Modellkonstruktion als ein Vehikel zur Bewusstmachung „latenter [Anm.: verborgener, ungenauer] Theorie“ (ebd.: 27f.) auf.

Zum Nutzen von Simulation schreibt Mayntz (1967: 24): „Der entscheidende Vorteil der Simulation ist, dass sie es erlaubt, dynamische Systeme darzustellen“ (zit. nach Kron und Lasarczyk 2006: 109).

Liest sich der Text von Renate Mayntz eher noch als Hinweis oder Ratschlag, sich zukünftig näher mit Modellierung und Simulation in der Soziologie zu beschäftigen, so werden fast 40 Jahre später Kron und Lasarczyk (2006) sehr viel deutlicher. Sie argumentieren, vor dem Hintergrund des gerade beendeten Sozionik-Schwerpunktprogramms im Jahr 2005, dass soziologischen Erklärungen „eine Notwendigkeit innewohnt, auf Computersimulationen zurückzugreifen“ (ebd.: 105). Ins-

<sup>3</sup>In jüngerer Zeit hat sich mit der *Analytischen Soziologie* ein Forschungsfeld entwickelt, das sich im Wesentlichen mit sozialen Mechanismen zur Erklärung von Makro-Mikro-Makro-Zusammenhängen befasst (vgl. Hedström 2008) und sich zunehmender Beliebtheit, auch im deutschsprachigen Raum, erfreut.

besondere die Verwendung von Zufallsprozessen zur Berücksichtigung von Vagheit<sup>4</sup> sei ein großer Vorteil dieser Methode (ebd.: 106).

Kron und Lasarczyk irritieren mit ihrer Einschätzung, dass mit Computersimulation grundsätzlich nichts Neues entdeckt werden könne, sondern nur eine neue Sicht auf Zusammenhänge möglich werde (ebd.: 110). Diese Aussage mag für solche Simulationen gelten, die lediglich bereits bekannte Ergebnisse mit Hilfe einer Computersimulation, die auf der Umsetzung bereits bekannter Wirkzusammenhänge basiert, replizieren. Die Aussage gilt schon nicht mehr, so lässt sich kritisch bemerken, wenn ein bekanntes sozialstrukturelles Phänomen mit relativ einfachen (bisher nicht bekannten Mechanismen) nachprogrammiert werden kann. So ist in dem Fall die neue Erkenntnis darin zu sehen, dass offenbar einfache Regeln ausreichen um bestimmte Strukturphänomene zu produzieren. Inwieweit sich diese neu gefundenen Regeln auch tatsächlich in der gesellschaftlichen Handlungspraxis wiederfinden, lässt sich nur als empirisch offene Frage stellen und durch rückkoppelnde Feldforschung unterstreichen oder widerlegen. Weiterhin lassen sich durch gezielte Variation von Simulationsparametern auch emergente, also unvorhergesehene, überraschende, Effekte erzeugen – s. Abschnitt 7.3 zur agentenbasierten Modellierung und Simulation. Zuletzt lässt sich entgegen, dass die Computersimulation nicht zwangsläufig selbstbezüglich arbeitet. Vielmehr bestehen auch Möglichkeiten, Computersimulationen mit Realdaten zu speisen oder auch menschliche Probanden mit Computersimulationen interagieren zu lassen. Eine systematische Auswertung der Simulationsläufe kann gerade im Fall von interaktiven Simulationen, wie sie für die vorliegende Arbeit programmiert wurden (vgl. u. a. Abschnitt 8.1 und 11), zu neuen Erkenntnissen führen und nicht nur zu einer anderen Sicht auf bereits Bekanntes.

Einen wesentlichen Vorteil der Computersimulation sehen Kron und Lasarczyk (2006: 117) hinsichtlich der Berücksichtigung „komplexer Aggregations-Dynamiken auf der Makro-Ebene“. Dabei steht nicht die *direkte* Nachbildung von Makro-Phänomenen im Vordergrund, die auch im Rahmen von Computersimulationen schwierig ist. Vielmehr wird es durch die Berücksichtigung des Verhaltens der Systemelemente auf der Mikro-Ebene möglich, komplexe Aggregationsphänomene auf der Makro-Ebene zu erzeugen. Die Autoren sehen (ebd.: 117f.) einen unmittelbaren Zusammenhang zwischen Multiagentensystemen<sup>5</sup> und dem methodologischen Individualismus: „Ohne Implementierung eines Handlungsmodells [Anm.: auf der Mikro-Ebene] läuft das Multiagentensystem nicht“ (ebd.: 119). Dabei handelt es sich um eine indirekte Simulation dynamischer Makro-Prozesse, da die Veränderungen, ganz im Sinne des

---

<sup>4</sup>Gemeint sind insbesondere Zufallsprozesse auf der Mikro-Ebene: Wird ein Agent zu zwei unterschiedlichen Situationszeitpunkten mit exakt den selben situativen Rahmenbedingungen konfrontiert, so bedeutet Vagheit, dass er sich trotz gleicher Situation je unterschiedlich entscheiden kann.

<sup>5</sup>Dieser Begriff ist weitgehend synonym mit dem Begriff der Agentenbasierten Modellierung und Simulation.

individualistischen Programms, stets über die Mikro-Ebene der beteiligten Systemelemente erzeugt werden.

Kron und Lasarczyk (2006: 132f.) plädieren schließlich dafür, die Methode der Computersimulation durch möglichst einfach zu bedienende Softwareumgebungen der sozialwissenschaftlichen Forschung und Lehre zugänglich zu machen, um die Grenzlinie zwischen Natur- und Sozialwissenschaften letztlich aufzulösen. Es lässt sich ergänzen, dass eine Simulationsumgebung, die auf der einen Seite von Sozialwissenschaftlern verstanden werden kann und anschlussfähig ist an etablierte Methodologien und auf der anderen Seite, die Programmierung auch komplizierter Zusammenhänge ermöglicht, als eine Art *Lingua franca* zwischen ebendiesen Disziplinen zu sehen ist.

### 7.3 Agentenbasierte Modellierung und Simulation

Gegenstand dieses Abschnitts ist die Methode der agentenbasierten Modellierung und Simulation (ABMS), die bereits im vorherigen Abschnitt unter dem Label Multiagentensystem Erwähnung fand. Die folgenden Ausführungen können eine lehrbuchartige Einführung in die Methode nicht ersetzen, sondern sollen vornehmlich einige Grundlagen und leitende Ideen dieses Ansatzes hervorheben.

Ausgehend von dem im vorherigen Abschnitt behandelten heuristischen Wert von Modellen stellt sich die Frage nach einem passenden Modellierungsansatz. Epstein und Axtell (1996: 1) argumentieren, dass soziale Prozesse und ihre Modellierung äußerst komplex sind und plädieren für einen Bottom-Up-Ansatz, bei dem ein (soziales) System durch die Aktivitätsprogramme der am System beteiligten Entitäten modelliert wird. Im Gegensatz zu Spieltheorie oder makroökonomischen Gleichgewichtstheorien, so die Autoren weiter (ebd.: 2), lassen sich mit agentenbasierten Modellen insbesondere dynamische Prozesse und heterogene Populationen untersuchen.

Das in den 1990er Jahren aufkeimende Forschungsfeld der komplexen adaptiven Systeme (Holland 1992) – *Complex Adaptive Systems* –, das als Vorläufer der ABMS angesehen werden kann, zeigte die Grenzen der Modellierbarkeit und Simulierbarkeit komplexer Systeme mittels globaler analytischer Modelle auf. Übertragen auf soziale Systeme, sieht Nigel Gilbert (1995) Vorteile simulativer, agentenbasierter Ansätze, vor allem darin begründet, dass Wechselwirkungen zwischen Mikro- und Makroebene untersucht werden können (ebd.: 144f.).

Eric Bonabeau (2002: 7280) stellt drei Vorteile agentenbasierter Modellierung (ABM) in der Vordergrund:

1. ABM macht emergente Phänomene fassbar bzw. erfahrbar.
2. ABM stellt eine natürliche Beschreibung eines Systems dar.
3. ABM ist flexibel.

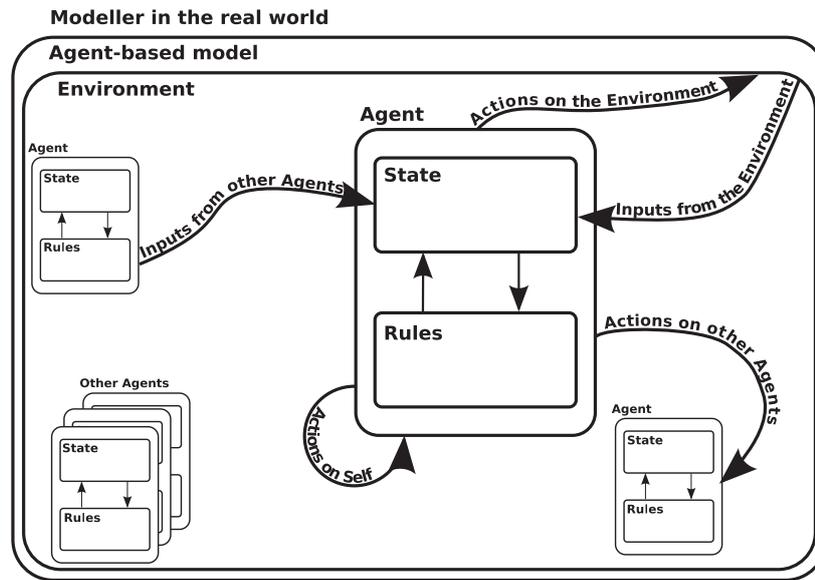
Als zentrale Triebkraft für ABM(S) benennt Bonabeau die Möglichkeit, mit dem an sich nicht fassbaren Phänomenen der *Emergenz* umgehen zu können (ebd.). Emergenz, so Bonabeau weiter, bezeichnet ein Systemverhalten, das sich nicht durch die einfache Verhaltensaddition der am System beteiligten Einheiten ergibt, sondern durch komplexe Interaktionen der Einheiten entsteht und losgelöst von den Einheiten zu neuen, überraschenden und nicht vorhersagbaren Resultaten führt. Unweigerlich schließt sich an das (durch Simulation) hervorgebrachte Phänomen die Frage an, wie dieses erklärt werden kann. Dass eine Erklärung eines emergenten Effekts keinesfalls trivial sein kann, folgt allein schon aus der Definition von Emergenz, denn „it should not be possible to derive analytically the global emergent behaviour solely from consideration of the properties of agents“ (Gilbert 1995: 150).

Joshua Epstein und Robert Axtell (1996: 19f.) schlagen in ihrem programmatischen Grundlagenwerk „Growing Artificial Societies. Social Science from the Bottom Up“ vor, die traditionellen wissenschaftlichen Erklärungsmuster zu überwinden. Sie argumentieren, dass nicht mehr gefragt werden soll „Can you explain it?“, sondern „Can you grow it?“. Ein analytischer Erklärungsanspruch wird dabei zu Gunsten einer generativen Wissenschaft weitgehend aufgegeben: Ein Phänomen gilt als erklärt, sofern (einfache) Mechanismen gefunden werden können, die fähig sind, dieses Phänomen Bottom-up hervorzubringen. Axelrod (1997: 3) spricht in dem Zusammenhang sogar von einem „third way of doing science“ (neben den klassischen erkenntnistheoretischen Methoden der Induktion und Deduktion).

### 7.3.1 Struktur einer ABMS

Van Dam u. a. (2013) fassen in ihrem Buch „Agent-Based Modelling of Socio-Technical Systems“ den aktuellen Forschungsstand zur ABMS zusammen und befassen sich aus theoretischer und praktischer Sicht mit der Thematik. Nikolic und Kasmire (2013) stellen in ihrem Beitrag in dem genannten Buch einige theoretische Überlegungen zur ABMS an und spezifizieren unterschiedliche Elemente agentenbasierter Modelle. Ein Tutorial zur ABMS haben Charles Macal und Michael North in verschiedenen Versionen vorgelegt und immer wieder aktualisiert (Macal und North 2005; 2010). Im Gegensatz zu Nikolic und Kasmire widmen sich die beiden Autoren eher der logischen Struktur agentenbasierter Modelle und beschäftigen sich ausführlich mit Interaktionen, während Nikolic und Kasmire eine eher informatisch-technische Sichtweise auf agentenbasierte Modelle und ihre Spezifizierung einnehmen. Eine Kombination beider Sichtweisen, gepaart mit eher grundlegenden Überlegungen von Epstein und Axtell (1996) und Gilbert (2008) soll einen möglichst facettenreichen Einblick in die Struktur agentenbasierter Modellierung und Simulation liefern.

In Abbildung 7.1 sind die wesentlichen Elemente einer ABMS nach Nikolic und Kasmire (2013: 58) dargestellt.



**Abbildung 7.1:** Grundstruktur einer ABMS  
Quelle: Nikolic und Kasmire (2013: 58)

Der Hauptbestandteil eines agentenbasierten Modells ist der *Agent*. Ein Agent ist eine gekapselte Einheit, die in eine *Umgebung* eingebettet ist. Ein Agent kann Aktionen auf der Umgebung vollziehen, kann Informationen aus der Umgebung verarbeiten sowie Aktionen auf sich selbst und auf anderen Agenten ausführen. Umsetzbar sind eine derartige Kapselung und Aktionsausführung durch Rückgriff auf das objektorientierte Programmierparadigma, auf dessen Grundlage agentenbasierte Simulationsumgebungen in der Regel aufbauen (vgl. Gilbert 2008: 64ff.).

## Agent

Nach Nikolic und Kasmire (2013) umfasst ein Agent insbesondere einen *Zustand* (ebd.: 58f.) sowie *Zustandsübergänge* (ebd.: 59ff.).

Ein Zustand beinhaltet dabei alle relevanten Informationen, um einen Agenten zu einer bestimmten Zeit zu charakterisieren (ebd.: 58). Programmiertechnisch lassen sich Zustände mit spezifischen Agentenvariablen realisieren, die sich im Laufe der Simulation stetig verändern können. Ein Zustand ist eine Momentaufnahme zu einem bestimmten Simulationszeitpunkt. Erweitert man die Betrachtungsebene des einzelnen Agenten, den Nikolic u. Kasmire als *internen Zustand* bezeichnen, umfasst der *lokale Zustand* den internen Zustand des Agenten zusammen mit den öffentlich beobachtbaren Zuständen der nah umgebenen Agenten, mit denen der Agent interagiert. Auf diese Weise lässt sich der interne Zustand eines Agenten in einen (lokalen)

Kontext einbetten. Bezieht man zusätzlich die über die unmittelbare Umgebung des Agenten einsehbaren Zustände mit ein, erhält man den *globalen Zustand*<sup>6</sup>.

Bei Macal und North (2010: 153) findet sich diese feingranulare Unterscheidung von Zustandstypen nicht, dort heißt es schlicht: „An agent has a *state* that varies over time“, „[...] that represents the essential variables associated with its [the agent’s] current situation“.

Sofern Agenten Zustände besitzen, stellt sich die Frage, wie sich die Zustände verändern können. Nikolic und Kasmire (2013: 59) nennen diese Änderungen *Zustandsübergänge*, die ihrerseits aus *Regeln*, *Aktionen* und *Verhalten* bestehen können.

Regeln, so die Autoren, geben an, wie Zustände in Aktionen und/oder neue Zustände überführt werden können. Sie unterscheiden dabei unterschiedliche Typen von Regeln, u. a. einfache If-Then-Else-Konstrukte aber auch komplexere multikriterielle Entscheidungsregeln und maschinelle Lernverfahren. Agentenregeln, so lässt sich festhalten, können je nach Anwendungsdomäne eine unterschiedliche Komplexität aufweisen – ganz im Gegensatz zu den Modellen bei Epstein und Axtell (1996), die auf der Agentenebene sehr einfache Regeln zum Aufbau ihrer künstlichen Gesellschaft verwendet haben.

*Aktionen* sind das Ergebnis der vorgelagerten Regelverarbeitung und können sich zustandsverändernd auf den Agenten selbst oder seine unmittelbare Umgebung auswirken (Nikolic und Kasmire 2013: 60).

Das *Verhalten* des Agenten ist schließlich das beobachtbare emergente Zusammenwirken aller Zustandsveränderungen und Aktionen. Das emergente Systemverhalten ergibt sich aus dem Zusammenwirken des Verhaltens aller am System beteiligten Agenten.

Zusammenfassend lassen sich aus der dargestellten architektonischen bzw. implementatorischen Sichtweise auf Agenten einige abstrakte Agenteneigenschaften benennen: In Anlehnung an die Definition von Wooldridge und Jennings (1995: 116) ist unter einem Agent ein softwaretechnisches System zu verstehen, das folgende Eigenschaften aufweist:

- *Autonomie*: Agenten agieren selbstständig ohne direkte Eingriffe von außen. Sie verfügen über einen internen Speicher und haben eine gewisse Kontrolle über ihre Aktionen.
- *Soziale Fähigkeiten*: Sie verfügen über die Möglichkeit, mit anderen Agenten zu interagieren (auch mit Menschen) und verwenden dabei ggf. eine Agentenkommunikationssprache.

---

<sup>6</sup>Die Bezeichnungen lokaler Zustand und globaler Zustand sind unglücklich gewählt, da selbst der globale Zustand, folgt man Nikolic und Kasmire (2013: 59), aus „all of the relevant states in the whole of the observable or influencing environment“ besteht und damit auch lokal ist.

- *Reaktivität*: Die Agenten haben die Fähigkeit, auf Änderungen der Umwelt zu reagieren und ihr eigenes Verhalten darauf auszurichten.
- *Proaktivität*: Agenten reagieren nicht nur, sondern verfolgen auch eigene Ziele und richten ihr Verhalten auf die Zielerreichung aus.

Diese Sammlung von Agenteneigenschaften wird bei Wooldridge und Jennings unter dem Label *schwacher Agentenbegriff* subsumiert. Der *starke Agentenbegriff* ergänzt die oben genannten Eigenschaften um zusätzliche Konzepte, die auch zur Konzeptualisierung menschlicher Eigenschaften verwendet werden (ebd. 117), z. B. BDI-Agenten, die explizit *beliefs*, *desires* und *intentions*, also genuin menschliche Eigenschaften (vgl. Rao und Georgeff 1995), bei der Modellierung berücksichtigen. Die Verwendung von SEU-Modellen auf Ebene der Agenten, wie sie im Rahmen der vorliegenden Arbeit erfolgt, fällt unter den starken Agentenbegriff. Zur Modellierung der Agenten greift diese Arbeit auf Modelle zurück, die in dieser Form auch in der empirischen Sozialforschung auf menschliche Akteure, etwa in der Survey-Forschung (Best 2007), angewendet werden.

### Umgebung

Agenten sind eingebettet in eine Umgebung. Diese Umgebung muss nicht zwangsläufig, wie bei den frühen agentenbasierten Modellen, eine räumliche, schachbrettartige Struktur aufweisen (Epstein und Axtell 1996: 21), sondern kann je nach Anwendungsfall unterschiedlich reichhaltig gestaltet sein (vgl. Macal und North 2010: 155) und z. B. auch eine netzwerkförmige Topologie aufweisen. Eine Umgebung, in die Agenten eingebettet sind, hat eine Doppelfunktion inne: Erstens dient sie der Bereitstellung bestimmter (lokaler) Informationen und zweitens dient sie als Infrastruktur, auf der sich die Agenten bewegen und miteinander interagieren. Die Umgebungsstruktur oder auch Topologie entscheidet darüber, welche Agenten miteinander interagieren können. Als mögliche Topologien werden zelluläre Automaten, euklidische Räume, Netzwerkstrukturen (z. B. Small-World-Netzwerke), Geoinformationssysteme sowie unstrukturierte Suppen genannt (ebd.). In der als Suppe (engl. *soup*) bezeichneten nicht-räumlichen Struktur kann zu jedem Zeitpunkt (im Prinzip) jeder Agent mit jedem anderen interagieren, feste Beziehungen zwischen den Agenten existieren nicht, Interaktionen entstehen spontan.

### Interaktionen

Interaktionen sind ein wesentlicher Bestandteil agentenbasierter Modelle. Für Bonabeau (2002: 7280) sind Interaktionen die wesentliche Ursache für Emergenz: „Emergent phenomena result from the interactions of individual entities“. Gilbert (2008: 6) fasst die explizite Berücksichtigung von Interaktionen als wesentliches Distinktionskriterium zu anderen computergestützten Modellierungstechniken auf: „The possibility

of modeling such agent-to-agent interactions is the main way in which agent-based modeling differs from other types of computational models.“

Grundsätzlich sind drei Arten von Interaktionen in agentenbasierten Modellen möglich: Agent-Agent-Interaktionen, Agent-Umgebung-Interaktionen sowie Umgebung-Umgebung-Interaktionen. Im letzten Fall kann man auch von zellulären Automaten sprechen, bei denen sich die Zustände von Zellen in Abhängigkeit von den jeweils benachbarten Zellen verändern. Ein Merkmal dieser Interaktionsart ist, dass die Interaktionen mehr oder minder automatisch nach festen Regeln erfolgen, die beschreiben, wie sich Zellen im zeitlichen Verlauf der Simulation in Abhängigkeit von ihren Nachbarzellen verändern.

Weitaus interessanter sind hingegen die von Agenten aktiv initiierten Interaktionen mit anderen Agenten oder solche mit ihrer jeweiligen Umgebung. Folgt man dem Vorschlag von Macal und North (2010: 154), so sind zwei Aspekte bei Agenteninteraktionen zu berücksichtigen:

1. Wer kann mit wem interagieren?
2. Welchem Mechanismus folgt die Interaktionsdynamik?

Wie schon bei den Ausführungen zur Umgebung angesprochen, restringiert die Topologie die Interaktionsmöglichkeiten eines Agenten. Typischerweise, so Macal u. North (ebd.), interagieren Agenten mit einer Agententeilmenge, genannt Nachbarn, die sich im zeitlichen Verlauf der Simulation stetig ändert. Wie genau sich diese Menge ändert, hängt wiederum von der jeweils implementierten Topologie ab.

Während die beiden Autoren dahingehend nicht weiter ins Detail gehen, wie die Interaktion vollzogen wird, ist Gilbert (2008: 15) präziser und argumentiert, dass je nach Anwendungsfall unterschiedlich komplexe Interaktionen berücksichtigt werden können: Im einfachsten Fall, so Gilbert, tauschen Agenten lediglich Daten aus. Es sind aber auch deutlich komplexere Szenarien denkbar, bei denen Agenten über eine Agentenkommunikationssprache Nachrichten austauschen, die von dem jeweiligen interaktiven Gegenpart erst interpretiert werden müssen.

### 7.3.2 ABMS-Anwendungen

Der ideengeschichtliche Ursprung der agentenbasierten Modellierung und Simulation, so wird zumindest übereinstimmend von wichtigen ABM(S)-Autoren (vgl. Epstein und Axtell 1996: 3; Gilbert und Troitzsch 2005: 23; Macal und North 2010: 156) berichtet, ist in den Segregationstudien von Schelling (1971) zu verorten, die in einem gesonderten Abschnitt behandelt werden.

Frühe Anwendungen von agentenbasierten Modellen finden sich zudem in dem Buch „The Complexity of Cooperation“ von Robert Axelrod (1997), in dem u. a. spieltheoretische Modelle mit Hilfe von ABMS analysiert werden und agentenbasierte Modelle

zur Erklärung von Normentstehung, sozio-kulturellem Wandel und aggregierten politischen Akteuren herangezogen werden.

Besonders eindrucksvoll haben zudem Joshua Epstein und Robert Axtell (1996) den Zusammenhang von einfachen Mikro-Regeln und emergenten Makro-Phänomenen präsentiert. So konnten die Autoren zeigen, dass sich aus nur einfachen und wenigen Agentenregeln, heterogenen Agentenpopulationen und Ressourcen (Zucker, Gewürze) komplexe künstliche Gesellschaften aufbauen lassen, in denen sich Strukturen wie Kultur, Konflikte und Handel wiederfinden lassen.

Aktuelle Anwendungsbereiche von ABMS erstrecken sich von ökonomischen Anwendungen wie der Modellierung noch nicht existierender Märkte, über Anwendungen in der Archäologie und Anthropologie zur Nachbildung vorzeitlicher Gesellschaften bis hin zur Simulation sozio-technischer Systeme der Flugsicherung (vgl. Macal und North 2010: 156f.).

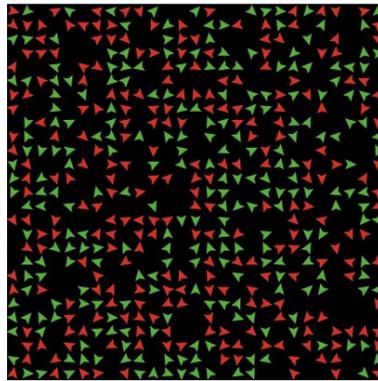
### Schellings-Segregationsstudien

Den Ausgangspunkt von Schellings Studie bildet ein empirisch beobachtetes Phänomen: die Segregation in Form von Ghetto-Bildung in Großstädten. Schelling versucht nun, dem generativen Mechanismus dieses Phänomens auf die Spur zu kommen. Dabei bedient er sich noch nicht der Methode der Computersimulation, sondern spielt seine Regeln anhand eines einfachen Brettspiels durch (ebd.: 156). Das Segregationsspiel hat folgende einfache Regeln: Es existieren zwei ethnische Gruppen. Die Mitglieder der jeweiligen Gruppe sind glücklich, wenn sich in ihrer unmittelbaren Nachbarschaft ein gewisser Prozentsatz ihrer eigenen ethnischen Gruppe befindet. Sie ziehen um, wenn sie unglücklich sind.

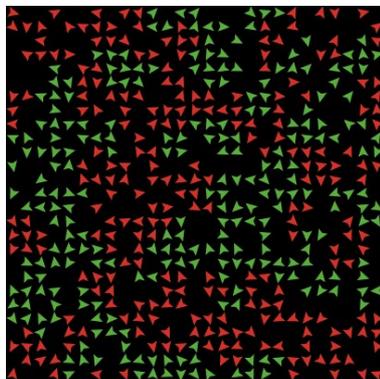
Schelling (1971: 158) hat herausgefunden, dass sich bereits bei einer Glücksschwelle von nur 30% (was bedeutet, dass bis zu 70% Fremde in der Nachbarschaft akzeptiert werden) nach wenigen Zügen eine Segregation ausbildet, bei einem Prozentsatz von 50% sogar eine starke Segregation. Bemerkenswert ist das Ergebnis, da die Regeln auf der Mikro-Ebene, die sogar eine hohe Toleranz gegenüber Fremden beinhalten, zu einem selbstverstärkenden Effekt auf der Makro-Ebene führen.

Die Segregations-Experimente von Schelling lassen sich mit der ABMS-Software Netlogo<sup>7</sup> nachvollziehen. Bei einer Startkonfiguration von 500 Agenten befinden sich *im Durchschnitt* etwa 50% Fremde in der Nachbarschaft (Abbildung 7.2a). Werden 30% der gleichen Ethnie in direkter Nachbarschaft gewünscht, ergibt sich nach nur 10 Schritten eine Segregation von 71% (Abbildung 7.2b), bei gewünschten 50% eine noch deutlichere Segregation von 88% in 18 Schritten (Abbildung 7.2c).

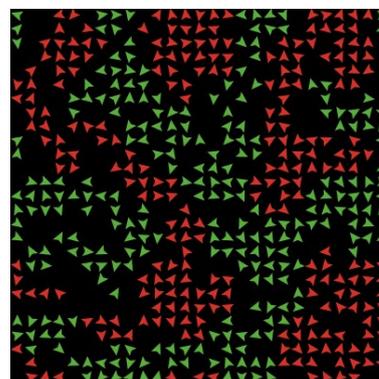
<sup>7</sup>Download: <http://http://cc1.northwestern.edu/netlogo/> (Abruf: 31. Januar 2014).



(a) Ausgangskonstellation mit 500 Agenten



(b) Gewünscht: 30% Gleiche,  
Resultat: 71% Gleiche



(c) Gewünscht: 50% Gleiche,  
Resultat: 88% Gleiche

**Abbildung 7.2:** Schellings-Segregations-Experiment: Vergleich von Ausgangs- und Endsituation als Netlogo-Simulation

### 7.3.3 Das KISS-Prinzip und seine Tücken

Agentenbasierte Modelle, die auf *einfachen* Mikro-Regeln basieren, bringen basierend auf diesen Regeln mehr oder weniger emergente Struktureffekte hervor. Diese Vorgehensweise wird als KISS-Prinzip (Keep it simple, stupid) der ABMS bezeichnet:

„When a surprising result occurs, it is very helpful to be confident that we can understand everything that went into the model. Although the topic being investigated may be complicated, the assumptions underlying the agent-based model should be simple. The complexity of agent-based modeling should be in the simulated results, not in the assumptions of the model.“ (Axelrod 1997: 5)

KISS ist in dem Fall als Aufforderung zu verstehen, die Agentenregeln so einfach wie möglich zu gestalten. Dabei geht es stets um die Entdeckung Sozialität hervorbringender Mechanismen.

Aus erkenntnistheoretischen Gesichtspunkten nimmt die ABMS eine Sonderstellung ein. So bezeichnet Robert Axelrod (1997: 3) sie neben der Induktion und Deduktion als „third way of doing science“. Auch Kron und Lasarczyk (2006: 116) bemerken, dass „[...] Mechanismen scheinbar einen anderen Anspruch an die soziologische Erklärung als die Forderung nach der Anwendung eines Gesetzes [stellen]“.

Dazu lässt sich kritisch bemerken, dass durch die Reklamation einer solchen Sonderstellung für die ABMS ihr Potenzial unnötig beschränkt wird. Die Entdeckung möglichst einfacher Mechanismen zum Nachvollziehen von Strukturphänomenen, dem „Züchten künstlicher Gesellschaften“ (Epstein und Axtell 1996), hat zweifelsfrei einen erklärenden Wert. Es handelt sich allerdings nur um eine soziologische Quasi-Erklärung, die zwar mit Agenten, allerdings ohne Akteure und damit ohne erklärende Elemente auskommt. Der Realitätsabgleich, also die Überprüfung inwieweit Simulation und Realität übereinstimmen, erfolgt bei der ABMS, sofern KISS-Regeln auf Mikro-Ebene verlangt werden, ausschließlich auf der Makro-Ebene.

Stellt sich ein bestimmtes Phänomen (z. B. Segregation) erst einmal ein und ist der Mechanismus entdeckt – im Falle der Segregation eine einfache Schwellenwertentscheidung –, der dieses Phänomen hervorbringen vermag, endet für den Agentenforscher die Arbeit. Sie endet dort, wo die eigentliche soziologische Arbeit beginnt – bei der Frage nach den Gründen für das Verhalten der Akteure/Agenten. Der nächste Abschnitt wird diese Feststellung zuspitzen und argumentieren, dass eine mikrofundierte ABMS, die anschlussfähig an Modelle aus der empirischen Sozialforschung ist, ihre Erklärungskraft nicht einbüßt, sondern durch systematische Berücksichtigung reichhaltiger Akteursmodelle zu einer *echten* soziologischen Makro-Mikro-Makro-Erklärung aufgewertet wird.

### 7.3.4 Für eine mikrofundierte ABMS

Selbst Kron und Lasarczyk (2006: 115), die eher als Befürworter der mechanistischen ABMS anzusehen sind, sind sich den Limitierungen eines KISS-Modells auf der Mikro-Ebene bewusst: „Wenn man nun nicht nur erklären möchte, dass eine bestimmte Form des handelnden Zusammenwirkens als Mechanismus zwischen einer Ursache und einer Wirkung fungiert, sondern auch zeigen möchte, wie der Mechanismus funktioniert, dann muss man zwangsläufig die einzelnen Handlungsselektionen, die auch hätten anders ausfallen können, erklären.“

Um diesem Erklärungsanspruch gerecht zu werden bietet sich ein Rückgriff auf das Makro-Mikro-Makro-Modell der soziologischen Erklärung an, das in Abschnitt 6.2 behandelt wurde. Hartmut Esser (1996: 136f.) schlägt für den Übergang von der Mikro-Ebene der Akteure zur Makro-Ebene explizit die Verwendung von Computersimulation vor. Das Modellierungsparadigma der ABMS drängt sich für eine Implementierung des Esser’schen Modells soziologischer Erklärung geradezu auf. Agenten werden dann zu simulierten Akteuren, die lokale Umgebung der Agenten zu situati-

ven Opportunitäten. Die Besonderheit der ABMS, dass durch das Zusammenwirken der Agenten auf der Mikro-Ebene emergente Effekte auf der Makro-Ebene generiert werden können, bleibt auch für ein reichhaltigeres Mikro-Modell erhalten. Das Modell soziologischer Erklärung, das für die Mikro-Ebene der Akteure auf SEU-Modelle zurückgreift, ist damit anschlussfähig an die empirische Sozialforschung Esser'scher Provenienz. Hennig Best (2007) hat gezeigt, dass, trotz einiger methodischer Schwierigkeiten, die Erhebung von Zielen und Eintrittswahrscheinlichkeiten mit Hilfe von Fragebögen möglich ist. Die so erhobenen Daten können in Form von typisierten Akteursmodellen Einzug in Mikro-Modelle einer agentenbasierten Modellierung und Simulation finden. Diese Überlegungen lassen sich zur forschungsleitenden Annahme (FLA 12) zusammenfassen.

#### FLA 12 (Mikrofundierung)

Eine mikrofundierte agentenbasierte Modellierung und Simulation hat das Potenzial eine soziologische Erklärung im Computer zu implementieren.

Anhand der Segregationsstudie von Schelling lässt sich der zusätzliche Erklärungswert von komplexeren Mikro-Modellen verdeutlichen: Schelling fand einen einfachen Mechanismus, der Segregation anhand eines auf der Mikro-Ebene wirksamen Schwellenwerts beschreiben konnte. Dieser Mechanismus sagt allerdings nichts über die *Gründe* für die Segregation aus Akteurssicht aus. So wäre z. B. denkbar, dass bestimmte religiöse Praktiken und Lebensgewohnheiten störend wirken und dazu führen, dass nur ein sehr geringer Anteil von Menschen fremder Ethnie in der Nachbarschaft geduldet wird. Die Verwendung reichhaltigerer Akteursmodelle auf der Mikro-Ebene eröffnet zumindest die prinzipielle Möglichkeit, solche Abwägungen in der ABMS konstitutiv zu berücksichtigen, statt durch Vorfestlegung auf KISS-Regeln diesen Erklärungsanspruch ex ante aufzugeben.

Ein besonders trauriges Ereignis, das die Grenzen von möglichst einfachen Mikro-Regeln in Computersimulationen aufzeigt, ist die Katastrophe bei der Duisburger Loveparade am 24. Juli 2010. Zur sicheren Durchführung dieses musikalischen Großereignisses hat der Duisburger Panikforscher Michael Schreckenberg ein Sicherheits- und Entfluchtungskonzept mitentwickelt und mittels Computersimulation validiert. Der völlig überfüllte Aufgang zur Veranstaltungsfläche hat eine Panik ausgelöst, infolgedessen 21 Menschen gestorben sind. Nur zwei Tage später gibt Michael Schreckenberg der Süddeutschen Zeitung<sup>8</sup> ein Interview und sagt dort folgenden bedenkenswerten Satz: „Auslöser war, dass sich einige nicht an die Spielregeln gehalten haben.“

Die von Schreckenberg durchgeführten Simulationen bedienen sich relativ einfacher Mikro-Regeln, die hauptsächlich auf physikalischen Flussmodellen und zellulären Au-

<sup>8</sup> <http://www.sueddeutsche.de/wissen/loveparade-experte-zur-ungluacksursache-die-treppe-haette-man-sprengen-muessen-1.979428> (Abruf: 31. Januar 2014).

tomaten basieren (Schreckenberg und Sharma 2002). Fußgänger werden dabei nicht als Akteure mit subjektivem Entscheidungsspielraum modelliert, sondern als einfache, sich regelhaft verhaltende Einheiten innerhalb eines Menschenflusses.

Dieses Beispiel zeigt, dass eine unterkomplexe Berücksichtigung subjektiver Entscheidungsspielräume fatale Folgen haben kann und sogar zu falschen Vorhersagen führen kann. Hat man einen generativen Mechanismus gefunden, der z. B. Verkehrsflüsse auf der Mikro-Ebene besonders elegant und einfach modellieren kann und weitgehend mit real beobachteten Flüssen von Menschenmengen übereinstimmt, so lässt eine solche Sichtweise außer acht, dass es einzelne Akteure geben kann, die auf Grund individueller Entscheidungen von diesem idealtypischen Flussverhalten abweichen. Durch die Modellannahme einer gewissen Uniformität der Menschenmenge kann den Einfluss und die möglichen Gefahren abweichenden Individualverhaltens auf die Makro-Ebene nicht näher bestimmt werden.

Abschließend zu den Überlegungen zur Mikrofundierung von Computersimulationen soll noch auf eine Simulation von Kroniger und Lücke (2010) verwiesen werden. Im Rahmen einer von mir betreuten Abschlussarbeit wurde das elaborierte Situationsdeutungs- und Entscheidungsmodell der Frame-Selektion (vgl. Abschnitt 6.3) in eine Verkehrssimulation integriert. An neuralgischen Punkten des Verkehrssystems wurden die simulierten Fahrertypen mit Entscheidungen konfrontiert, die typspezifisch und unter Berücksichtigung gegebener Situationsmerkmale subjektiv rational getroffen wurde. Ein solcher Ansatz kann nicht nur Effekte auf der Makro-Ebene, wie z. B. Staus, reproduzieren. Vielmehr wird es möglich, Effekte von Steuerungsimpulsen (z. B. das Sperren bestimmter Straßen oder das Schalten bestimmter Umleitungsempfehlungen) durch die systematische Berücksichtigung dieser neuen Situationsbedingungen auf der Mikro-Ebene der Fahrer im Sinne einer soziologischen Tiefenerklärung zu verstehen<sup>9</sup>.

---

<sup>9</sup>Diese Überlegungen lassen sich zusammen mit weiteren steuerungstheoretischen Fragen zu einem Mehrebenenmodell von Governance zusammenführen (vgl. auch Adelt u. a. 2014).



## 8 Vorschlag einer integrierenden Methodologie

Das folgende Kapitel ist in gewisser Weise als Synthese-Vorschlag zu betrachten. Dabei wird versucht, die vorangegangenen soziologischen Überlegungen zu autonomer Technik (insbesondere in Kapitel 5) in ein integratives Modell zu überführen.

Die Latour'sche Provokation einer radikal symmetrischen Ontologie von Natur, Technik und Gesellschaft (vgl. Abschnitt 5.1.1) hat den Blick geöffnet für eine Forschungsdisposition, die autonome Technik aus methodologischen Gesichtspunkten zunächst als weitgehend gleichberechtigten Partner anerkennt und ein Primat des Sozialen gegenüber dem Technischen nicht ex ante postuliert. Einige Fallbeispiele haben gezeigt, dass autonome Technik zunehmend handlungswirksam wird (Abschnitt 5.2), in höchstem Maße gesellschaftlich relevante Aktionen vollzieht und damit als soziologischer Untersuchungsgegenstand von Belang ist. Während die Handlungsfähigkeit von Technik in der soziologischen Theorie zumeist ignoriert bzw. ausgeschlossen wird, haben Werner Rammert und Ingo Schulz-Schaeffer ein attributionstheoretisches Modell vorgeschlagen, um die Frage nach der Handlungsträgerschaft als eine empirisch offene zu stellen (Abschnitt 5.3). Die ausführliche Würdigung ihrer Überlegungen hat schließlich auf einige Probleme aufmerksam gemacht, die sich im Zusammenhang mit der Anwendung ihres Modells ergeben.

Als besonders fruchtbare Grundlage für eine empirisch-experimentelle Untersuchung von Mensch-Maschine-Interaktion in soziologischer Perspektive dienen die in Abschnitt 6 vorgestellten Makro-Mikro-Makro-Modelle. Insbesondere das Esser'sche Modell soziologischer Erklärung mit seinem handlungstheoretischen Kern, der SEU-Theorie (Abschnitt 6.2), ist anschlussfähig an die experimentelle Methode und die agentenbasierte Modellierung und Simulation (Abschnitt 7.3), die u. a. als methodische Grundlage im Empirie-Teil der vorliegenden Arbeit verwendet wird. Eine wesentliche Erkenntnis aus den Überlegungen zur experimentellen Methode und zur Computersimulation bestand darin, dass eine Simulation mikrofundiert werden muss, zumindest dann, wenn die abgeleiteten Erkenntnisse denen einer soziologischen Tieferklärung genügen sollen.

Das *Modell soziologischer Erklärung hybrider Systeme* (HMSE), das in diesem Kapitel vorgestellt wird, versucht die Überlegungen aus den vorherigen Abschnitten auf eine Weise zusammenzuführen, die es ermöglicht, das Zusammenhandeln von Mensch und autonomer Technik im Rahmen von Laborexperimenten empirisch beobachtbar

zu machen und daraus soziologische Schlüsse zu ziehen, um schließlich Fragen des Vertrauens in autonome Technik handlungstheoretisch analysieren zu können.

Zunächst werden in Abschnitt 8.1 das HMSE und seine Modellbestandteile vorgestellt sowie die Modellierung eines hybriden Interaktionszusammenhangs beispielhaft demonstriert. Mit Abschnitt 8.2 schließen sich Überlegungen darüber an, welche unterschiedlichen Modellzuschreibungen bei Verwendung des HMSE im empirischen Forschungsprozess von Belang sind. Abschließend stellt Abschnitt 8.3 den methodologischen Nutzen des Modells heraus und führt aus, inwieweit das Modell den empirisch-experimentellen Teil der Arbeit anleiten kann und worin die Erklärungsstärke des Modells liegt.

## 8.1 Modell soziologischer Erklärung hybrider Systeme (HMSE)<sup>1</sup>

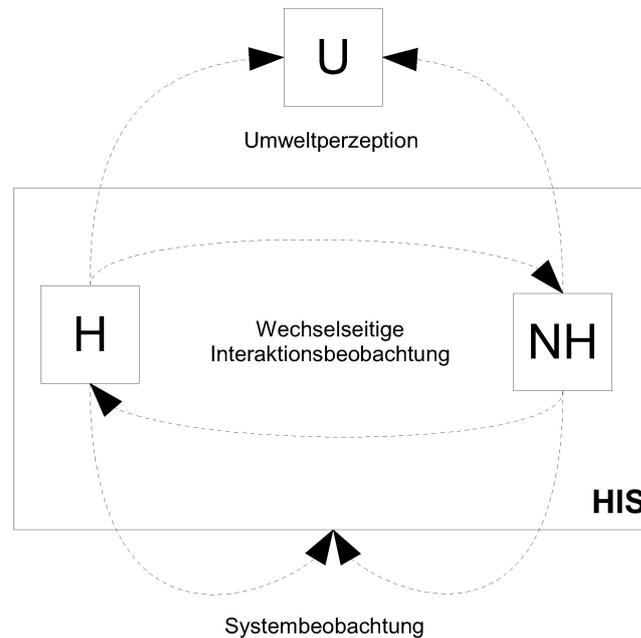
In Abschnitt 5.2 wurden einige Fallbeispiele für autonome Technik vorgestellt, bei denen Interaktivität und Handlungswirksamkeit wichtige Charakteristika ausmachten. So zeigte sich bei den Wikipedia-Bots (Abschnitt 5.2.3), dass sie ihre Umgebung in Form von Artikel-Historien beobachten und, falls notwendig, eingreifen. Ebenso zeigte sich bei der Betrachtung moderner Fahrerassistenzsysteme (Abschnitt 5.2.2), dass diese zunehmend eigenaktiv werden und mit dem menschlichen Fahrer zusammenhandeln bzw. -fahren und mit diesem in einem Interaktionszusammenhang stehen.

Vor dem Hintergrund erscheint es berechtigt, modelltheoretisch mit einem abstrahierten *hybriden Interaktionssystem* wie in Abbildung 8.1 dargestellt zu starten.

Die gestrichelten Pfeile in der Abbildung modellieren Beobachtungen (noch keine Aktionen). Ein *Hybrides Interaktionssystem (HIS)* konstituiert sich im Kern aus zwei Elementen: einem Mensch (H) und einem Nicht-Mensch (NH). Ein HIS hat klare Grenzen, die entweder physikalischer Natur sein können (etwa im Fall von Fahrerassistenzsystemen) oder auch virtueller Natur (wie im Fall der Wikipedia-Bots). Die Einziehung einer klaren Grenze für das hybride Interaktionssystem ist auch aus methodologischen Gesichtspunkten sinnvoll, da sie den Untersuchungsgegenstand damit deutlich umreist.

Während sich das hybride Interaktionssystem durch die wechselseitige Interaktionsbeziehung zwischen Mensch und Nicht-Mensch konstituiert, ist dieses System eingebettet in einen größeren soziotechnischen Zusammenhang, der in der Abbildung typisiert als Umwelt (U) aufgenommen ist. Die Elemente des HIS sind sich dieser Einbettung bewusst und beobachten die sie umgebende Umwelt. Innerhalb des HIS beobachten sich H und NH gegenseitig, sofern dies möglich ist und die Aktionen nicht so sehr gekapselt sind, dass eine direkte Beobachtung gar nicht mehr möglich wäre.

<sup>1</sup>Der folgende Abschnitt lehnt sich in Teilen an Fink (2009) und Fink und Weyer (2011) an.



**Abbildung 8.1:** Hybrides Interaktionssystem (HIS) – Abstrahierte Darstellung des Zusammenspiels von Mensch und autonomer Technik.

Weiterhin beobachten Mensch und Nicht-Mensch das durch ihre Interaktionen konstituierte HIS. Im Sinne des Zusammenhandelns (vgl. Abschnitt 5.3) kann das HIS von den internen Elementen des Interaktionssystems als agierende Einheit wahrgenommen werden und wirkt als solche auch nach außen. Würde das HIS ausschließlich selbstbezüglich agieren, wäre es als soziologischer Untersuchungsgegenstand uninteressant. Das HIS stellt sich bezüglich seiner soziotechnischen Einbettung als Einheit dar und kann von außen auch nur als solche adressiert und wahrgenommen werden. Den bisherigen Überlegungen zu dem hybriden Interaktionssystem fehlt noch eine wesentliche Eigenschaft: eine Handlungs- bzw. Prozesslogik. Dass sich H und NH in irgendeiner Weise wechselseitig beobachten, sagt noch nichts über die Auswirkungen dieser Beobachtungen und ihre Handlungswirksamkeit aus. Um diese näher zu umreißen wird zunächst die forschungsleitende Annahme (FLA 13) formuliert.

#### FLA 13 (Gemeinsame Aufgabe)

Das Zusammenhandeln von Mensch und autonomer Technik erfolgt stets im Kontext einer gemeinsam zu erfüllenden Aufgabe.

Diese Annahme umfasst zwei Unterannahmen, die sich jeweils auf den Aspekt der *Technikkonstruktion* und *Techniknutzung* beziehen. Technik wird nie ihrer selbst wegen konstruiert, sondern ihre Existenz leitet sich aus einem bestimmten Zweck ab,

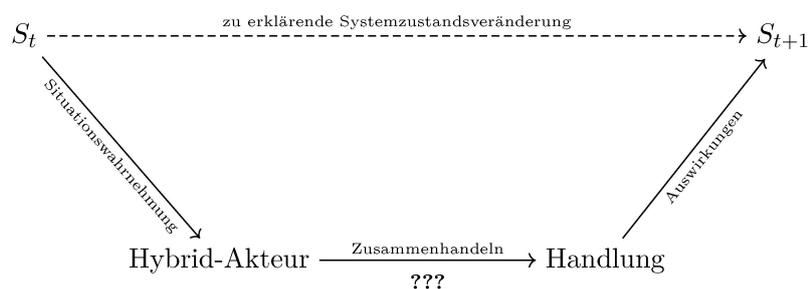
für den sie entworfen wurde – dies gilt gleichwohl für konventionelle Technik und autonome Technik. Wie in Abschnitt 5.3.3 gezeigt wurde, zeichnet sich autonome Technik dadurch aus, dass eine Identifikation von Einzelaktivitäten, die jeweils von Mensch oder autonomer Technik durchgeführt werden, schwierig ist. Allerdings haben die Nutzer von autonomer Technik zumindest ein grobes Bild davon, warum sie die Technik einsetzen und welche Aufgabe sie möglicherweise für sie erfüllen kann. Eine Kombination dieser beiden Aspekte führt zu der formulierten Annahme, dass Mensch und autonome Technik im Lichte einer bestimmten Aufgabe, eines Ziels, zusammenhandeln. Die Vorstellung darüber, welche gemeinsame Aufgabe zu erfüllen ist, kann zwischen H und NH durchaus differieren. Ob solche Unterschiede vorzufinden sind, kann im Zuge einer soziologischen bzw. handlungstheoretischen Analyse zum empirischen Untersuchungsgegenstand gemacht werden (vgl. Abschnitt 8.2).

Diese allgemeinen Überlegungen lassen sich an dem schon mehrfach angeführten Beispiel Adaptive Cruise Control (ACC) illustrieren: Die Entwickler und Ingenieure haben ACC zu einem bestimmten Zweck konstruiert: die Einhaltung einer bestimmten Wunschgeschwindigkeit unter Berücksichtigung vorausfahrender Fahrzeuge. Ein Fahrer, der das ACC-System nutzt, hat eine grobe Vorstellung darüber, welche Aufgaben das System erfüllt, allerdings weiß er möglicherweise nicht, welche Gewichtung das technische System für die Unterziele vornimmt. So wäre es z. B. denkbar, dass ein Fahrer davon ausgeht, dass das Ziel ausreichend Abstand zu einem vorausfahrenden Fahrzeug zu halten weniger wichtig ist, als eine bestimmte Geschwindigkeit einzuhalten. In Folge dessen könnte der Fahrer unnötigerweise selbst bremsen, obwohl die Vermeidung eines Zusammenstoßes in der internen Programmlogik des ACC-Systems höchste Priorität hat.

Zur weiteren Spezifizierung lässt sich das basale Modell eines hybriden Interaktionssystems (Abbildung 8.1) in einen größeren Erklärungszusammenhang einbetten. Aus (FLA 13) folgt, dass sich ein teleologisches Erklärungsmodell in besonderer Weise anbietet, das die Aktionen der beteiligten Akteure im Lichte bestimmter Ziele betrachtet.

Der Interaktionszusammenhang, der von den beteiligten Akteuren wechselseitig konstituiert wird, im Folgenden verkürzt *Hybrid-Akteur* genannt, führt über den Dreischritt von Situationswahrnehmung, (wie auch immer gearteter) Handlungsselektion (ausgehend von einer Anfangssituation) einen geänderten Systemzustand herbei.

Als erste Annäherung ist der beschriebene Zusammenhang in Abbildung 8.2 zusammengefasst. Das Explanandum, das zu Erklärende, ist die Zustandsveränderung, ausgehend vom Zustand zum Zeitpunkt  $t$ , also  $S_t$ , hin zum Zustand zum Zeitpunkt  $t+1$ , also  $S_{t+1}$ . Dabei wird die Situation bzw. der Ausgangszustand zum Zeitpunkt  $t$  von dem noch näher zu spezifizierenden Hybrid-Akteur wahrgenommen, das Zusammenhandeln ergibt schließlich eine Aktion, die ihrerseits den Systemzustand ändert und zum Folgezustand  $S_{t+1}$  führt.



**Abbildung 8.2:** Hybrid-Akteur im Makro-Mikro-Makro-Modell

Unklar bleibt bei der Darstellung, *wie* das Zusammenhandeln im Sinne des hybriden Interaktionssystems aus Abbildung 8.1 genau funktioniert. Ferner erscheint die Idealisierung, dass der Hybrid-Akteur die Situation direkt wahrnimmt, nicht angemessen, weil „der“ Hybrid-Akteur eigentlich aus zumindest zwei Teilen, einem menschlichen und einem nicht-menschlichen Part, besteht. H und NH, so lässt sich weiter argumentieren, interpretieren die Situation auf ihre jeweils eigene Art und Weise<sup>2</sup>. Wenn die Frage nach dem Zusammenhandeln von Mensch und autonomer Technik in den Mittelpunkt rückt, ist es unbedingt erforderlich, die Situationswahrnehmung des zusammengesetzten Hybrid-Akteurs und sein zusammengesetztes Handeln näher zu beschreiben.

Zur präziseren Modellierung dieses Zusammenhangs bietet sich ein genauerer Blick auf das Makro-Mikro-Makro-Modell der soziologischen Erklärung (MSE) an. In der Basisversion des MSE (s. Abschnitt 6.2) wird zunächst das Handeln eines einzelnen, typisierten Akteurs betrachtet. Das Zusammenhandeln mehrerer Akteure ist implizit über die Makro-Ebene enthalten, da diese, im Sinne des methodologischen Individualismus (Coleman 1991; Esser 1991; Opp 2009) als aggregierter Effekt des Handelns der Akteure auf der Mikro-Ebene anzusehen ist. Das zusammenwirkende Handeln mehrerer Akteure bzw. die Berücksichtigung ihres Handelns in Form von sozialen Prozessen lässt sich im individualistischen Programm auf zweierlei Weise modellieren:

Als sequentielle Verknüpfung von Einzelhandlungen einer oder mehrerer Akteure (Esser 1996: 102) oder als Interaktionssysteme (Esser 1996: 112ff.), bei denen parallele Entscheidungen mehrerer Akteure als soziales Gebilde zusammenwirken. Den ersten Fall nennt Esser „*horizontale* Erweiterung“ des Grundmodells, den zweiten Fall „*vertikale* Differenzierung von Mehr-Ebenen-Modellen der Erklärung des ‚Verhaltens‘ von sozialen Gebilden“ (ebd.: 102, Herv. im Original). Im Fall von sozialen Gebilden entsteht durch aufeinander bezogenes Handeln zweier oder mehrerer Akteure eine Art Zwischen- bzw. Meso-Ebene, die ihrerseits in einen größeren Erklärungszusammen-

<sup>2</sup> Autonome Technik nimmt die Umgebung in der Regel auf Basis digitaler Sensorinformationen oder Datenbankinformationen wahr, während menschliche Akteure sich ihres eigenen Wahrnehmungsapparats bedienen. Demnach generieren H und NH in ihrer jeweils eigenen Weise Lagebilder von der gemeinsamen Situation.

hang eingebettet ist (ebd.: 112). Als Beispiel für ein soziales Gebilde benennt Esser die Ehe, die als Meso-Ebene durch die Interaktion der Ehepartner hervorgeht und in einem größeren gesellschaftlichen Erklärungszusammenhang verortet ist.

Vor dem Hintergrund möglicher Erweiterungen des MSE stellt sich die Frage, welche der Erweiterungen anschlussfähig an die Modellierung des Zusammenhandelns von Mensch und autonomer Technik sind. Sofern das Zusammenhandeln und damit die Interaktion von Mensch und autonomer Technik als Erklärungsproblem thematisiert werden soll, bietet sich die vertikale Erweiterung des Modells als Grundlage für weitergehende Überlegungen an. Diese Überlegungen werden im folgenden Abschnitt zusammengefasst und münden schließlich in ein Modell soziologischer Erklärung hybrider Systeme (HMSE).

### 8.1.1 Symmetrische Modellierung

Hinsichtlich einer Übertragbarkeit der Esser'schen Mehr-Ebenen-Modelle ist zunächst festzustellen, dass bei Esser soziale Gebilde aus dem Zusammenhandeln *menschlicher* Akteure hervorgehen. Aus der Akteur-Netzwerk-Theorie (vgl. Abschnitt 5.1) wurde für die vorliegende Arbeit eine wichtige Forschungsdisposition abgeleitet. Autonomer Technik, so eine Schlussfolgerung aus der Gesamtwürdigung der ANT, sollte möglichst vorurteilsfrei begegnet werden. Insbesondere, so wurde dort argumentiert, verstelle ein Primat des Sozialen, der über das Technische gestellt wird, den Blick für die wesentlichen Fragestellungen des Zusammenhandelns von Mensch und autonomer Technik (ähnlich auch bei Rammert und Schulz-Schaeffer (2002b) / Abschnitt 5.3). Demnach scheint es als *methodologische* Annahme nicht unberechtigt, das Mehr-Ebenen-Modell von Esser auch auf nicht-menschliche Akteure anzuwenden. Die von Latour inspirierte Idee eines „Hybrid-Akteurs“ (vgl. Abschnitt 5.1.2) entspricht dabei dem sozialen Gebilde bei Esser.

In (FLA 13) wurde formuliert, dass das Zusammenhandeln von Mensch und autonomer Technik stets im Kontext einer gemeinsam zu erfüllenden Aufgabe zu sehen ist. Diese forschungsleitende Annahme kann für die Modellierung der Mensch-Maschine-Interaktion dahingehend zugespitzt werden, dass Mensch und autonome Technik mit der gleichen Modellierungssprache, nämlich der des Modells soziologischer Erklärung, beschrieben werden können. Im Sinne einer *methodologischen Symmetrie*, wie sie die ANT als Forschungsdisposition verordnet, erscheint dieser Schritt konsequent. Dabei ist es nicht erforderlich in einem ontologischen Sinne zu unterstellen, dass Nicht-Menschen Intentionen haben und rational-abwägend handeln. Vielmehr ermöglicht eine symmetrische Modellierung den Fokus auf Zuschreibungsprozesse zu verlagern (vgl. dazu ausführlicher Abschnitt 8.2) und das Zusammenhandeln von Mensch und autonomer Technik in der Perspektive wechselseitiger Zuschreibungen zum empirischen Untersuchungsgegenstand zu machen.

### Modellbestandteile

Das Modell soziologischer Erklärung hybrider Systeme (HMSE) besteht aus folgenden basalen Elementen, für die folgende Notation verwendet wird:

- menschlicher Akteur  $A_H$  mit zugehöriger Handlung  $H_H$
- nicht-menschlicher Akteur  $A_{NH}$  mit zugehöriger Handlung  $H_{NH}$
- ein aus der Interaktion von  $A_H$  und  $A_{NH}$  emergierender Hybrid-Akteur  $A_{Hy}$  mit zugehöriger zusammengesetzter Handlung  $H_{Hy}$
- Situation  $S$

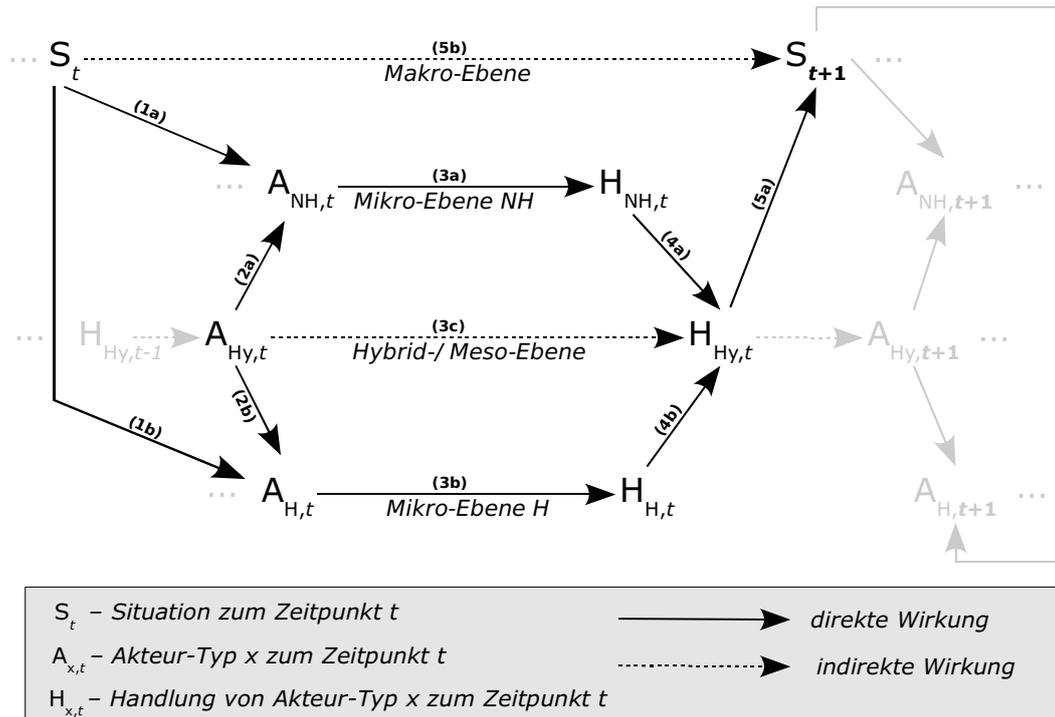
Das Modell ist als Prozessmodell anzusehen. Die Modellierung erfolgt jeweils für einen Übergang von Zeitpunkt  $t$  zum Zeitpunkt  $t+1$ . Die modellierten Akteure befinden sich dabei inmitten einer zuvor stattgefundenen Abfolge von Interaktionen. Um diesem zeitlichen Aspekt gerecht zu werden, werden die Modellbestandteile zusätzlich mit einem Zeitindex<sup>3</sup> versehen: So bezeichnet z. B.  $S_t$  die Situation zum Zeitpunkt  $t$ ,  $A_{H,t}$  den menschlichen Akteur zum Zeitpunkt  $t$  und  $H_{NH,t}$  die Handlung des nicht-menschlichen Akteurs zum Zeitpunkt  $t$ .

Der vollständige Interaktionszusammenhang ist in Abbildung 8.3 dargestellt und wird im Folgenden detailliert beschrieben.

In der Grafik ist zu sehen, dass es grau dargestellte und schwarz dargestellte Bereiche gibt. Die grau dargestellten Bereiche deuten an, dass es sich bei der betrachteten Situation  $S_t$  um einen *Ausschnitt* aus einer längeren Interaktionssequenz handelt und dass diese Sequenz auch nach Eintreten der Situation  $S_{t+1}$  weiter fortgesetzt wird. Im Schaubild sind drei Ebenen von Bedeutung:

- *Makro-Ebene* des übergeordneten soziotechnischen Systems: Der Zustand der Makro-Ebene ist beschrieben über die zeitlich indizierten Ausgangssituationen  $S_t$  bzw. die Nachfolgesituation  $S_{t+1}$ .
- *Mikro-Ebene* der Akteure: Der menschliche Akteur  $A_H$  sowie der nicht-menschliche Akteur  $A_{NH}$  agieren auf einer jeweils eigenen Mikro-Ebene, allerdings im Kontext eines durch sie wechselseitig konstituierten Interaktionssystems und einer gemeinsamen Situation.
- *Meso- bzw. Hybrid-Ebene*: Durch das Zusammenwirken von  $A_H$  und  $A_{NH}$  wird eine Hybrid- bzw. Meso-Ebene konstituiert. Diese Ebene bildet das nach außen sichtbare bzw. zustandsverändernde Handeln des Hybrid-Akteurs, das Auswirkungen auf die Makro-Ebene hat.

<sup>3</sup>Dieser Zeitindex ist nicht im Sinne einer bestimmten *Zeiteinheit* zu verstehen. Ein Übergang von  $t$  zu  $t + 1$  kann in Bruchteilen einer Sekunde stattfinden oder auch mehrere Tage auseinander liegen. Welche Taktung sich für das Prozessmodell anbietet, hängt vom empirischen Fall und von der Untersuchungsfrage ab.



**Abbildung 8.3:** Modell soziologischer Erklärung hybrider Systeme  
Quelle: Fink und Weyer (2011: 97)

Zur Beschreibung der einzelnen Wirkzusammenhänge wird das Schaubild von links nach rechts und unter Beachtung des zeitlichen Ablaufs besprochen.

Das übergeordnete Erklärungsproblem ist der Übergang vom Makro-Systemzustand  $S_t$  zum Nachfolgezustand  $S_{t+1}$ . Diese Zustandstransition ist im Sinne einer soziologischen Tiefenerklärung nicht erklärbar ohne die an diesem Übergang beteiligten Akteure zu berücksichtigen. Der gestrichelte Pfeil (5b) deutet an, dass die Ausgangssituation auf die Folgesituation wirkt, diese Wirkung allerdings indirekten Charakter hat. Die Ausgangssituation wirkt hingegen direkt auf die beteiligten Einheiten des hybriden Interaktionssystems –  $A_H$  und  $A_{NH}$  definieren die Ausgangssituation auf ihre je eigene Weise (Pfeile 1a und 1b). Neben der soziotechnischen Ausgangssituation auf der Makro-Ebene geht in die Situationsdefinition zusätzlich die als gemeinsame Interaktionsebene wahrgenommene Hybrid-Ebene ein (Pfeile 2a und 2b).

Nach subjektiver Definition der Situation erfolgt die Handlungsselektion von  $A_H$  und  $A_{NH}$ . Die Handlungsselektion auf den jeweiligen Mikro-Ebenen der Akteure erfolgt gemäß der SEU-Theorie (vgl. Abschnitt 6.2), d. h. die Handlungsalternative mit dem je höchsten subjektiven Erwartungsnutzen wird selektiert (Pfeile 3a und 3b) und bewirkt die im Schaubild dargestellte Handlung des menschlichen Akteurs  $H_{H,t}$  sowie die Handlung des nicht-menschlichen Akteurs  $H_{NH,t}$ . Die beiden Einzelhandlungen aggregieren sich zum Handeln des Hybrid-Akteurs  $H_{Hy,t}$ , in dem die Einzelhandlungen

gen durch das sie umgebende Interaktionssystem zusammengeführt werden (Pfeile 4a und 4b). Die Zusammenführung der Einzelhandlungen zu dem nach außen wirksamen Handlungsaggregat  $H_{Hy,t}$  kann je nach Interaktionszusammenhang sehr unterschiedlich erfolgen und muss für den jeweiligen empirischen Fall beschrieben werden.

Schließlich fehlt der Schaubildbeschreibung der letzte Schritt, in dem die aggregierte Handlung auf die Makro-Ebene des umgebenden soziotechnischen Systems zurückwirkt (Pfeil 5a) und einen neuen Folgezustand  $S_{t+1}$  herbeiführt. Wie genau die Aggregation von der Hybrid-Akteur-Aktion zum Nachfolgesystemzustand führt, hängt vom modellierten soziotechnischen System ab.

Der Übergang von der Ausgangssituation  $S_t$  zur Nachfolgesituation  $S_{t+1}$  ist das zu erklärende Ausgangsphänomen. Dieser Übergang lässt sich wie beschrieben soziologisch tiefenerklären über das Handeln eines aus wechselseitigen Interaktionen hervorgehenden Hybrid-Akteurs. Auf der Hybrid-Ebene ist ein zweites Erklärungsproblem zu verorten, die Erklärung des Zusammenhandelns von  $A_H$  und  $A_{NH}$  (Pfeil 3c).

Im Vergleich zu den sozialen Gebilden bei Esser sind im Rahmen des HMSE für die Situationsdefinition leicht modifizierte Wirkzusammenhänge berücksichtigt. Die Ursprungssituation  $S_t$  wirkt nicht direkt auf den Hybrid-Akteur, sondern nur auf die Akteure  $A_H$  und  $A_{NH}$ , die den hybriden Interaktionszusammenhang konstituieren. Zwar findet sich bei Esser (1993: 114) eine Anmerkung, dass nur Akteure (und keine sozialen Gebilde) Situationen wahrnehmen können. Dennoch findet sich bei Esser ein gestrichelter Pfeil, der von der Ausgangssituation zum sozialen Gebilde führt. Dem HMSE fehlt dieser gestrichelte Pfeil, da  $A_H$  und  $A_{NH}$  die Situation völlig unabhängig voneinander definieren und innerhalb des hybriden Interaktionssystems keine erkennbare Zusammenführung der Situationsdeutungen erfolgt.

Nach den theoretischen Ausführungen folgt eine beispielhafte Anwendung des HMSE auf einen hybriden Interaktionszusammenhang. Dabei soll deutlich werden, dass ein heuristischer Modellierungsmehrwert vom HMSE ausgeht, der zu einem handlungstheoretisch fundierten Verständnis von Mensch-Maschine-Interaktion beiträgt.

### 8.1.2 Illustration des HMSE am Beispiel von Adaptive Cruise Control

Adaptive Cruise Control (ACC) hat sich schon bisher als eine Art „Running Gag“ durch verschiedene Abschnitte der vorliegenden Arbeit gezogen. Zum Zwecke der Illustration ist ACC gut geeignet, da es auf der einen Seite übersichtlich und in seiner Funktionsweise gut nachvollziehbar ist, auf der anderen Seite verschiedene Sichtweisen auf das Zusammenspiel von Mensch und autonomer Technik ermöglicht.

In Abbildung 8.4 ist eine typische Fahrsituation, wie sie bei Verwendung von ACC vorkommt, symbolhaft dargestellt. Der Autofahrer stellt über die Benutzerschnittstelle seines Fahrzeugs die gewünschte Geschwindigkeit ein. ACC sorgt für die Einhaltung der Wunschgeschwindigkeit, erfasst allerdings permanent per Radarsensor die Umge-



**Abbildung 8.4:** Symbolische Darstellung von ACC – gemeinsame Situationswahrnehmung von Fahrzeug und Fahrer (Quelle: Daimler AG, <http://www.daimler.com/dccom/0-5-1210218-1-1210321-1-0-0-1210228-0-0-135-0-0-0-0-0-0-0.html>, Abruf: 31. Januar 2014)

bung des Fahrzeugs und passt die Geschwindigkeit bei Bedarf an ein vorausfahrendes Fahrzeug an, um einen Zusammenstoß zu vermeiden.

Diese Konstellation des Zusammenfahrens wird nun mit Hilfe des HMSE modelliert und näher beschrieben. Die Beschreibung orientiert sich dabei an den drei Esser'schen Logiken der Situation, Selektion und Aggregation. Der Gesamtzusammenhang besteht aus dem hybriden Interaktionssystem, das sich seinerseits zusammensetzt aus einem menschlichen Akteur  $A_H$ , dem Autofahrer, und einem nicht-menschlichen Akteur  $A_{NH}$ , dem ACC-System. Das hybride Interaktionssystem ist eingebettet in das soziotechnische System Straßenverkehr, präziser: in die umgebende Verkehrssituation auf der gerade befahrenen Straße.

### Logik der Situation

Der Fahrer  $A_H$  nimmt zunächst eine subjektive Definition der Situation vor. In einer Situation, wie in Abbildung 8.4 dargestellt (unabhängig davon ob ACC gerade aktiviert ist oder nicht), beobachtet er den vorausfahrenden Verkehr und schätzt auf Grundlage subjektiver Parameter, wie z. B. des Wetters, des Straßenzustandes oder der eigenen Wachsamkeit, ein, ob der aktuelle Abstand zum vorausfahrenden Fahrzeug ausreichend ist. Das ermittelte Lagebild umfasst eine Beurteilung, inwiefern die zur Verfügung stehenden Aktionen wie Bremsen, Beschleunigen oder Ausweichen eine Auswirkung auf die Erreichung bestimmter Folgezustände wie z. B. Unfallvermeidung oder Abstandseinhaltung haben.

Das ACC-System bzw. der nicht-menschliche Akteur  $A_{NH}$  erfasst die Situation ebenfalls. ACC greift zur Erfassung der umgebenden Verkehrssituation auf die Daten von Radarsensoren zurück. Aus den Sensordaten generiert ACC ein Lagebild und be-

rechnet im Hinblick auf einen erwünschten Nachfolgezustand (z. B. Einhaltung eines Mindestabstands) die Konsequenzen der bereitstehenden Aktionen (vgl. Winner u. a. 2002; Stanton und Young 2005).

Das beschriebene Handlungs- und Interaktionssystem umfasst im vorliegenden Fall zwei Akteure, die eine Verkehrssituation wahrnehmen und diese getrennt voneinander definieren (Pfeile 1a und 1b in Abbildung 8.3 auf Seite 196). Für den Fall aus Abbildung 8.4 könnte es zu unterschiedlichen Situationsdeutungen kommen. Angenommen, das Fahrzeug auf der linken Spur würde plötzlich nach rechts blinken, so könnte der menschliche Fahrer dieses Fahrzeug auf Grund des Blinkens bereits als möglichen Konfliktfall antizipieren. Für das ACC-System spielt das blinkende Fahrzeug noch keine wesentliche Rolle, da es sich nicht auf der gleichen Spur befindet. Würde sich gar kein Fahrzeug auf der eigenen Spur befinden, könnte ACC die Spur für frei halten und damit die Situation falsch einschätzen und ggf. im nächsten Schritt eine Beschleunigungsaktion einleiten.

### Logik der Selektion

Die im HMSE berücksichtigte Logik der Selektion folgt, wie bereits angedeutet, keinem ontologischen Ansatz, der Intentionalität von Technik unterstellt. Die symmetrische Konzeption von  $A_H$  und  $A_{NH}$  als intentional handelnde Akteure ist vielmehr ein methodologischer „Trick“, der es ermöglicht, Zuschreibungsprozesse zu analysieren und Handlungsträgerschaft empirisch zu fassen (Details dazu in Abschnitt 8.2). Bei der Entscheidung über eine auszuführende Handlung können beide Akteure aus einer Menge von Aktionen auswählen. In Kombination mit den zu verfolgenden Zielen (z. B. einen Unfall zu vermeiden) handeln beide Akteure nach dem Prinzip der subjektiven Nutzenmaximierung: Der menschliche Akteur  $A_H$  handelt intentional im ursprünglichen Sinne des Wortes (Pfeil 3b), der nicht-menschliche Akteur  $A_{NH}$  auf Grundlage seiner Software-Implementierung (Pfeil 3a), die aus den bereitgestellten Sensordaten die Aktion auswählt, die am ehesten zum erwünschten Nachfolgezustand führen wird.

### Logik der Aggregation I / Das Zusammenhandeln

Aus dem Zusammenwirken von  $A_H$  und  $A_{NH}$  ergibt sich das Gesamtverhalten des hybriden Akteurs (Pfeil 3c), wobei es durchaus zu überraschenden Effekten kommen kann. In dem oben beschriebenen Fall einer divergierenden Situationsdeutung würde der Fahrer bremsen, das Fahrerassistenzsystem hingegen beschleunigen. Über die Hybrid-Ebene werden diese Aktionen jedoch wechselseitig wahrgenommen (Pfeile 2a, 2b). Dadurch wird das Verhalten der Akteure in den folgenden Schritten der Handlungssequenz so lange beeinflusst, bis das Gesamtsystem zu einem adäquaten Verhalten gelangt ist, also keine gegenläufigen Steuerungsaktionen mehr notwendig

sind. Für außen stehende Beobachter ist jedoch nur das Verhalten des Hybrid-Akteurs  $A_{Hy}$  beobachtbar. Ein am Straßenrand stehender Fußgänger würde ein Fahrzeug beobachten können, das seine Geschwindigkeit an das jeweils vorausfahrende Fahrzeug anpasst und ansonsten eine konstante Geschwindigkeit fährt. Welche Aktionen *innerhalb* des hybriden Interaktionssystems dafür notwendig sind, bleibt einem Beobachter verborgen.

### **Logik der Aggregation II / Die Auswirkungen**

An diesem Punkt unterscheidet sich das HMSE nicht vom Esser'schen Modell soziologischer Erklärung. Die Transformations-Mechanismen, die eine Vielzahl von Einzelhandlungen von (rein) menschlichen Akteuren einerseits und hybriden Akteuren andererseits zu kollektiven Effekten zum Zeitpunkt  $S_{t+1}$  aggregieren (Pfeil 5a), sind identisch mit den von Esser genannten. Im Fall des HMSE können wir insbesondere die Möglichkeiten der Computersimulation nutzen, deren Stellenwert Esser (1991: 48) bereits 1991 betonte. Sowohl zelluläre Automaten als auch die Verfahren der agentenbasierten Modellierung und Simulation (vgl. Abschnitt 7.3) eignen sich für eine Darstellung und Analyse der Aggregation der Handlungen von Agenten (sogar in sehr großer Zahl).

Die Anwendung des HMSE auf das Fahrerassistenzsystem ACC hat gezeigt, dass es möglich ist, hybride Interaktionszusammenhänge mit diesem Modell zu modellieren. Durch die symmetrische Modellierung von Mensch und Nicht-Mensch können die Situationsdeutungen und Handlungsselektionen systematisch miteinander verglichen werden und mögliche Probleme der Mensch-Maschine-Interaktion aufgedeckt werden. So wäre es z. B. denkbar, dass die je eigene Vorstellung über erwünschte Folgezustände zwischen  $A_H$  und  $A_{NH}$  differiert. In der Sprache der SEU-Theorie würden sich solche Differenzen in unterschiedlichen Zielbewertungen niederschlagen.

## **8.2 Modellzuschreibungen**

Ausgehend vom vorgestellten HMSE ist die Frage zu stellen, worin genau der heuristische Modellierungsmehrwert liegt. Dazu wird in diesem Abschnitt der Begriff der Modellzuschreibung eingeführt und argumentiert, dass durch Kombination der experimentellen Methode und simulativer Ansätze aus der agentenbasierten Modellierung und Simulation Differenzen in wechselseitigen Modellzuschreibungen in den Blick genommen werden können. Diese Differenzen, so eine These der vorliegenden Arbeit, wirken sich auf Vertrauen in autonome Technik aus.

Der Abschnitt unterteilt sich in zwei Teilabschnitte. Zunächst wird ein Vorschlag zum systematischen Vergleich von SEU-Modellen unterbreitet (Abschnitt 8.2.1), der Modelle auf Basis ihrer Struktur und Ausprägung vergleicht. Im zweiten Teil (Abschnitt

8.2.2) werden diese Unterscheidungskriterien angeführt, um das Konzept der wechselseitigen Modellzuschreibungen einzuführen, mit dem hybride Interaktionszusammenhänge mit Hilfe der Modellierungsmöglichkeiten des HMSE vergleichend analysiert werden können.

### 8.2.1 Systematischer Vergleich von SEU-Modellen

Auf Seite 163 in Abschnitt 6.2 ist bereits angeklungen, dass der nomologische Kern des Modells soziologischer Erklärung (MSE), die SEU-Theorie, weit mehr ist als eine bloße Verrechnung von Handlungsalternativen zur Alternative mit dem größten Erwartungsnutzen. Der besondere Gewinn der SEU-Theorie für den Forschungsprozess, so meine Lesart, besteht darin, eine soziologische lingua franca zur Hand zu haben, mit der Sachverhalte einheitlich modelliert werden können und sich zueinander in Beziehung setzen lassen.

Mit den Notationen der SEU-Modellbestandteile von Seite 163 kann ein SEU-Modell beschrieben werden als Fünf-Tupel<sup>4</sup>:

$$(A, O, \mathbf{U}^{1 \times m}, \mathbf{P}^{m \times n}, s)$$

Sollen nun zwei SEU-Modelle  $M_1$  und  $M_2$  miteinander verglichen werden, bietet es sich an, diesen Vergleich auf Basis der Modellbestandteile aus dem Fünf-Tupel zu systematisieren. Zunächst ist festzustellen, dass die aufgeführten Modellbestandteile nicht unabhängig voneinander existieren. Insbesondere geben die betrachteten Handlungsalternativen  $A$  und die betrachteten Ziele  $O$  die *Struktur* des SEU-Modells vor. Sie beantworten die Frage, welche Handlungsoptionen für die betrachteten Akteure in der betrachteten Situation relevant sind. Somit geben die Ziele  $O$  die Dimensionalität (Anzahl der Einträge) des Zielbewertungsvektors  $\mathbf{U}$  und ihre inhaltliche Interpretation vor. Die Handlungsalternativen  $A$  bestimmen zusammen mit den Zielen die Dimensionalität der Erwartungsmatrix  $\mathbf{P}$  und die inhaltliche Interpretation der einzelnen Matrix-Einträge.

Auf Basis dieser Überlegungen können  $M_1$  und  $M_2$  auf zwei Ebenen vergleichend betrachtet werden: Auf Ebene der *Modellstruktur* und auf Ebene der *Modellausprägung*. Unterschiede in der Modellstruktur manifestieren sich darin, dass sich die zwei zu vergleichenden Modelle hinsichtlich ihrer Ziele und Handlungsalternativen unterscheiden. Unterschiede in der Modellausprägung liegen vor, wenn die modellierten Ziele und Handlungsalternativen zwar identisch, die Wertausprägungen des U-Vektors und der P-Matrix allerdings unterschiedlich sind.

<sup>4</sup>Zur Erinnerung:  $A$  – Handlungsalternativen,  $O$  – Ziele,  $\mathbf{U}$  – bewertete Ziele,  $\mathbf{P}$  – Erwartungen, die Handlungsalternativen und Ziele bzgl. ihrer erwarteten Eintrittswahrscheinlichkeiten bewerten sowie  $s$  – die Selektionsfunktion.

SEU-Modelle auf Basis von Modellausprägungen zu vergleichen ist der reguläre Vergleichsfall, der z. B. dann verwendet wird, wenn begründet werden soll, warum sich Akteursgruppen für bestimmte Handlungsalternativen entscheiden.

### Typen von Modelldifferenzen

Sofern die Selektionsfunktion als fester Modellbestandteil angesehen wird, ergeben sich aus der obigen Definition eines SEU-Modells insgesamt vier Modelldifferenztypen, die jeweils für sich betrachtet werden können:

**Alternativendifferenz:** Ein struktureller Unterschied zwischen zwei Modellen besteht, wenn sich die Alternativmengen voneinander unterscheiden, also  $A_1 \neq A_2$  gilt. In dem Fall unterscheiden sich die Modelle hinsichtlich der berücksichtigten Handlungsalternativen. Die relevanten Handlungsalternativen eines Fußballspielers und eines Handballspielers zum Erzielen eines Tors für die eigene Mannschaft unterscheiden sich voneinander. Demnach unterscheiden sich die SEU-Modelle bereits auf modellstruktureller Ebene.

**Zielmengendifferenz:** Analog zur Alternativendifferenz kann von einer Zielmengendifferenz gesprochen werden, wenn  $O_1 \neq O_2$  gilt. Ein solcher modellstruktureller Unterschied kann eintreten, wenn für zwei Akteursgruppen unterschiedliche Relevanzstrukturen wirksam sind. So wird beispielsweise ein Hochschulprofessor andere Zielvorstellungen hinsichtlich seines akademischen Handelns haben als ein Studierender.

**Zielintensitätsdifferenz:** Im Fall von Zielintensitätsdifferenzen unterscheiden sich die bewerteten Ziele der miteinander zu vergleichenden SEU-Modelle. Mathematisch bedeutet dies, dass die Anzahl der Einträge im U-Vektor beider Modelle zwar identisch ist, die Werte sich allerdings voneinander unterscheiden, also  $U_1 \neq U_2$  gilt. In dem Fall besteht der Unterschied zwischen beiden Modellen in der konkreten Modellausprägung. Dieser Fall tritt typischerweise beim Vergleich zweier Akteursgruppen auf und wird häufig als Unterscheidungskriterium herangezogen. Werden beispielsweise die Zielintensitäten zwischen Öko-Bauern und konventionellen Bauern miteinander verglichen, so bewerten konventionelle Bauern die Wichtigkeit effektiver Schädlingsbekämpfung deutlich höher als Öko-Bauern (vgl. Best 2007: 201).

**Eintrittswahrscheinlichkeitsdifferenz:** Sofern sich die P-Matrizen hinsichtlich ihrer Matrix-Einträge voneinander unterscheiden, lässt sich der Zusammenhang mathematisch als  $P_1 \neq P_2$  formulieren. Unterschiedliche Bewertungen der Handlungsalternativen hinsichtlich der Erreichung eines bestimmten Ziels bedeuten, dass die situativen Opportunitäten je unterschiedlich bewertet werden. Um beim Beispiel

von Best zu bleiben: Öko-Bauern schätzen die Wahrscheinlichkeit mit konventioneller Landwirtschaft zu einer umweltfreundlichen Produktionsweise zu kommen deutlich niedriger ein als konventionelle Landwirte (ebd.)

Inwieweit modellstrukturelle Differenzen inhaltlich interpretiert werden können, hängt vom empirischen Fall ab. Sofern beispielsweise eine identische Situation von zwei Akteursgruppen hinsichtlich der bereitstehenden Handlungsalternativen unterschiedlich antizipiert wird, so lassen sich durch die Aufdeckung solcher modellstruktureller Differenzen Unterschiede zwischen den beiden Akteursgruppen ausmachen, die inhaltlich interpretiert werden können. Übertragen auf die Interaktion von Mensch und autonomer Technik können modellstrukturelle Differenzen fatale Folgen haben. Wird beispielsweise eine bestimmte Handlungsalternative von einem menschlichen Akteur gar nicht erst antizipiert, können Erwartungen, die von dem nicht-menschlichen Akteur gestellt werden, unter keinen Umständen erfüllt werden.

Aus forschungspragmatischer Sicht ist man allerdings eher bestrebt, Modelle mit unterschiedlichen Ausprägungen miteinander zu vergleichen, da solche Vergleiche leichter zu bewerkstelligen sind und mit den üblichen mathematischen Vergleichsoperatoren auskommen. Bei einem Vergleich von strukturell unterschiedlichen Modellen müssten hingegen zusätzliche Annahmen getroffen und überprüft werden, ob ein Vergleich überhaupt sinnvoll möglich ist.

In Hinblick auf das HMSE ist zu untersuchen, inwieweit die spezifizierten Typen von Modelldifferenzen systematisch in Form von wechselseitigen Modellzuschreibungen untersucht werden können und welche Implikationen sich für ein Experimentaldesign ergeben, das diese Differenzen in den Blick nimmt.

### 8.2.2 Das Konzept der wechselseitigen Modellzuschreibungen

Die Grundidee des Konzepts der wechselseitigen Modellzuschreibungen besteht darin, dass sich  $A_H$  und  $A_{NH}$  jeweils ein Bild ihres Interaktionspartners machen und das eigene Handeln u. a. an diesem Bild orientieren. Lebensweltlich gesprochen hat ein menschlicher Akteur gewisse Vorstellungen über die Funktions- und Wirkungsweise von autonomer Technik, mit der er interagiert<sup>5</sup>. Analog kann man für autonome Technik argumentieren, dass diese auch gewisse Vorstellungen bzw. Modelle von denjenigen Akteuren hat, mit denen sie interagiert. Diese Vorstellungen entstehen im Gegensatz zu denen des menschlichen Akteurs nicht in einem kognitiven oder sinnstiftenden Akt der Technik selbst, sondern sind von Entwicklern der autonomen Technik in die Prozeduren und Abläufen eingeschrieben. Es ist auch denkbar, dass autonome Technik derart programmiert ist, dass sie die Vorstellungen ihres Gegen-

<sup>5</sup>Es wird unterstellt, dass die Interaktion nicht zufällig zu Stande kommt, sondern dass sich  $A_H$  bewusst dafür entscheidet, mit  $A_{NH}$  in einen Interaktionszusammenhang einzutreten.

übers in Abhängigkeit von stattgefundenen Interaktionen laufend aktualisiert, dass sie also in gewisser Weise lernfähig ist.

Dass sich Techniknutzer ein Bild von (interaktiver) Technik machen, geht auf die Überlegungen zu *mentalen Modellen* des Kognitionspsychologen Donald A. Norman (1983; 1986) zurück. Norman (1986) begründete das Feld des sogenannten *Cognitive Engineering* (ebd.: 31). Seine kognitionswissenschaftlichen Arbeiten sind im Kontext der zu dieser Zeit zunehmenden Verbreitung des Personal Computers und der damit einhergehenden Bedienungsprobleme zu sehen. Das Design von Benutzerschnittstellen stimmte in der Regel nicht mit den Zielvorstellungen der Nutzer überein. Es orientierte sich zumeist an den internen technischen Abläufen des jeweiligen Programms und nicht an den Vorstellungen der Nutzer der Programme. Als zentrale Aufgabe des Cognitive Engineering identifiziert Norman (1986: 43) die Überwindung der Lücke zwischen System und Nutzervorstellung und fordert eine Hinwendung zum „user-centered design“ (ebd.: 59f.). Zentral ist dabei der Begriff des *mentalen Modells*, das sich ein menschlicher Akteur im Zuge der Interaktion mit einem nicht-menschlichen Interaktionspartner von ebendiesem bildet:

„Mental models seem a pervasive property of humans. I believe that people form internal, mental models of themselves and of the things and people with whom they interact. [...] Mental models *evolve naturally through interaction* with the world and with the particular system under consideration [...]. These models are *highly affected by the nature of the interaction*, coupled with the person's prior knowledge and understanding.“ (Norman 1986: 46, Hervorhebungen R. F.)

Ausgehend von dieser Definition leitet Norman drei Modellkonzepte ab – zwei kognitive und ein physisches. Er unterscheidet zwischen *Design Model*, *User's Model* und *System Image* (ebd.: 47). Das Design-Modell ist das Modell, das die Designer bei der Entwicklung des Systems berücksichtigt haben. Das mentale Modell der Systemnutzer, das User's Modell, ist das Modell, das Anwender im Verlaufe der Interaktion mit dem System herausbilden. Grundlage für die Herausbildung des User's Model ist das System Image, also die Art wie sich das System nach außen darstellt – beispielsweise die Elemente der Benutzerschnittstelle und die Dokumentation des Systems. Im Gegensatz zu den anderen beiden Modellen ist das System Image kein mentales Modell, sondern umfasst alle sichtbaren physischen Elemente und Ausgaben des Systems.

Neben den mentalen Modellen, die ein menschlicher Nutzer für das System, mit dem er interagiert, herausbildet, spezifiziert Norman (1986: 41f.) einen Handlungskreislauf, der die Interaktion mit dem technischen System beschreibt. Er unterscheidet dabei zwischen physischen Aktivitäten und mentalen Aktivitäten. Physische Aktivitäten sind dabei die unmittelbaren Interaktionen mit dem technischen System (z. B. die Betätigung einer bestimmten Taste). Ausgehend von bestimmten Zielen („Goals“) bildet der Nutzer eine bestimmte Absicht („Intention“) aus, spezifiziert diese in einem Handlungsprogramm („Action Specification“) und führt diese Handlung, die auf

Interaktion mit dem technischen System gerichtet ist, schließlich aus („Execution“). Nachdem der Computer gemäß der gegebenen Anweisungen Instruktionen ausgeführt hat, werden die Wirkungen vom Benutzer wahrgenommen („Perception“), interpretiert und schließlich mit den ursprünglichen Absichten abgeglichen. Von dort beginnt ein erneuter kognitiver Kreislauf der Mensch-Maschine-Interaktion.

Die Überlegungen von Norman lassen sich, wenn auch nicht vollständig, auf die attributionstheoretischen Überlegungen zum HMSE übertragen. Sofern  $A_H$  und  $A_{NH}$  methodologisch-symmetrisch konzipiert sind, lassen sich zumindest vier Modelle differenzieren, von denen es sich bei zweien um zugeschriebene Modelle handelt:

- $SEU_{H \rightarrow NH}$ : Das Modell, das der menschliche Akteur  $A_H$  dem nicht-menschlichen Akteur  $A_{NH}$  zuschreibt.
- $SEU_{NH \rightarrow H}$ : Das Modell, das der nicht-menschliche Akteur dem menschlichen Akteur zuschreibt. Dabei handelt es sich um eine Vorstellung über einen typischen interaktiven Mitspieler, die in der Implementierung des nicht-menschlichen Akteurs eingeschrieben ist. Die Zuschreibung erfolgt in der Regel nicht explizit, es sei denn es handelt sich um einen lernenden nicht-menschlichen Akteur, der sein eigenes Handeln auf jeweils aktualisierte Vorstellungen über den menschlichen Mitspieler ausrichtet.
- $SEU_H$ : Das SEU-Modell des menschlichen Akteurs. Wobei dieses nicht objektiv als *das* SEU-Modell vorliegt, sondern in methodisch sinnvoller Weise erhoben werden muss.
- $SEU_{NH}$ : Das SEU-Modell des nicht-menschlichen Akteurs stellt die Wirk- und Funktionsweise von  $A_{NH}$  in Form eines abstrakten SEU-Modells dar.

Zwei dieser Modelle lassen sich mit den konzeptionellen Modellen bei Norman identifizieren:  $SEU_{H \rightarrow NH}$  ist weitgehend identisch mit dem User's Model bei Norman. Der wichtige Unterschied, und dabei kommt der Aspekt des teleologisch-interaktiven Charakters autonomer Technik ins Spiel, besteht darin, dass eine Angabe darüber gemacht wird, *wie* dieses zugeschriebene Modell strukturell aufgebaut ist – *nach Zielen und Aktionen*. Erhoben werden kann dieses zugeschriebene Modell über typische Erhebungsmethoden der empirischen Sozialforschung, z. B. via Fragebogen. Dabei werden die menschlichen Interaktionspartner befragt, welche Ziel- und Handlungskonsequenzbewertungen dem nicht-menschlichen Akteur zugeschrieben werden. Ein solcher Ansatz findet auch im Rahmen der vorliegenden Arbeit Verwendung.

Weiterhin ist  $SEU_{NH}$  mit dem Design-Modell bei Norman zu identifizieren.  $SEU_{NH}$  umfasst eine abstrakte Beschreibung mit Hilfe der Modellierungsmöglichkeiten der SEU-Theorie, nach welcher Logik die autonome Technik, die beschrieben wird, agiert. In der Regel wird man, um ein solches Modell zu erhalten, die Entwickler des Systems

darüber befragen müssen, welche Design-Ziele in  $A_{NH}$  umgesetzt sind. Ist dieses Modell expliziert, kann es ggf. mit  $SEU_{H \rightarrow NH}$  verglichen werden, um Differenzen zwischen Zuschreibung und Implementierung aufzudecken. Eine weitere Möglichkeit zur Explikation von  $SEU_{NH}$  besteht darin, die Implementierung selbst vorzunehmen, nachdem das abstrakte Modell spezifiziert worden ist. Eine solche Möglichkeit bietet sich in der Regel nur, wenn im Rahmen der experimentellen Methode die Implementierung von  $A_{NH}$  durch den Forscher selbst vorgenommen wird. Sofern die Prozessierung eines SEU-Modells unmittelbar implementiert ist, ist kein weiterer Abstraktionsschritt erforderlich, um aus der konkreten Implementierung das abstrahierte  $SEU_{NH}$ -Modell abzuleiten. In jedem Fall erfordert eine adäquate Modellbildung für  $SEU_{NH}$  eine mehr oder minder präzise Systemkenntnis.

$SEU_{NH \rightarrow H}$  bezeichnet die Zuschreibung einer bestimmten Handlungslogik, die dem menschlichen Akteur von der autonomen Technik unterstellt wird. Eine Erhebung dieser Zuschreibung ist methodisch anspruchsvoll, da sich die einer Implementierung eingeschriebenen menschlichen Handlungslogiken in der Regel nur implizit in der Software (selbst im Quellcode) wiederfinden. Am ehesten lassen sich diese dort finden, wo bestimmte Erwartungen, ein bestimmtes Mithandeln des menschlichen Akteurs zur gemeinsamen Problemlösung erwartet wird. Im Falle des Assistenzsystems ACC dürfte  $A_{NH}$  dem menschlichen Akteur unterstellen, dass dieser in einer Gefahrensituation das Ziel, einen Zusammenstoß zu vermeiden, ebenso hoch bewertet wie das Assistenzsystem. Eine Entsprechung dieser Zuschreibungsrichtung findet sich bei Norman nicht.

Ein Modell für den menschlichen Akteur<sup>6</sup>  $SEU_H$  findet bei Norman (1986: 41) den Niederschlag in der Spezifizierung von Nutzeraktivitäten („Stages of User Activities“). Dabei kommen die einzelnen Stationen, die durchlaufen werden, dem Modell soziologischer Erklärung strukturell relativ nahe. Allerdings fehlt eine operationalisierbare Beschreibung der einzelnen Stationen und eine Einordnung in einen größeren Makro-Mikro-Makro-Zusammenhang. Ferner ist zu bemerken, dass das Modell der Mensch-Maschine-Interaktion bei Norman asymmetrisch konzipiert ist. Der Nutzer ist der alleinige Handlungsträger, der durch Auswahl bestimmter Aktionen auf der Benutzerschnittstelle den Computer für seine Zwecke nutzt. Der Computer als entsprechender Gegenpart ist nicht mit einem solchen intentionalen Modell beschrieben, sondern operiert nach weitgehend festen Ablaufregeln gemäß der Benutzereingaben.

---

<sup>6</sup>Das Modell für den menschlichen Akteur ist freilich, zumindest nicht in positivistischer Sicht, angebar. Vielmehr handelt es sich um eine Annäherung, die mit Hilfe geschickter Erhebungsmethoden erfolgt. Trotz dieser terminologischen Feinheit wird weiterhin von dem Modell des menschlichen Akteurs die Rede sein, wohl wissend, dass dieses auch nur eine annähernde indirekte Modellierung sein kann, die von der Qualität der Erhebungsmethode abhängt.

Ausgehend von den SEU-Modellen,  $SEU_H$  und  $SEU_{NH}$ , und zugeschriebenen SEU-Modellen,  $SEU_{H \rightarrow NH}$  und  $SEU_{NH \rightarrow H}$ , lassen sich, im Hinblick auf Mensch-Maschine-Interaktion, zwei Arten von Modelldifferenzen unterscheiden:

- $SEU_{H \rightarrow NH} \neq SEU_{NH}$ : Unterscheidet sich das vom menschlichen Akteur zugeschriebene SEU-Modell von dem tatsächlich implementierten SEU-Modell, kann es im Zuge der Mensch-Maschine-Interaktion zu Problemen bzw. Unstimmigkeiten kommen. Unterscheiden sich beispielsweise die Zielbewertungen voneinander, können  $A_H$  und  $A_{NH}$  unterschiedliche Schlussfolgerungen aus einem gemeinsamen situativen Kontext ziehen. Sind die Differenzen bereits auf modellstruktureller Ebene zu verorten, so kann es dazu kommen, dass  $A_H$  den Vollzug einer bestimmten Aktion von  $A_{NH}$  erwartet, obwohl diese Option gar nicht als Handlungsalternative für  $A_{NH}$  zur Verfügung steht. Analog kann für die zugeschriebenen Ziele argumentiert werden, dass der menschliche Akteur dem nicht-menschlichen Akteur bestimmte Ziele zuschreibt, die in der Handlungslogik von  $A_{NH}$  gar keine Berücksichtigung finden.
- $SEU_{NH \rightarrow H} \neq SEU_H$ : Eine solche Modelldifferenz zielt insbesondere auf eine inadäquate Konzeption der Mensch-Maschine-Interaktion hin. Dem Interaktionspartner autonomer Technik werden dabei bestimmte Beweggründe unterstellt, die so gar nicht vorhanden sind. Ein unangemessenes Nutzermodell, das seitens  $A_{NH}$  unterstellt wird, kann dazu führen, dass bestimmte Beweggründe und Aktionen erwartet werden, die  $A_H$  in der Form nicht antizipiert. Typischerweise treten solche Missverständnisse besonders deutlich auf der Ebene der Benutzerschnittstelle zu Tage. Das Design einer Benutzerschnittstelle unterstellt situations- bzw. zustandsabhängig bestimmte Interaktionsmuster, die von  $A_H$  ausgehen. Bleiben diese Interaktionsmuster aus, weil das zugeschriebene Handlungsmodell mit dem von  $A_H$  realisierten nicht übereinstimmt, kommt es schlimmstenfalls dazu, dass eine gemeinsam zu erfüllende Aufgabe nicht erfüllt werden kann.

Wie gezeigt, können die zwei mentalen Modelltypen von Norman den entsprechenden SEU-Modellen bzw. zugeschriebenen SEU-Modellen zugeordnet werden. Ein Äquivalent zum System Image findet sich beim HMSE nicht. Implizit ist ein System Image in den Situationsdefinitionen enthalten. Mit Bezug auf die HMSE-Grafik auf Seite 196 beinhalten die Pfeile zur Situationslogik (1b) und (2b) implizit eine Vorstellung darüber, wie die äußeren Bedingungen der Situation und die wahrgenommene Interaktion die subjektive Definition der Situation von  $A_H$  beeinflussen.

### 8.3 Erklärungsanspruch und methodologischer Nutzen des HMSE

In Anschluss an die konzeptionellen Betrachtungen zum HMSE und der Definition unterschiedlicher SEU-Modelle bzw. zugeschriebener SEU-Modelle stellt sich die Frage nach der Erklärungsstärke und dem methodologischen Nutzen des HMSE. Weiter wird kurz geklärt, inwieweit Aspekte des HMSE, insbesondere der Ansatz der Modellzuschreibungen, Einzug in den empirisch-experimentellen Teil der vorliegenden Arbeit finden. Weiterhin wird geklärt, inwieweit sich aus den konzeptionellen Überlegungen zum HMSE forschungsleitende Annahmen für die Behandlung des Untersuchungsgegenstandes Vertrauen in autonome Technik ableiten lassen.

In den obigen Ausführungen klang bereits an, dass eine implementatorische Kontrolle über  $SEU_{NH}$  im Rahmen eines experimentellen Settings genutzt werden kann, um implementierte Modelle und zugeschriebene Modelle systematisch miteinander in Beziehung zu setzen. Eine solche implementatorische Kontrolle macht sich das Experiment SIMHYBS (vgl. Kapitel 11) zu Nutze. In Vorgriff auf das Kapitel kann gesagt werden, dass dort eine interaktive agentenbasierte Fahrsimulation auf Basis eines SEU-Modells implementiert wird. Durch Kenntnis der für  $A_{NH}$  relevanten Ziele und Zielbewertungen kann ein Abgleich zwischen dem implementierten Modell und dem von den Probanden zugeschriebenen SEU-Modell erfolgen. Aus festgestellten Differenzen lassen sich ggf. Ursachen und Gründe feststellen, die das Vertrauen in dieses System beeinflussen.

Analog lässt sich auch das Befragungsexperiment zum Entscheidungshelfer MensaAssist (Kapitel 10) auffassen. Dort wird im Rahmen eines webbasierten Fragebogens ein fiktiver Entscheidungshelfer von den Befragten hinsichtlich der Erfüllung / Nicht-Erfüllung bestimmter Ziele bewertet.

In beiden Fällen besteht der methodologische Trick darin, über das Experimentaldesign die jeweiligen SEU-Modellstrukturen weitgehend vorzugeben. Die Analysen können dann auf einen einheitlichen Modellierungsrahmen zurückgreifen und vergleichende Schlüsse daraus ableiten. Das Konzept der Modellzuschreibungen wird insoweit berücksichtigt, dass Zuschreibungen erhoben werden und mit den jeweiligen Implementierungen in Beziehung gesetzt werden. Auf diese Weise wird es möglich, die Zuschreibungsprozesse zu beobachten und zu analysieren, die sich bei der Interaktion von Mensch und autonomer Technik ergeben. Zum anderen lässt sich auf diese Weise nachvollziehen, inwiefern das Maß an Handlungsträgerschaft, das autonomer Technik von Menschen zugeschrieben wird, mit dem technisch implementierten übereinstimmt bzw. wie stark es davon abweicht. Die Verwendung einer intentionalistischen Handlungstheorie, der SEU-Theorie, auf der Mikro-Ebene als Grundlage für das HMSE hat den Vorteil, dass Handlungsträgerschaft nicht bloß auf Aktivitätsebene untersucht werden kann, wie dies bei Rammert und Schulz-Schaeffer (2002a) der

Fall ist. Ausgehend von den Modellbestandteilen der SEU-Theorie lassen sich Ziele und Handlungen getrennt voneinander beobachten und vermessen.

Im Sinne des Kroneberg'schen Modulierbarkeitskriteriums für Handlungstheorien (vgl. Kroneberg 2009: 240) ist es für eine Analyse des Zusammenspiels von Mensch und autonomer Technik nicht zwangsläufig notwendig, alle vier im Kontext des HMSE relevanten SEU-Modelle bzw. -Modellzuschreibungen zu berücksichtigen. Je nach Untersuchungsfrage können unterschiedliche Modelle und Modellzuschreibungen miteinander verglichen werden. Hinzu kommt, dass mit dem HMSE eine Art idealtypische soziologische bzw. handlungstheoretische Sicht auf Mensch-Maschine-Interaktion vorliegt. Ob das Modell in dieser Weise auf jede mögliche Interaktion von Mensch und autonomer Technik anwendbar ist, kann ex ante nicht behauptet werden. Vielmehr steht im Vordergrund, eine hybride Konstellation mit den Modellierungsmöglichkeiten des HMSE so angemessen wie möglich darzustellen, um damit Vertrauensproblemen auf den Grund zu gehen und diese besser zu verstehen. Ein Schlüssel zum Verständnis des Vertrauens in autonome Technik, so lässt sich als forschungsleitende Annahme (FLA 14) formulieren, ist die Berücksichtigung soziologischer bzw. handlungstheoretischer Charakteristika der Mensch-Maschine-Interaktion.

**FLA 14 (Zugeschriebene vs. wirksame SEU-Modelle)**

Der systematische Vergleich wirksamer und zugeschriebener SEU-Modelle trägt zu einem handlungstheoretischen und damit genuin soziologischen Verständnis von Vertrauen in autonome Technik bei.



**Teil III**

**Empirie**



# 9 Vorstellung der Experimente und Hypothesen

Zur besseren Strukturierung des empirischen Teils wird ein Hypothesenrahmen vorgestellt, an dem sich die Auswertungen der durchgeführten Experimente in den Kapiteln 10 und 11 orientieren werden.

In den vorherigen Kapiteln wurden jeweils forschungsleitende Annahmen (FLAs) abgeleitet, die drei Funktionen haben: Erstens haben diese einen heuristischen Erkenntniswert und umfassen Aspekte, die bei der Untersuchung von Vertrauen speziell in Bezug auf autonome Technik von Bedeutung sind. Auf diesen heuristischen Erkenntniswert wird im Schlusskapitel 12 eingegangen. Zweitens dienen sie als Vorüberlegungen zur Formulierung anwendungsbezogener Hypothesen für den folgenden empirischen Teil. Drittens leiten sie die Experimentaldesigns an, die erforderlich sind, um Vertrauen in autonome Technik im Rahmen von Laborexperimenten bzw. Befragungen zu untersuchen.

Im Idealfall stehen am Ende des empirischen Teils mehrere bestätigte (oder widerlegte) Hypothesen, die ihrerseits rückbezogen werden können auf die forschungsleitenden Annahmen.

## 9.1 Überblick über die bisherigen forschungsleitenden Annahmen

Um einen guten Überblick zu ermöglichen, sind im Folgenden noch einmal sämtliche forschungsleitenden Annahmen aufgelistet. In der ersten Spalte befindet sich das Kapitel, in dem die Annahme formuliert wurde. Die zweite Spalte enthält eine Kurzbeschreibung der Annahme inkl. der zugehörigen Nummer. Die letzte Spalte gibt die Seite an, auf der die Annahme ausformuliert und erklärt ist.

<i>Kapitel</i>	<i>Kurzbeschreibung der forschungsleitenden Annahme</i>	<i>Seite</i>
2	Multidimensionalität (FLA 1) . . . . .	36
2	Verhalten und Vertrauen (FLA 2) . . . . .	36
2	Zuschreibung (FLA 3) . . . . .	38
3	Laborexperimente (FLA 4) . . . . .	71
3	Interaktiver Prozess (FLA 5) . . . . .	71

3	Erwartungen (FLA 6) . . . . .	71
4	Gründe (FLA 7) . . . . .	95
4	Ausgangsvertrauen (FLA 8) . . . . .	96
5	Eigenaktivität / Gleichberechtigte Partner (FLA 9) . . . . .	107
5	Handlungswirksamkeit (FLA 10) . . . . .	136
6	MSE als modellierungstheoretische Leitidee (FLA 11) . . . . .	167
7	Mikrofundierung (FLA 12) . . . . .	186
8	Gemeinsame Aufgabe (FLA 13) . . . . .	191
8	Zugeschriebene vs. wirksame SEU-Modelle (FLA 14) . . . . .	209

## 9.2 Kurzzvorstellung der Experimente und zugehörigen Hypothesen

Zunächst wird auf die forschungsleitenden Annahmen eingegangen, die für *beide* Experimente relevant sind. Daran schließt sich eine Kurzbeschreibung der Experimente *MensaAssist* (vgl. ausführlich Kapitel 10) sowie *SIMHYBS* (Kapitel 11) an. Im Kontext der jeweiligen Experimente wird auf weitere relevante FLAs eingegangen sowie auf Hypothesen, die für die entsprechenden Experimente untersucht werden sollen. Eine Operationalisierung der Hypothesen erfolgt im zugehörigen Auswertungskapitel.

Gemäß (FLA 2) stehen Vertrauen und Verhalten in einem interdependenten Zusammenhang. Daran schließt die Überlegung an, Vertrauen in autonome Technik im Rahmen eines handlungstheoretisch fundierten Makro-Mikro-Makro-Paradigmas gemäß (FLA 11) zu untersuchen. Autonome Technik weist einen gewissen Grad an Eigenaktivität auf (FLA 9) und ist, zumindest auf methodologischer Ebene, als gleichberechtigter, sich (teil-)autonom verhaltener Partner zu berücksichtigen.

Vertrauen wird in beiden Experimenten anhand einer ins Deutsche übersetzten Vertrauensskala (Jian u. a. 2000) konzeptualisiert. Die Vertrauensskala enthält verschiedene Items, die der Multidimensionalität von Vertrauen (FLA 1) gerecht werden. Es werden gerade solche Vertrauensaspekte formalisiert, die im Zusammenhang mit autonomer Technik relevant sind (vgl. Abschnitt 10.3.2). Wenngleich im Sinne einer symmetrischen Sichtweise (vgl. Abschnitt 5.1) auch autonome Technik als Treugeber agieren und menschlichen Akteuren Vertrauen zuschreiben könnte, so wird Vertrauen bei beiden Experimenten doch als gerichtete Zuschreibung, ausgehend von dem menschlichen Akteur und bezogen auf das Verhalten und die Gründe des Verhaltens der autonomen Technik, aufgefasst (FLA 14).

Beiden Experimenten ist zudem gemein, dass sie Vertrauen im Kontext von Laborexperimenten gemäß (FLA 4) untersuchen. Bei dem Experiment *SIMHYBS* handelt es sich um ein Laborexperiment im engeren Sinne, das unter Aufsicht eines Versuchs-

leiters durchgeführt wird. MensaAssist kann als Laborexperiment im weiteren Sinne aufgefasst werden. Die Untersuchung ist umgesetzt mit einem Online-Befragungstool und Szenarien, die so angelegt sind, dass die Probanden davon ausgehen, mit einem Entscheidungshelfer zu interagieren. Auch dort können die Bedingungen, insofern als die durchgespielten Runden für alle Probanden identisch sind, als *kontrolliert* angesehen werden.

Schließlich setzen beide Experimente die in (FLA 13) formulierte Annahme um, dass sich das Zusammenhandeln von Mensch und autonomer Technik im Kontext einer gemeinsam zu erfüllenden Aufgabe abspielt. Im Fall von MensaAssist besteht die gemeinsame Aufgabe darin, ein optimales Essen unter gegebenen Randbedingungen zu finden. Bei SIMHYBS besteht die Aufgabe darin, eine spielerisch umgesetzte Fahr-situation, die aus definierten Unteraufgaben (Geschwindigkeitsbegrenzung einhalten, Unfälle vermeiden etc.) besteht, zu absolvieren. Gründe bzw. die Unterstellung von Erwartungen und Handlungsmustern (FLA 6) sind in beiden Experimenten berücksichtigt und untersuchbar, da ihre modellierungstheoretische Grundlage, das HMSE bzw. MSE (vgl. Kapitel 6.2, 8.1), Erwartungen ebenso wie Gründe (FLA 7) konstitutiv berücksichtigt.

### 9.2.1 MensaAssist

Das Experiment MensaAssist simuliert mit Hilfe eines Online-Umfrage-Tools insgesamt sechs Mensa-Tage, bei denen die Befragten von dem automatischen Entscheidungshelfer bei der Essenswahl unterstützt werden. An jedem der Tage wird den Befragten eine Auswahl mit bewerteten Gerichten präsentiert. Gemäß ihrer eigenen Vorlieben müssen sich die Probanden für eines der Gerichte entscheiden. MensaAssist empfiehlt daraufhin ein Essen, das am ehesten dem vorgewählten Essentyp entspricht, den die Probanden zu Anfang wählen konnten. Die Probanden haben die Möglichkeit, die Entscheidung von MensaAssist zu übernehmen oder ihre eigene Entscheidung beizubehalten. Nach insgesamt sechs Runden bzw. Mensa-Tagen wird nach dem Vertrauen in MensaAssist gefragt.

Im Gegensatz zum zweiten Experiment SIMHYBS wird Vertrauen im ersten Experiment nicht interaktiv gemessen, sondern als eine Art Gesamtbewertung am Ende aller Experimente. Unter Bezug auf die definierten Modellzuschreibungen und -differenzen in Abschnitt 8.2.2 wird mit dem MensaAssist die Auswirkung von

$$SEU_{NH \rightarrow H} \neq SEU_H$$

auf Vertrauen untersucht. Durch die Vorgabe bestimmter Essensempfehlungen schreibt der nicht-menschliche Akteur (NH) MensaAssist dem menschlichen Akteur (H) ein bestimmtes SEU-Modell  $SEU_{NH \rightarrow H}$  zu. Die Hypothesen, die sich auf Modell-differenzen beziehen, rekurren dabei auf diesen Zusammenhang.

Wie erläutert, haben die Probanden zu Anfang die Möglichkeit, aus vorgegebenen Essenstypen zu wählen, an denen MensaAssist seine Entscheidungen orientieren soll. Daher ist zunächst zu prüfen, ob die vorausgewählten Typen (vgl. ausführlich 10.3.3) mit den am Ende des Fragebogens abgefragten Zielbewertungen übereinstimmen und sich in einer Clusterung nach U-Werten wiederfinden lassen. Dieser Zusammenhang ist in These (MA 1) formuliert.

#### **MA 1 (Essenstyp vs. Essensvorlieben)**

Die zu Anfang vorgenommene Auswahl eines bestimmten Essenstyps findet sich in einer automatischen Clusterung der Essensvorlieben (U-Werte) wieder.

Findet sich dieser Zusammenhang nicht in dem Datenmaterial, könnte dieses Auswirkungen auf das Vertrauen in MensaAssist haben und hinsichtlich der gemessenen Vertrauenswerte eine größere Streuung bedeuten.

Im Hinblick auf die SEU-Modell-differenzen ist zu prüfen, ob ein negativer Zusammenhang besteht zwischen kumulierten (also für alle Tage aufsummierten) Modell-differenzen und Vertrauen. Diese Zusammenhang ist in Hypothese (MA 2) formuliert.

#### **MA 2 (Kumulierte SEU-Differenzen und Vertrauen)**

Es besteht ein negativer Zusammenhang zwischen kumulierten SEU-Differenzen und Vertrauen.

In Bezug auf die *Gründe* für geringes oder starkes Vertrauen ist zu prüfen, ob sich die SEU-Modellausprägungen (s. Abschnitt 8.2.1), insbesondere im Hinblick auf die durchschnittlichen Zielbewertungen voneinander unterscheiden (MA 3).

#### **MA 3 (Vertrauen und SEU-Modellausprägung)**

Stark und gering Vertrauende unterscheiden sich hinsichtlich ihrer SEU-Modellausprägung.

Wie eingangs genannt erfolgt eine Operationalisierung der Hypothesen in den entsprechenden Auswertungskapiteln.

### **9.2.2 Simhybs**

Im Gegensatz zu MensaAssist bedient sich SIMHYBS gemäß (FLA 12) der Methode der mikrofundierten agentenbasierten Modellierung und Simulation. Es handelt sich dabei um eine interaktive Fahrsimulation, in der ein menschlicher Fahrer (H) mit einem nicht-menschlichen (NH) Assistenzsystem in unterschiedlichen Modi der

Handlungsverteilung über die Tastatur interagiert und sich auf einem Spielfeld mit weiteren Autofahrern bewegt.

Das Design des Experiments sieht eine mehrfache Messung von Vertrauenswerten in unterschiedlichen Modi der Handlungsverteilung vor. Damit eröffnet sich gemäß (FLA 5) eine interaktiv-prozessförmige Sichtweise auf Vertrauen. So wird es, wie in (FLA 8) formuliert, möglich, die Auswirkungen von anfänglichem Vertrauen bzw. von Vertrauen nach erstem Kontakt mit der Automation auf spätere Vertrauensmessungen zu ermitteln.

Im Gegensatz zu *MensaAssist* ist die relevante Zuschreibungsrichtung bei der Ermittlung von SEU-Modell differenzen (FLA 14) umgekehrt. Der nicht-menschliche Akteur NH ist gemäß SEU-Logik in Form eines softwareseitig implementierten Agenten bzw. Fahrerassistenzsystems umgesetzt. Demnach können Modell differenzen zwischen dem tatsächlich implementierten SEU-Modell  $SEU_{NH}$  und dem SEU-Modell, das dem nicht-menschlichen Akteur vom menschlichen Akteur zugeschrieben wird, also  $SEU_{H \rightarrow NH}$  ermittelt werden. Formal sind also die Auswirkungen von

$$SEU_{H \rightarrow NH} \neq SEU_{NH}$$

auf Vertrauen zu untersuchen. Im Gegensatz zum *MensaAssist*-Experiment können im Fall von *SIMHYBS* keine kumulierten SEU-Differenzen ermittelt werden, da die Abfolge des Zusammenhandelns so kurz ist, dass nach jedem steuernden Tastaturanschlag nach den Zielbewertungen und Situationseinschätzungen gefragt werden müsste. Stattdessen wird nach jedem Durchlauf gefragt, welche Ziele das Assistenzsystem in welcher Intensität verfolgt hat. Demnach ist zu untersuchen (SH 1), ob es einen positiven Zusammenhang zwischen Zielintensitäten und Vertrauen gibt, der gemäß (FLA 6) und (FLA 7) plausibel wäre.

### SH 1 (Vertrauen und zugeschriebene Zielintensitäten)

Vertrauen und zugeschriebene Zielintensitäten sind positiv korreliert.

Da in unterschiedlichen Modi der Handlungsverteilung experimentiert wird, ist zu prüfen, ob die Probanden mit bestimmten Modi der Handlungsverteilung besser oder schlechter umgehen können. In der Automationsliteratur (vgl. Abschnitt 4.3) wird ein enger Zusammenhang zwischen systembedingten Automationsfehlern und Vertrauen gesehen. Gemäß (FLA 2) und (FLA 3) ist Vertrauen u. a. als eine auf Verhalten bezogene Disposition aufzufassen. Mit (SH 2) ist folglich zu prüfen, wie sehr Vertrauen von subjektsspezifischen Zuschreibungen und Dispositionen abhängt und inwieweit von systemimmanenten Schwächen der Automation, oder, wie in der These formuliert, von einer Kombination aus beidem.

**SH 2 (Vertrauen system- und subjektspezifisch)**

Vertrauen hängt ab von systembedingten Schwächen und subjektspezifischen Zuschreibungen und Dispositionen.

Durch die mehrfache Messung von Vertrauen während des Experiments ist es möglich, unterschiedliche Vertrauensentwicklungen in den Blick zu nehmen. Mit (FLA 8) wurde bereits formuliert, dass ein Zusammenhang zwischen dem anfänglichen Vertrauen und späteren Vertrauensmessungen vermutet wird. Demnach wird mit (SH 3) die These formuliert, dass Vertrauensmessungen jeweils von den vorangegangenen Messungen beeinflusst werden.

**SH 3 (Einfluss vorheriger Vertrauensmessungen)**

Vertrauensmessungen werden von den vorangegangenen Messungen beeinflusst.

In nochmaligem Bezug zu SEU-Modellen lässt sich untersuchen, ob sich Unterschiede in Vertrauenstrajektorien (also unterschiedliche Entwicklungspfade des Vertrauens), die durch die mehrfache Messung von Vertrauen ermittelt werden können, ebenso in trajektorienspezifisch zugeschriebenen SEU-Modellen wiederfinden. Dieser Zusammenhang ist in (SH 4) beschrieben und im Verlaufe der Auswertung zu prüfen.

**SH 4 (Vertrauenstrajektorien und SEU-Modelle)**

Es lassen sich typische Vertrauenstrajektorien ermitteln und anhand zugeschriebener SEU-Modelle voneinander unterscheiden.

Diese Hypothese schließt an (SH 2) an, da sich in Vertrauenstrajektorien sowohl subjektspezifische, etwa eine unterschiedlich starke Reaktion auf Automationsprobleme, wie auch systemimmanente Eigenschaften, z. B. Automationschwächen, finden lassen.

Abschließend lässt sich durch die mehrfache Messung von Vertrauen noch eine weitere These untersuchen, die an Beobachtungen aus der Spieltheorie anschließt (vgl. Abschnitt 3.2.5), dass sich nämlich Vertrauen bei Nutzung, also bei bewusster Entscheidung für eine Vertrauenshandlung, mehrt. Mit (SH 5) wird zu prüfen sein, ob dieser Zusammenhang auch für Vertrauen in autonome Technik feststellbar ist.

**SH 5 (Nutzung von Vertrauen)**

Vertrauen mehrt sich bei Nutzung.

## Notationen

Zwecks einer besseren Lesbarkeit wird jeweils am Rand vermerkt, dass eine bestimmte empirische These im zugehörigen Textabschnitt behandelt wird. Sofern sich Indizien zur Bestätigung oder Widerlegung der These ergeben, wird dies ebenfalls mit einem entsprechenden Symbol gekennzeichnet.

(MA 1) ?	Bedeutet, dass ab dem so gekennzeichneten Textabschnitt die These (MA 1) untersucht wird.
(MA 1) ✗	Bedeutet, dass sich aus der vorangegangenen Auswertung folgern lässt, dass die These (MA 1) falsch ist.
(MA 2) ✓	Bedeutet, dass sich aus der vorangegangenen Auswertung folgern lässt, dass die These (MA 2) korrekt ist.
(SH 5) – ? –	Bedeutet, dass sich aus der vorangegangenen Auswertung weder folgern lässt, dass die These (SH 5) korrekt, noch dass sie falsch ist.



## 10 Experiment: MensaAssist<sup>1</sup>

In diesem Abschnitt wird ein fiktiver<sup>2</sup> Mensa-Assistent, genannt MensaAssist, vorgestellt: ein (teil-)autonomer Entscheidungshelfer für die Essenswahlentscheidung in der Universitätsmensa, der Mensa-Besucher bei der Wahl eines optimalen Essens gemäß ihrer individuellen Präferenzen unterstützten möchte. Mit Hilfe von MensaAssist, der sich gegenüber den Befragten als mehr oder weniger „intelligente“ autonome Technik präsentiert sollen vornehmlich die Hypothesen 1 bis 3 untersucht werden. Auf dem Auswertungsweg werden sich darüber hinaus weitere Teilergebnisse ergeben, die nicht unmittelbar über das Hypothesensystem abgebildet sind.

Da in Anschluss an (FLA 3) und (FLA 14) die Untersuchung von Zuschreibungen und zugeschriebenen Gründen im Vordergrund steht, ist es unerheblich, ob sich MensaAssist als intelligent im algorithmischen Sinne verhält oder ob der Entscheidungshelfer nur vorgibt, nach intelligenter Entscheidungslogik zu agieren. Im Vordergrund steht gemäß (FLA 13) eine gemeinsam zu erfüllende Aufgabe.

Als Vorstudie für die Konzeption der MensaAssist-Befragung und als prinzipielle Machbarkeitsstudie, die Essenswahl in der Mensa als subjektiv-rationale Wahlhandlung zu modellieren, dienen zwei von mir betreute Abschlussarbeiten: die Diplomarbeit von Zeynep Yilmaz (2012) und die Bachelor-Arbeit von Fabian Winkler (2011). Die Arbeit von Winkler hat das Problem der Essenswahl in der Mensa zunächst in die Variablen der SEU-Theorie übersetzt. Via Papierfragebogen wurden Personen direkt in der Mensa nach ihren Präferenzen und Essgewohnheiten befragt ( $N = 100$ ) und aus den Ergebnissen typisierte SEU-Modelle (vgl. Abschnitt 10.1.1) erstellt. Yilmaz nutzte in ihrer Arbeit das Online-Befragungstool LimeSurvey<sup>3</sup> ( $N = 163$ ) und berechnete ebenfalls typisierte SEU-Modelle.

Wenngleich die Annahme einer rational-abwägenden Wahlhandlung bei der Essenswahl in den Ernährungsverhaltenswissenschaften (behavioural nutrition) umstritten ist (Scheibehenne u. a. 2007) und stattdessen vorgeschlagen wird auf einfache Heuristiken, die nur ein Hauptziel berücksichtigen, zurückzugreifen (Schulte-Mecklenbeck u. a. 2013) erscheint die Modellannahme im vorliegenden Fall dennoch adäquat. In

---

<sup>1</sup>Ein herzlicher Dank gilt an dieser Stelle den Teilnehmerinnen und Teilnehmern des Master-Projektseminars des Fachgebiets Techniksoziologie im WS 11/12, die unter meiner Anleitung Experimente und Auswertungen durchgeführt haben sowie zwei Absolventen, die einige empirische Erhebungen zum Essenswahlproblem in der Mensa durchgeführt haben.

<sup>2</sup>Der vorgestellte Entscheidungshelfer MensaAssist ist fiktiv, weil dieser als Endgerät (noch) nicht existiert, sondern lediglich über eine elektronische Fragebogensoftware simuliert wird.

<sup>3</sup><http://www.limesurvey.org> (Abruf: 31. Januar 2014).

den zwei empirischen Vorstudien wird nicht nach konkreten Speisen gefragt, sondern abstrakt nach unterschiedlichen Gruppen von Speisen (z. B. Tagesgericht, Wahllessen), die in der Mensa zur Verfügung stehen. Zudem wurden die Einflussfaktoren (in der Sprache der SEU-Theorie: Ziele) auf insgesamt fünf beschränkt, im Gegensatz etwa zum umfänglichen Food Choice Questionnaire von Steptoe u. a. (1995). Bei der späteren Verwendung des Modells im Rahmen der Konzeption und Auswertung der MensaAssist-Befragung lassen sich unterschiedlich stark gewichtete Ziele berücksichtigen um z. B. für bestimmte Gruppen festzustellen, dass ein bestimmtes Ziel dominiert. Damit wird implizit die Kritik aus der Ernährungsverhaltenswissenschaft berücksichtigt. Weiterhin steht weniger ein möglichst realitätsgetreues Essenswahlentscheidungsmodell im Vordergrund, sondern die laborexperimentelle Frage, inwieweit sich Vertrauen in autonome Technik mit Hilfe der soziologischen Handlungstheorie erklären lässt.

In den folgenden zwei Abschnitten werden zunächst die Kernergebnisse der empirischen Vorstudien zusammengefasst und miteinander in Beziehung gesetzt. Daran schließt sich die Hauptuntersuchung mit dem fiktiven Entscheidungshelfer MensaAssist an.

## 10.1 Modellierungsgrundlage: Zwei empirische Erhebungen

Wie im Einführungstext beschrieben erfolgten zwei Erhebungen zur Essenswahl in der Mensa. Dabei wurde bewusst auf unterschiedliche Methoden (Papier vs. Online) zurückgegriffen, um eine hohe Modellvalidität und Stabilität des Modells zu erreichen. Es folgt zunächst eine Beschreibung der Papierbefragung sowie eine Zusammenfassung wichtiger Ergebnisse. Daran anschließend werden einige Aspekte der Online-Befragung ausgewertet und zentrale Ergebnisse notiert<sup>4</sup>.

### 10.1.1 Papierbefragung in der Mensa

Als Modellierungsvorbild dient die Mensa der Technischen Universität Dortmund. Um das Modell möglichst einfach zu halten, wurden die Handlungsalternativen (bzw. Essensalternativen) auf vier Gerichte reduziert, die jeden Tag in der Mensa angeboten werden. Die angebotenen Gerichte lassen sich wie folgt charakterisieren:

- *Wahllessen*: Ein tägliches wechselndes Angebot eines Fleischgerichts mit meist drei Beilagen, ca. 2,00 €;
- *Vegetarisches Menü*: Ein täglich wechselndes vegetarisches Menü mit meist drei Beilagen, ca. 2,40 €;

---

<sup>4</sup>Die beiden folgenden Abschnitte rekurren zwar auf die Abschlussarbeiten von Winkler und Yilmaz, allerdings wurden lediglich die dort erhobenen Daten verwendet. Im Rahmen der vorliegenden Arbeit wurden die Daten neu ausgewertet.

- *Tagesgericht*: Ein besonders günstiges Essen mit festen Beilagen. ca. 1,50 €;
- *Aktionsteller*: Ein etwas teureres Sättigungsgericht, meist sehr fleischlastig und mit üppigen Sättigungsbeilagen (Pommes Frites, Kartoffelecken), ca. 3,30 €.

In den späteren Auswertungen hat sich gezeigt, dass insbesondere Essen, die durch trennscharfe Eigenschaften charakterisiert sind, zu besonders aussagekräftigen Ergebnissen führen.

Methodisch war die Befragung wie folgt aufgebaut: In qualitativen Fokusgruppen-Interviews wurde versucht, die für das SEU-Modell relevanten Variablen zu identifizieren. Als mögliche Ziele für die Wahl eines bestimmten Essens wurden schließlich „satt werden“, „Essen schnell erhalten“, „etwas Leckeressens“, „wenig bezahlen“ und „etwas Gesundes essen“ gemeinsam mit den befragten Personen identifiziert. Diese Ziele fanden zusammen mit den Handlungsalternativen Eingang in den endgültigen Fragebogen.

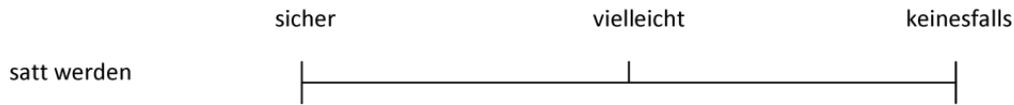
Im Folgenden werden für die Ziele Formulierungen in Langform, Kurzform oder als Kurzform in Kombination mit dem *u*-Präfix aus der SEU-Theorie synonym verwendet:

Langform	Kurzform	U-Form
„etwas Leckeressens“	lecker	<i>ulecker</i>
„etwas Gesundes essen“	gesund	<i>ugesund</i>
„Essen schnell erhalten“	schnell	<i>uschnell</i>
„satt werden“	satt	<i>usatt</i>
„wenig bezahlen“	zahlen	<i>uzahlen</i>

Der Fragebogen besteht dabei im Wesentlichen aus vier Teilen: Im ersten Teil werden soziodemografische Faktoren (Alter, Geschlecht, Studiengang etc.) abgefragt, im zweiten Teil Zielbewertungen („Wie wichtig ist es Ihnen satt zu werden“, „wenig zu bezahlen?“ etc.), im dritten Teil eine Bewertung von Handlungsalternativen zu den jeweiligen Konsequenzen, im vierten Teil wird schließlich gefragt, welches Essen die Befragten in letzter Zeit am häufigsten gewählt haben. Methodisch interessant ist die Verwendung einer visuellen Analogskala (vgl. Reips und Funke 2008) zur Abfrage von Zielbewertungen bzgl. bestimmter Handlungsalternativen (s. Abbildung 10.1). Da es sich bei den P-Werten aus der SEU-Theorie um kontinuierliche Werte (Wahrscheinlichkeiten) handelt, erscheint diese Abfrageart adäquater als die Möglichkeit nur aus wenigen Wahrscheinlichkeiten (sehr wahrscheinlich, wahrscheinlich, etc.) auswählen zu können.

Die Verteilung der Fragebögen wurde davon abhängig gemacht, für welches Essen sich die Befragten am Tag der Befragung entschieden haben<sup>5</sup>. Bei insgesamt 100 Befragten ergaben sich für jedes Gericht auf dem Fragebogen also genau 25 befragte

<sup>5</sup>Dies war leicht möglich, da die Befragung in der Mensa stattfand.



**Abbildung 10.1:** Je Essen wurden den Befragten insgesamt fünf Ziele zur Bewertung vorgelegt. Die Ermittlung der Wahrscheinlichkeit erfolgte durch Ausmessen.

Personen. Die Fragebögen waren (für die Befragten nicht sichtbar) so markiert, dass bei der Auswertung bekannt war, welches Essen tatsächlich gewählt wurde. So war es möglich, die tatsächliche Essenswahl mit der Voraussage aus der SEU-Theorie zu vergleichen. Es stellt sich allerdings heraus, dass nicht der Abgleich mit der aktuellen Essenswahl, sondern der mit der häufigsten Essenswahl zu besonders passgenauen, in Hinblick auf Übereinstimmung von Theorie und Empirie, SEU-Modellen führte. In 52% der Fälle stimmte die Angabe der Befragten zur häufigsten Essenswahl mit der Vorhersage aus der direkten Anwendung<sup>6</sup> der SEU-Theorie überein.

Die direkte Anwendung der SEU-Theorie auf primär erhobene Umfragedaten ist in der Literatur umstritten (vgl. Friedrichs u. a. 1993; Braun und Franzen 1995: 235; Opp 1998). Während der Auswertung der Umfragedaten stellte sich heraus, dass die Berechnung typisierter SEU-Modelle aussagekräftigere Ergebnisse ergab.

### Exkurs: Typisierte SEU-Modelle

Wie Abschnitt 6.2 zu entnehmen ist, können als subjektive Bestandteile eines SEU-Modells die Erwartungen  $\mathbf{P}$  und die Zielintensitäten  $\mathbf{U}$  angenommen werden. Würde man davon ausgehen, dass je Akteur ein jeweils unterschiedliches SEU-Modell prozessiert würde, ergäben sich für  $N$  Akteure insgesamt  $N$  verschiedene SEU-Modelle mit den jeweiligen subjektiven Komponenten  $(\mathbf{P}_1, \mathbf{U}_1), (\mathbf{P}_2, \mathbf{U}_2), \dots, (\mathbf{P}_N, \mathbf{U}_N)$ . Da die Soziologie praktisch nie einzelne Akteure in den Blick nimmt, sondern insbesondere Strukturphänomene erklären möchte, ist die Berechnung von SEU-Modellen je Akteur weder sinnvoll noch von besonderem soziologischen Interesse. Anzustreben ist hingegen eine Angabe von typisierten SEU-Modellen, die für eine bestimmte Gruppe von Akteuren hinsichtlich statistischer Qualitätsmaße (z. B. Varianz, Unterschiedlichkeit) besonders gut funktionieren.

Formal werden dabei die befragten  $N$ -Akteure in  $K$  ( $K < N$ ) Akteursgruppen unterteilt. Wie diese Gruppen gebildet werden, hängt vom empirischen Fall ab. Die Gruppenbildung kann entweder automatisch erfolgen, z. B. mit Cluster-Verfahren, oder auf Basis abgefragter Soziodemografika weiterer unabhängiger Variablen.

Für die jeweiligen Gruppen werden dann „durschnittliche“ SEU-Modelle berechnet. Angenommen in einer wie auch immer gebildeten Gruppe befinden sich  $m$  Ak-

<sup>6</sup>Direkte Anwendung bedeutet, dass gemäß der *pro Proband* erhobenen U- und P-Werte die Handlungsalternative mit dem höchsten SEU-Wert berechnet wird.

häuf. Wahl		satt	schnell	lecker	zahlen	gesund	max. SEU
<i>Vegetarisch</i>	$p_{veg, \_}$	0.7352	0.8637	0.7504	0.6107	0.8193	Vegetarisch
$N = 27$	$u_{\_}$	2.44	1.22	2.48	1.59	1.96	(5.687)
<i>Wahlessen</i>	$p_{wahl, \_}$	0.7532	0.5528	0.5785	0.7247	0.4434	Wahlessen
$N = 47$	$u_{\_}$	2.28	1.09	2.06	1.43	0.91	(4.5415)
<i>Aktionsteller</i>	$p_{akt, \_}$	0.7915	0.6785	0.7915	0.5823	0.4500	Aktionsteller
$N = 13$	$u_{\_}$	2.38	2.00	2.46	1.54	1.00	(6.0887)
<i>Tagesgericht</i>	$p_{tag, \_}$	0.7831	0.7138	0.6115	0.9315	0.5500	Tagesgericht
$N = 13$	$u_{\_}$	2.23	1.23	2.09	2.31	1.85	(4.9916)

**Tabelle 10.1:** Durchschnittliche SEU-Modelle für unabhängige Variable häufigste Wahl. (Für sämtliche Werte s. Anhang A.3.)

teure. Weiter sei angenommen, dass  $\mathbf{P}_i$  und  $\mathbf{U}_i$  die je Akteur via Fragebogen erhobenen einzelnen subjektiven Werte für das SEU-Modell bezeichnen. Dann lässt sich ein gruppenspezifisches SEU-Modell berechnen:

$$\mathbf{U}_{gruppe} = \frac{1}{m} \sum_{i=1}^m \mathbf{U}_i \quad \mathbf{P}_{gruppe} = \frac{1}{m} \sum_{i=1}^m \mathbf{P}_i$$

### Kurzzusammenfassung der Ergebnisse

Bei der Kurzzusammenfassung der Ergebnisse aus der ersten Mensa-Befragung, die auf Papier stattfand, wird auf die oben erläuterte Berechnung durchschnittlicher SEU-Modelle zurückgegriffen. Im vorliegenden Fall wurde die Angabe der häufigsten Essenswahl als Typisierungskriterium verwendet.

Die Ergebnisse sind Tabelle 10.1 zu entnehmen. Dabei wurden aus Gründen der besseren Übersichtlichkeit nur die P-Werte für die Handlungsalternative mit dem höchsten SEU-Wert angegeben.

Bemerkenswert ist das Ergebnis, dass für die durchschnittlichen SEU-Modelle, die nach dem Kriterium häufigste Essenswahl berechnet wurden, in allen vier Fällen die Vorhersage aus der Theorie mit der Empirie übereinstimmt. So ergibt sich z. B. für diejenigen Befragten, die als häufigste Wahl das vegetarische Essen angegeben haben, für ebendieses auch der höchste SEU-Wert (für diese Gruppe 5.687). Eine so hohe Übereinstimmung spricht für die Validität des Modells.

Trotz dieser Übereinstimmung stellt sich die Frage, ob die Eigenschaften der so gebildeten Gruppen überhaupt eine Aussagekraft im Vergleich zu den Durchschnittsmodellen der jeweiligen Gegengruppe (die aus allen Anderen besteht) haben.

Um die aus der unabhängigen Variable „häufigste Wahl“ abgeleiteten Gruppen über die jeweiligen P- und U-Werte näher zu charakterisieren, wurden zahlreiche t-Tests

durchgeführt. So haben beispielsweise 27 Befragte angegeben, dass sie am häufigsten das vegetarische Menü wählen. Damit ergibt sich eine Fallzahl von  $N_{veg} = 27$  und  $N_{\neg veg} = N - N_{veg} = 73$ .<sup>7</sup> Tabelle 10.2 gibt in vereinfachter Form an, für welche P- und U-Werte das Unterscheidungskriterium häufigste Wahl einen signifikanten Unterschied ausmacht.

An der abstrahierten Vergleichstabelle ist gut abzulesen, dass diejenigen, die ein bestimmtes Essen besonders häufig wählen, ebendieses hinsichtlich der abgefragten Ziele besonders gut bewerten (vgl. auch Fußnoten bei der Tabelle zu den Zelleinträgen). So bewerten etwa die häufig vegetarisch Essenden ihr „eigenes“ Essen hinsichtlich der Ziele satt zu werden, das Essen schnell zu erhalten, etwas Leckeres zu essen und wenig zu bezahlen besser, als die restlichen Befragten das vegetarische Essen bewerten. Von denjenigen, die das Tagesgericht am häufigsten essen, werden die Ziele satt zu werden und wenig zu bezahlen für das eigene Essen besser bewertet als von den restlichen Befragten, die ein anderes Essen besonders häufig essen.

Eine andere Situation ergibt sich beim Wahlessen: Die Befragten, die am häufigsten das Wahlessen wählen, zeichnen sich besonders dadurch aus, dass sie Eigenschaften der jeweils anderen Essen (Vegetarisch, Aktionsteller und Tagesgericht) signifikant schlechter bewerten als die anderen Befragten.

Grundsätzlich scheint die Modellierung der Essenswahlentscheidung als SEU-Modell praktikabel. Die Essen  $x$ , die von den Befragten am häufigsten gewählt werden, führen für die jeweiligen  $p_{x,j}$ -Werte zu trennscharfen Bewertungen. Die nach dem Kriterium der häufigsten Essenswahl gruppierten Befragten unterscheiden sich bezüglich ihrer Zielintensitäten nur geringfügig. Lediglich für die Gruppe der am häufigsten das Tagesgericht Wählenden ist das Ziel wenig zu bezahlen wichtiger als für alle anderen Befragten. Diejenigen, die am häufigsten den Aktionsteller wählten, war es wichtiger als allen anderen Gruppen, ihr Essen schnell zu erhalten.

---

<sup>7</sup>Das Zeichen  $\neg$  steht für die logische Negation.  $N_{\neg veg}$  gibt somit die Fallzahl derjenigen an, die nicht das vegetarische Menü am häufigsten wählen.

Variable	häufigste Wahl			
	Vegetarisch	Wahlessen	Aktionsteller	Tagesgericht
$P_{veg,-}$	satt	+ <sup>1</sup>		-
	schnell	+		
	lecker	+	-	
	zahlen	+	-	
$P_{akt,-}$	lecker			+
	zahlen		- <sup>2</sup>	
$P_{tag,-}$	satt	+		+
	schnell		-	
	zahlen			+
	gesund		-	
$u_{-}$	schnell		+	
	zahlen			+

<sup>1</sup> Interpretation: Diejenigen, die am häufigsten das vegetarische Menü essen, ist der Mittelwert der Variable  $p_{veg,satt}$  (Geschätzte W'keit, dass die Auswahl des vegetarischen Essens zur Erfüllung des Ziels „satt“ führt) laut t-Test ( $\alpha = 0.95$ ) signifikant *höher* als für diejenigen, die *nicht* das vegetarische Menü am häufigsten essen.

<sup>2</sup> Interpretation: Befragte mit Wahlessen als häufigste Wahl schätzen das Ziel wenig zu bezahlen für den Aktionsteller schlechter ein als die restlichen Befragten.

**Tabelle 10.2:** Variablen für das Unterscheidungskriterium „häufigste Wahl“, die sich gemäß t-Test mit  $\alpha = 0.95$  signifikant vom Durchschnitt der jeweiligen Gegengruppe unterscheiden.

### 10.1.2 Online-Befragung zur Essenswahl in der Mensa

Die zweite Befragung zur Essenswahl in der Mensa wurde im Rahmen der Abschlussarbeit von Zeynep Yilmaz (2012) durchgeführt. Yilmaz hat die selbe Thematik wie die Papier-Befragung von Winkler behandelt, sich dabei allerdings einer anderen Erhebungsform bedient, um zusätzlich die Validität der Modellierung zu überprüfen. Die Bitte an der Befragung teilzunehmen wurde über verschiedene Email-Verteiler und über Facebook gezielt an Studierende geschickt. Die Stichprobe umfasst insgesamt  $N = 163$  gültige Fragebögen.

Die meisten Übereinstimmungen von häufigster Wahl und Vorhersage aus der SEU-Theorie ergeben sich für das vegetarische Essen (66.7%). Für den Aktionsteller ergibt sich zumindest noch eine Übereinstimmung von 52.9%. Für das Tagesgericht (37.0%) und das Wahlessen (27.2%) fallen die Übereinstimmungen deutlich schlechter aus.

Der Grund für die schlechte Übereinstimmung bei den beiden zuletzt genannten Essen ist nicht in der mangelnden Validität des SEU-Modells zu suchen, sondern in der Problematik, dass das Wahlessen und das Tagesgericht hinsichtlich ihrer Eigenschaften stark variieren und deswegen nicht trennscharf beurteilt werden können<sup>8</sup>.

Auch bei der Befragung von Yilmaz (2012) zeigt sich, dass die direkte Anwendung der SEU-Theorie auf einzelne Fälle problematisch ist. Betrachtet man, gemittelt über alle 163 Befragten, die sich aus der Befragung ergebende Präferenzreihenfolge, ergibt sich:

$$\underbrace{u_{satt} \geq u_{lecker}}_{\text{Hauptziele}} > \underbrace{u_{zahlen} \geq u_{schnell} \geq u_{gesund}}_{\text{Nebenziele}}$$

$$\underbrace{2.190 \geq 2.126} > \underbrace{1.040 \geq 0.902 \geq 0.810}$$

Erkennbar ist, dass sich zwei Typen von Zielen unterscheiden lassen: Die Ziele „satt werden“ und „etwas Leckerer essen“ haben deutlich (laut T-Test mit  $\alpha = 0.95$ ) stärkere Intensitäten als die übrigen drei Ziele. Die Präferenzreihenfolgen lassen sich nach dem schon zuvor verwendeten Gruppierungskriterium „häufigste Wahl“ weiter unterstrukturieren. Tabelle 10.3 listet durchschnittliche SEU-Modelle je Gruppe vergleichend auf. Zu erkennen ist zunächst die schon zuvor beschriebene qualitative Unterscheidung zwischen Haupt- und Nebenzielen. Interessant ist, dass sich die Präferenzreihenfolgen zwischen den Gruppen unterscheiden: So ist es denjenigen, die beispielsweise am häufigsten den Aktionsteller wählen, wichtiger satt zu werden als etwas Leckerer zu essen. Denjenigen dagegen, die am häufigsten das vegetarische Gericht wählen, ist es wichtiger, etwas Leckerer zu essen als satt zu werden. Die am häufigsten vegetarisch Essenden unterscheiden sich insbesondere bzgl. des Ziels etwas Gesundes zu essen deutlich von den anderen Gruppen.

Überraschend und aus Modellvaliditätsgesichtspunkten unschön ist die Tatsache, dass sich für diejenigen, die am häufigsten das Wahlessen wählen, der höchste SEU-Wert nicht beim Wahlessen, sondern beim Aktionsteller ergibt. Da die beiden SEU-Werte allerdings sehr nah zusammenliegen, können sie statistisch als gleich angenommen werden (ein t-Test bestätigt dies). Für die anderen drei Gerichte stimmen häufigste Wahl und das Essen mit dem größten vorhergesagten SEU-Wert hingegen überein.

<sup>8</sup>Dieser Aspekt wird bei der Konzeption von MensaAssist zu berücksichtigen sein.

häuf. Wahl		satt	schnell	lecker	zahlen	gesund	max. SEU
<i>Vegetarisch</i>	$p_{veg, \_}$	0.7507	0.7911	0.7078	0.6407	0.9593	Vegetarisch
$N = 27$	$u_{\_}$	2.22	0.50	2.39	1.00	1.78	(5.5708)
<i>Wahlessen</i>	$p_{wahl, \_}$	0.7709	0.5972	0.5815	0.7847	0.443	Wahlessen <sup>1</sup>
$N = 92$	$u_{\_}$	2.15	0.96	2.10	0.88	0.49	(4.3635)
<i>Aktionsteller</i>	$p_{akt, \_}$	0.7879	0.632	0.7047	0.4526	0.404	Aktionsteller
$N = 17$	$u_{\_}$	2.29	0.88	2.12	0.88	0.79	(4.8834)
<i>Tagesgericht</i>	$p_{tag, \_}$	0.6704	0.6178	0.642	0.8089	0.506	Tagesgericht
$N = 27$	$u_{\_}$	2.22	1.11	1.94	1.72	0.94	(5.2946)

<sup>1</sup> Der größte SEU-Wert ergab sich eigentlich für den Aktionsteller (4.3814). Da beide Werte jedoch nahe zusammenliegen wurde zur besseren Lesbarkeit auch hier  $p_{wahl, \_}$  angegeben.

**Tabelle 10.3:** Durchschnittliche SEU-Modelle für unabhängige Variable häufigste Wahl. (für sämtliche Werte s. A.4.1)

Abschließend soll nun, wie schon bei der Auswertung der anderen Befragung, gezielt nach solchen Variablen aus dem SEU-Modell gesucht werden, die die jeweils betrachtete Gruppe (Kriterium: häufigste Wahl) besonders gut charakterisieren können. Ein  $p_{x,y}$  bzw. ein  $u_z$ -Wert kann eine Gruppe besonders gut charakterisieren, wenn sich statistisch mittels t-Test nachweisen lässt, dass sich der Mittelwert für die Variable, eingeschränkt auf die betrachtete Gruppe, signifikant vom Mittelwert der jeweiligen Gegengruppe unterscheidet. Tabelle 10.4 listet das Ergebnis zusammenfassend auf.

Auch in der hier betrachteten Befragung zeichnen sich die am häufigsten vegetarisch essenden Befragten dadurch aus, dass sie sehr trennscharfe P- und U-Werte angegeben haben, die sie von den anderen Gruppen unterscheiden. Auffällig ist auch, dass das jeweils „eigene“<sup>9</sup> Essen besonders gut bewertet wird. So wird das eigene Essen für mindestens ein Ziel  $x$  (formal  $p_{eigen,x}$ ) besser bewertet als von der jeweiligen Gegengruppe. Besonders augenfällig ist dies für die Gruppe der am häufigsten den Aktionsteller Essenden: Der Aktionsteller wird von dieser Gruppe hinsichtlich der Ziele „lecker“, „zahlen“ und „gesund“ besser bewertet als von allen anderen Befragten. Konsistent erscheinen auch die Unterschiede zwischen den Gruppen hinsichtlich der Zielangaben: So bewerten diejenigen, die am häufigsten das günstige Tagesgericht wählen das Ziel „zahlen“ höher als alle anderen Befragten. Die am häufigsten vegetarisch Essenden bewerten erwartungsgemäß das Ziel „gesund“ höher als alle anderen Befragten.

<sup>9</sup>Wenn von dem „eigenen“ Essen gesprochen wird, dann steht dies abkürzend für diejenigen, die als häufigste Wahl das vorhergesagte Essen angegeben haben.

Variable		häufigste Wahl			
		Vegetarisch	Wahlessen	Aktionsteller	Tagesgericht
<b>P<sub>veg,</sub></b>	satt	+			-
	lecker	+			
<b>P<sub>wahl,</sub></b>	satt		+		
	lecker	-			
<b>P<sub>akt,</sub></b>	lecker	-	+	+	
	zahlen		-	+	
	gesund	-		+	
<b>P<sub>tag,</sub></b>	schnell	-	+		
	lecker	-			+
<b>u<sub></sub></b>	schnell	-			
	zahlen		-		+
	gesund	+	-		

**Tabelle 10.4:** Vergleich der SEU-Modelle für häufigste Wahl mit den Durchschnittsmodellen der jeweils anderen Gruppen. Für die Bedeutung von + und - vgl. Fußnoten bei Tabelle 10.2

### 10.1.3 Fazit aus beiden Befragungen und Implikationen für MensaAssist

Die hier vorgestellten Ergebnisse, die sich aus den Daten von Winkler (2011) und Yilmaz (2012) ableiten lassen, haben gezeigt, dass sich das Problem der Essenswahl in der Mensa als Entscheidung im Modellierungsparadigma der SEU-Theorie darstellen lässt. Die hier erneut vorgenommene Auswertung der Daten hat allerdings auch einige Modellierungsprobleme aufgedeckt: Durch das wechselnde Angebot an Essen in der Mensa variieren die tatsächlichen Gerichte, die z. B. unter dem allgemeinen Label „Wahlessen“ angeboten werden, sehr stark, so dass eine präzise Bewertung hinsichtlich der untersuchten Ziele schwierig ist. Statistisch besonders gutmütig (d. h. besonders trennscharfe Ergebnisse liefernd) sind solche Handlungsalternativen, die recht eindeutige Eigenschaften aufweisen, z. B. als besonders gesund oder als besonders günstig gelten. Ein systematischer Vergleich beider Befragungen ist nicht sinnvoll, da insbesondere bei der Winkler-Befragung gezielt Einfluss auf die Zusammensetzung der Stichprobe genommen wurde.

Für die Konzeption des im folgenden Abschnitts vorgestellten Mensa-Assistenten (MensaAssist) werden die Erkenntnisse aus den beiden Befragungen konstitutiv berücksichtigt. Insbesondere wird MensaAssist von der konkreten Essenswahl in der Mensa der TU Dortmund abstrahiert. Statt nach „Wahlessen“, „Aktionsteller“ etc. zu fragen, werden Eigenschaften wie Preis, Geschmack etc. angegeben. MensaAssist

schlägt dann ein bestimmtes Essen für den Probanden vor. Die befragten Personen sollen dabei die Leistungsfähigkeit von (und das Vertrauen in) MensaAssist bewerten.

## 10.2 Design von MensaAssist-Experiment und Fragebogenelementen

Der vollständige Fragebogen zum MensaAssist-Experiment ist in Anhang A.5.1 zu finden.<sup>10</sup>

Der Ablauf der unterschiedlichen Fragenblöcke, auf die im späteren Verlauf genauer eingegangen wird, lässt sich in der Reihenfolge ihres Auftretens in Kurzform wie folgt zusammenfassen:

1. Messung der Technikaffinität
2. Auswahl eines Essenstyps, der zum Probanden besonders gut passt: *Sparfuchs*, *Gesundheitsbewusster*, *Eiliger*
3. Sechs Mensatage: Wunschessen auswählen und Entscheidung von MensaAssist ggf. übernehmen
4. Angaben zur eigenen Entscheidungsgrundlage anhand individueller U-Werte
5. Zufriedenheit mit MensaAssist bewerten
6. Vertrauen in MensaAssist nach Jian u. a. (2000) bewerten
7. Abfrage von Soziodemografika

Die Beschreibung des Experiments orientiert sich an den Position des Auftretens der unterschiedliche Fragebatterien.

### 10.2.1 Elemente des Fragebogens

In diesem Abschnitt wird auf einige für die Auswertung besonders relevante Teile des Fragebogens eingegangen. Der vollständige Fragebogen findet sich, wie eingangs genannt, im Anhang.

---

<sup>10</sup>Der Fragebogen ist im Rahmen einer von mir durchgeführten Lehrveranstaltung entstanden. Da es bei der Lehrveranstaltung nicht ausschließlich um „Vertrauen in autonome Technik“ ging, enthält der Fragebogen auch solche Fragen, die hier nicht im Detail ausgewertet werden. Der Vollständigkeit halber wird im Anhang dennoch der komplette Fragebogen wiedergegeben.

## Typauswahl

Vor Beginn der Befragung zu dem simulierten Tagen müssen sich die Befragten für einen Typ entscheiden, der ihren eigenen Vorlieben an gutes Mensa-Essen möglichst nahe kommt. Dadurch soll die Aufmerksamkeit auf die Qualität der Entscheidungen von MensaAssist gelenkt werden. Ebenso dient die Typauswahl der Vorbereitung einer weiter hinten im Fragebogen befindlichen Frage. Dort wird gezielt nach der Entscheidungsgrundlage der Probanden für ihre Essenswahl gefragt wird, bei denen den unterschiedlichen Zielen feingranulare Relevanzwerte (U-Werte) zugewiesen werden können.

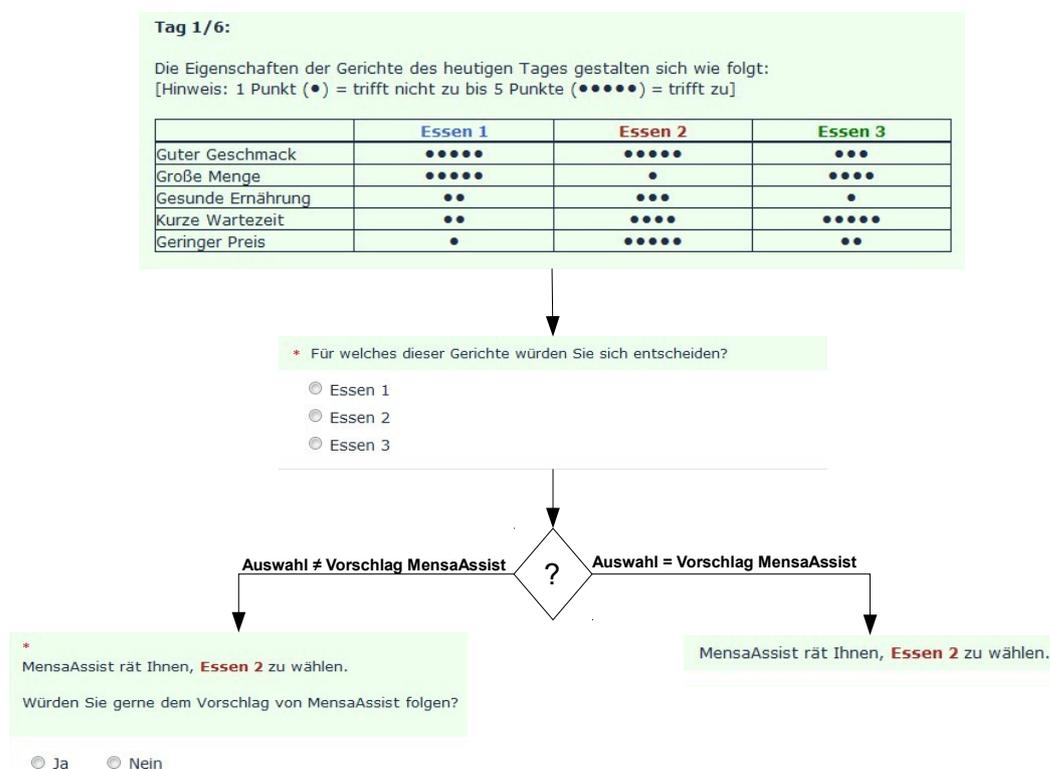
Die Beschreibung der Typen ist inspiriert von der Ergebnissen aus den Tabellen 10.1 und 10.3 sowie der auf Seite 228 angegebenen Präferenzreihenfolge, die sich in den vorherigen empirischen Befragungen zur Essenswahl in der Mensa gezeigt haben. Die vorgegebenen Typen sind als *Sparfuchs*, *Gesundheitsbewusster* und *Eiliger* benannt und werden im Fragebogen jeweils kurz charakterisiert (vgl. Tabelle 10.5). In der Tabelle ist zusätzlich eine mögliche Operationalisierung dieser Typen dargestellt.

Typ	Beschreibung	Zielintensitäten <sup>1</sup>												
<i>Sparfuchs</i>	„Sehr wichtig ist mir, dass ich Geld spare. Danach achte ich zuerst auf die Menge und dann auf den Geschmack. Ich habe im Prinzip keinen Stress (Wartezeit) und brauche absolut kein gesundes Essen (Gesundheit).“	<table style="display: inline-table; vertical-align: middle;"> <tr> <td style="text-align: center;">5</td> <td style="text-align: center;">4</td> </tr> <tr> <td style="text-align: center;">zahlen,</td> <td style="text-align: center;">satt,</td> </tr> <tr> <td style="text-align: center;">3</td> <td style="text-align: center;">1</td> </tr> <tr> <td style="text-align: center;">lecker,</td> <td style="text-align: center;">schnell,</td> </tr> <tr> <td style="text-align: center;">1</td> <td style="text-align: center;">1</td> </tr> <tr> <td style="text-align: center;">gesund</td> <td></td> </tr> </table>	5	4	zahlen,	satt,	3	1	lecker,	schnell,	1	1	gesund	
5	4													
zahlen,	satt,													
3	1													
lecker,	schnell,													
1	1													
gesund														
<i>Gesundheitsbewusster</i>	„Sehr wichtig ist mir, dass ich satt werde UND es gut schmeckt. Danach achte ich zuerst darauf Gesundes zu wählen und dann auf die Wartezeit. Ich habe im Prinzip keine Geldsorgen (Preis).“	<table style="display: inline-table; vertical-align: middle;"> <tr> <td style="text-align: center;">5</td> <td style="text-align: center;">5</td> </tr> <tr> <td style="text-align: center;">satt,</td> <td style="text-align: center;">lecker,</td> </tr> <tr> <td style="text-align: center;">3</td> <td style="text-align: center;">3</td> </tr> <tr> <td style="text-align: center;">gesund,</td> <td style="text-align: center;">schnell,</td> </tr> <tr> <td style="text-align: center;">1</td> <td style="text-align: center;">1</td> </tr> <tr> <td style="text-align: center;">zahlen</td> <td></td> </tr> </table>	5	5	satt,	lecker,	3	3	gesund,	schnell,	1	1	zahlen	
5	5													
satt,	lecker,													
3	3													
gesund,	schnell,													
1	1													
zahlen														
<i>Eiliger</i>	„Sehr wichtig ist mir, dass es gut schmeckt und dann, dass ich satt werde. Dann achte ich zuerst darauf, dass die Wartezeit gering ist und ich Geld spare. Ich verschwende keine Gedanken an meine Gesundheit.“	<table style="display: inline-table; vertical-align: middle;"> <tr> <td style="text-align: center;">5</td> <td style="text-align: center;">4</td> </tr> <tr> <td style="text-align: center;">lecker,</td> <td style="text-align: center;">satt,</td> </tr> <tr> <td style="text-align: center;">3</td> <td style="text-align: center;">2</td> </tr> <tr> <td style="text-align: center;">schnell,</td> <td style="text-align: center;">zahlen,</td> </tr> <tr> <td style="text-align: center;">1</td> <td style="text-align: center;">1</td> </tr> <tr> <td style="text-align: center;">gesund</td> <td></td> </tr> </table>	5	4	lecker,	satt,	3	2	schnell,	zahlen,	1	1	gesund	
5	4													
lecker,	satt,													
3	2													
schnell,	zahlen,													
1	1													
gesund														

<sup>1</sup> Für die Intensitäten bedeutet ein Wert von 5 „sehr wichtig“ und 1 „sehr unwichtig“

**Tabelle 10.5:** Beschreibung der Essenstypen, für die sich die Befragten am Anfang des Fragebogens entscheiden mussten

## Die Mensatage



**Abbildung 10.2:** Fragesequenz im Online-Fragebogen zu MensaAssist: Stimmt das gewählte Gericht nicht mit der Entscheidung von MensaAssist überein, wird der Befragte gefragt, ob er seine eigene Entscheidung revidieren möchte.

Insgesamt werden sechs Mensatage mit dem Fragebogen „durchgespielt“. Der Ablauf der Fragen je Tag ist in Abbildung 10.2 dargestellt (die Ziele heißen auf dem Fragebogen marginal anders, sind aber mit den modelltheoretisch berücksichtigten Zielen identisch). Zunächst werden den Befragten drei unterschiedliche Gerichte mit ihren jeweiligen Bewertungen präsentiert. Modelltheoretisch (und damit später für die Auswertung wichtig) handelt es sich um drei Handlungsalternativen mit ihren jeweils zugehörigen P-Werten. Die unterschiedlichen Essen sind mit Punkten bewertet, wobei ein Punkt (●) gleichbedeutend ist mit „Essen erfüllt die Eigenschaft nicht“ und fünf Punkte (●●●●●) heißt „Eigenschaft ist voll erfüllt“. In späteren Auswertungen werden den Essensbewertungen entsprechende P-Werte zugewiesen (vgl. Tabelle 10.6).

Wird ein Essen beispielsweise für das Ziel „Kurze Wartezeit“ mit fünf Punkten bewertet („ist voll erfüllt“) wird entsprechend der Wert  $p_{Essen,Wartezeit} = 1$  gesetzt. Damit wird das Ziel möglichst kurz zu warten bei der Wahl des entsprechenden Essens zu 100 % erfüllt.

$p_{i,j} =$	0	0.25	0.5	0.75	1
	•	••	•••	••••	•••••

**Tabelle 10.6:** MensaAssist – Operationalisierung von Essensbewertungen

### Abschließende Fragen / Soziodemografika

Abschließend enthält der Fragebogen einige Fragen zu Soziodemografika. Dort wurden das Alter, das Geschlecht, der höchste Bildungsabschluss sowie die derzeitige Beschäftigung abgefragt.

### 10.2.2 Arbeitsweise von MensaAssist und methodische Einschränkungen

Obwohl in der Einleitung zum Fragebogen MensaAssist als ein „noch in der Entwicklung befindliches Assistenzsystem, das Sie in Zukunft bei der für Sie passenden Essenswahl [...] unterstützen könnte“ beschrieben wird, ist das Verhalten innerhalb der Online-Fragebogen-Software statisch implementiert. Statisch bedeutet, dass MensaAssist unabhängig von der Typenauswahl, die zu Anfang der Befragung getroffen wird, für alle Befragten die gleichen Gerichte empfiehlt. Die Entscheidung, MensaAssist innerhalb des Fragebogens so umzusetzen, liegt forschungspragmatisch darin begründet, dass ein größeres Maß an Adaptivität in der Online-Fragebogensoftware LimeSurvey nur schwer umzusetzen gewesen wäre. Für die spätere Auswertung, bei der Differenzen zwischen MensaAssist und den Befragten ermittelt werden sollen, ist eine statische Funktionsweise zu bevorzugen, weil ein zusätzlicher Einfluss durch Adaptivität von MensaAssist eine Auswertung deutlich erschwert hätte bzw. eine Isolation einzelner Effekte nahezu unmöglich gemacht hätte.

Die Essenswahlentscheidung von MensaAssist ist so umgesetzt, dass sie jeweils unterschiedlich oft mit den Essenstypen aus der Vorauswahl übereinstimmt. So stimmt die Entscheidung in vier (von sechs) Fällen mit dem Typ „Der Eilige“ überein, in drei Fällen mit dem Typ „Der Gesundheitsbewusste“ und in zwei Fällen mit dem Typ „Der Sparfuchs“. Auch wenn die Essenstypen bei der späteren Auswertung keine Rolle mehr spielten (s.o.), so sollte durch die variable Zahl von Übereinstimmungen gewährleistet sein, dass die Qualität der Entscheidungen von MensaAssist von den Befragten, je nach ihren eigenen Präferenzen, die ex ante noch nicht bekannt sind, unterschiedlich bewertet wird. Die detaillierten Berechnungen zur Essenswahl von MensaAssist und der Übereinstimmung der Entscheidung mit den unterschiedlichen Essenstypen findet sich im Anhang A.5.2.

Aus der statischen Implementierung von MensaAssist ergeben sich methodische *Einschränkungen*, die eine Verallgemeinerbarkeit der abgeleiteten Erkenntnisse auf den Gegenstandsbereich der autonomen Technik möglicherweise einschränken. Dadurch, dass es sich nicht um tatsächlich (teil-)autonom agierende Technik handelt, kann

nicht automatisch gefolgert werden, dass sich die Erkenntnisse in ähnlicher Weise bei einem tatsächlich autonom agierenden Mensa-Assistenten wiederfinden würden. Im Rahmen der Befragung wurde zudem nicht danach gefragt, ob die Probanden bemerkt haben, dass MensaAssist ein statisches Verhalten aufweist und auf die Bewertungen der Probanden gar nicht reagiert.

### 10.2.3 Entscheidungsgrundlage der Befragten / Erfassung von U-Werten

0% 100%

**Ihre Entscheidungsgrundlage**

\* Wie wichtig sind Ihnen die untenstehenden Kriterien bei Ihrer Entscheidung für eines der Essen gewesen?

Kriterium	unwichtig	Wichtigkeit	wichtig
Guter Geschmack		40	
Große Menge		70	
Gesunde Ernährung		50	
Kurze Wartezeit		77	
Geringer Preis		17	

Später Fortfahren Weiter >> Umfrage verlassen und löschen

**Abbildung 10.3:** Die Abfrage der Zielintensitäten erfolgt mit einem frei einstellbaren Slider mit möglichen Werten zwischen 1 und 100. Diese Art der Abfrage entspricht visuellen Analogskalen aus Papierfragebögen.

Zur Berechnung von SEU-Differenzen ist es erforderlich, die Entscheidungsgrundlage der Befragten in Erfahrung zu bringen. Die Abfrage erfolgt am Ende des Experiments, weil davon auszugehen ist, dass die Befragten während der sechs simulierten Mensa-Tage entsprechend oft über ihre Essenswahlentscheidung reflektiert haben und sich ihrer Zielintensitäten, zumindest einigermaßen, bewusst geworden sind. Abbildung 10.3 zeigt die Umsetzung dieser Abfrage im Online-Fragebogen. Für jedes Ziel können mit einem Slider Werte von 0 bis 100 frei eingestellt werden. Diese Art der Abfrage hat den Vorteil, dass eine Vielzahl von Auswahlmöglichkeiten zur Verfügung steht und die resultierenden Werte in der Auswertung als kontinuierlich behandelt werden können.

### 10.2.4 Vertrauensmessung

Zur Messung von Vertrauen wird auf die Vertrauensskala von Jian u. a. (2000) zurückgegriffen (s. Tabelle 10.7 auf Seite 236), die in der Automationsforschung eine breite Anwendung findet (vgl. z. B. Bisantz und Seong 2001; Dzindolet u. a. 2003;

Nr.	Englische Originalformulierung von Jian u. a. (2000)	Deutsche Übersetzung
1	The system is deceptive. <i>Ehrlichkeitsvertrauen, Gutmütigkeitsvertrauen</i>	MensaAssist leitet mich fehl.
2	I am suspicious of the system's intent, action, or output. <i>Intentionales Vertrauen, Verhaltensvertrauen</i>	Ich bin misstrauisch, was die Empfehlungen von MensaAssist angeht.
3	I am wary of the system. <i>Kompetenzvertrauen</i>	Ich stehe MensaAssist insgesamt skeptisch gegenüber.
4	I am confident in the system. <i>Kompetenzvertrauen</i>	Insgesamt bin ich überzeugt von MensaAssist.
5	The system provides security. <i>Konditionalvertrauen</i>	MensaAssist gibt mir Sicherheit (bzgl. meiner Essenswahl).
6	The system is reliable. <i>Verhaltensvertrauen, Kompetenzvertrauen</i>	MensaAssist ist zuverlässig.
7	I can trust the system. <i>Intentionales Vertrauen</i>	Ich kann MensaAssist vertrauen.
8	I am familiar with the system. <i>Intentionales Vertrauen, Verhaltensvertrauen</i>	Die Funktionsweise von MensaAssist ist mir vertraut.

**Tabelle 10.7:** Übersetzung der Items von Jian u. a. (2000: 69) zur Vertrauensmessung und Anpassung der Formulierungen auf MensaAssist sowie Zuordnung zu Vertrauensformen gemäß Abschnitt 2.2

Lees und Lee 2007). Sie wurde für die vorliegende Arbeit ins Deutsche übersetzt und leicht angepasst<sup>11</sup>.

Aufgrund der hohen Verbreitung und des äußerst aufwendigen Verfahrens, mit dem die Skala empirisch ermittelt wurde (vgl. Jian u. a. 2000: 55ff.), ist von einer hohen Reliabilität auszugehen. Zudem lassen sich die einzelnen Items mit verschiedenen Elementen des Vertrauens gemäß der Vertrauentypologie von Nooteboom (2002: 50, s.

<sup>11</sup>Es wurden die Fragen „The system behaves in an underhanded manner“, „The system's actions will have a harmful or injurious outcome“ und „The system is dependable“ weggelassen. Weder MensaAssist, noch das später vorgestellte System SIMHYBS können sich hinterhältig oder verletzend verhalten. Beide sind vielmehr als assistierende Systeme konzipiert. Die Experimente sind zudem zu kurz um langfristige „dependability“ zu messen zu können.

Zuordnungsvorschlag in Tabelle 10.7 / vgl. Abschnitt 2.2) in Verbindung bringen. Die Elemente des Vertrauens, auf die sich die Items der Skala beziehen, finden sich ebenso in den bisher formulierten forschungsleitenden Annahmen wieder, die sich auf die Vertrauenskonzepte aus den Kapitel 2, 3 und 4 beziehen: Vertrauen wird operationalisiert als multidimensionales Konzept (FLA 1). Die Formulierungen von Jian u. a. beziehen sich auf Zuschreibungen, die auf das Verhalten des Treuhänders gerichtet sind (FLA 3), im Fall von MensaAssist also das richtige Essen auszuwählen. Weiterhin richten sich einige Aspekte auf das Verhalten des Systems (FLA 2) und adressieren bestimmte Erwartungen an ebendieses (FLA 6). Letztlich geht auch das Verständnis für bestimmte Gründe des Systemverhaltens (FLA 7) in die Vertrauensmessung ein. Insgesamt zeigt sich also, dass die Skala von Jian u. a. zahlreiche Vertrauenselemente abfragt und inhaltlich gut geeignet ist, um Vertrauen in autonome Technik vor dem Hintergrund der formulierten forschungsleitenden Annahmen zu messen.

Jian u. a. (2000: 70) bemerken, dass mit der Vertrauensmetrik ebenso Misstrauen in Form von negativen Vertrauenswerten gemessen werden kann, da es sich bei Vertrauen und Misstrauen um je gegenteilige Konzepte handelt. Weiterhin haben die Autoren herausgefunden, dass Vertrauen unabhängig vom Typ der Vertrauensbeziehung (Mensch/Mensch, Mensch/Nicht-Mensch und Vertrauen im Allgemeinen) nahezu identisch zugeschrieben wird.

Eine Schwierigkeit für die Verwendung im Rahmen des Experiments MensaAssist bestand darin, die in englischer Sprache formulierten Aussagen ins Deutsche zu übersetzen. Eine Übersetzung muss besonders sorgfältig erfolgen, damit hohe Reliabilität der Vertrauensmessung, die für die englische Version der Skala nachgewiesen wurde, gewährleistet bleibt. Tabelle 10.7 stellt die Formulierungen aus der Originalskala den im Fragebogen verwendeten Formulierungen gegenüber.

Die Messung von Vertrauen dient als Referenzgröße, um diese später mit anderen Items oder zusammenfassenden Konstrukten (z. B. ausgewählten Elementen von SEU-Modellen bzw. Modelldifferenzen) zu korrelieren.

Getrennt von der Vertrauensmessung wurde auch die Zufriedenheit mit MensaAssist abgefragt. In der späteren Auswertung wird sich zeigen, ob zwischen der Messung der Zufriedenheit und dem Konstrukt „Vertrauen“ nach Jian u. a. ein Zusammenhang besteht.

## 10.3 Auswertung

In Abschnitt 10.3.1 erfolgt zunächst ein Überblick über die Stichprobe anhand relevanter Soziodemografika. Es schließt sich in Abschnitt 10.3.2 eine Erläuterung zur Vertrauensmessung an, die auf der etablierten Skala von Jian u. a. (2000) beruht. In 10.3.3 werden auf Basis der angegebenen U-Werte Essenstypcluster gebildet, um zu zeigen, dass U-Werte im Gegensatz zur Typvorauswahl eine höhere Aussagekraft

besitzen. In Abschnitt 10.3.4 werden kumulierte SEU-Differenzen formal definiert um diese dann in 10.3.5 mit Befragten unterschiedlicher Vertrauensstärke in Verbindung zu bringen um diese dann in Abschnitt 10.3.5 anhand ihrer U-Werte inhaltlich interpretieren zu können. An passender Stelle erfolgt ein Bezug auf das zuvor spezifizierte Hypothesensystem.

### 10.3.1 Überblick über die Stichprobe

Variable	N	Statistiken	
<i>Alter</i>	131	Mittelwert	25.351
		Standardabweichung	7.403
		Median	23
		Min/Max	14/57
		25%/50%/75%-Quartil	21/23/26
<i>Geschlecht</i>	145	51.0 % weiblich, 49.0 % männlich	
<i>Abschluss</i>	143	Abitur/Fachabitur	53.8 %
		FH/Universität	36.4 %
		Berufsausbildung/Lehre	5.6 %
		Sonstige	4.3 %
<i>Beruf</i>	143	Studierende	75.5 %
		Sonstige	24.5 %
<i>Studiengang</i>	98	Wirtschaftswissenschaften	78.6 %
		Informatik	8.2 %
		Sonstige	13.3 %

**Tabelle 10.8:** Soziodemografika zum Experiment MensaAssist

Ein zusammenfassender Überblick über die Stichprobe findet sich in Tabelle 10.8. Insgesamt haben 146 Befragte den Fragebogen bis zum Ende ausgefüllt und alle Pflichtfragen<sup>12</sup> beantwortet.

Die Stichprobe<sup>13</sup> umfasst vornehmlich jüngere Personen (Median 23, Durchschnittsalter 25.35) und überwiegend (75.5 %) Studierende. Ferner ist die Stichprobe nahezu geschlechterparitätisch (51 % weibliche Personen, 49 % männliche). Das Bildungsniveau der Befragten ist hoch – gut 90 % verfügen über ein Abitur oder Fachabitur oder einen Fachhochschul- oder Universitätsabschluss (auch Bachelor). Von denjenigen, die studieren, weist die Stichprobe einen starken Bias hinsichtlich des Studiengangs auf –

<sup>12</sup>Der Fragebogen enthielt auch optionale Angaben. So durften die Befragten bei den Soziodemografika auch auf nähere Angaben verzichten. Damit sollte die Akzeptanz erhöht und die Probanden bewegt werden, den Fragebogen trotzdem abzuschicken. Für die Variable Alter machten beispielsweise nur 131 der 146 Befragten eine Angabe.

<sup>13</sup>Wenn von „der Stichprobe“ die Rede ist müsste es eigentlich im Rahmen des soziodemografischen Teils heißen: Diejenigen, die bei dieser freiwilligen Frage eine Angabe gemacht haben.

78.6 % haben angegeben einen wirtschaftswissenschaftlichen Studiengang zu studieren.

Insgesamt kann die Stichprobe weder als repräsentativ für die deutsche Bevölkerung noch als repräsentativ für die Studierendenschaft der TU Dortmund angesehen werden. Für die Aussagekraft der folgenden Auswertung stellt dies jedoch kein größeres Problem dar, weil das Hauptanliegen der vorliegenden Arbeit vor allem darin besteht, Vertrauen in autonome Technik mit Möglichkeiten der soziologischen Handlungstheorie näher zu beschreiben und zu verstehen. Sofern sich Vertrauenskonstrukte in handlungstheoretischen Strukturen wiederfinden lassen, sollten diese einen hohen Allgemeinheitsgrad besitzen und weitgehend unabhängig vom Vorliegen bestimmter soziodemografischer Merkmale konstatierbar sein.

### 10.3.2 Vertrauensmessung

Wie bereits zuvor erläutert, wurde bei der Messung von Vertrauen auf eine leicht abgewandelte Fassung der Vertrauensskala von Jian u. a. (2000) zurückgegriffen.

In diesem Abschnitt wird die Güte der Vertrauensmessung für die vorliegende Stichprobe überprüft. Eine solche Überprüfung scheint schon deswegen notwendig, weil die ursprüngliche Skala vom Englischen ins Deutsche übersetzt wurde und weil sich die Fragen ursprünglich auf physikalisch eingebettete automatisierte Systeme bezogen haben und nicht auf elektronische Entscheidungshelfer.

Die einzelnen Variablen `vertrauen_<nr>`, mit denen Vertrauen gemessen werden soll, sind in Tabelle 10.7 aufgelistet. Würde die Vertrauensskala von Jian u. a. strikt angewendet werden, würde sich die Frage nach der optimalen Anzahl von Faktoren nicht stellen, da eine Faktorenberechnung lediglich die acht vertrauensmessenden Variablen zu einem Faktor zusammenfassen und die entsprechenden Faktorladungen berechnen müsste. Ich habe mich bewusst gegen dieses vereinfachte Vorgehen entschieden um zu überprüfen, ob ggf. eine oder mehrere Variablen für den gegebenen Anwendungsfall keinen Einfluss auf die Messung nehmen.

Methodisch kommt eine explorative Faktorenanalyse mit Varimax-Rotation zum Einsatz. Es ergibt sich ein KMO<sup>14</sup> von 0.858, der als gut bis sehr gut bezeichnet werden kann und zeigt, dass eine Faktorenanalyse im Falle der betrachteten Variablen völlig unproblematisch durchführbar ist.

Wendet man das Kaiser-Kriterium (vgl. Backhaus u. a. 2006: 295) auf die Variablen `vertrauen_1` bis `vertrauen_8` an, ergibt sich überraschenderweise eine optimale Faktorenzahl von *zwei* und nicht etwa von eins, wie dies bei einem bereits erprobten Konstrukt zu erwarten wäre. Eine visuelle Überprüfung anhand des Screeplots nach

<sup>14</sup>KMO bezeichnet das sogenannten *Kaiser-Meyer-Olkin*-Kriterium (vgl. Backhaus u. a. 2006: 276), das den Grad der Zusammengehörigkeit von Variablen hinsichtlich ihrer Korrelation bezeichnet. Je größer der KMO-Wert, desto besser sind die betrachteten Variablen für eine Faktorenanalyse geeignet. Werte ab 0.5 gelten als „kläglich“, gerade aber noch akzeptabel, Werte  $\geq 0.7$  als ziemlich gut und Werte über 0.9 als erstaunlich (vgl. Kaiser und Rice 1974: 112; Backhaus u. a. 2006: 276).

dem Elbow-Kriterium (vgl. Backhaus u. a. 2006: 296f.) bestätigt, dass an der Stelle einer Zweifaktorlösung ein Knick erkennbar ist und damit ebenfalls auf eine optimale Faktorenzahl von zwei zu schließen ist (für Details s. Anhang A.5.3.1).

Ein Blick auf die Faktorladungen in Tabelle 10.9 zeigt, dass die zweifaktorielle Lösung lediglich durch den Einfluss der Variable `vertrauen_08`, „Die Funktionsweise von MensaAssist ist mir vertraut“ zustande gekommen ist. Setzt man die übliche Anforderung für relevante Faktorladungen mit Werten  $\geq 0.5$  (die noch geringere Anforderung  $\geq 0.3$  würden nichts an der Aussage ändern) an (vgl. Backhaus u. a. 2006: 299), so zeigt sich, dass sich Faktor 1 durch starke (allesamt  $\geq 0.7$ )<sup>15</sup> Faktorladungen auf den Variablen `vertrauen_1` bis `vertrauen_7` auszeichnet.

Für die weiteren Auswertungen, insbesondere hinsichtlich der Zusammenhänge zwischen kumulierten SEU-Differenzen, wird ein Faktor `FAC_Vertrauen` gebildet, der sich aus den Variablen `vertrauen_1` bis `vertrauen_7` zusammensetzt.

Für die einfaktorielle Lösung ergibt sich ein guter bis sehr guter KMO-Wert von 0.865 und nach dem Kaiser-Kriterium die Bestätigung einer einfaktoriellen Lösung des Variablenbündels `vertrauen_1...vertrauen_7`. Ebenso zeigt der Cronbachs-Alpha-Wert (vgl. Cronbach 1951) von 0.89 eine hohe interne Reliabilität der so gebildeten Vertrauensskala an. Für `FAC_Vertrauen` ergeben sich Faktorladungen gemäß Tabelle 10.10. Die negativen und positiven Vorzeichen sind, wie zu erwarten war, konsistent mit den entsprechend negativ und positiv formulierten Items. Für den Fall von `MensaAssist` kann man, wie es Jian u. a. (2000: 71) auch für ihre Vertrauensskala festgestellt haben, folgern, dass Vertrauen und Misstrauen das gleiche messen, nur mit jeweils unterschiedlichen Vorzeichen.

Die Faktorwerte für den so gebildeten Faktor `Vertrauen` können laut Kolmogorov-Smirnov-Anpassungstest mit Signifikanzniveau  $\alpha = 0.05$  als annähernd normalverteilt angenommen werden. Weitere statistische Kenngrößen für die Faktorwerte lau-

**Rotierte Komponentenmatrix<sup>a</sup>**

	Komponente	
	1	2
<code>vertrauen_1</code>	-,752	
<code>vertrauen_2</code>	-,753	
<code>vertrauen_3</code>	-,738	
<code>vertrauen_4</code>	,782	
<code>vertrauen_5</code>	,738	
<code>vertrauen_6</code>	,831	
<code>vertrauen_7</code>	,839	
<code>vertrauen_8</code>		,981

Extraktionsmethode:  
Hauptkomponentenanalyse.  
Rotationsmethode: Varimax mit  
Kaiser-Normalisierung.<sup>a</sup>

a. Die Rotation ist in 3 Iterationen konvergiert.

**Tabelle 10.9:** Faktorladungen ( $\geq 0.5$ ) für eine Zweifaktorlösung: Die Variable `vertrauen_8` bildet einen eigenen Faktor und „verursacht“ eine Zweifaktorenlösung statt einer Einfaktorlösung.

<sup>15</sup>Die Größer-Gleich-Relation bezieht sich immer auf den Betrag der Faktorladungen, demnach wäre der Einfluss einer Variablen für eine Faktorladung von  $-0.7$  ebenfalls bedeutsam, da  $|-0.7| \geq 0.5$  gilt.

Variable	Formulierung	Ladung
vertrauen_1	MensaAssist leitet mich fehl.	-0.735
vertrauen_2	Ich bin misstrauisch, was die Empfehlungen von MensaAssist angeht.	-0.777
vertrauen_3	Ich stehe MensaAssist insgesamt skeptisch gegenüber.	-0.760
vertrauen_4	Insgesamt bin ich überzeugt von MensaAssist.	0.796
vertrauen_5	MensaAssist gibt mir Sicherheit (bzgl. meiner Essenswahl).	0.729
vertrauen_6	MensaAssist ist zuverlässig.	0.819
vertrauen_7	Ich kann MensaAssist vertrauen.	0.851

**Tabelle 10.10:** MensaAssist – Faktorladungen für den Faktor Vertrauen ( $N = 131$ )

ten: Minimum:  $-2.076$ , Maximum  $2.7133$ , Median:  $0.00158$ , Mittelwert:  $0$ , Standardabweichung  $1$ .

Die Faktorwerte lassen sich unter Vorgabe von vier gewünschten Gruppen gemäß ihrer Perzentile gleichmäßig aufteilen (Histogramm s. Anhang A.5.3.1):

- Vertrauen *sehr gering*:  $-2.076$  bis  $-0.712$  ( $N_1 = 31$ )
- Vertrauen *gering*:  $> -0.712 \leq 0.00158$  ( $N_2 = 34$ )
- Vertrauen *mittel*:  $> 0.00158 \leq 0.8031$  ( $N_3 = 33$ )
- Vertrauen *stark*:  $> 0.8031$  ( $N_4 = 33$ )

### 10.3.3 SEU-Modelle und Essenstypen

Direkt zu Anfang des Fragebogens mussten die Probanden eine Vorentscheidung treffen, welcher Essenstyp mit seinen Eigenschaften am ehesten ihren Vorlieben entspricht (vgl. Abschnitt 10.2). Diese Vorauswahl diente zur Rahmung der Befragten, dass MensaAssist sich mit seinen Vorschlägen fortan an dieser Vorauswahl orientieren werde (was nicht der Fall war). Am Ende des Fragebogens wurde nochmals nach der Entscheidungsgrundlage der Essenswahl gefragt und U-Werte für die fünf relevanten Ziele erhoben (vgl. Seite 235). Gemäß Hypothese (MA 1) ist nun zu prüfen, ob die Vorauswahl der Essenstypen weitgehend mit Essenstypen übereinstimmt, die sich aus den erhobenen U-Werten ergeben.

Die Methode der Clusteranalyse hilft, diese These untersuchbar zu machen. In diesem Abschnitt wird nun anhand einer Clusteranalyse mit einer vorgegebenen Zahl von drei Clustern<sup>16</sup> überprüft, ob sich die vorgewählten Typen in der Stichprobe wiederfinden.

<sup>16</sup>Eine Drei-Cluster-Lösung wäre auch bei einer automatisch ermittelten optimalen Clusterzahl herausgekommen.

Am Ende des Fragebogens war es möglich Freitextanmerkungen zu verfassen. Dort äußerten von 15 Befragten, die das Freitextfeld ausgefüllt haben, zehn Personen die Kritik, dass die vorgegebenen Typen nicht ihren persönlichen Essenspräferenzen entsprechen. Die folgende Clusteranalyse bestätigt diese Kritik in der Weise, dass sich nur einer von drei vorgegebenen Essenstypen in der Stichprobe wiederfindet. Daraus lässt sich folgern, dass die Probanden MensaAssist recht kritisch begegneten bzw. zumindest aufmerksam hinsichtlich der Vorschläge von MensaAssist waren.

Eine Clusteranalyse nach Ward-Methode mit vorgegebener Clusterzahl und den Input-Variablen  $u_{lecker}$ ,  $u_{satt}$ ,  $u_{gesund}$ ,  $u_{schnell}$  und  $u_{zahlen}$  ergibt die in Tabelle 10.11 gezeigten Mittelwerte und 95 %-Konfidenzintervalle für die abgefragten Ziele. Zusätzlich ist zu den jeweiligen Clustern in der Spalte *Typwahl* die Anzahl der je vorausgewählten Essenstypen zu finden. Dem Cluster 3 wurden beispielsweise 45 Personen zugeordnet, die sich bei der anfänglichen Frage selbst dem Typ des Gesundheitsbewussten zugeordnet haben.

Cluster	$u_{lecker}$	$u_{satt}$	$u_{gesund}$	$u_{schnell}$	$u_{zahlen}$	Typwahl
<b>1</b> $N = 59$	90.64 [88.24, 93.05]	66.54 [62.08, 71.01]	56.76 [49.53, 64.00]	61.88 [57.28, 66.48]	58.34 [50.95, 65.73]	5x Sparfuchs 28x Gesunder 26x Eiliger
<b>2</b> $N = 36$	67.72 [63.95, 71.49]	63.67 [59.77, 67.57]	37.11 [30.11, 44.11]	39.83 [32.70, 46.97]	68.36 [61.46, 75.26]	16x Sparfuchs 6x Gesunder 14x Eiliger
<b>3</b> $N = 50$	80.76 [77.40, 84.12]	46.30 [39.73, 52.87]	75.28 [69.63, 80.93]	27.58 [23.94, 31.22]	38.50 [31.91, 45.09]	1x Sparfuchs 45x Gesunder 4x Eiliger

**Tabelle 10.11:** Zielbewertungen für drei Essenstypcluster

Werte in eckigen Klammern [ ] geben zu den jeweiligen Mittelwerten die 95 %-Konfidenzintervalle an.

### Cluster 1: Lecker

Cluster 1, mit insgesamt 59 Befragten, lässt sich anhand der Zielbewertungen besonders gut beschreiben. Es dominiert klar das Ziel etwas Leckeres zu essen. Nicht nur, dass der Mittelwert mit 90.64 (maximal ist 100 möglich) sehr hoch ist, sondern das Konfidenzintervall zu diesem Mittelwert ist auch sehr schmal [88.24, 93.05], was auf eine einheitliche Einschätzung von  $u_{lecker}$  innerhalb des Clusters hindeutet. Die anderen Zielbewertungen bewegen sich, auch erkennbar an den überlappenden Konfidenzintervallen, hinsichtlich ihrer Größenordnung auf sehr ähnlichem Niveau. Die Typwahl, die von den Befragten selbst getroffen wurde, lässt sich nicht direkt mit

den Mittelwerten der Zielbewertungen zusammenbringen. Da im Fragebogen angenommen wurde, dass allen Befragten wichtig ist, etwas Leckerer zu essen, wurde ein Typ „Leckermaul“ nicht in die Auswahl aufgenommen. Ebendiese Vorauswahl hätte für die Befragten im hier betrachteten Cluster besonders gut gepasst. Stattdessen entschieden sich 28 Personen für den Typ „Gesunder“ und 26 für den Typ „Eiliger“, wobei die entsprechenden Zielbewertungen im hiesigen Cluster den geringsten Wert einnehmen<sup>17</sup>.

Für das erste Cluster kann festgehalten werden, dass die Typvorauswahl nicht zu den Zielbewertungen passt.

### Cluster 2: Sparsam und lecker

Cluster 2, mit einem Umfang von 36 Befragten, zeichnet sich vor allem dadurch aus, dass sich Haupt- und Nebenziele sehr deutlich voneinander unterscheiden lassen. So ist es den Befragten in diesem Cluster ähnlich wichtig wenig zu zahlen, etwas Leckerer zu essen und satt zu werden. Deutlich weniger wichtig ist es ihnen das Essen schnell zu erhalten und etwas Gesundes zu essen. Vergleicht man die bewerteten Ziele auch hier mit der Typvorauswahl, findet sich allenfalls der Sparfuchs in den Zielbewertungen wieder: 16 Befragte entschieden sich zu Anfang für den Typ Sparfuchs, das Ziel  $u_{zahlen}$  wird, vergleicht man (ohne Beachtung der Konfidenzintervalle) nur die Mittelwerte, am höchsten bewertet. Für die anderen zwei als wichtig bewerteten Ziele,  $u_{lecker}$  und  $u_{satt}$ , gibt es in der Typvorauswahl keinen korrespondierenden Typ, so dass auch keine Typvorauswahlentsprechung zu erwarten ist.

### Cluster 3: Lecker und gesund

Cluster 3 ist schließlich mit „lecker und gesund“ überschrieben. Den in diesem Cluster befindlichen Befragten ist es besonders wichtig, etwas Leckerer ( $u_{lecker}$ ) und etwas Gesundes ( $u_{gesund}$ ) zu essen. Diese beiden Ziele haben mit 80.76 bzw. 75.28 die jeweils höchsten U-Werte. Deutlicher weniger wichtig ist den Befragten satt zu werden (46.30) oder wenig zu bezahlen (38.50). Mit Abstand am unwichtigsten ist es den Personen in Cluster 3 ihr Essen schnell zu erhalten (27.58)<sup>18</sup>.

Cluster 3 ist das einzige Cluster, bei dem sich die Typvorauswahl in der Ausprägung der U-Werte wiederfindet. In diesem Cluster haben von insgesamt 50 Befragten immerhin 45 den Typ „Gesunder“ ausgewählt. Wie beschrieben, gehört das Ziel etwas

<sup>17</sup>Dies gilt zumindest, wenn man sich auf einen bloßen Vergleich der Mittelwerte beschränkt. Die zusätzliche Angabe der Konfidenzintervalle zeigt allerdings, dass der Unterschied *nicht* signifikant ist, da sich die Konfidenzintervalle z. T. überschneiden.

<sup>18</sup>Methodische Anmerkung: Durch die zusätzliche Angabe der Konfidenzintervalle in Ergänzung zu den Mittelwerten fällt die Interpretation deutlich leichter als nur auf Basis der Mittelwerte zu argumentieren. So kann beispielsweise auf Grund der disjunkten Konfidenzintervalle von  $u_{zahlen}$  und  $u_{satt}$  mit  $u_{schnell}$  begründet werden, warum für die Befragten  $u_{schnell}$  am unwichtigsten ist.

Gesundes zu essen zu den wichtigsten Zielen der in diesem Cluster befindlichen Personen. Diese Übereinstimmung zeigt, dass die Essenstypvorauswahl um so besser passt, je näher ebendiese an die typisierten U-Werte heranreicht.

### Fazit Essenstypen

Fassen wir noch einmal zusammen, welche Erkenntnisse sich aus dem Vergleich der Essenstypen aus der Clusteranalyse mit der Vorauswahl der Essenstypen und damit für (MA 1) ergeben haben:

Der Vergleich der Essenstypen aus der Vorauswahl mit der inhaltlichen Benennung der Cluster anhand ihrer U-Werte hat die eingangs formulierte These *nicht* bestätigt: Die zur Auswahl stehenden Essenstypen stimmen nur unzureichend mit den tatsächlich vorgefundenen Typen in der Stichprobe überein. Lediglich für Cluster 3 lässt die inhaltliche Deutung der Cluster anhand ihrer durchschnittlichen U-Werte eine große Übereinstimmung mit dem Vorauswahltyp „Gesunder“ erkennen.

Neben dem eher enttäuschenden Ergebnis hinsichtlich der Essenstyp-Vorauswahl stimmt das Ergebnis der Clusteranalyse dennoch optimistisch bzgl. der inhaltlichen Reichhaltigkeit der U-Werte und damit der typisierten SEU-Modelle, die sich in der Stichprobe finden. Es lassen sich insgesamt drei Cluster ausmachen, die sich in der Ausprägung ihrer jeweiligen Zielbewertungen *deutlich* voneinander unterscheiden. Die weitere Auswertung wird die Essenstyp-Vorauswahl, die von den Probanden selbst vorgenommen wurde, allenfalls am Rande berücksichtigen und stattdessen den Fokus auf SEU-Modelle und -Modell differenzen sowie deren Zusammenhang zu Vertrauen legen. Die Berücksichtigung von SEU-Modell differenzen schließt implizit an das Ergebnis aus der vorgestellten Clusteranalyse an, da U-Werte die Stichprobe besser beschreiben und der Heterogenität der Akteure eher gerecht werden als die offenbar unpassende zur Auswahl stehenden Essenstypen vom Anfang des Fragebogens.

(MA 1)  
X

#### 10.3.4 SEU-Modelle und Modelldifferenzen

Mit den in Abschnitt 8.2.1 eingeführten Begriffen zum SEU-Modellvergleich (für die Begriffe SEU-Modellstruktur und -Ausprägung vgl. ebenfalls Abschnitt 8.2.1) stellt sich für den vorliegenden Anwendungsfall die Frage, *welche* SEU-Modelle miteinander verglichen werden sollen. Da das Experimentaldesign die Modellstruktur vorgegeben hat, kann ein Vergleich von SEU-Modellen nur auf Basis unterschiedlicher Modellausprägungen erfolgen. Führen wir uns noch mal vor Augen, wie das MensaAssist-Experiment aufgebaut war: An jedem Mensatag wurden den Befragten hinsichtlich bestimmter Eigenschaften/Ziele bewertete Menüs vorgestellt. Sie sollten sich daraufhin für ein bestimmtes Menü entscheiden, bekamen danach jeweils angezeigt, ob ihre eigene Entscheidung mit der Entscheidung von MensaAssist übereingestimmt hat, und konnten ihre eigene Entscheidung ggf. revidieren. Nach Abschluss der sechs

Mensatage mussten die Probanden Angaben bzgl. ihrer Entscheidungsgrundlage machen (s. Abbildung 10.3 auf Seite 235). Dabei konnten sie die in der Untersuchung relevanten Ziele („etwas Leckereres essen“, „satt werden“, „etwas Gesundes essen“, „Essen schnell erhalten“ und „wenig bezahlen“) hinsichtlich ihrer individuellen Bedeutsamkeit bewerten. Damit ergibt sich für jeden Probanden  $i$  ein subjektsspezifischer  $\mathbf{U}_i$ -Vektor. Durch diesen subjektsspezifischen Vektor lässt sich ex post für jedes Essen an jedem Tag ein entsprechender SEU-Wert berechnen. Beispielhaft soll diese Berechnung für Essen 1 an Tag 1 (vgl. Abbildung 10.2 auf Seite 233) für den Befragten mit der laufenden Nummer 118 vorgeführt werden. Essen 1 ist wie folgt bewertet: ●●●●● Geschmack, ●●●●● Menge, ●● Gesund, ●● Wartezeit, ● Preis. Gemäß der bereits vorgestellten Tabelle 10.6 (Seite 234) lassen sich den Essensbewertungen entsprechende P-Werte, die für alle Befragten identisch sind, zuweisen. Für Essen 1 ergibt sich z. B.:

$$\mathbf{p}_{Essen\ 1}(Tag\ 1) = (1, 1, 0.25, 0.25, 0)$$

Für Essen 2, das MensaAssist an Tag 1 empfiehlt, ergibt sich folgender Vektor:

$$\mathbf{p}_{Essen\ 2}(Tag\ 1) = (1, 0, 0.5, 0.75, 1)$$

Bei der Angabe der Entscheidungsgrundlage hat der Befragte 118 folgende Angaben gemacht:

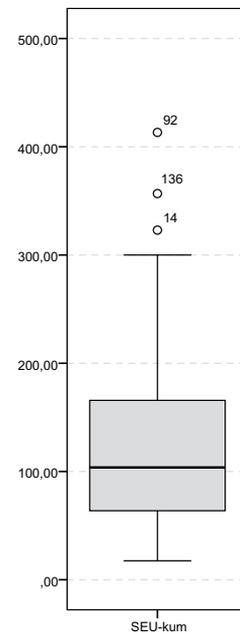
$$\mathbf{U}_{118} = (90, 70, 85, 30, 20)$$

Mit diesen Angaben lässt sich nun der SEU-Wert für das entsprechende Essen berechnen. Es ergibt sich:

$$SEU(Essen\ 1, Tag\ 1, Fall\ 118) = \mathbf{p}_{Essen\ 1}(Tag\ 1) \cdot \mathbf{U}_{118}^T = 188.75$$

MensaAssist empfiehlt jedem Befragten an Tag 1 das Essen Nr. 2. Es lässt sich nun, da die subjektive Bewertung der unterschiedlichen Ziele bekannt ist, für jeden Befragten der SEU-Wert für das Essen 2 berechnen. Beachtenswert ist, dass hier explizit der subjektiven Bewertung des Probanden Rechnung getragen wird. Für den hier betrachteten Fall ergibt sich folgender SEU-Wert:

$$SEU(Essen\ 2, Tag\ 1, Fall\ 118) = \mathbf{p}_{Essen\ 2}(Tag\ 1) \cdot \mathbf{U}_{118}^T = 175.00$$



**Abbildung 10.4:** Boxplot für kumulierte SEU-Differenzen

Es soll nun These (MA 2) untersucht werden, ob ein Zusammenhang zwischen kumulierten SEU-Differenzen und Vertrauen besteht. Dazu sind zunächst einige Vorbereitungsarbeiten notwendig.

Wie an Fallnummer 118 vorgeführt, lässt sich für jeden Mensatag und jedes Essen ein SEU-Wert berechnen. Bildet man nun die Differenz der SEU-Werte mit dem höchst bewerteten Essen und dem Essen, das MensaAssist vorschlägt, so erhält man für jeden Tag einen Abweichungswert, der von den Zielintensitäten der befragten Person abhängt. Summiert man diese Abweichungswerte für alle Tage auf, so erhält man den Abweichungsgrad aller Entscheidungen von MensaAssist von der Entscheidungsgrundlage des betrachteten Befragten. Im folgenden bezeichnet  $SEU\text{-max}_{i,j}$  den jeweils höchsten SEU-Wert eines Essens für den Befragten  $i$  an Tag  $j$  und  $SEU\text{-MA}_{i,j}$  den subjektspezifischen SEU-Wert für den Befragten  $i$  des Essens, das MensaAssist an Tag  $j$  wählen würde.

Es lässt sich nun für jeden Probanden  $i$  eine **kumulierte SEU-Differenz**,  $SEU\text{-kum}(i)$ , berechnen, die eine Maßzahl für den Grad der Abweichung zwischen MensaAssist und dem Befragten  $i$  angibt:

$$SEU\text{-kum}(i) = \sum_{Tage\ j \in \{1, \dots, 6\}} SEU\text{-max}_{i,j} - SEU\text{-MA}_{i,j}$$

Schaut man sich für die gegebene Stichprobe die Verteilung der kumulierten SEU-Differenzen an (vgl. Abbildung 10.4), so zeigt sich, dass sich für die befragten Personen sehr unterschiedliche kumulierte SEU-Differenzen ergeben und die neu berechnete Maßzahl gut geeignet ist, um aussagekräftige Korrelationen mit anderen Variablen zu erhalten. Es zeigen sich dabei nur wenige Ausreißer, wobei die Verteilung nach oben hin stärker streut als nach unten.

Der Vollständigkeit halber seien auch noch weitere statistische Kenngrößen genannt: Mittelwert 119.357, Median 103.75, Minimum 17.50, Maximum 413.25, Standardabweichung 74.712.

### 10.3.5 Zusammenhang von Vertrauen und SEU-Modellstruktur

In diesem Abschnitt wird nun These (MA 2) behandelt, nämlich der Zusammenhang zwischen Unterschieden in SEU-Modellen und dem Konstrukt Vertrauen, behandelt. Hinsichtlich der in Abschnitt 8.2.2 spezifizierten Modelldifferenzen wird hier  $SEU_{NH \rightarrow H} \neq SEU_H$  untersucht. Der nicht-menschliche Akteur (NH, hier: MensaAssist) schreibt dem menschlichen Akteur (H) bestimmte Ziele und damit ein entsprechendes SEU-Modell zu ( $SEU_{NH \rightarrow H}$ ) und schlägt entsprechende Speisen vor. Dieses Modell unterscheidet sich vom „tatsächlichen“ SEU-Modell, das durch die abgefragten U-Werte des menschlichen Akteurs erhoben wurde ( $SEU_H$ ). Es gilt nun zu überprüfen, ob ein Zusammenhang zwischen Vertrauen und ebendiesen SEU-Modelldifferenzen besteht.

(MA 2) ?
-------------

**Korrelationen**

		SEU-kum	FAC-Vertrauen	Zufriedenheit
SEU-kum	Korrelation nach Pearson	1	-,244**	-,307**
	Signifikanz (2-seitig)		,005	,000
	N	149	131	148
FAC_Vertrauen	Korrelation nach Pearson	-,244**	1	,733**
	Signifikanz (2-seitig)	,005		,000
	N	131	131	131
Zufriedenheit	Korrelation nach Pearson	-,307**	,733**	1
	Signifikanz (2-seitig)	,000	,000	
	N	148	131	148

\*\* . Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

**Abbildung 10.5:** MensaAssist – Zusammenhang von kumulierten SEU-Differenzen und Vertrauen: Für das Signifikanzniveau  $p < 0.01$  lässt sich ein negativer Zusammenhang feststellen.

Im vorherigen Abschnitt wurden die kumulierten SEU-Differenzen als Maßzahl eingeführt und für das MensaAssist-Experiment berechnet. Beachtenswert ist, dass die Maßzahl zunächst in keinerlei Zusammenhang mit den Vertrauensfragen, aus denen der Faktor Vertrauen berechnet wurde, steht. Die abgefragten Zielbewertungen bezogen sich dabei ausschließlich auf die im Experiment relevanten Ziele und zwar aus der jeweils individuellen Sichtweise der Befragten. Die Vertrauensfragen bezogen sich dagegen ausschließlich auf die Bewertung unterschiedlicher Eigenschaften des Entscheidungshelfers MensaAssist.

Sollte sich nun ein statistisch signifikanter Zusammenhang zwischen SEU-Modelldifferenzen und Vertrauen feststellen lassen, so wäre es möglich, Gründe für mangelndes Vertrauen in autonome Technik zu benennen. Die Vertrauensskala von Jian u. a. (2000) ermöglicht zwar eine Messung von Vertrauen, diese Messung ist allerdings unabhängig von dem konkreten System. Übertragen auf das in Abschnitt 8.2.1 eingeführte abstrakte Konzept des Vergleichs von SEU-Modellen und der Operationalisierung in Form von kumulierten SEU-Differenzen für den vorliegenden Fall (Abschnitt 10.3.4), könnte man herausfinden, aus welchen *Gründen* MensaAssist mehr oder weniger stark vertraut wird. Sind die Gründe für mangelndes Vertrauen erst einmal bekannt, wäre es für einen Systemkonstrukteur möglich, ebendiesen Gründen durch punktuelle Modifikationen der Systemarchitektur zu begegnen.

Zur Überprüfung der Hypothese (MA 2) werden Korrelationen zwischen dem Faktor FAC\_Vertrauen und der Maßzahl SEU-kum berechnet. Zur Kontrolle wird zusätzlich die Variable Zufriedenheit mit in die Korrelationsrechnung aufgenommen, da aus der Automationsliteratur (vgl. Kapitel 4, exemplarisch Lee und See (2004: 59) bekannt ist, dass Vertrauen und Performance bzw. Zufriedenheit in einem engen Zusammenhang stehen. Das Ergebnis der Korrelationsrechnung ist in Abbildung 10.5 dargestellt.

Es zeigt sich, dass sowohl eine statistisch signifikante, wenn auch schwache, negative Korrelation zwischen kumulierten SEU-Differenzen und Vertrauen besteht ( $-0.244$ ) wie auch ein negativer Zusammenhang zwischen kumulierten SEU-Differenzen und der Zufriedenheit mit MensaAssist ( $-0.307$ ). Ebenso bestätigt die Korrelationsrechnung einen starken Zusammenhang zwischen dem Faktor Vertrauen und der Zufriedenheit mit MensaAssist ( $0.733$ ). Demnach kann These (MA 2) als zumindest teilweise bestätigt gelten.

(MA 2)  
 ✓

### 10.3.6 Inhaltliche Interpretation unterschiedlicher Vertrauensstärke

Die im vorherigen Abschnitt vorgenommene Korrelationsanalyse deutet auf Zusammenhänge zwischen Vertrauen, kumulierten SEU-Differenzen und Zufriedenheit hin. Diese Zusammenhänge sollen für zwei Gruppen unterschiedlicher Vertrauensstärke (für die Gruppierung s. Seite 242), nämlich die sehr gering Vertrauenden und die stark Vertrauenden, gemäß der zuvor formulierten Hypothese (MA 3), in diesem Abschnitt näher betrachtet werden, da für diese beiden Gruppen die stärksten und damit inhaltlich gehaltvollsten Aussagen zu erwarten sind.

Dazu wird im ersten Schritt überprüft, ob sich bereits beim direkten Vergleich zwischen beiden Gruppen deutliche Unterschiede ergeben. Im zweiten Schritt erfolgt eine gruppenspezifische Betrachtung. Gibt es z. B. für die Befragten, die sehr wenig vertrauen, typische Charakterisierungen z. B. anhand ihrer Zielintensitäten<sup>19</sup>?

#### Vergleich zwischen sehr gering und stark Vertrauenden

(MA 3)  
 ?

Ein direkter Vergleich zwischen der Gruppe der sehr gering Vertrauenden und den stark Vertrauenden ist zunächst ernüchternd. Weder zwischen den kumulierten SEU-Differenzen noch zwischen den Zielintensitäten lassen sich deutliche Unterschiede feststellen. Allerdings deutet die Betrachtung der Zufriedenheit an, dass sich dennoch Unterschiede zwischen den unterschiedlichen Vertrauensgruppen ausmachen lassen. Für die Zufriedenheit  $\bar{X}$  in den beiden Gruppen lassen sich folgende Mittelwerte und 95%-Konfidenzintervalle berechnen:

- sehr gering Vertrauende ( $N = 31$ ):  
 $\bar{X} = 31.7097$ , 95 %-Konfidenzintervall: [24.1069, 39.3124]
- stark Vertrauende ( $N = 33$ ):  
 $\bar{X} = 78.8182$ , 95 %-Konfidenzintervall: [74.2847; 83.3517]

Die Konfidenzintervalle für die Zufriedenheit in den beiden betrachteten Gruppen überschneiden sich nicht, so dass gefolgert werden kann, dass die Befragten in der

<sup>19</sup>Es sei angemerkt, dass Zielintensitäten hier nicht als *Gründe* im strikt kausalen Sinne für schwaches oder starkes Vertrauen interpretiert werden können. Sie können die jeweilige Gruppe nur näher beschreiben. Vergleicht man dann die Charakterisierung der sehr gering Vertrauenden mit der der stark Vertrauenden, lassen sich Unterschiede in der Gruppenbeschreibung herausstellen.

Gruppe der stark Vertrauenden deutlich zufriedener mit MensaAssist waren als diejenigen in der Gruppe der sehr gering Vertrauenden.

Dieses Ergebnis ist wenig überraschend, da die Items in der verwendeten Vertrauensskala teilweise auch den Aspekt Zufriedenheit enthielten, so dass eine Vertrauensclustering auch eine teilweise Clustering von Zufriedenheit umfasst.

Nach der eher groben Betrachtung der beiden Vertrauensstärkegruppen sollen diese in den folgenden Abschnitten noch näher beschrieben und hinsichtlich unterschiedlicher Zielintensitätsmuster in vergleichender Perspektive und unter Zuhilfenahme der SEU-Theorie analysiert werden.

### **Teil 1: Charakterisierung der sehr gering Vertrauenden**

Sehr gering Vertrauende, gemäß der Klassifizierung auf Seite 242, sind solche Befragte, die einen Faktorwert für `FAC_Vertrauen` von weniger als  $-0.712$  haben. Diese Eigenschaft trifft auf 31 Befragte zu.

Da nicht zu erwarten ist, dass sich diese Gruppe durch homogene Zielintensitäten auszeichnet, sondern ggf. durch relativ extreme Zielintensitäten, d. h. der überragenden Relevanz von jeweils einem Ziel, soll mit Hilfe einer Clusteranalyse zunächst geklärt werden, ob sich eine sinnvolle Anzahl Cluster finden lässt, die die Gruppe der sehr gering Vertrauenden trennscharf anhand der Zielintensitäten unterteilen kann. Zur Anwendung kommt eine hierarchische Clusteranalyse nach Ward-Methode mit der euklidischen Distanz als Ähnlichkeitsmaß. Es wird wie bei einer explorativen Clusteranalyse üblich zweischrittig vorgegangen: Im ersten Schritt wird zunächst die optimale Anzahl an Clustern ermittelt. Als Kriterium dienen dabei besonders große Sprünge der Koeffizientenwerte bei den Verschmelzungsschritten und ein damit einhergehender Knick gemäß des Elbow-Kriteriums (vgl. Backhaus u. a. 2006: 535f.). Im vorliegenden Fall ist insbesondere im letzten Verschmelzungsschritt ein großer Sprung festzustellen, so dass eine Zwei-Cluster-Lösung angestrebt wird (für Details s. A.5.3.3).

Nach Bildung der Cluster werden diese anhand von durchschnittlichen Zielintensitäten charakterisiert und somit ein inhaltlicher Zusammenhang zwischen dem gemessenen Vertrauen und den Zielintensitäten der Befragten hergestellt. Die Mittelwerte der beiden Cluster hinsichtlich der betrachteten Ziele sind in Abbildung 10.6 dargestellt. Anhand der durchschnittlichen Zielintensitäten ist zu sehen, dass sich beide Cluster vor allem dadurch auszeichnen, dass jeweils zwei Ziele besonders ausgeprägt sind. Cluster 1 zeichnet sich dadurch aus, dass die beiden Ziele „lecker“ und „satt“ besonders ausgeprägt sind. Das Ziel „zahlen“ ist mit einem Wert von 63.588 (über 50) zwar noch wichtig, jedoch weit weniger wichtig als die beiden anderen Ziele. Die Ziele „gesund“ und „schnell“ sind eher unwichtig.

Cluster 2 zeichnet sich ebenfalls durch zwei sehr stark bewertete Ziele aus. Für dieses Cluster ist allerdings das Ziel „gesund“ mit einem Mittelwert von 88.5 mit Abstand

cluster sehr geringes vertrauen		u_lecker	u_satt	u_gesund	u_schnell	u_zahlen
1	Mittelwert	83,5882	74,7647	31,1176	41,882	63,5882
	N	17	17	17	17	17
	Standardabweichung	12,12466	16,52093	22,85247	23,3984	29,29390
2	Mittelwert	69,2143	48,9286	88,5000	47,071	59,5714
	N	14	14	14	14	14
	Standardabweichung	28,75675	26,28677	16,07554	25,8352	27,70647
Insgesamt	Mittelwert	77,0968	63,0968	57,0323	44,226	61,7742
	N	31	31	31	31	31
	Standardabweichung	22,12744	24,81579	35,11646	24,2511	28,18594

**Abbildung 10.6:** Mittelwerte der Zielintensitäten für die Zwei-Cluster-Lösung. Betrachtet wurden bei der Clusterung nur die Gruppe der sehr gering Vertrauenden  
 $N = 31$

am wichtigsten. Das Ziel „lecker“ ist mit einem Wert von 69.214 ebenfalls wichtig. Wenig zu bezahlen ist noch eher wichtig, während die beiden Ziele „schnell“ und „satt“ weniger wichtig sind.<sup>20</sup>

Was bedeutet die inhaltliche Interpretation nun für MensaAssist? Stelle man sich vor, MensaAssist wäre bereits ein im Einsatz befindliches System, so wäre es möglich über die Charakterisierung der sehr wenig Vertrauenden durch Designänderungen an MensaAssist ein möglicherweise höheres Maß an Vertrauen zu bewirken, wenngleich keinesfalls klar ist, ob die ausschließlichen Gründe für mangelndes Vertrauen in den spezifischen Ausprägungen der Zielintensitäten zu suchen sind. Zumindest für Cluster 2 mit dem stark ausgeprägten Ziel „gesund“ unterscheidet sich dieses stark von den Zielintensitäten von MensaAssist. Wie Tabelle 10.5 zu entnehmen ist, spielt das Ziel „gesund“ selbst für den Typ *Gesundheitsbewusster* nicht die entscheidende Rolle bei der Essenswahl. Bei den zwei anderen Typen ist das Ziel sogar am unwichtigsten. Berechnet man für alle sechs Tage die hypothetische Essenswahl auf Basis der durchschnittlichen Zielintensitäten, ergeben sich für Cluster 1 insgesamt drei Übereinstimmungen, für Cluster 2 nur zwei Übereinstimmungen mit den Vorschlägen von MensaAssist.

## Teil 2: Charakterisierung der stark Vertrauenden

Stark Vertrauende, gemäß der Klassifizierung auf Seite 242 sind solche Befragte, die einen Faktorwert für *FAC\_Vertrauen* mehr als 0.8031 haben. Diese Eigenschaft trifft auf 33 Befragte zu.

Analog zur Fragestellung nach der inhaltlichen Interpretation für die Gruppe der sehr gering Vertrauenden kann man auch untersuchen, welche Eigenschaften hinsichtlich

<sup>20</sup>Führt man sich vor Augen, dass die primäre Funktion des Essens, zumindest früher, das Sattwerden war, ist es durchaus überraschend, dass einigen Befragten das Sattwerden und das Essen schnell zu erhalten sogar eher unwichtig sind.

der betrachteten Ziele die Gruppe der stark Vertrauenden aufweist. Berechnet man im hiesigen Fall ebenfalls eine Zwei-Cluster-Lösung<sup>21</sup>, ergeben sich für die beiden Cluster die in Abbildung 10.7 dargestellten Mittelwerte.

cluster	hohes vertrauen	u_lecker	u_satt	u_gesund	u_schnell	u_zahlen
1	Mittelwert	91,1667	63,6111	60,1111	60,833	60,2222
	N	18	18	18	18	18
	Standardabweichung	8,29777	10,69436	24,61202	17,9616	19,26458
2	Mittelwert	78,5333	52,7333	64,4000	40,667	36,1333
	N	15	15	15	15	15
	Standardabweichung	10,56860	22,28217	24,00536	22,8087	17,78790
Insgesamt	Mittelwert	85,4242	58,6667	62,0606	51,667	49,2727
	N	33	33	33	33	33
	Standardabweichung	11,23619	17,55646	24,05455	22,4272	21,99897

**Abbildung 10.7:** Mittelwerte der Zielintensitäten für die Zwei-Cluster-Lösung. Betrachtet wurden bei der Clustering nur die Gruppe der stark Vertrauenden  
 $N = 33$

Auffällig ist, dass in beiden Clustern *ein* Ziel besonders wichtig ist – das Ziel „lecker“. In Cluster 1 beträgt der Mittelwert für dieses Ziel 91.167 und ist damit sehr wichtig, in Cluster 2 beträgt der Wert immerhin noch 78.533 und ist damit immer noch wichtig. Die anderen Ziele folgen in weitem Abstand: Für Cluster 1 sind alle anderen Ziele eher wichtig mit Mittwerten um 60, während in Cluster 2 die Ziele „satt“ und „gesund“ mit Werten von 52.733 bzw. 64.4 gerade noch in den Bereich von eher wichtig bzw. wichtig kommen und die beiden restlichen Ziele „schnell“ und „zahlen“ als eher unwichtig angesehen werden.

Bezogen auf MensaAssist bedeutet das Ergebnis, dass gerade diejenigen Befragten MensaAssist in hohem Maße vertrauen, denen ein Ziel, nämlich „lecker“, besonders wichtig ist. Für diese Befragten liegen ihre jeweiligen internen SEU-Modelle und die von MensaAssist realisierten Zielintensitäten dicht beieinander. Ein Blick auf Tabelle 10.5 (Seite 232), in der die Bewertungen der Ziele für die unterschiedlichen Essenstypen aufgelistet sind, zeigt, dass die höchste Bewertung für  $u_{lecker}$  verzeichnet ist. Rechnet man auch für die Cluster der stark Vertrauenden zurück, an wie vielen der sechs Tage die Cluster-Entscheidung mit der von MensaAssist übereinstimmt, ergeben sich für Cluster 1 drei Übereinstimmungen, für Cluster 2 sogar fünf.

(MA 3)



Bei aller Vorsicht, die wegen der geringen Fallzahl durch mehrfache Clustering geboten ist, kann festgestellt werden, dass sich die SEU-Modelle von stark Vertrauenden und sehr gering Vertrauenden unterscheiden und damit These (MA 3) weitgehend bestätigt ist. Die Unterscheidung schlägt sich in einer unterschiedlichen Anzahl an

<sup>21</sup>Die Entscheidung für zwei Cluster fällt hier allerdings etwas knapper aus. Dennoch ist der größte Koeffizientensprung im vorletzten Schritt vorzufinden. Für Details s. A.5.3.3.

Übereinstimmungen mit den Vorschlägen von MensaAssist nieder. Die SEU-Modelle der stark Vertrauenden stimmen stärker mit den Entscheidungen von MensaAssist (und damit implizit mit dem SEU-Modell) überein als die der sehr gering Vertrauenden.

## 10.4 Zwischenfazit: MensaAssist

In diesem Abschnitt werden die zentralen Ergebnisse zum MensaAssist-Experiment noch mal zusammengefasst und in den Gesamtzusammenhang der vorliegenden Arbeit eingeordnet.

Die zwei vorgestellten Vorstudien (Abschnitt 10.1) im Rahmen zweier von mir betreuten Abschlussarbeiten haben gezeigt, dass sich die Essenswahl in der Mensa mit Hilfe eines relativ einfachen SEU-Modells modellieren lässt, das mit nur fünf Zielen auskommt. Berücksichtigt wurden die Ziele etwas *Leckeres* essen, wenig *bezahlen*, das Essen *schnell* erhalten, *satt* werden und etwas *Gesundes* essen. Dabei kamen unterschiedliche Erhebungsmethoden zum Einsatz – eine Papierbefragung in der Mensa und eine Online-Befragung.

Auf Grundlage der Vorarbeiten wurde ein Online-Fragebogen entwickelt (10.2), der einen elektronischen Entscheidungshelfer, MensaAssist, simulierte. Mittels der damit erhobenen Daten wurde der Zusammenhang von Vertrauen und SEU-Modellunterschieden untersucht. Die Probanden hatten verschiedene Essenstypen zur Auswahl, an denen sich, so die Instruktion im Fragebogen, MensaAssist bei seinen Vorschlägen orientieren würde. Da am Ende des Fragebogens zusätzlich die individuelle Entscheidungsgrundlage der Probanden anhand einer feingranularen Zielintensitätsskala erhoben wurde, war es möglich die vorgegebenen Essenstypen mit den Typen, die sich tatsächlich in der Stichprobe fanden, zu vergleichen (10.3.3). Dabei stellte sich heraus, dass die vorgegebenen Typen mit den Zielgewichtungen der Befragten z. T. nur unzureichend übereinstimmten. Dies stellte für die weitere Analyse jedoch kein Problem dar. Das Gegenteil ist der Fall, da es dadurch ggf. zu einem Vertrauensproblem kommen konnte.

Um den Zusammenhang von Vertrauen und SEU-Modellen näher in den Blick zu bekommen, wurde der in Abschnitt 8.2.1 eingeführte Ansatz des systematischen SEU-Modellvergleichs auf den hiesigen Fall angewendet. Dazu wurden kumulierte SEU-Differenzen (10.3.4) definiert, die für jeden Probanden die Abweichung der MensaAssist-Entscheidungen von dem eigenen Entscheidungsmodell messen. Es hat sich in Bestätigung von These (MA 2) ein statistisch signifikanter *negativer* Zusammenhang zwischen kumulierten SEU-Differenzen und Vertrauen gezeigt. Je stärker sich die aufsummierten Abweichungen des eigenen Entscheidungsmodells von denen des Entscheidungshelfers MensaAssist unterscheiden, desto weniger Vertrauen haben die Probanden in MensaAssist (10.3.4).

Abschnitt 10.3.6 hat schließlich zwecks Untersuchung von These (MA 3) versucht, Gruppen unterschiedlicher Vertrauensstärke anhand typisch auftretender Zielbewertungen zu charakterisieren. Insgesamt sind dabei nicht so trennscharfe Ergebnisse herausgekommen wie erwartet, was u. a. mit der durch mehrfache Gruppenbildung reduzierten Probandenzahl erklärbar wird. Dennoch hat dieser Ansatz gezeigt, dass ein differenzierter Vergleich von Zielbewertungen zwischen Probanden und MensaAssist Anhaltspunkte für Gründe sehr geringen Vertrauens aufspüren kann. So hat sich z. B. gezeigt, dass eine Teilgruppe der sehr gering Vertrauenden, gemäß ihrer Selbsteinschätzung am Ende des Fragebogens, in hohem Maße darauf achtet, gesundheitsbewusst zu essen. Verständlich wird das sehr geringe Vertrauen dieser Gruppe dadurch, dass der Aspekt des gesundheitsbewussten Essens bei den Vorschlägen von MensaAssist kaum eine Rolle spielte (vgl. Abschnitt 10.2.2).

Durch die Berücksichtigung relevanter Ziele und ihrem Abgleich mit MensaAssist sind die „Stellschrauben“ bekannt, an denen gedreht werden kann um die Modelle der Probanden und von MensaAssist näher zusammenzubringen. Für zukünftige Untersuchungen könnte es interessant sein zu überprüfen, ob ein lernendes System, das die U-Werte im Laufe des Experiments an die des Probanden anpasst, zu einem höheren Maß an Vertrauen führt.

# 11 Experiment: Simhybs<sup>1</sup>

In diesem Kapitel wird das Experiment SIMHYBS<sup>2</sup> vorgestellt. Bei SIMHYBS handelt es sich um eine interaktive Computersimulation, bei der Probanden zusammen mit einem Fahrerassistenzsystem in unterschiedlichen Modi der Handlungsverteilung eine Fahraufgabe bewältigen müssen. Der Aufbau folgt damit (FLA 13) – das Zusammenhandeln von Mensch und autonomer Technik lässt sich besonders gut im Kontext einer gemeinsam zu erfüllenden Aufgabe untersuchen.

Die selbst entwickelte Software SIMHYBS wurde in einer älteren Version bereits in vorherigen Arbeiten verwendet (vgl. Fink 2008; 2009; Fink und Weyer 2011) und hat dort das grundsätzliche Potenzial nachgewiesen, handlungstheoretisch fundierte Aussagen über die Mensch-Maschine-Interaktion in hybriden Systemen generieren zu können.

In der vorliegenden Arbeit soll an dem Grundkonzept von SIMHYBS festgehalten werden, allerdings erfolgt eine inhaltliche Verschiebung des Fokus auf Fragen nach dem Vertrauen in autonome Technik (hier: das Fahrerassistenzsystem) in einer handlungstheoretischen Perspektive.

Aus der Automationsliteratur ist bekannt, dass ein enger Zusammenhang zwischen dem Vertrauen in eine Automation, dem Allokationsverhalten und damit der Nutzungsintensität bzw. dem Ausbleiben von manuellen Eingriffen besteht (vgl. Abschnitt 4). Dieser Zusammenhang soll auch mit SIMHYBS nachgewiesen werden. Dabei soll es nicht bei der bloßen Diagnose eines solchen Zusammenhangs bleiben. Es werden weitergehende Überlegungen angestellt, in welchem Wechselwirkungsverhältnis Allokationsverhalten, Vertrauen und wechselseitige Modellzuschreibungen stehen.

Im folgenden Abschnitt 11.1 werden zunächst die Benutzeroberfläche vom SIMHYBS und Grundzüge des implementierten Modells vorgestellt. Daran schließt sich in 11.2 eine Beschreibung des Experimentaldesigns sowie eine Operationalisierung der zu untersuchenden Hypothesen an. Bei der Entwicklung des Experimentaldesigns wird es darum gehen, das in Abschnitt 8.2 beschriebene Konzept der wechselseitigen Modellzuschreibungen untersuchbar zu machen. Im Auswertungsabschnitt 11.3 werden die

---

<sup>1</sup>Ein herzlicher Dank gilt auch hier den Teilnehmerinnen und Teilnehmern des Master-Projektseminars des Fachgebiets Techniksoziologie im WS 11/12, die unter meiner Anleitung Experimente mit Simhybs geplant und durchgeführt haben.

<sup>2</sup>SIMHYBS steht für Simulation hybrider Systeme. Genauer gesagt heißt die Software SIMHYBS und dient als softwaretechnische Grundlage zur Durchführung von Beobachtungen der Mensch-Maschine-Interaktion.

bereits formulierten Thesen (SH 1) bis (SH 4) operationalisiert und gezielt ausgewertet.

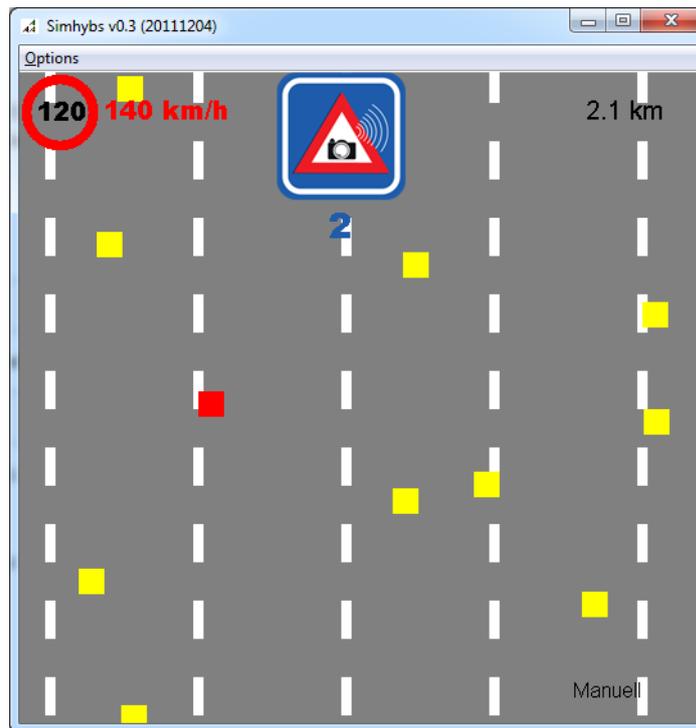
## 11.1 Simhybs-Oberfläche und Handlungsmodell

Wie im Abschnitt zuvor erwähnt, wurde SIMHYBS bereits für vorherige Forschungsarbeiten verwendet. Die folgenden Erläuterungen zu SIMHYBS und der zugehörigen Oberfläche beziehen sich auf die neue Version (v0.3), die für die vorliegende Arbeit erstellt wurde. Die Neuerungen umfassen vornehmlich eine Verschönerung der Benutzeroberfläche, deswegen wird auf einen systematischen Vergleich der Programmversionen verzichtet und stattdessen direkt die neue Version vorgestellt und erläutert.

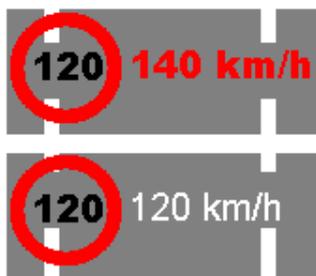
### 11.1.1 Oberfläche

In Abbildung 11.1 ist die Oberfläche dargestellt, wie sie sich den Probanden präsentiert. Der Hauptbildschirm 11.1a besteht aus einer an eine mehrspurige Straße erinnernde Struktur mit grauer Hintergrundfarbe und gestrichelten weißen Linien, die Spuren andeuten. Die Spuren haben innerhalb der Simulation keine Semantik, sie dienen nur der visuellen Unterstützung. Auf der Straßenstruktur bewegen sich unterschiedlich gefärbte Autos. Autos in gelber Farbe stellen erratische Fahrer dar, die sich unter Berücksichtigung der erlaubten Höchstgeschwindigkeit über die Oberfläche bewegen. Sie dienen gegenüber dem hybriden Fahrer, der als rotes Quadrat dargestellt ist, als Hindernis und sollen die Fahraufgabe erschweren. Das als rotes Quadrat dargestellte Fahrzeug kann von den menschlichen Probanden zusammen mit einem Fahrerassistenzsystem gesteuert werden. Die konkrete Handlungsverteilung hängt von dem eingestellten Modus (s. nächster Abschnitt) ab. Weiter findet sich auf der Oberfläche links oben eine in rot eingekreiste Zahl für die aktuell *zulässige Höchstgeschwindigkeit*. Im vorliegenden Beispiel beträgt diese 120 km/h.

Neben der aktuell zulässigen Höchstgeschwindigkeit wird rechts daneben die *aktuelle Geschwindigkeit* des roten Fahrzeugs angezeigt. Ist die momentane Geschwindigkeit höher als die zulässige Höchstgeschwindigkeit, so färbt sich die Geschwindigkeitsangabe rot, ansonsten werden die Zahlen in weißer Schrift angezeigt (Abbildung 11.1b). Im rechten oberen Bereich ist in schwarzer Schrift die bisher zurückgelegte Strecke hinterlegt. Im mittleren oberen Bereich der Oberfläche sind aktuelle Ereignisse verzeichnet. Sofern der Proband für längere Zeit die zulässige Höchstgeschwindigkeit überschreitet, kann er „geblitzt“ werden. Dies wird durch ein Radarkontrollsymbol (s. Abbildung 11.1a) angezeigt. Direkt unterhalb des Radarkontrollsymbols befindet sich die Anzahl der bisher geahndeten Geschwindigkeitsüberschreitungen. Als weiteres Ereignissymbol wird im Falle eines Zusammenstoßes mit einem anderen Fahrzeug ein *Unfall-Symbol* angezeigt (s. Abbildung 11.1d). Direkt unterhalb des Symbols befindet sich die Anzahl der bisher erfolgten Zusammenstöße mit anderen Fahrzeugen.



(a) Hauptbildschirm



(b) Anzeige für Geschwindigkeitsüberschreitung (rot) und Einhaltung des Geschwindigkeit (weiß)



(c) Hinweis auf eine Radarkontrolle



(d) Hinweis auf einen Unfall

**Abbildung 11.1:** Elemente der SIMHYBS-Oberfläche

Schließlich wird im rechten unteren Teil der Oberfläche der aktuelle Betriebsmodus des Fahrerassistenzsystems bzw. der *Modus der Handlungsverteilung* angezeigt. Im hiesigen Fall befindet sich die Simulation im Modus „Manuell“. Als weitere Modi sind noch „Geschwindigkeits-Auto“, „Lenk-Auto“ und „Voll-Auto“ möglich (vgl. Tabelle 11.1 auf Seite 259).

Abschließend werden in diesem Abschnitt noch das Handlungsmodell und die zur Verfügung stehenden Handlungsalternativen erläutert, die der interaktiven Computersimulation zu Grunde liegen.

### 11.1.2 Handlungsmodell

Auf Details zur internen Programmierung von SIMHYBS wird hier nicht näher eingegangen, da sich diese im Vergleich zu vorherigen Forschungsarbeiten (Fink 2009; Fink und Weyer 2011), trotz verbesserter Benutzeroberfläche, nicht geändert hat. Es handelt sich um eine agentenbasierte Simulation, die auf Mikro-Ebene der Agenten nach SEU-Logik agiert. Das SEU-Modell ist mit der Aufgabe der Probanden identisch. Diese besteht darin, unter Berücksichtigung der Ziele<sup>3</sup>

- Unfallvermeidung – **Cr**
- Geschwindigkeitseinhaltung – **Gw**
- Strecke machen – **Rnd**

in Zusammenarbeit mit einem Fahrerassistenzsystem ein Fahrzeug zu steuern. Dabei stehen den Probanden folgende Aktionen zur Verfügung, die über Tastaturkommandos gesteuert werden können:

- Auslenkung nach links – **L**
- Auslenkung nach rechts – **R**
- Keine Auslenkung – **Geradeaus**
- Geschwindigkeit erhöhen – **G+**
- Geschwindigkeit verringern – **G-**

$$A = \{L, R, G, G+, G-\}$$

Die Ziele der am Spiel beteiligten Akteure sind durch die Spielregeln vorgegeben. Es sind gemäß Instruierung der Probanden folgende Ziele zu verfolgen:

$$O = \{Cr, Gw, Rnd\}$$

Die Bewertungen der Ziele richten sich nach deren Relevanz aus subjektiver Akteursicht. Es wird dabei unterstellt, dass die Crashvermeidung ( $u_{Cr}$ ) als wichtigstes Ziel angesehen wird, gefolgt von der Einhaltung der Geschwindigkeitsbegrenzung ( $u_{Gw}$ ) und schließlich von dem Ziel Runden zu machen ( $u_{Rnd}$ ). Da im Rahmen dieser Arbeit keine umfangreichen Vorstudien hinsichtlich der Zielbewertungen eines typischen Akteurs durchgeführt wurden, wird folgende methodologische Annahme hinsichtlich der Ziele und ihrer Bewertungen getroffen:

$$u_{Cr}^3 > u_{Gw}^2 > u_{Rnd}^1$$

<sup>3</sup>Zu Konsistenz Zwecken wird auf die Abkürzungen aus vorherigen Veröffentlichungen zurückgegriffen. Dabei ist Unfallvermeidung mit *Cr* (Crash-Vermeidung) und Strecke machen mit *Rnd* (Runden machen) abgekürzt.

Ob die Zielgewichtungen in dieser Form haltbar sind, lässt sich durch entsprechende Befragungsmethoden ex post klären. Auf diese Weise können dann z. B. Differenzen zwischen Erwartung und Implementierung als empirische Frage in den Blick genommen werden.

Gemäß Tabelle 11.1 sind vier Modi der Handlungsverteilung möglich. Die Fahrmodi bedeuten nur, dass sich die Aktionen auf Mensch und Technik verteilen. Eine Berücksichtigung aller Ziele gilt unabhängig vom eingestellten Modus der Handlungsverteilung.

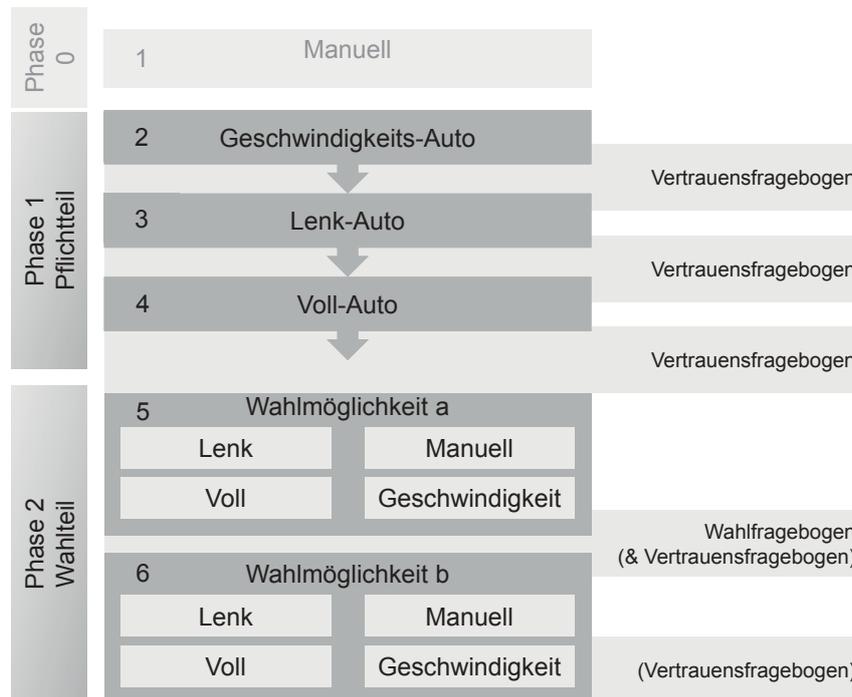
<i>Geschwindigkeits-Auto</i>	–	Das Assistenzsystem übernimmt die Erhöhung und Verringerung der Geschwindigkeit. {G+, G–}
<i>Lenk-Auto</i>	–	Das Assistenzsystem übernimmt die Aktionen Auslenkung nach links, rechts und keine Auslenkung. {L, R, G}
<i>Manuell</i>	–	Das Assistenzsystem greift nicht aktiv ein, sondern warnt lediglich durch ein rot gefärbtes Cockpit bei einer Geschwindigkeitsüberschreitung. {}
<i>Voll-Auto</i>	–	Das Assistenzsystem übernimmt alle Aktionen selbst, der Proband hat aber die Möglichkeit einzugreifen, damit die Aktionsübernahme durch das Assistenzsystems für einige Zeit außer Kraft zu setzen und die Aktionen selbst zu übernehmen. {L, R, G, G+, G–}

**Tabelle 11.1:** Mögliche Fahrmodi. Die geschweiften Klammern {} geben die Aktionen an, die in dem jeweiligen Modus vom Assistenzsystem übernommen werden.

## 11.2 Design des Simhybs-Experiments

Der gesamte Ablauf des Experiments ist in Abbildung 11.2 dargestellt. Nach einem Eingewöhnungsfahren (Phase 0) im manuellen Modus zum Kennenlernen der Simulation mussten die Probanden zwei weitere Versuchsphasen durchlaufen: einen *Pflichtteil* und einen *Wahlteil*. Im Pflichtteil mussten zunächst alle drei möglichen Automationsmodi durchlaufen werden: „Geschwindigkeits-Auto“, in dem das Fahrerassistenzsystem (FAS) die Beschleunigungs- und Verlangsamungsaktionen übernimmt, „Lenk-Auto“, in dem das FAS die Links-, Rechts- und Geradeaus-Aktionen übernimmt, sowie schließlich „Voll-Auto“, in dem das FAS sämtliche Aktionen übernimmt, der Proband aber die Möglichkeit hat einzugreifen.

Nach jedem Modus, bei dem Unterstützung durch das FAS gegeben wurde, wurde mit einem Fragebogen („Vertrauensfragebogen“) Vertrauen in Anlehnung an die Skala von Jian u. a. (2000) gemessen sowie zugeschriebene *U*-Werte für die im



**Abbildung 11.2:** Simhybs Experimentaldesign

Spiel relevanten Ziele ermittelt. Zwischen den beiden Wahlmodi wurde ein „Wahlfragebogen“ eingebracht, mit dem die Entscheidung für den im ersten Wahlteil gewählten Modus ergründet werden sollte. Sofern bei Wahlmöglichkeit a oder b ein (teil-)automatischer Modus gewählt wurde, also Voll-Auto, Lenk-Auto oder Geschwindigkeits-Auto, wurden in diesen Fällen zusätzlich Vertrauen und  $U$ -Werte per „Vertrauensfragebogen“ ermittelt. Am Ende der insgesamt sechs Versuchsdurchläufe mussten die Probanden noch Angaben zu Soziodemografika machen.

Tabelle 11.2 zeigt die ins Deutsche übersetzte Vertrauensskala von Jian u. a. (2000), die im Rahmen der SIMHYBS-Experimente verwendet wurde (vgl. Abschnitt 10.2.4). Die Richtung der unterschiedlichen Items entspricht der gleichen Richtung (positive vs. negative Aussagen), die auch bei dem zuvor vorgestellten Experiment MensaAssist verwendet wurde (vgl. Abschnitt 10.3.2).

### Datenaufzeichnung

Neben dem Fragebogen, den die Probanden nach den Versuchsläufen ausfüllen mussten, wurden während jeden Experimentdurchlaufs Daten aufgezeichnet. Beispielfhaft finden sich in Listing 11.1 Ausschnitte einer solchen Datei.

Nr.	Originalformulierung von Jian u. a. (2000)	Deutsche Übersetzung
1	The system is deceptive.	Das System leitet mich fehl.
2	I am suspicious of the system's intent, action, or output.	Ich bin misstrauisch, was die Ergebnisse des Systems angeht.
3	I am wary of the system.	Ich stehe dem System insgesamt skeptisch gegenüber.
4	I am confident in the system.	Insgesamt bin ich überzeugt von dem Fahrassistenzenzmodus.
5	The system provides security.	Das System gibt mir Sicherheit (bzgl. meinen verfolgten Zielen).
6	The system is reliable.	Das System ist zuverlässig.
7	I can trust the system.	Ich kann dem System vertrauen.
8	I am familiar with the system.	Die Funktionsweise des Systems ist mir vertraut.

**Tabelle 11.2:** Übersetzung der Items von Jian u. a. (2000) zur Vertrauensmessung. Die Nummern beziehen sich auf die Variablen `vertrauen_<nr>` aus dem Fragebogen.

**Listing 11.1:** Ausschnitt aus der Datenaufzeichnung

```

<statistics>
  <experiment>J,2,2</experiment>
  <text key="intervention" value="false"/>
  <text key="agency" value="FAS_STEERING"/>
  <summary>
    <distanceTravelled>4.534350000000025</distanceTravelled>
    <crashes>2</crashes>
    <speedings>1</speedings>
    <strokes>26</strokes>
  </summary>
  <strokes>
    <stroke key="down" time="20.0"/>
    ...
    <stroke key="up" time="142.0"/>
    ...
    <stroke key="left" time="431.0"/>
    ...
  </strokes>
  <crashes>
    <crash time="894.0"/>
    <crash time="2371.0"/>
  </crashes>
</statistics>

```

Für die spätere Auswertung sind insbesondere folgende Elemente von Bedeutung:

- Modus: `<text key="agency">` – zeichnet den eingestellten Modus der Handlungsverteilung auf. `FAS_STEERING` steht dabei für den Modus Lenk-Auto.
- Zurückgelegte Strecke: `<distanceTravelled>` – zeichnet die zurückgelegte Strecke je Durchlauf auf.
- Anzahl Crashes: `<crashes>` – zeichnet die Anzahl der Zusammenstöße mit anderen Fahrzeugen auf.
- Anzahl Geschwindigkeitsüberschreitungen: `<speedings>` – enthält die Anzahl geahndeter Geschwindigkeitsüberschreitungen.
- Anzahl Tastaturanschläge: `<strokes>` – die Gesamtzahl aller Tastaturanschläge

Weiterhin umfasste die Datenaufzeichnung eine Speicherung *jeder* Tastaturaktion mit ihrem zugehörigen Zeitpunkt (`<stroke key="x" time="y" />`) sowie eine Aufzeichnung von Zusammenstößen mit ihren zugehörigen Zeitpunkten (`<crash time="2371.0"/>`).

## 11.3 Auswertung

In Abschnitt 11.3.1 wird zunächst ein Überblick über die Zusammensetzung der Stichprobe gegeben. Es folgt (11.3.2) die Erläuterung zur Vertrauensmessung für SIMHYBS. Es werden unterschiedliche Modi der Handlungsverteilung in vergleichender Perspektive dargestellt (11.3.3) sowie Zusammenhänge zwischen Vertrauen und Zielintensitäten berechnet (11.3.4). Es schließen sich mit Abschnitt 11.3.5 Überlegungen zur Entwicklung von Vertrauen sowie ihrer Operationalisierung an, um das Auswertungskapitel mit einer umfangreichen Auswertung zu unterschiedlichen Vertrauensentwicklungstypen (11.3.6) zu beschließen, die ihrerseits handlungstheoretisch beschrieben werden.

### 11.3.1 Überblick über die Stichprobe

Die Untersuchung wurde mit insgesamt 100 Probanden durchgeführt. Gemäß des Ablaufplans in Abbildung 11.2 ist unmittelbar ersichtlich, dass sich für die unterschiedlichen Modi der Handlungsverteilung unterschiedliche Fallzahlen ergeben können, da die Probanden nach Ablauf des Pflichtteils die Wahl zwischen vier Modi der Handlungsverteilung hatten – selbst der manuelle Modus konnte ausgewählt werden.

Fahr-Modus	Pflichtteil	Wahlteil		$\sum_{\text{Modus}}$
		1.	2.	
Geschwindigkeit-Auto	100	21	28	149
Lenk-Auto	100	9	13	122
Voll-Auto	100	12	21	133
Manuell	0	58	38	96
$\Sigma$	300	200		500

**Tabelle 11.3:** Fallzahlen für die jeweiligen Modi, gruppiert nach Pflicht- und Wahlteil

### 11.3.2 Vertrauensmessung

Analog zur Auswertung des MensaAssist-Experiments (vgl. Abschnitt 10.3.2) wird als Methode auch hier eine explorative Faktoranalyse mit Varimax-Rotation angewendet, um zu überprüfen, ob die Zusammenfassung der acht Vertrauensfragen zu einem Faktor „Vertrauen“ zulässig ist.

Auch hier stellt sich die Variable `vertrauen_08` („Die Funktionsweise des Systems ist mir vertraut.“) als weitgehend ungeeignet dar. Zwar bildet sie nicht wie im Falle von MensaAssist einen eigenen Faktor, weist allerdings bei einer einfaktoriellen Lösung eine sehr geringe Faktorladung auf, so dass entschieden wurde, den Faktor *Vertrauen* auch für SIMHYBS nur aus den Variablen `vertrauen_1...vertrauen_7` zu bilden (für Details zur Faktorenanalyse s. Anhang A.6.2). Damit sind die Faktoren aus den beiden Experimenten aus den gleichen Variablen gebildet, wenn auch mit leicht unterschiedlichen Faktorladungen.

Die Variablen `vertrauen_1` bis `vertrauen_7` eignet sich mit einem KMO-Wert von 0.888 gut bis sehr gut für eine Faktorenanalyse. Der Cronbachs-Alpha-Wert von 0.924 deutet zudem auf eine sehr gute interne Reliabilität der Vertrauensskala hin. Es ergeben sich die in Tabelle 11.5 dargestellten Faktorladungen.

Die Faktorwerte für den so gebildeten Faktor Vertrauen können laut Kolmogorov-Smirnov-Anpassungstest mit Signifikanzniveau  $\alpha = 0.05$  als annähernd normalver-

Variable	Ladung
<code>vertrauen_1</code>	-0.744
<code>vertrauen_2</code>	-0.806
<code>vertrauen_3</code>	-0.812
<code>vertrauen_4</code>	0.836
<code>vertrauen_5</code>	0.858
<code>vertrauen_6</code>	0.856
<code>vertrauen_7</code>	0.895

**Tabelle 11.5:** SIMHYBS- Faktorladungen für den Faktor Vertrauen ( $N = 404$ )

Variable	N	Statistiken	
<i>Alter</i>	100	Mittelwert	29.28
		Standardabweichung	8.728
		Median	26
		Min/Max	16/54
		25%/50%/75%-Quartil	24/26/31
<i>Geschlecht</i>	100	62 % weiblich, 38 % männlich	
<i>Abschluss</i>	143	Abitur/Fachabitur	15 %
		FH/Universität	56 %
		Berufsausbildung/Lehre	23 %
		Sonstige	6 %

**Tabelle 11.4:** Soziodemografika zum Experiment SIMHYBS

teilt angenommen werden<sup>4</sup>. Weitere statistische Kenngrößen für die Faktorwerte lauten: Minimum:  $-2.095$ , Maximum  $1.665$ , Median:  $0.0574$ , Mittelwert:  $0$ , Standardabweichung  $1$ .

Die Faktorwerte lassen sich unter Vorgabe von vier gewünschten Gruppen auch hier gemäß ihrer Perzentile gleichmäßig aufteilen (Histogramm s. Anhang A.6.2):

- Vertrauen *sehr gering*:  $< -0.7587$  ( $N_1 = 99$ )
- Vertrauen *gering*:  $> 0.7587 \leq 0.0574$  ( $N_2 = 103$ )
- Vertrauen *mittel*:  $> 0.0574 \leq 0.8195$  ( $N_3 = 101$ )
- Vertrauen *stark*:  $> 1.665$  ( $N_4 = 101$ )

Ob sich der Faktor Vertrauen zwischen unterschiedlichen Modi der Handlungsverteilung unterscheidet, wird der folgende Abschnitt zeigen. Da im Pflichtteil jeder Modus nur einmal vorkam, lassen sich mit einer differenzierenden Untersuchung der Modi auch Aussagen über die Entwicklung von Vertrauen treffen (vgl. Abschnitt 11.3.5).

### 11.3.3 Automationsmodi im Vergleich

Wie aus dem Abschnitt zum Experimentaldesign deutlich geworden ist, erfolgten die Experimente in unterschiedlichen Modi der Handlungsverteilung. Zur Gewährleistung

<sup>4</sup>Es ergibt sich für die Nullhypothese „Die Verteilung der Werte ist *nicht* normalverteilt“ eine asymptotische Signifikanz von  $p = 0.141$ . Damit ist die Nullhypothese für die übliche Forderung  $p < 0.05$  zu verwerfen, es gilt die Alternativhypothese der annähernden Normalverteilung.

Modus	$\bar{X}$	Min	Max	95%-CI Vertrauen	$N$
Geschwindigkeits-Auto	0.1878	-1.8502	1.6650	[0.0118, 0.3638]	100
Lenk-Auto	-0.6007	-2.0950	1.6650	[-0.8098, -0.3916]	100
Voll-Auto	0.2312	-1.9477	1.6650	[0.0523, 0.4101]	100
Wahlmodus 1	0.3201	-1.5534	1.6650	[0.0452, 0.5950]	42
Wahlmodus 2	0.0762	-2.0950	1.6650	[-0.1610, 0.3133]	62

**Tabelle 11.6:** Mittelwerte und 95%-Konfidenzintervalle (CI) je Modus für den Faktor *Vertrauen*

einer validen Interpretation der Vertrauenskonstrukte und SEU-Modelle ist es daher zunächst erforderlich, darzustellen, inwieweit sich die unterschiedlichen Automationsmodi hinsichtlich verschiedener Kenngrößen voneinander unterscheiden.

### Vertrauen in unterschiedlichen Modi der Handlungsverteilung

(SH 2)  
?

Sofern sich die unterschiedlichen Modi der Handlungsverteilung hinsichtlich des Vertrauenswertes (genauer: Faktorwert für den zuvor beschriebenen Faktor Vertrauen) voneinander unterscheiden, bedeutet dies zunächst, dass die zugeschriebenen Vertrauenswerte im Sinne der These (SH 2) nicht ausschließlich auf unterschiedliche Dispositionen der betrachteten menschlichen Akteure zurückzuführen sind. Sinnvoll erscheint eine Kombination von Akteurs- und Systemeigenschaften zur Erklärung von Vertrauen in autonome Technik. Diese These wird bei den folgenden Auswertungen stets mit berücksichtigt und kann erst am Ende als weitgehend überprüft oder widerlegt gelten, da sich die jeweiligen system- und akteurspezifischen Einflüsse im Verlauf der weiteren Auswertungen herauskristallisieren werden.

In Tabelle 11.6 sind die mittleren Faktorwerte für die jeweiligen Modi aufgelistet.

Neben der Angabe von Mittelwerten sind zusätzlich 95%-Konfidenzintervalle angegeben. Die Berechnung von Konfidenzintervalle ist vor allem zur besseren Interpretation von Mittelwerten erforderlich. Überschneiden sich Konfidenzintervalle nicht, kann mit einer gewissen Irrtumswahrscheinlichkeit (meist 5%) davon ausgegangen werden, dass sich die Mittelwerte zweier Variablen tatsächlich voneinander unterscheiden.

In Bezug auf die Pflichtmodi lässt sich feststellen, dass der Modus Lenk-Auto sehr deutlich aus dem Rahmen fällt. Der Vertrauenswert ist in diesem Modus mit einem Mittelwert von  $-0.6007$  deutlich niedriger als in den anderen Modi und unterscheidet sich auch unter Beachtung der überschneidungsfreien Konfidenzintervalle von den anderen Modi. Voll-Auto fällt zwar hinsichtlich des Mittelwertes im Pflichtteil mit einem Vertrauenswert von  $0.2312$  am besten aus, dieser Unterschied ist im Vergleich zu Geschwindigkeits-Auto jedoch gering, gerade wenn man die stark überschneidenden

Konfidenzintervalle betrachtet. Der erste Wahlmodus<sup>5</sup> erhält mit einem Mittelwert von 0.3201 den höchsten Vertrauenswert, wobei ein echter Unterschied gemäß überschneidungsfreien Konfidenzintervalle nur zu dem Modus Lenk-Auto konstatiert werden kann. Trotz sich überschneidender Konfidenzintervalle erscheint der Unterschied zwischen Wahlmodus 1 und Wahlmodus 2 recht deutlich. Eine Überprüfung dieser Unterschiedlichkeitshypothese mittels T-Test mit Irrtumswahrscheinlichkeit  $\alpha = 0.05$  kann dieses Ergebnis allerdings nicht bestätigen: Die Signifikanz des T-Test beträgt 0.184 (erforderlich wären mindestens 0.1), das 95%-Konfidenzintervall für die Mittelwertdifferenzen  $[-0.14312, 0.60564]$  enthält die Null, somit ist die Nullhypothese, dass die Mittelwerte unterschiedlich sind, zu verwerfen.

Für weitere Auswertungen ist somit ein besonderes Augenmerk auf den Modus Lenk-Auto zu werfen, da der Automation in diesem Modus deutlich weniger vertraut wird als in den anderen Modi. Da dieser Zusammenhang für alle Akteure gleichermaßen gilt, kann mit Blick auf (SH 2) vorläufig festgehalten werden, dass systemspezifische Eigenschaften die Vertrauensmessung beeinflussen. Ob und in welchem Maße subjektspezifische Eigenschaften Einfluss nehmen, wird noch zu klären sein.

### Unterscheidung der Modi anhand des semantischen Differenzials

Die Methode des semantischen Differenzials (vgl. Osgood 1952: 222ff.) wurde vor allem zur Messung der Bedeutung von Begriffen entwickelt. Im Fall der vorliegenden Untersuchung wird dieses ergänzend verwendet, um die Einstellung der Probanden zur Automation über den gesamten Experimentverlauf zu beobachten. Eingeleitet werden die Bewertungen zum semantischen Differenzial mit der Formulierung „Bitte beurteilen Sie den soeben verwendeten Modus mit seinen spezifischen Aktionen nach den folgenden Kriterien“ (s. Abbildung 11.3).

Mittels Faktorenanalyse<sup>6</sup> (für Details s. Anhang A.6.3) lassen sich die einzelnen Items zu einem Faktor zusammenfassen. Betrachtet man nun die Werte für das semantische Differenzial für die unterschiedlichen Modi der Handlungsverteilung kann damit eine Aussage über die wahrgenommene Qualität der Automation getroffen werden. Die sich ergebenden Werte sind in Tabelle 11.7 aufgelistet.

Für das semantische Differenzial (im Folgenden vereinfachend *Bewertung* genannt) ergibt sich ein ganz ähnliches Bild wie für den Faktor Vertrauen, was auf einen engen Zusammenhang zwischen Modus-Bewertung und Vertrauen hindeutet (s. nächster Abschnitt). Der Modus Lenk-Auto wird mit einer mittleren Bewertung von  $-0.6726$

<sup>5</sup>Zu beachten ist, dass die Mittelwerte zu den beiden Wahlmodi nur eine bedingte Aussagekraft haben, da die Probanden sich dort für einen der vier Handlungsverteilungsmodi entscheiden konnten. Die Berechnung eines Mittelwerts *ohne* Berücksichtigung der wählbaren Modi lässt mögliche Unterschiede verschwimmen. In späteren Auswertungen werden die beiden Wahlmodi weiter differenziert.

<sup>6</sup>U. a. hat sich dabei gezeigt, dass sich aus den Items idealerweise genau ein Faktor extrahieren lässt. Die Stichprobeneignung kann mit einem KMO von 0.918 kann als hervorragend angesehen werden.

		3. Bitte beurteilen Sie den soeben verwendeten Modus mit seinen spezifischen Aktionen nach den folgenden Kriterien.					
		(Bitte machen Sie nur ein Kreuz pro Zeile)					
		1	2	3	4	5	
3.1	positiv	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	negativ
3.2	interessant	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	uninteressant
3.3	leicht bedienbar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	schwer bedienbar
3.4	vertrauenswürdig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	nicht vertrauenswürdig
3.5	hilfreich	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	nicht hilfreich
3.6	hohe Qualität	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	niedrige Qualität
3.7	nachvollziehbar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	undurchschaubar
3.8	zuverlässig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unzuverlässig
3.9	sicher	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unsicher

Abbildung 11.3: Fragebatterie zum semantischen Differenzial

Modus	$\bar{X}$	Min	Max	95%-CI Sem. Dif.	$N$
Geschwindigkeits-Auto	0.1119	-1.9488	1.5659	[-0.0516, 0.2755]	100
Lenk-Auto	-0.6726	-2.4088	1.5659	[-0.8775, -0.4678]	100
Voll-Auto	0.3404	-1.5498	1.5659	[0.1675, 0.5133]	100
Wahlmodus 1	0.3497	-1.6033	1.5659	[0.0979, 0.6015]	42
Wahlmodus 2	0.1185	-2.6941	1.5659	[-0.1421, 0.3790]	62

Tabelle 11.7: Mittelwerte und 95%-Konfidenzintervalle (CI) je Modus für das semantische Differenzial (Sem. Dif.)

mit Abstand am schlechtesten bewertet – auch unter Berücksichtigung der überschneidungsfreien Konfidenzintervalle. Weitere überschneidungsfreie Konfidenzintervalle sind nicht vorhanden. Allerdings ergibt ein T-Test mit  $\alpha = 0.05$ , dass sich die Bewertung für den Modus Geschwindigkeits-Auto signifikant von der Bewertung für Voll-Auto unterscheidet<sup>7</sup>: Der Modus Geschwindigkeits-Auto wird zwar mit einem Wert von 0.1119 knapp über Null noch positiv bewertet, allerdings deutlich schlechter als der Modus Voll-Auto mit 0.3404.

Die Betrachtung des semantischen Differenzials je Modus deutet ebenso wie die Auswertung des Faktors Vertrauen darauf hin, dass auf den Modus Lenk-Auto ein besonderes Augenmerk zu richten ist, da dieser auch bei der Bewertung mittels des semantischen Differenzial am schlechtesten abschneidet. Demnach lässt sich durch die kombinierte Betrachtung von Vertrauen und semantischem Differenzial ein Zusammenhang von Automationsqualität und Vertrauen untersuchen. Mit Blick auf (SH 2)

<sup>7</sup>Signifikanz  $p = 0.058$  ( $p < 0.1$  oder weniger wird in der Regel für die Ablehnung der Nullhypothese „Die Mittelwerte aus den empirischen Verteilungen sind identisch“ verlangt), das 95%-Konfidenzintervall für die Mittelwertunterschiede lautet [-0.46494, 0.00810] – auch wenn die Null knapp enthalten ist, scheint es berechtigt, einen Unterschied zwischen beiden Modi anzunehmen.

kann auch hier festgestellt werden, dass es offenbar systemspezifische Einflüsse auf Vertrauen gibt, die unabhängig von den betrachteten Akteuren sind.

### Der Zusammenhang von Vertrauen und semantischem Differenzial

Ein erster struktureller Vergleich der Tabellen 11.6 und 11.7 legt auf Grund der ähnlichen Faktorwerte je Modus einen Zusammenhang beider Größen nahe.

Überprüft man zunächst die Korrelation über alle Modi ergibt sich eine auf dem Niveau 0.01 signifikante Korrelation von 0.830 – ein enorm hoher Wert. Eine separate Überprüfung der Korrelationen je Modus ergibt für *jeden* Modus eine signifikante Korrelation von Vertrauen und dem semantischen Differenzial. Die stärkste Korrelation von 0.886 ergibt sich für Wahlmodus 2, die schwächste mit 0.746 für den Modus Lenk-Auto.

Inhaltlich ist dieser Zusammenhang wenig überraschend, da sich die Items in ihrer Formulierung sehr ähneln. Betrachtet man die Stärke der Korrelationen für die untersuchten Modi getrennt, lassen sich zwei geringfügig unterschiedliche Korrelationsstärken feststellen, allerdings sind die Unterschiede so gering, dass diese nicht weiter inhaltlich interpretiert werden können – für alle Modi besteht eine starke Korrelation zwischen Vertrauen und semantischem Differenzial.

Die Korrelation zwischen Vertrauen und semantischem Differenzial kann so gedeutet werden, dass insbesondere Kompetenzvertrauen im Sinne von Nooteboom (2002)<sup>8</sup> beim Zusammenspiel von Mensch und autonomer Technik eine wichtige Rolle spielt.

### SEU-Modelle in unterschiedlichen Modi der Handlungsverteilung

Nach jedem gefahrenen Modus wurden die Probanden nach ihrer Einschätzung befragt, welche Ziele das Assistenzsystem in dem jeweiligen Modus verfolgt hat. Es wurde also das von den Probanden (H) für das Assistenzsystem (NH) zugeschriebene SEU-Modell,  $SEU_{H \rightarrow NH}$  (vgl. Abschnitt 8.2.2), abgefragt. Eine Betrachtung der Zielintensitäten je Modus macht deutlich, dass die Probanden bestimmten Modi der Handlungsverteilung eine charakteristische Präferenzreihenfolge für die Ziele zuordneten. Tabelle 11.8 fasst die zentralen Ergebnisse zusammen.

Deutlich zu sehen ist, dass sich die von Fink und Weyer (2011: 116f.) nachgewiesene Vermischung von Aktions- und Zieldelagation auch in der vorliegenden Stichprobe wiederzufinden ist. Im Modus Geschwindigkeits-Auto dominiert das Ziel  $u_{gw}$  (Geschwindigkeitseinhaltung) mit einer durchschnittlichen Zielintensität von 4.45, im Vergleich zu 2.43 für  $u_{cr}$  und 2.43 für  $u_{rnd}$ . Dies ist deswegen überraschend, da auch mit einer Verlangsamung oder Beschleunigung Zusammenstöße vermieden werden können. Eine ähnliche Beobachtung kann bei Lenk-Auto gemacht werden: Dort dominiert das zugeschriebene Ziel  $u_{cr}$ , obwohl bei einer Auslenkung des Fahrzeugs nach

<sup>8</sup>Vgl. dazu ausführlich Abschnitt 2.2.3.

<b>Modus</b>	$u_{cr}$	$u_{gw}$	$u_{rnd}$	$N$
Geschwindigkeits-Auto	<b>2.43</b> [2.14, 2.71] <sup>1</sup>	<b>4.45</b> [4.26, 4.64]	<b>2.42</b> [2.17, 2.67]	100
Lenk-Auto	<b>3.51</b> [3.22, 3.80]	<b>1.81</b> [1.54, 2.08]	<b>2.53</b> [2.29, 2.77]	100
Voll-Auto	<b>4.54</b> [4.39, 4.69]	<b>4.63</b> [4.49, 4.77]	<b>3.32</b> [3.05, 3.59]	100
Wahlmodus 1	<b>3.40</b> [2.94, 3.45]	<b>4.21</b> [3.83, 4.59]	<b>3.05</b> [2.61, 3.49]	42
Wahlmodus 2	<b>3.39</b> [2.99, 3.78]	<b>4.05</b> [3.70, 4.40]	<b>3.10</b> [2.75, 3.44]	62
WM 1, Geschwindigkeits-Auto	<b>2.57</b> [1.97, 3.17]	<b>4.38</b> [4.01, 4.75]	<b>2.86</b> [2.31, 3.40]	21
WM 1, Lenk-Auto	<b>3.22</b> [2.38, 4.06]	<b>2.89</b> [1.53, 4.24]	<b>3.33</b> [2.32, 4.35]	9
WM 1, Voll-Auto	<b>5.00</b> [5.00, 5.00]	<b>4.92</b> [4.73, 5.10]	<b>3.17</b> [1.99, 4.34]	12
WM 2, Geschwindigkeits-Auto	<b>2.54</b> [1.95, 3.09]	<b>4.56</b> [4.16, 4.77]	<b>3.07</b> [2.53, 3.61]	13
WM 2, Lenk-Auto	<b>3.54</b> [2.45, 4.63]	<b>2.38</b> [1.24, 3.53]	<b>2.69</b> [1.79, 3.59]	9
WM 2, Voll-Auto	<b>4.43</b> [4.12, 4.74]	<b>4.52</b> [4.21, 4.83]	<b>3.38</b> [2.83, 3.93]	21

Einige Probanden haben den manuellen Modus gewählt und sind nicht aufgeführt.

**Tabelle 11.8:** Dem Assistenzsystem zugeschriebene U-Werte je Modus

<sup>1</sup> Die Werte in eckigen Klammern bezeichnen die jeweiligen 95%-Konfidenzintervalle zu den darüberstehenden Mittelwerten.

links oder rechts auch die Geschwindigkeit geringfügig langsamer wird. Im direkten Vergleich der Modi Geschwindigkeits-Auto und Lenk-Auto ist auch festzustellen, dass die Probanden trotz der vertauschten Handlungsverteilung keine exakt getauschten Zielintensitäten zuschreiben. Im Modus Lenk-Auto wird dem Assistenzsystem mit einem Wert von 3.51 für  $u_{cr}$  eine deutlich geringere Zielintensität zugeschrieben als mit 4.45 im Modus Geschwindigkeits-Auto. Diese Beobachtung ist konsistent mit den Modus-Unterschieden in Bezug auf das semantische Differenzial (vgl. Tabelle 11.7 auf Seite 267). Betrachtet man den Modus Voll-Auto, in dem das Assistenzsystem sämtliche Aktionen übernimmt, so zeigt sich, dass sowohl  $u_{cr}$  mit 4.54 als auch  $u_{gw}$  mit 4.63 sehr hoch bewertet wird (bei entsprechend schmalen Konfidenzintervallen). Eine Interpretation der U-Werte aus Tabelle 11.8 für die beiden Wahlmodi ist nur begrenzt sinnvoll, da dort unterschiedliche Handlungsverteilungsmodi gewählt werden konnten und sich somit die Zielintensitäten stark vermischen dürften. Aus dem Grund wurden die beiden Wahlmodi noch einmal aufgesplittet in die tatsächlich gewählten Modi und ihre Zielintensitäten vergleichend aufgeführt. Verglichen mit den Modi aus dem Pflichtteil ergeben sich sehr ähnliche Zielintensitätsmuster, wenngleich die Konfidenzintervalle wegen der reduzierten Probandenzahl deutlich breiter ausfallen. Besonders bemerkenswert ist die enorm hohe Bewertung des Modus Voll-Auto im ersten Wahlmodus („WM 1, Voll-Auto“): Dort haben alle zwölf Probanden, die diesen Modus gewählt haben, das Ziel der Geschwindigkeitseinhaltung mit dem höchstmöglichen Wert von 5 bewertet. Eine ähnlich Bewertung ergibt sich im selbigen Modus für die Crashvermeidung  $u_{cr}$ : Mit einem Intensitätswert von 4.92 bei einem für diese geringe Probandenzahl sehr schmalen Konfidenzintervall von [4.73, 5.10] ist dieses Ziel sehr hoch bewertet.

Wir haben bereits gesehen, dass die Probanden je nach Modus unterschiedlich stark vertrauen (vgl. Tabelle 11.6). Nimmt man beide Beobachtungen mit den Unterschieden der Zielintensitäten zusammen, so kann man folgern, dass intentionales Vertrauen im Sinne von Nooteboom (vgl. Abschnitt 2.2.3) neben dem bereits genannten Kompetenzvertrauen ein wichtigstes Vertrauenselement in Bezug auf autonome Technik darstellt. Die Zuschreibung von bestimmten Intentionen und ihren Intensitäten hängt, wie schon in These (SH 1) formuliert, eng mit dem Vertrauenskonzept zusammen. Dieser Zusammenhang wird im anschließenden Abschnitt 11.3.4 näher untersucht.

(SH 1)

?

### 11.3.4 Zusammenhang von Vertrauen und Zielintensitäten

Gemäß These (SH 1) ist zu prüfen, ob Vertrauen und zugeschriebene Zielintensitäten in einem positiven Zusammenhang stehen. Tabelle 11.9 führt alle signifikanten Korrelationen zwischen Vertrauen und Zielintensitäten (U-Werte) auf. Die Korrelationen werden nicht über die gesamte Stichprobe betrachtet, sondern differenziert nach den drei Pflichtmodi Geschwindigkeits-Auto, Lenk-Auto und Voll-Auto sowie

dem Wahlmodus 2<sup>9</sup>. Diese Unterscheidung erscheint sinnvoll, da im vorherigen Abschnitt (vgl. Tabelle 11.8) gezeigt wurde, dass sich die Einschätzungen der Probanden je nach Modus z. T. drastisch unterscheiden. Die Korrelationsstärken werden anhand der von Cohen (1992: 157) vorgeschlagenen Klassifizierung für Effektstärken bei Pearson-Korrelation gedeutet<sup>10</sup>.

		Korrelationen					
		FAC_Vertrauen.2	FAC_Vertrauen.3	FAC_Vertrauen.4	FAC_Vertrauen.5	FAC_Vertrauen.6	
Geschwindigkeit-Auto	FAC_Vertrauen.2	Pearson-Korrelation Sig. (2-seitig) N	1 .001 100	<b>,340**</b> .001 100	<b>,360**</b> .000 100	<b>,480**</b> .001 42	<b>,295**</b> .020 62
	u_gw.2	Pearson-Korrelation Sig. (2-seitig) N	<b>,275**</b> .006 100	-.097 .337 100	.016 .878 100	.050 .754 42	.223 .081 62
	u_rnd.2	Pearson-Korrelation Sig. (2-seitig) N	.210 .036 100	.046 .646 100	.027 .788 100	-.262 .094 42	.175 .175 62
Lenk-Auto	FAC_Vertrauen.3	Pearson-Korrelation Sig. (2-seitig) N	<b>,340**</b> .001 100	1 100	<b>,281**</b> .005 100	.254 .104 42	.167 .194 62
	u_gw.3	Pearson-Korrelation Sig. (2-seitig) N	.144 .152 100	<b>,286**</b> .004 100	.087 .390 100	.011 .947 42	-.090 .485 62
	u_cr.3	Pearson-Korrelation Sig. (2-seitig) N	.055 .589 100	<b>,470**</b> .000 100	.071 .485 100	.065 .683 42	.269 .034 62
Voll-Auto	FAC_Vertrauen.4	Pearson-Korrelation Sig. (2-seitig) N	<b>,360**</b> .000 100	<b>,281**</b> .005 100	1 100	<b>,577**</b> .000 42	.250 .050 62
	u_cr.4	Pearson-Korrelation Sig. (2-seitig) N	.086 .398 100	.125 .215 100	<b>,329**</b> .001 100	.132 .405 42	.066 .612 62
	FAC_Vertrauen.5	Pearson-Korrelation Sig. (2-seitig) N	<b>,480**</b> .001 42	.254 .104 42	<b>,577**</b> .000 42	1 42	.171 .404 26
Wahlmodus 1	FAC_Vertrauen.6	Pearson-Korrelation Sig. (2-seitig) N	.295 .020 62	.167 .194 62	.250 .050 62	.171 .404 26	1 62
	u_gw.6	Pearson-Korrelation Sig. (2-seitig) N	.161 .212 62	.037 .776 62	.041 .752 62	.173 .399 26	<b>,502**</b> .000 62
	u_rnd.6	Pearson-Korrelation Sig. (2-seitig) N	.116 .368 62	.047 .716 62	.104 .419 62	-.082 .692 26	.313 .013 62

\*\* . Korrelation ist bei Niveau 0,01 signifikant (zweiseitig).

\* . Korrelation ist bei Niveau 0,05 signifikant (zweiseitig).

**Tabelle 11.9:** Korrelation zwischen Vertrauensmessungen und U-Werten

Die Bezeichnungen FAC\_Vertrauen.x beziehen sich auf die Vertrauensmessungen in den jeweiligen Modi beginnend mit FAC\_Vertrauen.2 für Geschwindigkeits-Auto im Pflichtteil und endend mit FAC\_Vertrauen.6 für den Wahlmodus 2.

<sup>9</sup>Wahlmodus 1 kommt nicht vor, da für diesen keine Korrelationen zwischen Zielintensitäten und Vertrauen festgestellt wurden.

<sup>10</sup>Jacob Cohen (1992) schlägt in seinem breit rezipierten Artikel „A power primer“ vor, die Effektstärken einer Produktmomentkorrelation  $r$  wie folgt zu unterscheiden:  $0.1 < |r| \leq 0.3$ : gering,  $0.3 < |r| \leq 0.5$ : mittel,  $|r| > 0.5$ : hoch.

(SH 1)



Die Korrelationen zwischen den einzelnen Vertrauensmessungen werden in der dargestellten Tabelle ebenfalls deutlich und werden in Abschnitt 11.3.5 ausführlich behandelt. Hier werden zunächst Zusammenhänge zwischen Vertrauen und ausgewählten Zielbewertungen qualitativ bewertet:

Alle Korrelationen zwischen Vertrauen und Zielintensitäten sind positiv, was konsistent mit These (SH 1) ist, dass Vertrauen und zugeschriebene Zielintensitäten in einem positiven Zusammenhang stehen.

Hoch korreliert sind lediglich  $u_{gw}$  und Vertrauen im Wahlmodus 2 mit einer Korrelation von 0.502. Mittelmäßig stark korreliert ist Vertrauen mit  $u_{cr}$  im Modus Lenk-Auto mit 0.470 sowie mit  $u_{cr}$  im Modus Voll-Auto mit 0.329. Gering korreliert mit dem Faktor Vertrauen ist schließlich  $u_{gw}$  im Modus Geschwindigkeits-Auto (0.275) und in Modus Lenk-Auto (0.286).

Interessant an den gefundenen Korrelationen ist, dass sich diese allesamt auf die Ziele Crashvermeidung und Geschwindigkeitsbegrenzung einhalten beziehen. Beide Ziele sind eng mit der Benutzeroberfläche von SIMHYBS verknüpft: Bei Überschreitung der zulässigen Höchstgeschwindigkeit wird ein Warnsymbol angezeigt, so dass diese schnell von den Probanden feststellbar ist. Fällt die Erhöhung und Verringerung der Geschwindigkeit wie im Modus Geschwindigkeits-Auto und Voll-Auto in den Zuständigkeitsbereich des Assistenzsystems, so ist die Verantwortung für eine Geschwindigkeitsüberschreitung leicht dem Assistenzsystem zuschreibbar. Analog kann für die Einhaltung der zulässigen Höchstgeschwindigkeit argumentiert werden: Passt das Assistenzsystem die Geschwindigkeit gemäß der zulässigen Höchstgeschwindigkeit an, kann dies als Zielerfüllung gedeutet werden. Gleiches gilt für das Ziel Crashvermeidung: Zielerfüllungen und insbesondere Ziel-Nicht-Erfüllungen (Unfälle) sind leicht von den Probanden feststellbar.

Aus der Automationsforschung sind die negativen Auswirkungen von leicht erkennbaren Fehlfunktionen auf Vertrauen bekannt (vgl. Abschnitt 4.3.2), so dass gefolgert werden kann, dass vor allem die Ziele mit Vertrauen korrelieren, deren Erfüllung und Nicht-Erfüllung leicht überprüfbar ist. Fehlfunktionen im Sinne von unzureichend erfüllten Erwartungen und damit niedrigen U-Werten wirken sich negativ auf das Vertrauen in die Automation aus.

Welches Fazit ergibt sich aus dieser Korrelationsanalyse für die weitere Auswertung? Vertrauen und Zielintensitäten sind keine unabhängigen Konzepte, sondern es existieren je nach Automationsmodus mehr oder weniger starke Zusammenhänge. Dies bedeutet, ohne dadurch eine bestimmte Kausalitätsrichtung unterstellen zu wollen, dass sich Gruppen, die auf Grundlage unterschiedlicher Vertrauenswerte gebildet werden, anhand ihrer durchschnittlichen Zielintensitäten näher charakterisieren lassen (s. Abschnitt 11.3.6).

### 11.3.5 Entwicklung von Vertrauen

Im Gegensatz zum MensaAssist-Experiment, bei dem Vertrauen nur einmal am Ende gemessen wurde, ist es im Falle der SIMHYBS-Experimente durch die Messung von Vertrauen nach jedem Versuchsdurchlauf<sup>11</sup> möglich, Vertrauensentwicklungen in den Blick zu nehmen.

Im theoretischen Teil der Arbeit wurden unterschiedliche Vertrauensbegriffe untersucht. In praktisch allen betrachteten Forschungsrichtungen spielt anfängliches Vertrauen – Initialvertrauen – gemäß (FLA 8) eine mehr oder weniger entscheidende Rolle. Beim probabilistischen Vertrauensmodell (Abschnitt 3.1.2) als Standard einschätzung  $p^*$ , beim spieltheoretischen Vertrauensbegriff durch stärkere Berücksichtigung von Ausgangsvertrauen (Abschnitt 3.2.4) im Rahmen des iterierten Vertrauensspiels und schließlich auch in der Automationsforschung (Abschnitt 4.3.1). Als zusammenfassende Synthese, insbesondere der spieltheoretischen Untersuchungen zu Vertrauen sowie der behandelten Vertrauensstudien aus der Automationsforschung, hat sich gemäß (FLA 5) ergeben, dass für eine soziologische Analyse von Vertrauen in autonome Technik der interaktive Charakter zu berücksichtigen ist.

Den Überlegungen hinsichtlich der Berücksichtigung von Vertrauen in interaktiven Zusammenhängen schließen sich zwei Untersuchungsfragen, die in (SH 3) und (SH 4) formuliert sind, an: Welchen Einfluss hat das Initialvertrauen auf spätere Vertrauensmessungen? Allgemein gilt es also zu untersuchen, in wie weit sich vorangegangene Vertrauensmessungen auf spätere Messungen auswirken (SH 3). Zweitens wird geklärt, ob sich im Datenmaterial typische Vertrauenstrajektorien finden lassen, die sich gemäß These (SH 4) anhand typischer SEU-Modelle voneinander unterscheiden lassen.

Folgt man Gambetta (1988a: 234), s. Abschnitt 3.2.3, so ist einmal zerstörtes Vertrauen nur schwer wiederherstellbar. Dieser Annahme soll in Form typischer Vertrauenstrajektorien (s. Abschnitt 11.3.6) ebenfalls nachgegangen werden. Sofern sich ein Vertrauenseinbruch messen lässt, dürften sich die Probanden von ebendiesem im weiteren Verlauf nicht mehr erholen.

#### **Einfluss von Initialvertrauen auf spätere Messungen**

Gemäß (SH 3) wird zunächst der Einfluss von Vertrauensmessungen auf spätere Messungen untersucht.

In einem ersten Schritt ist es zunächst erforderlich den vorhandenen Datensatz zu aggregieren, so dass für jeden Probanden genau ein Fall erstellt wird. Damit ergeben

(SH 3) ?
-------------

<sup>11</sup>Vertrauen wurde genauer genommen nach jedem Versuchsdurchlauf gemessen, bei dem das FAS ein Teil der möglichen Aktionen übernimmt. So wurde in der Eingewöhnungsphase, in der das Fahrzeug manuell gesteuert wurde, kein Vertrauenswerte erhoben, ebenso nicht im Wahlteil, für den Fall, dass sich Probanden für den manuellen Modus entschieden haben.

sich dann exakt  $N = 100$  Fälle, bei denen jede Variable mit dem entsprechenden Durchlauf indiziert wird.

Zunächst einige Erklärungen zur Notation, die z. T. bereits vorher en passant verwendet wurde: Da der erste, in der Auswertung nicht näher beachtete Durchlauf, zunächst eine Eingewöhnungsphase im manuellen Modus umfasst, beginnt die Nummerierung mit .1. So bezeichnet z. B. die Variable `distance_travelled.1` die zurückgelegte Distanz im ersten Eingewöhnungsdurchlauf im manuellen Modus. Für die Auswertung relevant sind insbesondere die Variablen `FAC_Vertrauen.2` bis `FAC_Vertrauen.6`, wobei `FAC_Vertrauen.2` den Faktorwert für Vertrauen im ersten Modus des Pflichtteils (Geschwindigkeits-Auto) bezeichnet und `FAC_Vertrauen.6` den Faktorwert für den letzten Modus im Wahlteil. Die beiden Faktorwerte des Wahlteils `FAC_Vertrauen.5` und `FAC_Vertrauen.6` können für einige Probanden auch gar keinen Wert enthalten, falls diese diese für den manuellen Modus entschieden haben, in dem kein Fragebogen zum Vertrauen ausgeteilt wurde.

	FAC_v.2	FAC_v.3	FAC_v.4	FAC_v.5	FAC_v.6
FAC_v.2	1				
FAC_v.3	0.340**	1			
FAC_v.4	0.360**	0.281**	1		
FAC_v.5	0.480**		0.577**	1	
FAC_v.6	0.295*				1

**Tabelle 11.10:** Signifikante Korrelationen für Vertrauensmessungen in den jeweiligen Durchläufen.  
Signifikanzniveau \*\* -  $p < 0.01$ , \* -  $p < 0.05$

Für ein prinzipielles Verständnis der Entwicklung von Vertrauen während der Versuchsläufe werden zunächst Korrelationen zwischen den Faktorwerten berechnet. Die signifikanten Korrelationen (mindestens  $p < 0.05$ ) sind in Tabelle 11.10 eingetragen. Die gefundenen Zusammenhänge sind bemerkenswert und bestätigen die zuvor formulierte Annahme, dass Vertrauensmessungen mit den jeweils vorherigen Vertrauensmessungen in Zusammenhang stehen: Für jeden der Durchläufe drei bis sechs ist der Faktorwert für Vertrauen stets korreliert mit dem Faktorwert aus dem zweiten Durchlauf. Dieser Zusammenhang ist unabhängig vom jeweiligen Fahrerassistenzmodus. Das bedeutet, dass der erste Eindruck, den ein Proband von dem Fahrerassistenzsystem bei seiner ersten Interaktion mit diesem gewinnt, prägend für alle weiteren Abläufe ist. Ist das Vertrauen bei diesem „Erstkontakt“ am Ende der Simulationsdauer hoch, so wirkt sich dies auf die Vertrauensmessung für alle weiteren Abläufe aus. Umgekehrt gilt: Ist der erste Eindruck erst mal ruiniert und das Vertrauen dementsprechend niedrig, strahlt dieser Effekt auf alle weiteren Durchläufe aus.

Die Tabelle lässt sich noch weiter dahingehend interpretieren, dass neben dem ersten Eindruck zusätzlich die unmittelbare Erfahrung aus dem vorherigen Durchlauf zählt, während Vertrauensmessungen in anderen Zwischendurchläufen keine signifikanten Auswirkungen haben. Das Vertrauen in Durchlauf 4 ist mit dem Vertrauen vom Initialdurchlauf 2 und dem vorangegangenen Durchlauf 3 korreliert (mit 0.360 für 2 und 0.281 für den vorherigen Durchlauf 3). Besonders beachtenswert sind die Korrelationen in Durchlauf 5, dem ersten Wahlmodus. Das Vertrauen in diesem Durchlauf ist stark korreliert mit dem Initialdurchlauf 2 (0.480) und mit dem vorangegangenen Durchlauf 4 (0.577). Die Messung ist hingegen nicht korreliert mit dem Vertrauen im zwischenliegenden Durchlauf 3. Der letzte Durchlauf nimmt eine Sonderstellung ein, da dieser nur mit dem Initialdurchlauf korreliert ist. Eine Erklärung könnte sein, dass Probanden im letzten Durchlauf bewusst einen anderen Handlungsverteilungsmodus als im vorherigen Durchlauf 4 ausprobieren wollten.

Bezugnehmend auf (SH 3) kann konstatiert werden, dass der zuerst gemessene Vertrauenswert Einfluss auf alle folgenden Durchläufe hat. Weiter kann festgehalten werden, dass ein Zusammenhang besteht zwischen dem in Durchgang  $t$  und dem unmittelbar davor liegenden Vertrauenswert  $t - 1$ . Das Ergebnis ist deswegen bemerkenswert, weil sich die gemessenen Vertrauenskorrelationen auf den jeweils vorherigen Modus mit einer anderen Handlungsverteilung beziehen. These (SH 3) kann damit als bestätigt gelten, wie das folgende Regressionsmodell zusätzlich unterstreicht.

(SH 3)



Die gefundenen Zusammenhänge lassen sich auch in multiplen linearen Regressionsmodellen<sup>12</sup> darstellen:

$$\text{FAC\_Vertrauen.3} = -0.677 + 0.401 \cdot \text{FAC\_Vertrauen.2}$$

$$R^2 = 0.116$$

$$\text{FAC\_Vertrauen.4} = 0.266 + 0.304 \cdot \text{FAC\_Vertrauen.2} + 0.153 \cdot \text{FAC\_Vertrauen.3}$$

$$R^2 = 0.158$$

$$\text{FAC\_Vertrauen.5} = 0.203 + 0.303 \cdot \text{FAC\_Vertrauen.2} + 0.484 \cdot \text{FAC\_Vertrauen.4}$$

$$R^2 = 0.412$$

<sup>12</sup>Falls nicht anders angegeben sind die präsentierten Regressionsmodelle laut F-Test signifikant mit  $p < 0.001$ , ebenso wie die jeweiligen Einflussgrößen, die mit einem T-Test mit  $p < 0.05$  abgesichert werden. Zusätzlich wird zu jeder Regression  $R^2$  mit angegeben, das den Anteil der durch die Regressionsrechnung erklärten Varianz darstellt. Ein Wert von 1 bedeutet, dass 100 % der Varianz der abhängigen Variable durch die Regression erklärt werden können.

Für **FAC\_Vertrauen.6** lässt sich kein signifikantes Regressionsmodell berechnen. Die Ergebnisse decken sich mit der Betrachtung der Korrelationen.

Sämtliche Details zu den Regressionsrechnungen sind im Anhang A.6.2.1 zu finden.

### Beeinflussung von Vertrauen gleicher Fahrmodi

Wir haben gesehen, dass die Vertrauenswerte untereinander korreliert sind, obwohl sich die miteinander korrelierten Messungen auf potenziell unterschiedliche Modi der Handlungsverteilung beziehen<sup>13</sup>. Abschließend zur Untersuchung der Vertrauensentwicklung für die gesamte Stichprobe wird gezeigt, dass die Korrelationen von Vertrauensmessungen zwischen gleichen Modi noch stärker korreliert sind als Vertrauensmessungen zwischen unterschiedlichen Modi.

Im Pflichtteil wiederholen sich die Modi nicht, so dass dort keine Korrelationen zwischen gleichen Modi ausgerechnet werden können. Im Wahlteil hingegen haben die Probanden die Möglichkeit, zweimal einen selbst gewählten Modus (auch manuell) zu verwenden. Eine Berechnung der Einflussname zweier gleicher Fahrmodi kann folglich nur zwischen dem Pflicht- und dem Wahlteil erfolgen. Erachtet man Fallzahlen (vgl. Tabelle 11.3) von  $N > 20$  als für eine Korrelationsrechnung sinnvoll, so lassen sich folgende Vertrauenskonstellationen miteinander vergleichen:

- Pflichtteil Geschwindigkeits-Auto (GW)  $\sim$  Wahlteile (WT1, WT2) im Modus Geschwindigkeits-Auto:  
 $N_{GW \sim WT1} = 21$  (Überschneidungen mit Wahlteil 1) /  $N_{GW \sim WT2} = 28$  (Überschneidungen mit Wahlteil 2)
- Pflichtteil Voll-Auto (VA)  $\sim$  2. Wahlteil (WT2) im Modus Voll-Auto:  
 $N_{VA \sim WT2} = 21$

Für die Korrelation von **FAC\_Vertrauen.2** (Geschwindigkeits-Auto, Pflichtteil) und **FAC\_Vertrauen.5** (Geschwindigkeits-Auto, Wahlteil 1) mit  $N_{GW \sim WT1} = 21$  ergibt sich eine starke Korrelation von 0.876 auf einem Signifikanzniveau von  $p = 0.01$ . Für Wahlteil 2 ergibt sich eine Korrelation von 0.616 mit gleichem Signifikanzniveau und  $N_{GW \sim WT2} = 28$ .

Für den vollautomatischen Modus „Voll-Auto“ ergibt sich zwischen **FAC\_Vertrauen.4** (Voll-Auto, Pflichtteil) und **FAC\_Vertrauen.6** (Voll-Auto, Wahlteil) eine starke Korrelation von 0.847 mit  $N_{VA \sim WT2} = 21$ .

Welche Interpretation ergibt sich nun aus dem zunächst relativ technisch anmutenden Ergebnis? Eine Betrachtung von Korrelationen zwischen *gleichen* Modi, also z. B. Geschwindigkeits-Auto im Wahl- und im Pflichtteil, hat gezeigt, dass ein starker Zusammenhang der Vertrauensmessung im ersten Durchlauf des betrachteten Modus zu

<sup>13</sup>Korrelationen zwischen gleichen Modi sind implizit eingeflossen, da in den Wahlteilen aus den bereits zuvor absolvierten Modi gewählt werden konnte.

der zweiten Messung im selben Modus existiert. Verglichen mit der Korrelationen über alle Fälle (s. Tabelle 11.10) sind die Wirkzusammenhänge mit dieser Einschränkung noch stärker. Für die vollständigen Korrelationstabellen sei auch hier auf Anhang A.6.2.1 verwiesen.

### 11.3.6 Vertrauensentwicklungstypen

Im vorherigen Abschnitt ging es vor allem darum, einige globale Analysen hinsichtlich der Vertrauensentwicklung durchzuführen. Man kann diese Art der Analyse auch als *technikzentrierte* oder *automationszentrierte* Vertrauensanalyse bezeichnen, da sich diese an den technisch vorgegebenen Automationsmodi orientiert hat und modusspezifische Unterschiede herausgestellt hat.

Die in diesem Abschnitt vorgestellte Analyse stellt hingegen unterschiedliche Akteuertypen in den Vordergrund und untersucht, ob sich für ebendiese spezifische Muster finden lassen. Völlig losgelöst von den Automationsmodi kann diese Analyse freilich nicht erfolgen, da der je spezifische Automationsmodus, wie im vorherigen Abschnitt gezeigt, einen Einfluss auf das Vertrauen hat.

Ausgehend von der bestätigten These (SH 3), dass das Initialvertrauen spätere Vertrauensmessungen beeinflusst, stellt sich die Frage, ob sich typische Vertrauenstrajektorien gemäß These (SH 4) finden lassen, die beispielsweise einen kaum wiederherstellbaren Vertrauensverlust nachweisen können. Da zuvor festgestellt wurde, dass die Vertrauensmessungen auch stark von dem jeweiligen Automationsmodus abhängen, können zur Untersuchung dieser Fragestellung nur die Messungen der Durchläufe 2, 3 und 4 herangezogen werden. Zur Erinnerung: Der erste Durchlauf umfasste eine Eingewöhnungsphase im manuellen Modus, ab Durchlauf 5 begannen die Wahlmöglichkeiten für die Probanden, so dass eine Berücksichtigung der Wahlmodi bei der Ermittlung von Vertrauensentwicklungstypen nur zusätzliche Streuung hinzufügen würde und durch die mangelnde Überdeckung von Probanden in allen Modi die Probandenzahl radikal reduziert würde.

Zur automatisierten Ermittlung von Vertrauensentwicklungstypen wurde eine hierarchische Clusteranalyse mit den Faktorwerten `FAC_Vertrauen.2` bis `FAC_Vertrauen.4` durchgeführt (für Details s. Anhang A.6.5). Anhand des Elbow-Kriteriums stellte sich eine optimale Clusterzahl von vier heraus. In Abbildung 11.4 sind die Mittelwerte der vier Cluster für die Vertrauensmessungen in den Experimentdurchläufen 2, 3 und 4 grafisch und numerisch dargestellt.

Die Trajektorien der aus der Clusteranalyse ermittelten Vertrauensentwicklungstypen lassen sich, wie aus der Grafik abgeleitet werden kann, anhand von drei Aspekten charakterisieren:

1. *Ausgangsvertrauen*, das nach dem ersten Durchlauf in einem teilautomatisierten Modus gemessen wird.

(SH 4) ?
-------------

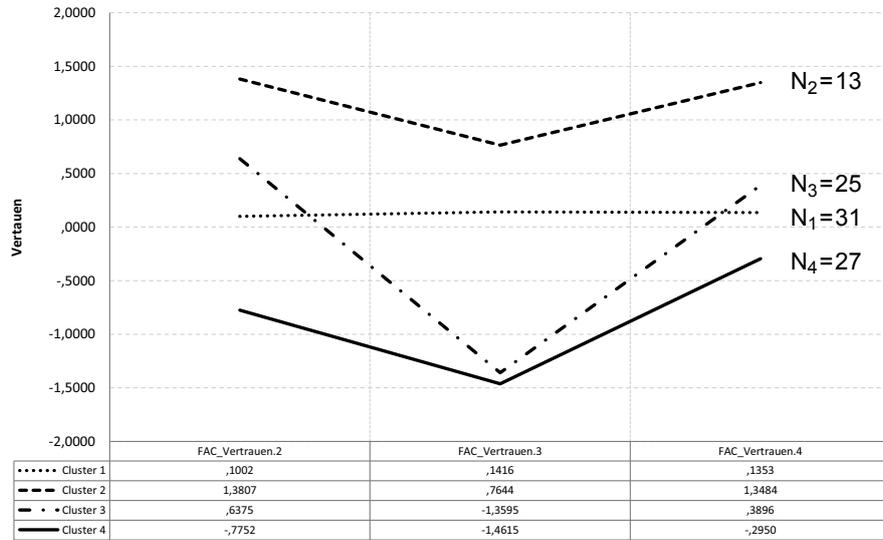


Abbildung 11.4: Grafische Darstellung der Vertrauensentwicklungstypen

2. *Vertrauenserschütterung* als Reaktion auf den dritten Durchlauf (Lenk-Auto), der von allen Probanden (ausgenommen Cluster 1) deutlich schlechter bewertet wird als die anderen Automationsmodi.
3. *Vertrauenswiederherstellung* nach dem schlecht bewerteten dritten Modus.

Anhand dieser drei Charakterisierungsmerkmale lassen sich die gefundenen Cluster näher beschreiben und benennen:

### Cluster 1: Gleichbleibende

Mit einem Umfang von  $N_1 = 31$  ist Cluster 1 das größte Cluster. Der Durchschnittsproband in diesem Cluster zeichnet sich dadurch aus, dass sich über die drei Versuchsläufe mit geänderten Automationsmodi keinerlei Änderungen bzgl. des gemessenen Vertrauens ausmachen lassen und daher als *Gleichbleibender* bezeichnet wird. Er hat zwar eine leicht positive Vertrauendenz (Vertrauenswerte knapp über Null), allerdings kann die Nähe zur Null und das nahezu vollständige Fehlen von Reaktionen auf unterschiedliche Automationsmodi so gedeutet werden, dass dem Gleichbleibenden entweder die Unterschiede zwischen den Modi nicht bewusst sind, oder ihm diese schlichtweg egal sind.

**Cluster 2: Unerschütterlich stark Vertrauende**

Die gemessenen Vertrauenswerte in allen betrachteten Clustern und Modi changieren zwischen  $+1.5$  und  $-1.5$ , so dass Cluster 2, dem insgesamt dreizehn Probanden zugeordnet wurden, sofort als Gruppe der stark Vertrauenden gedeutet werden kann. Mit einem Wert von  $1.3807$  nach dem ersten Automationskontakt im Modus Geschwindigkeits-Auto zeichnet sich das Cluster durch ein hohes Ausgangsvertrauen aus. Für Durchlauf 3 (Lenk-Auto) ist zwar ein geringer Vertrauenseinbruch festzustellen, allerdings stellt sich das Vertrauen nach dem insgesamt besser bewerteten Durchlauf vier wieder fast auf dem Ausgangsniveau, das nach dem zweiten Durchlauf gemessen wurde, ein. Für diese Gruppe erscheint daher der Begriff *unerschütterlich stark Vertrauende* gerechtfertigt.

**Cluster 3: Vertrauende und schnell Verzeihende**

Im Vergleich zu Cluster 2 ist das durchschnittliche Ausgangsvertrauen der 25 Probanden in Cluster 3 mit einem Wert von  $0.6375$  deutlich niedriger. Der Vertrauensverlust nach Durchlauf 3 ist in diesem Cluster enorm und sinkt auf den zweitniedrigsten überhaupt vorkommenden Wert von  $-1.3595$  ab. Dies kann so gedeutet werden, dass eine negativ bewertete Automation für diese Gruppe zu einem enormen Vertrauensverlust führt. Nach dem Verlust stellt sich das Vertrauen nach Durchlauf 4 wieder fast auf dem Niveau von Durchlauf 2 ein. Es lässt sich also folgern, dass das Vertrauen der Probanden in diesem Cluster zwar schnell erschüttert werden kann, jedoch ebenso schnell wiederhergestellt ist – *Vertrauende und schnell Verzeihende* erscheint als kurze Beschreibung des Clusters angemessen.

**Cluster 4: Generell Misstrauende**

Cluster 4 mit einem Umfang von 27 Probanden zeichnet sich durch prinzipielles Misstrauen aus. Unabhängig vom Automationsmodus werden ausschließlich negative Vertrauenswerte gemessen. Mit einem Ausgangsvertrauen von  $-0.7752$  nach dem zweiten Durchlauf startet diese Gruppe bereits mit einem hohen Maß an Misstrauen. Dieses wird dann noch einmal verstärkt, so dass nach Durchlauf drei mit einem Vertrauenswert von  $-1.4615$  der niedrigste Wert über alle Durchläufe und alle Cluster ermittelt werden kann. Trotz des offenbaren *generellen Misstrauens* wird im vierten Durchlauf in hohem Maße Misstrauen abgebaut, so dass mit einem Wert von  $-0.2950$  zwar immer noch Misstrauen vorherrscht, dieses jedoch geringer ausfällt als das gemessene Misstrauen nach Durchlauf zwei. Auf längere Sicht betrachtet wäre es denkbar, dass auch die Gruppe der Misstrauenden Vertrauen gewinnt (und in den positiven Wertebereich gelangt), sofern die autonome Technik, mit der sie interagiert, ihnen keinen Grund für eine gegenteilige Einschätzung wie in Modus drei gibt.

Welches Fazit kann aus der Deutung der unterschiedlichen Vertrauensentwicklungstypen gezogen werden?

1. Für alle Cluster (mit Ausnahme der indifferent Gleichbleibenden) lässt sich ein enormer Vertrauensverlust nach Durchlauf 3 (Lenk-Auto) feststellen.
2. Das jeweilige Ausgangsvertrauen (Vertrauenswert nach Modus 2) ist zwischen den Clustern stark unterschiedlich.
3. Vertrauen wird, trotz Vertrauensverlust, in etwa auf dem Niveau des Ausgangsvertrauens wiederhergestellt – für Cluster 4 sogar auf höherem Niveau.

Die Trajektorien des Vertrauensverlustes und der Vertrauenswiederherstellung nehmen für alle Cluster (mit Ausnahme der Gleichbleibenden) einen ähnlichen Verlauf, auf jeweils unterschiedlichem, vom Ausgangsvertrauen abhängigen, Niveau. Es kann konstatiert werden, dass im hier betrachteten Fall Vertrauen zwar anfällig für Zerstörung ist, eine Wiederherstellung, eine höher wahrgenommene Qualität der Automation und Erfüllung der Erwartungen vorausgesetzt, jedoch möglich ist. Dieses Ergebnis liegt in gewisser Weise quer zur Feststellung von Diego Gambetta (1988a: 234), die in der Vertrauensforschung weit verbreitet ist: Ist erst einmal Misstrauen eingetreten, bleibt dieses hartnäckig bestehen, da Vertrauen nicht mehr genutzt wird und kein Gegenbeweis (mehr) erbracht werden kann. Die Situation im Fall vom SIMHYBS stellt sich etwas anders dar, da die Probanden in Modus 4, der dem in Hinblick auf Vertrauen schlecht bewerteten Modus 3 folgte, keine Möglichkeit hatten, die Automation *nicht* zu nutzen. Sie wurden in gewisser Weise zu ihrem *Glück gezwungen* und haben im Ergebnis ihr vorheriges Vertrauensniveau weitgehend wiederhergestellt.

Weitergehend kann daraus gefolgert werden, dass Vertrauen in autonome Technik eine dispositionell-subjektive (vgl. Abschnitt 2.2.3) Komponente umfasst, die sich im starken Einfluss des Ausgangsvertrauens manifestiert. Darüber hinaus haben systemimmanente Eigenschaften der Automation, z. B. die wahrgenommene Qualität, ebenso Einfluss auf das Vertrauen, da andernfalls die Trajektorien über fast alle Cluster nicht ähnlich verlaufen würden.

### **Beschreibung der Cluster anhand ihrer U-Werte**

Im Gegensatz zum MensaAssist-Experiment wurden die p-Werte im Falle von SIMHYBS nicht erhoben, da diese stark von der jeweiligen Verkehrssituation abhängen. Aus dem Grund wird in diesem Abschnitt lediglich untersucht, ob sich die U-Werte als zugeschriebene Zielintensitäten zwischen den verschiedenen Clustern unterscheiden (vgl. Abbildung 11.5). Mit dieser Analyse kann herausgefunden werden, ob ein Zusammenhang zwischen zugeschriebenen Zielen und Vertrauen besteht.

Unerschütterlich Vertrauende, Cluster 2, haben in fast jeder Spalte die höchsten U-Werte, für  $u_{gw,2}$  sogar den höchstmöglichen Wert von fünf. Die hohen U-Werte sind

Ward-Methode CLU_Trust_Development	u_cr_2	u_cr_3	u_cr_4	u_gw_2	u_gw_3	u_gw_4	u_rnd_2	u_rnd_3	u_rnd_4
Cluster 1: Gleichbleibend Indifferente	2,68	4,03	4,42	4,16	2,29	4,45	2,23	2,77	3,32
Cluster 2: Unerschütterlich Vertrauende	2,46	4,31	4,77	5,00	2,23	4,69	2,85	2,46	3,23
Cluster 3: Vertrauende und schnell Verzeihende	2,20	2,76	4,64	4,60	1,28	4,76	2,60	2,40	3,20
Cluster 4: Generell Misstrauende	2,15	3,11	4,41	4,41	1,37	4,63	2,19	2,30	3,52
Gesamtsumme	2,38	3,48	4,52	4,46	1,76	4,61	2,40	2,50	3,33

**Abbildung 11.5:** U-Werte für Vertrauensentwicklungscluster. Die markierten Einträge werden im Text gesondert besprochen.

konsistent mit der Beschreibung des Typs aus der Clusteranalyse der Vertrauensentwicklungstypen. Auf Grund der geringen Fallzahl von  $N_2 = 13$  (dies gilt auch für die anderen Cluster) kann diese Unterschiedlichkeit der unterschütterlich Vertrauenden nicht statistisch signifikant, etwa mittels T-Test, abgesichert werden, sondern lediglich als Unterschiedlichkeitstendenz benannt werden.

Interessant sind darüber hinaus einige Beobachtungen für Cluster 3, die Vertrauenden und schnell Verzeihenden: Dort finden sich im Fall von Durchlauf 3 (Lenk-Auto) zwei jeweils niedrigste Zielintensitäten, für  $u_{cr,3}$  und  $u_{gw,3}$ . Das bedeutet, dass die Probanden aus diesem Cluster für Durchlauf 3 die Zielverfolgung durch das Assistenzsystem mit 2.76 für das Ziel „Crashvermeidung“ und 1.28 für das Ziel „Geschwindigkeitsbegrenzung einhalten“ am geringsten einschätzen.

Bemerkenswert ist auch, dass die Indifferenten durchweg hohe U-Werte haben, obwohl sich hinsichtlich der Vertrauensmessung keine starken Ausschläge für diese Gruppe messen lassen.

Insgesamt lassen sich Tendenzen hinsichtlich der Unterschiede in den zugeschriebenen Zielintensitäten zwischen den verschiedenen Clustern feststellen. Besonders deutlich werden diese Tendenzen in Durchlauf 3, der z. T. auf Widerstand bei den Probanden stieß und insgesamt deutlicher schlechter als andere Modi bewertet wurde. Mit Bezug auf (SH 4) kann festgehalten werden, dass diese als schwach bestätigt gelten kann. Dennoch sind weitere Experimente erforderlich, um den Zusammenhang von Vertrauensentwicklungstrajektorien und zugeschriebenen SEU-Modellen näher in den Blick zu bekommen.

(SH 4)



Eine Korrelationsanalyse von Vertrauen und U-Werten getrennt nach den ermittelten Gruppen bringt wenig neue Informationen (s. Anhang A.6.5), zeigt allerdings, dass es besonders viele und starke Korrelationen für die Gruppe der unerschütterlich Vertrauenden gibt. Gerade für diese Gruppe können U-Werte besonders gut zur näheren Beschreibung der Gründe ihres Vertrauens herangezogen werden.

### Vertrauenscluster und Tastaturanschläge

 (SH 5)  
?

Mit Blick auf (SH 5) soll die These untersucht werden, ob sich Vertrauen bei Nutzung mehrt. Zur Untersuchung dieser These wird die Differenzierung von Vertrauensentwicklungsclustern beibehalten und für diese überprüft, ob die inhaltliche Beschreibung der Cluster mit unterschiedlichen Anzahlen von Tastaturanschlägen und Gegensteuerungsaktionen in Verbindung gebracht werden kann. Werden im Vergleich wenige Tastaturanschläge und Gegensteuerungsaktionen getätigt, spricht dies für die *Nutzung* von Vertrauen (nicht für Vertrauen selbst) bzw. für den Vollzug einer Vertrauenshandlung (vgl. Abschnitt 3.1).

In der Automationsforschung ist der Zusammenhang zwischen Vertrauen und Allokationsverhalten recht gut belegt (vgl. Abschnitt 4.3, exemplarisch die Korrelation von Allokationsverhalten und Vertrauen bei Muir und Moray 1996). Zwar hatten die Probanden in den Durchläufen 2 bis 4 keine Möglichkeit, die Automation auszuschalten, allerdings konnten Sie durch eigene Aktionen in das Fahrgeschehen eingreifen. Je nach Modus der Handlungsverteilung hatten ihre Aktionen Auswirkungen. Im Modus Lenk-Auto konnten die Probanden mit der Pfeiltaste nach oben  $\uparrow$  und nach unten  $\downarrow$  die Geschwindigkeit erhöhen bzw. verringern. Eine Betätigung der Pfeiltaste nach links  $\leftarrow$  oder rechts  $\rightarrow$  hatte hingegen keine Auswirkungen, da die Automation diese Aktionen im Modus Lenk-Auto übernahm. Im Modus Voll-Auto hatte keine Aktion Auswirkungen auf das Fahrgeschehen, dennoch wurde z. T. versucht aktiv gegenzusteuern.

Cluster		Gw-Auto	Lenk-Auto	Voll-Auto	tastHR.2	tastLR.3
		tast.2	tast.3	tast.4		
1 N=31	$\bar{X}$	72,2903	88,0968	14,5806	27,6452	11,6129
	$\sigma$	143,3760	232,3607	31,2802	118,6756	42,0196
2 N=13	$\bar{X}$	18,2308	31,3846	5,6923	4,8462	4,8462
	$\sigma$	17,5317	13,3200	8,6928	8,3850	6,8416
3 N=25	$\bar{X}$	60,9200	46,4800	14,7200	4,7200	3,2800
	$\sigma$	96,5116	49,6958	23,7671	7,0505	7,2544
4 N=27	$\bar{X}$	43,8889	51,4074	12,7037	4,2222	4,8519
	$\sigma$	51,6433	48,6323	21,0691	6,2224	19,1124
$\sum$ N=96	$\bar{X}$	54,0208	59,2604	12,8854	12,0000	6,6250
	$\sigma$	99,6964	137,0465	24,3236	67,8056	26,2515

Abbildung 11.6: Zusammenhang von Vertrauensentwicklungstypen und Tastaturanschlägen

Abbildung 11.6 präsentiert Mittelwerte  $\bar{X}$  und Standardabweichung  $\sigma$  für die Anzahl aufgezeichneter Tastaturaktionen je Vertrauensentwicklungscluster. Dabei bezeichnet *tast.X* die Anzahl der Tastaturanschläge in Experimentdurchlauf *X*. Zur besseren Lesbarkeit sind zusätzlich die zugehörigen Modi verzeichnet. Eine Besonderheit bilden die beiden Einträge *tastHR.2* (Hoch-Runter-Aktionen im Modus Geschwindigkeits-

Auto, die entsprechend ohne Wirkung blieben) und *tastLR.3* (Links-Rechts-Aktionen im Modus Lenk-Auto, die qua Modus ebenfalls ohne Wirkung blieben). Demnach sind *tastHR.2* und *tastLR.3* Maßzahlen für wirkungslose Gegensteuerungsaktionen.

Auf die gekennzeichneten Einträge ❶–❺ wird gesondert eingegangen.

Zunächst fällt auf, dass Cluster 1 (Indifferente) gewissermaßen aus dem Rahmen fällt. Auffällig ist die im Vergleich zu anderen Clustern enorm hohe Standardabweichung (❶, ❸). Nach manueller Überprüfung der einzelnen Daten in dieser Gruppe kann ausgeschlossen werden, dass die hohe Standardabweichung von nur einem einzelnen Probanden verursacht wurde. Die Vertrauensentwicklungsgruppe der Indifferenten lässt sich tatsächlich durch eine besonders hohe Anzahl an Tastaturanschlägen sowie eine hohe Varianz charakterisieren.

Die unerschütterlich Vertrauenden, Cluster 2, weisen hingegen besonders wenige Tastaturanschläge *und* eine sehr geringe Standardabweichung bzw. Varianz auf. Besonders auffällig wird dies im Modus Voll-Auto (❹), in dem die Probanden in Cluster 2 mit durchschnittlich 5.69 Tastaturanschlägen offenbar ohne großes Eingriffsinteresse die vollautomatische Fahrt beobachten. Die unerschütterlich Vertrauenden haben auch im Modus Geschwindigkeits-Auto im Vergleich zu den anderen Clustern seltener die Tastatur betätigt (❷).

Für die Gegensteuerungsaktionen kann keine clusterunterscheidende Aussage getroffen werden. In allen Clustern gibt es punktuell Gegensteuerungsaktionen (trotz ausbleibender Wirkung). Allenfalls bemerkenswert ist, dass die Probanden aus Cluster 4, die generell Misstrauenden, nicht mehr, sondern tendenziell sogar weniger gegensteuert haben. So gab es etwa im Modus Geschwindigkeits-Auto im Durchschnitt 4.22 Gegensteuerungsaktionen (❺) bei geringer Standardabweichung.

Überraschend ist der Befund, dass im Modus Voll-Auto von allen Probanden gegensteuert wurde. Überraschend ist dies, weil der Modus Voll-Auto sowohl hinsichtlich des zugeschriebenen Vertrauens wie auch bei der Bewertung mittels des semantischen Differenzials besonders gut bewertet wurde (vgl. Abschnitt 11.3.3).

Neben den Tastaturanschlägen wurden noch weitere Performance-Daten im Hintergrund aufgezeichnet, z. B. die zurückgelegte Strecke. Da diese allerdings keine weiteren Besonderheiten aufweisen, insbesondere im Hinblick auf die Vertrauensentwicklungscluster, sei für die vollständigen Performance-Daten auf den Anhang A.6.7 verwiesen. Im Hinblick auf These (SH 5) ergibt sich aus den vorangegangenen Betrachtungen kein einheitliches Bild. Für Cluster 2 der stark Vertrauenden lässt sich feststellen, dass diese im Modus Lenk-Auto und Voll-Auto vergleichsweise wenigen Tastaturanschläge vollziehen und dementsprechend Vertrauen nutzen. Um diese Erkenntnis allerdings zu generalisieren, sind die anderen Entwicklungen zu uneindeutig und es liegen zu wenige Messungen vor. Allenfalls sind Tendenzen erkennbar, die dafür sprechen, dass die Überlegung der Vermehrung von Vertrauen durch Nutzung auch für autonome

(SH 5)
– ? –

Technik zutrifft. Für (SH 5) ergibt sich somit ein indifferentes Bild – diese These kann weder als bestätigt noch als widerlegt gelten.

## 11.4 Zwischenfazit

Die abschließenden Betrachtungen zu einer These stehen noch aus. Mit These (SH 2) sollte untersucht werden, ob sowohl systemimmanente Schwächen als auch subjekt-spezifische Dispositionen mit Vertrauen zusammenhängen.

Es hat sich auf der einen Seite gezeigt, dass bestimmte Systemcharakteristika Einfluss auf Vertrauen haben. So wurde etwa nachgewiesen, dass sich für alle Probanden eine schlechtere Vertrauensbewertung für den Modus Lenk-Auto ergeben hat (vgl. Tabelle 11.6 auf Seite 265). Dieser Modus ist auch, wie Tabelle 11.7 auf Seite 267 zeigt, generell schlecht bewertet. Betrachtet man dazu allerdings vergleichend subjektspezifische Vertrauensstrajektorien (s. Abbildung 11.4 auf Seite 278), so ist festzustellen, dass die verschiedenen Vertrauensentwicklungstypen je unterschiedlich auf den Modus Lenk-Auto reagieren. Ebenso ist ersichtlich, dass die Vertrauensstrajektorien für die verschiedenen Akteursgruppen von unterschiedlichem Ausgangsvertrauen ausgehen. Demnach kann These (SH 2) als bestätigt angesehen werden, da systemspezifische Eigenschaften ebenso Einfluss auf Vertrauensmessungen nehmen wie subjektspezifische Eigenschaften.

(SH 2)



Für die weiteren Thesen, die in Zusammenhang mit SIMHYBS untersucht wurden, werden die zentralen Ergebnisse nochmals zusammengefasst. These (SH 1) hat sich bestätigt. Es konnte ein Zusammenhang zwischen Zielintensitäten und Vertrauen festgestellt werden. Dies bedeutet für SEU-Modell-differenzen, dass sich Unterschiede zwischen SEU-Modellen, die der autonomen Technik von einem menschlichen Akteur unterstellt werden ( $SEU_{H \rightarrow NH}$ ) und solchen, die tatsächlich implementiert sind ( $SEU_{NH}$ ), auf das Vertrauen in diese Technik auswirken.

Gemäß (SH 3) konnte gezeigt werden, dass sich Vertrauensmessungen auf Messungen zu späteren Zeitpunkten auswirken. Besonders wichtig ist dabei der erste Eindruck bzw. die erste Vertrauensmessung, da sich diese besonders stark auf alle weitere Messungen auswirkt.

In Bezug auf (SH 4) konnte gezeigt werden, dass sich unterschiedliche Vertrauensentwicklungstypen anhand charakteristischer, dem nicht-menschlichen Akteur zugeschriebenen SEU-Modelle, voneinander unterscheiden lassen. Folglich kann die Methode der SEU-Modellzuschreibung, die im Kontext der eigenen integrierenden Methodologie (Abschnitt 8) entwickelt wurde, auf Fragestellungen des Vertrauens in autonome Technik angewendet werden.

Abschließend ergab sich für die These, dass sich Vertrauen bei Nutzung mehrt (SH 5) ein indifferentes Bild. Es konnten zwar erste Hinweise gefunden werden, die auf ei-

---

ne Bestätigung der These hindeuten, allerdings sind weitere Messungen erforderlich, um einen entsprechenden Zusammenhang zweifelsfrei nachzuweisen. Durch Verknüpfung der agentenbasierten Modellierung und Simulation mit interaktiven Laborexperimenten konnten Daten wie Tastaturanschläge aufgezeichnet werden, die eine Untersuchung von Vertrauenshandlungen ermöglichen und derartig formulierte Thesen zumindest prinzipiell untersuchbar machen.



## 12 Fazit

Den Ausgangspunkt der vorliegenden Arbeit bildete die Einsicht, dass avancierte Technik sich nicht mehr auf das einfache „Funktionieren“ beschränkt, sondern mit-handelt und aktiv in Mensch-Maschine-Interaktionen eingreift und diese mitgestaltet. Die Selbsttätigkeit als Charakteristikum autonomer Technik beinhaltet die Gefahr, dass diese, gegen die Intention des menschlichen Interaktionspartners, fehlerhaft agieren kann. In dem Zusammenhang stellten sich Fragen nach den Bedingungen und Elementen des Vertrauens in autonome Technik.

Die zusammenfassende Darstellung der Ergebnisse wird nicht jede bestätigte oder widerlegte Einzelannahme repetieren, sondern stattdessen versuchen, die zentralen Einsichten und Ideen zu kondensieren und einen Ausblick auf weitere Forschungsperspektiven zu geben. Die Ergebnisdarstellung unterscheidet dabei zwischen *allgemeinen theoretischen Einsichten*, *forschungsheuristischen Erkenntnissen* und *empiriebezogenen Resultaten*.

### Allgemeine theoretische Einsichten

Die Beschäftigung mit unterschiedlichen metaanalytischen Studien zu Vertrauen hat gezeigt, dass Vertrauen als multidimensionales Konzept aufzufassen ist. Es gibt nicht *das eine* Vertrauenskonzept, sondern vielmehr handelt es sich bei Vertrauen um ein äußerst facettenreiches Phänomen, das sehr unterschiedliche Vertrauentypen umfasst. Je nach Untersuchungsgegenstand sind bestimmte Aspekte des Vertrauens mal stärker, mal weniger stark zu berücksichtigen.

Eine Grobunterscheidung kann zwischen solchen Konzepten vorgenommen werden, die Vertrauen als *Disposition* auffassen und solchen, die eine *Vertrauenshandlung* in den Blick nehmen. Eine *Vertrauenshandlung* stellt die binäre Entscheidung zwischen den Handlungsalternativen „vertrauen“ und „misstrauen“ in den Vordergrund. Vertrauenshandlungen sind insbesondere in probabilistischen und spieltheoretischen Vertrauenskonzepten wiederzufinden. Probabilistische Vertrauenskonzepte befassen sich mit der Abwägung von Risiken und Konsequenzen eines Vertrauensmissbrauchs, während spieltheoretische Konzepte Vertrauen als *strategische* Interdependenzsituation konzeptualisieren. Für den Gegenstandsbereich der autonomen Technik und der Frage des Vertrauens, die sich aus dem Zusammenhandeln mit der Technik stellt, sind Konzepte, die eine *explizite* Vertrauenshandlung zum Gegenstand haben, wenig

geeignet. In der Arbeit wurde argumentiert, dass sich das Zusammenhandeln von Mensch und autonomer Technik in einem interaktiven Prozess und im Kontext einer gemeinsam zu erfüllenden Aufgabe abspielt. Im Fall der Interaktion mit dieser Technik ist Vertrauen deswegen als Disposition aufzufassen, die sich auf bestimmte Erwartungen und Handlungsmuster des Interaktionspartners richtet.

Im Gegensatz zu einer expliziten Vertrauenshandlung impliziert *auf Vertrauen bezogenes Verhalten*, dass sich der dispositionelle Einfluss von Vertrauen im eigenen Verhalten widerspiegelt und sich auf die Mensch-Maschine-Interaktion auswirkt. Kurz zusammengefasst, kann Vertrauen angesehen werden als eine gerichtete Zuschreibung, die sich ausgehend von einem Treugeber auf (zukünftiges) Verhalten bzw. dem Verhalten zugrundeliegende Motive des Treuhänders bezieht. Eine feste Zuordnung von Mensch als Treugeber und Technik als Treuhänder ist dabei nicht zwingend erforderlich. Aus methodologischen Gesichtspunkten wurden allerdings in der Hauptsache menschliche Treugeber in den Blick genommen.

Ausgehend von der Überlegung, dass Vertrauen in engem Wechselverhältnis zu Verhalten und damit Handlungen steht, ist eine soziologisch fundierte Sichtweise auf Aspekte der Mensch-Maschine-Interaktion von wesentlicher Bedeutung. Unter Rekurs auf handlungstheoretische Überlegungen hat sich gezeigt, dass sich das Zusammenhandeln von Mensch und autonomer Technik in Form eines operationalisierbaren Makro-Mikro-Makro-Modells soziologischer Erklärung rekonstruieren lässt. Ausgehend von diesem Modell wurden Überlegungen zu wechselseitigen Modellzuschreibungen, also Handlungsmodellen, die dem jeweiligen Interaktionspartner unterstellt werden, angestellt. In dieser Arbeit hat sich gezeigt, dass ein systematischer Vergleich zwischen wirksamen und zugeschriebenen Handlungsmodellen (und damit auch Handlungsmotiven) sowie eine Untersuchung von Modelldifferenzen zu einem genuin soziologischen Verständnis von Vertrauen in autonome Technik beitragen kann.

## **Forschungsheuristische Erkenntnisse**

Inspiziert von den laborexperimentellen Vertrauensstudien der Spieltheorie konnte herausgestellt werden, dass sich Vertrauen in Form von Laborexperimenten in iterierten prozessförmigen Konstellationen untersuchen lässt.

Eine nähere Betrachtung von Vertrauensstudien aus der Automationsforschung bestätigten den Mehrwert einer laborexperimentellen Herangehensweise zur Untersuchung von Fragen des Vertrauens. Aus der Automationsforschung stammte auch die Anregung, laborexperimentelle Situationen mit Computersimulationen bzw. elektronischen Entscheidungshelfern zu kombinieren.

Aus der Betrachtung einiger Praxisbeispiele autonomer Technik, u. a. Autonomous Trading Agents (autonom agierende Börsencomputer), Wikipedia-Bots und fortgeschrittenen Fahrerassistenzsysteme bis hin zur Vision des vollständig autonomen

---

Fahrens, hat sich herausgestellt, dass Technik handlungswirksam ist und die Mensch-Maschine-Interaktion aktiv mitgestaltet. Auf Grund dieser Beobachtung ist autonome Technik, zumindest methodologisch, als gleichberechtigter Partner und Mitgestalter von Mensch-Technik-Konstellationen aufzufassen und im Lichte einer soziologischen Analyse konstitutiv zu berücksichtigen.

Zu diesem Zweck wurde eine integrierende Methodologie vorgeschlagen, um Mensch-Maschine-Interaktion im Rahmen eines einheitlichen Bezugsrahmens untersuchbar zu machen. Dabei wurde das Modell soziologischer Erklärung hybrider Systeme (HMSE) vorgestellt, das die Grundannahme eines Interaktionszusammenhangs im Lichte einer gemeinsam zu erfüllenden Aufgabe (s.o.) aufgreift und mit Hilfe eines handlungstheoretisch fundierten Makro-Mikro-Makro-Modells umsetzt. Das Handlungsmodell auf der Mikro-Ebene der Akteure (Mensch *und* autonome Technik) fußt auf einem teleologischen Kern, der Handlungswahlen im Kontext bestimmter Ziele und situativer Opportunitäten operationalisiert. Die Verwendung einer einheitlichen Modellierungssprache (SEU-Theorie) für Mensch und autonome Technik ermöglichte eine systematische Untersuchung von Modelldifferenzen und wechselseitigen Modellzuschreibungen. Der Vergleich wirksamer und zugeschriebener Modelle trägt zu einem handlungstheoretischen und damit genuin soziologischen Verständnis von Vertrauen in autonome Technik bei.

Auf diesen Überlegungen gründet sich die Konzeption der Experimentalstudien im Empirieteil. Insbesondere hat sich gezeigt, dass durch das operationalisierbare Modell der Mensch-Maschine-Interaktion Anschlussfähigkeit an die Methode der agentenbasierten Modellierung und Simulation besteht. Die Verwendung einer mikrofundierten Computersimulation im Rahmen laborexperimenteller Szenarien ermöglichte die Untersuchung von Vertrauen in autonome Technik in interaktiver Perspektive. Zugespitzt könnte man formulieren, dass eine soziologische Erklärung als Computersimulation implementiert wurde.

## Empiriebezogene Resultate

Vor dem Hintergrund der theoretischen Überlegungen zu Vertrauen sowie forschungsheuristischen Einsichten wurden zwei Experimentalstudien vorgestellt, die Vertrauen in autonome Technik in Form von Laborexperimenten bzw. Befragungen untersuchten.

Im ersten Experiment *MensaAssist* wurden Befragte gebeten, einen Entscheidungshelfer, der sie bei der Auswahl bestimmter Speisen gemäß eigener Vorlieben unterstützt, zu bewerten. Durch eine Vorauswahl bestimmter Essenstypen konnten die Befragten zunächst eine Grobeinordnung vornehmen, die *MensaAssist* als (vorgebliche) Vorschlagsgrundlage für besonders passende Speisen diente. Wie sich zeigte, stimmte diese Vorauswahl nur eingeschränkt mit den tatsächlichen Vorlieben der

Befragten überein, so dass zu erwarten war, dass die Befragten, je nach Übereinstimmung ihres eigenen Essensauswahlmodells mit dem Modell des Entscheidungshelfers unterschiedlich stark vertrauen würden. Tatsächlich zeigte sich ein negativer Zusammenhang zwischen aufsummierten Modellunterschieden und Vertrauen. Modellunterschiede, die mit diesem Experimentaldesign in den Blick kamen, beziehen sich auf Differenzen zwischen dem „tatsächlichen“ Entscheidungsmodell der Befragten und dem Modell, das den Befragten von dem Entscheidungshelfer zugeschrieben wurde. Es hat sich außerdem ergeben, dass Befragte, die unterschiedlich stark vertrauen, sich auch bzgl. ihrer Entscheidungsmodellprägungen unterscheiden. Folglich besteht ein Zusammenhang zwischen Modelldifferenzen und Vertrauen.

Im zweiten Experiment SIMHYBS rückte vor allem der interaktive Aspekt des Vertrauens in den Vordergrund, da mehrfache Vertrauensmessungen erfolgten. SIMHYBS ist eine Art Fahrsimulation, in der Probanden zusammen mit einem Fahrerassistenzsystem in unterschiedlichen Modi der Handlungsverteilung ein Fahrzeug steuern müssen. Für die jeweiligen Modi wurden Vertrauenswerte gemessen sowie die der Automation zugeschriebenen Handlungsmotive abgefragt. Im Zuge des Experiments hat sich ein Zusammenhang von Vertrauen und zugeschriebenen Zielintensitäten (also der Stärke von Handlungsmotiven) ergeben. Als weiteres interessantes Ergebnis konnte nachgewiesen werden, dass Vertrauen in autonome Technik sowohl von systembedingten Schwächen der Automation als auch von subjektspezifischen Zuschreibungen und Dispositionen abhängt. Die mehrfachen Vertrauensmessungen haben ergeben, dass sich vorheriges Vertrauen auf spätere Vertrauensmessungen auswirkt. Insbesondere das initiale Vertrauen in die Automation hat wesentliche Auswirkungen auf spätere Messungen. Weiterhin konnte gezeigt werden, dass sich typische Vertrauenstrajektorien ermitteln lassen, die sich anhand von zugeschriebenen Entscheidungsmodellen voneinander unterscheiden lassen und damit handlungstheoretisch interpretierbar sind.

## **Ausblick und weitere Forschungsperspektiven**

In Anschluss an die theoretischen, forschungsheuristischen und empirischen Erkenntnissen sollen einige Überlegungen skizziert werden, an die weitere Forschungsarbeiten anschließen können.

Die vorgeschlagene Methode zum systematischen Abgleich von Differenzen zwischen zugeschriebenen Modellen und tatsächlich wirksamen Modellen steht noch am Anfang und ist ausbaufähig. Im Rahmen der Laborexperimente haben sich zwar Zusammenhänge von Modelldifferenzen und Vertrauen gezeigt, diese könnten allerdings im Rahmen umfangreicherer Experimente noch systematischer untersucht werden. Dabei wäre eine Kombination der vorgestellten Experimente interessant, bei der komplexere Aufgaben von Mensch und autonomer Technik in einem interaktiven Zusammenhang betrachtet werden.

---

Ausgehend von der vorgeschlagenen integrierenden Methodologie einer soziologisch fundierten Mensch-Maschine-Interaktion ist es denkbar, dass diese nicht nur auf laborexperimentelle Settings angewendet wird, sondern auch im Rahmen von Feldforschung Verwendung findet. So wäre zu überlegen, ob sich die vorgeschlagene integrierende Methodologie eignet, um Anwender und Entwickler autonomer Technik gemeinsam damit modellieren zu lassen. Dabei sollte die Zielsetzung darin bestehen, zu einem übereinstimmenden Bild der gemeinsam von Mensch und Technik zu lösenden Aufgabe zu gelangen. Ein übereinstimmendes Bild würde die Modelldifferenzen zwischen menschlichem Akteur (Anwender) und autonomer Technik auf ein Minimalmaß reduzieren und möglicherweise zu höherem Vertrauen und damit ggf. auch zu höherer Akzeptanz der Technik beitragen.

Schlussendlich wäre zu prüfen, ob sich Überlegungen zu einem teleologisch-symmetrischen Verständnis von Mensch-Maschine-Interaktion innerhalb der Software-Implementierung konstitutiv berücksichtigen ließen. Sofern autonome Technik über eine Benutzerschnittstelle verfügt, die einen permanenten Abgleich von Mensch-Technik-Modellen ermöglicht und die Erwartungen beider Interaktionspartner aufeinander abstimmt, würde dies ggf. zu einem vertrauensvolleren Umgang mit autonomer Technik beitragen.



## **Teil IV**

# **Literaturverzeichnis und Anhang**



# Literaturverzeichnis

## **Adelt u. a. 2014**

ADELDT, Fabian ; WEYER, Johannes ; FINK, Robin D.: Steuerung komplexer Systeme: Ergebnisse einer soziologischen Simulationsstudie (im Ersch.). In: *Soziale Welt* 65 (2014), Nr. 1

## **Amber und Amber 1962**

AMBER, G. H. ; AMBER, P. S.: *Anatomy of automation*. Prentice-Hall, 1962

## **Aronson u. a. 2011**

ARONSON, Elliot ; WILSON, Timothy D. ; AKERT, Robin M.: *Sozialpsychologie*. 6., aktualisierte Aufl. München : Pearson Studium, 2011 (PS - Psychologie)

## **Axelrod 1997**

AXELROD, Robert M.: *The complexity of cooperation: Agent based models of competition and collaboration*. Princeton and NJ : Princeton Univ. Press, 1997 (Princeton studies in complexity)

## **Backhaus u. a. 2006**

BACKHAUS, Klaus ; ERICHSON, Bernd ; PLINKE, Wulff ; WEIBER, Rolf: *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. 11., überarb. Aufl. Berlin : Springer, 2006

## **Beck 2012**

BECK, Ulrich: *Risikogesellschaft: Auf dem Weg in eine andere Moderne*. 21. Aufl. Frankfurt am Main : Suhrkamp, 2012

## **Best 2007**

BEST, Henning: Die Messung von Nutzen und subjektiven Wahrscheinlichkeiten: Ein Vorschlag zur Operationalisierung der Rational Choice Theorie. In: *Methoden - Daten - Analysen* 1 (2007), Nr. 2, S. 183-212

## **Bijker u. a. 1987**

BIJKER, Wiebe E. (Hrsg.) ; HUGHES, Thomas P. (Hrsg.) ; PINCH, Trevor J. (Hrsg.): *The Social Construction of Technological Systems*. Cambridge/Mass : MIT Press, 1987

**Bisantz und Seong 2001**

BISANTZ, Ann M. ; SEONG, Younho: Assessment of operator trust in and utilization of automated decision-aids under different framing conditions. In: *International journal of industrial ergonomics* 28 (2001), S. 85–97

**Bonabeau 2002**

BONABEAU, Eric: Agent-based modeling: Methods and techniques for simulating human systems. In: *Proceedings of the National Academy of Sciences (PNAS)* 99 (2002), Nr. 3, S. 7280–7287

**Bortz und Döring 2006**

BORTZ, Jürgen ; DÖRING, Nicola: *Forschungsmethoden und Evaluation: Für Human- und Sozialwissenschaftler*. 4., überarbeitete Aufl. Heidelberg : Springer, 2006

**Bortz und Schuster 2010**

BORTZ, Jürgen ; SCHUSTER, Christof: *Statistik für Human- und Sozialwissenschaftler*. 7., vollst. überarb. und erw. Aufl. Berlin [u.a.] : Springer, 2010

**Bratman 1987**

BRATMAN, M.: *Intention, plans, and practical reason*. Cambridge : Harvard University Press, 1987

**Braun und Franzen 1995**

BRAUN, Normann ; FRANZEN, Axel: Umweltverhalten und Rationalität. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 47 (1995), Nr. 2, S. 231–248

**Braun-Thürmann 2007**

BRAUN-THÜRMAN, Holger: Künstliche Interaktion. In: RAMMERT, Werner (Hrsg.): *Technik - Handeln - Wissen*. Wiesbaden : VS Verl. für Sozialwiss., 2007, S. 221–243

**Breznitz 1983**

BREZNITZ, Shlomo: *Cry-wolf: the psychology of false alarms*. Hillsdale : Lawrence Erlbaum, 1983

**Bright 1958**

BRIGHT, James R.: *Automation and management*. Boston and Massachusetts : Division of Research, Havard Business School, 1958

**Buskens 1995**

BUSKENS, Vincent: *Social Networks and the Effect of Reputation on Cooperation*. 1995

**Buskens 1998**

BUSKENS, Vincent: The social structure of trust. In: *Social Networks* 20 (1998), Nr. 3, S. 265–289

**Callon 1986**

CALLON, Michel: Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. In: LAW, John (Hrsg.): *Power, action and belief: a new sociology of knowledge?* London : Routledge, 1986, S. 196–223

**Callon und Latour 1992**

CALLON, Michel ; LATOUR, Bruno: Don't Throw the Baby Out with the Bath School! A Reply to Collins and Yearley. In: PICKERING, Andrew (Hrsg.): *Science as Practise and Culture*. Chicago : Chicago UP, 1992, S. 343–368

**Carr 2010**

CARR, Nicholas G.: *Wer bin ich, wenn ich online bin ... und was macht mein Gehirn solange? Wie das Internet unser Denken verändert*. 1. München : Blessing, 2010

**Charette 2009**

CHARETTE, Robert N.: This car runs on code. In: *IEEE Spectrum* 46 (2009), Nr. 3, S. 3

**Cohen 1992**

COHEN, Jacob: A power primer. In: *Psychological Bulletin* 112 (1992), Nr. 1, S. 155

**Coleman 1991**

COLEMAN, James S.: *Grundlagen der Sozialtheorie*. Bd. 1: Handlungen und Handlungssysteme. München : Oldenbourg, 1991

**Collins und Yearley 1992**

COLLINS, Harry M. ; YEARLEY, Steven: Epistemological chicken. In: PICKERING, Andrew (Hrsg.): *Science as Practise and Culture*. Chicago : Chicago UP, 1992, S. 301–326

**Cronbach 1951**

CRONBACH, Lee J.: Coefficient alpha and the internal structure of tests. In: *Psychometrika* 16 (1951), Nr. 3, S. 297–334

**Cross 2010**

CROSS, Richard: Medieval Theories of Haecceity. In: EDWARD N. ZALTA (Hrsg.): *The Stanford Encyclopedia of Philosophy*. URL <http://plato.stanford.edu/entries/medieval-haecceity/>, 2010

**Cummings und Bruni 2009**

CUMMINGS, Mary L. ; BRUNI, Sylvain: Collaborative Human-Automation Decision Making. In: NOF, Shimon Y. (Hrsg.): *Handbook of Automation*. Heidelberg : Springer, 2009, S. 437–447

**Dasgupta 1988**

DASGUPTA, Partha: Trust as a Commodity. In: GAMBETTA, Diego (Hrsg.): *Trust: making and breaking cooperative relations*. New York and NY [u.a.] : Basil Blackwell, 1988, S. 49–71

**Deutsch 1958**

DEUTSCH, Morton: Trust and suspicion. In: *Journal of Conflict Resolution* 2 (1958), Nr. 4, S. 265–279

**Deutsch 1976**

DEUTSCH, Morton: *Konfliktregelung: Konstruktive und destruktive Prozesse*. München : Reinhardt, 1976

**Diekmann 2009**

DIEKMANN, Andreas: *Spieltheorie: Einführung, Beispiele, Experimente*. Reinbek : Rowohlt-Taschenbuch-Verl, 2009

**Diekmann und Voss 2004**

DIEKMANN, Andreas ; VOSS, Thomas: Die Theorie rationalen Handelns. Stand und Perspektiven. In: DIEKMANN, Andreas (Hrsg.) ; VOSS, Thomas (Hrsg.): *Rational-Choice-Theorie in den Sozialwissenschaften*. München : Oldenbourg, 2004, S. 13–29

**Dzindolet u. a. 2003**

DZINDOLET, Mary T. ; PETERSON, Scott A. ; POMRANKY, Regina A. ; PIERCE, Linda G. ; BECK, Hall P.: The role of trust in automation reliance: Trust and Technology. In: *International Journal of Human-Computer Studies* 58 (2003), Nr. 6, S. 697–718

**Epstein und Axtell 1996**

EPSTEIN, Joshua M. ; AXTELL, Robert: *Growing artificial societies: social science from the bottom up*. Cambridge : MIT Press, 1996

**Esser 1991**

ESSER, Hartmut: *Alltagshandeln und Verstehen. Zum Verhältnis erklärender und verstehender Soziologie am Beispiel von Alfred Schütz und "Rational Choice"*. Tübingen : Mohr Siebeck, 1991

**Esser 1993**

ESSER, Hartmut: *Soziologie. Allgemeine Grundlagen*. Frankfurt/M. : Campus, 1993

**Esser 1996**

ESSER, Hartmut: *Soziologie - Allgemeine Grundlagen. 2*. Frankfurt/M. : Campus, 1996

**Esser 1999**

ESSER, Hartmut: *Soziologie - Spezielle Grundlagen. Bd. 1: Situationslogik und Handeln*. Frankfurt/M. : Campus, 1999

**Esser 2000a**

ESSER, Hartmut: *Soziologie - Spezielle Grundlagen. Bd. 2: Die Konstruktion der Gesellschaft*. Frankfurt/M. : Campus, 2000

**Esser 2000b**

ESSER, Hartmut: *Soziologie - Spezielle Grundlagen. Bd. 3: Soziales Handeln*. Frankfurt/M. : Campus, 2000

**Esser 2000c**

ESSER, Hartmut: *Soziologie. Spezielle Grundlagen, Bd. 3: Soziales Handeln*. Frankfurt/M. : Campus, 2000

**Esser 2001**

ESSER, Hartmut: *Soziologie - Spezielle Grundlagen. Bd. 6: Sinn und Kultur*. Frankfurt/M. : Campus Verlag, 2001

**Etzrodt 2000**

ETZRODT, Christian: Alfred Schütz - Ökonom und/oder Soziologe? Eine Kritik an Hartmut Essers Interpretation der Theorie von Alfred Schütz und an seiner Definition der Situation. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 52 (2000), Nr. 4, S. 761–782

**Falcon 2011**

FALCON, Andrea: Aristotle on Causality. In: EDWARD N. ZALTA (Hrsg.): *The Stanford Encyclopedia of Philosophy*. 2011

**Fink 2008**

FINK, Robin D.: *Untersuchung hybrider Akteurskonstellationen mittels Computersimulation: Diplomarbeit*. 2008

**Fink 2009**

FINK, Robin D.: *Experimentelle Untersuchung von Handeln in hybriden Systemen: Soziologisches Arbeitspapier Nr. 23*. 2009

**Fink und Liboschik 2010**

FINK, Robin D. ; LIBOSCHIK, Tobias: *Bots – Nicht-menschliche Mitglieder der Wikipedia-Gemeinschaft: Soziologisches Arbeitspapier Nr. 28*. 2010

**Fink und Liboschik 2012**

FINK, Robin D. ; LIBOSCHIK, Tobias: Bots – Nicht-menschliche Mitglieder der Wikipedia-Gemeinschaft. In: SOEFFNER, Hans-Georg (Hrsg.): *Transnationale Vergesellschaftungen. Verhandlungen des 35. Kongresses der Deutschen Gesellschaft für Soziologie in Frankfurt am Main 2010*. Wiesbaden : VS Verlag, 2012

**Fink und Weyer 2011**

FINK, Robin D. ; WEYER, Johannes: Autonome technische Systeme als Herausforderung der soziologischen Handlungstheorie. In: *Zeitschrift für Soziologie* 40 (2011), Nr. 2, S. 91–111

**Friedrichs u. a. 1993**

FRIEDRICHS, Jürgen ; STOLLE, Martin ; ENGELBRECHT, Gudrun: Rational Choice-Theorie: Probleme der Operationalisierung. In: *Zeitschrift für Soziologie* 22 (1993), Nr. 1, S. 2–15

**Fuller 1994**

FULLER, Steve: Making Agency Count : A Brief Foray Into the Foundations of Social Theory. In: *American Behavioral Scientist* 37 (1994), Nr. 6, S. 741–753

**Gad und Bruun Jensen 2010**

GAD, Christopher ; BRUUN JENSEN, Caspar: On the consequences of post-ANT. In: *Science, Technology & Human Values* 35 (2010), S. 55–80

**Gambetta 1988a**

GAMBETTA, Diego: Can We Trust Trust? In: GAMBETTA, Diego (Hrsg.): *Trust: making and breaking cooperative relations*. New York and NY [u.a.] : Basil Blackwell, 1988, S. 213–237

**Gambetta 1988b**

GAMBETTA, Diego (Hrsg.): *Trust: making and breaking cooperative relations*. New York and NY [u.a.] : Basil Blackwell, 1988

**Garfinkel 2005**

GARFINKEL, Harold: A conception of and experiments with 'trust' as a condition of concerted stable actions. In: O'BRIEN, Jodi (Hrsg.): *The production of reality*. Thousand Oaks and Calif : Pine Forge Press, 2005, S. 370–380

**Geser 1989**

GESER, Hans: Der PC als Interaktionspartner. In: *Zeitschrift für Soziologie* 18 (1989), S. 230–243

**Gilbert 2008**

GILBERT, G. N.: *Agent based models*. Los Angeles : Sage, 2008 (Quantitative Applications in the Social Sciences)

**Gilbert und Troitzsch 2005**

GILBERT, G. N. ; TROITZSCH, Klaus G.: *Simulation for the social scientist*. 2. Maidenhead and Berkshire : Open Univ. Press, 2005

**Gilbert 1995**

GILBERT, Nigel: Emergence in social simulation. In: GILBERT, Nigel (Hrsg.) ; CONTE, Rosaria (Hrsg.): *Artificial societies: The computer simulation of social life*. London : UCL Press, 1995

**Gilbert und Terna 2000**

GILBERT, Nigel ; TERNA, Pietro: How to Build and Use Agent-Based Models in Social Science. In: *Mind & Society* 1 (2000), Nr. 1, S. 57–72

**Grabowski und Sanborn 2003**

GRABOWSKI, Martha ; SANBORN, Stephen D.: Human performance and embedded intelligent technology in safety-critical systems: Trust and Technology. In: *International Journal of Human-Computer Studies* 58 (2003), Nr. 6, S. 637–670

**Grunewald 2007**

GRUNEWALD, Armin: Technikdeterminismus oder Sozialdeterminismus: Zeitbezüge und Kausalverhältnisse aus der Sicht des "Technology Assessment". In: DOLATA, Ulrich (Hrsg.) ; WERLE, Raymund (Hrsg.): *Gesellschaft und die Macht der Technik*. Frankfurt am Main : Campus, 2007, S. 63–82

**Hahne u. a. 2006**

HAHNE, Michael ; LETTKEMANN, Eric ; LIEB, Renate ; MEISTER, Martin: Going Data in Interaktivitätsexperimenten: Neue Methoden zur Analyse der Interaktivität zwischen Mensch und Maschine. In: RAMMERT, Werner (Hrsg.) ; SCHUBERT, Cornelius (Hrsg.): *Technografie: Zur Mikrosoziologie der Technik*. Frankfurt/M. : Campus, 2006, S. 275–309

**Hartmann 2012**

HARTMANN, Peter: Logik. In: *Mathematik für Informatiker*. Vieweg+Teubner Verlag, 2012, S. 28–49. – URL [http://dx.doi.org/10.1007/978-3-8348-2002-0\\_2](http://dx.doi.org/10.1007/978-3-8348-2002-0_2)

**Hedström 2008**

HEDSTRÖM, Peter: *Anatomie des Sozialen - Prinzipien der analytischen Soziologie*. Wiesbaden : VS Verlag, 2008

**Hempel und Oppenheim 1948**

HEMPEL, Carl G. ; OPPENHEIM, Paul: Studies in the Logic of Explanation. In: *Philosophy of Science* 15 (1948), Nr. 2, S. 135–175

**Holland 1992**

HOLLAND, John H.: Complex Adaptive Systems. In: *Daedalus* 1 (1992), Nr. 121, S. 17–30

**Inagaki 2008**

INAGAKI, Toshiyuki: Smart collaboration between humans and machines based on mutual understanding. In: *Annual Review in Control* 32 (2008), S. 253–261

**Inagaki und Itoh 2013**

INAGAKI, Toshiyuki ; ITOH, Makoto: Human's Overtrust in and Overreliance on Advanced Driver Assistance Systems: A Theoretical Framework. In: *International Journal of Vehicular Technology* 2013 (2013), S. 1–8. – URL <http://dx.doi.org/10.1155/2013/951762>

**Jansen 2006**

JANSEN, Dorothea: *Einführung in die Netzwerkanalyse: Grundlagen, Methoden, Forschungsbeispiele*. 3. Wiesbaden : VS Verlag, 2006

**Jian u. a. 2000**

JIAN, Jiun-Yin ; BISANTZ, Ann M. ; DRURY, Colin G.: Foundations for an Empirically Determined Scale of Trust in Automated System. In: *International Journal of Cognitive Ergonomics* 4 (2000), Nr. 1, S. 53–71

**Jordan 1963**

JORDAN, Nehemiah: Allocation of functions between man and machines in automated systems. In: *Journal of Applied Psychology* 47 (1963), Nr. 3, S. 161–165

**Kaiser und Rice 1974**

KAISER, Henry F. ; RICE, John: Little Jiffy, Mark Iv. In: *Educational and Psychological Measurement* 34 (1974), Nr. 1, S. 111–117

**Kantowitz und Sorkin 1987**

KANTOWITZ, Barry H. ; SORKIN, Robert D.: Allocation of functions. In: SALVENDY, Gavriel (Hrsg.): *Handbook of human factors*. Oxford : John Wiley & Sons, 1987, S. 355–369

**Keppel 1973**

KEPPEL, Geoffrey: *Design and analysis: A researcher's handbook*. Englewood Cliffs and N.J. : Prentice-Hall, 1973

**Keppel und Wickens 2004**

KEPPEL, Geoffrey ; WICKENS, Thomas D.: *Design and analysis: A researcher's handbook*. 4. Upper Saddle River and N.J : Pearson Prentice Hall, 2004

**Knoblauch und Heath 1999**

KNOBLAUCH, Hubert ; HEATH, Christian: Technologie, Interaktion und Organisation: die Workplace Studies. In: *Schweizer Zeitschrift für Soziologie* 25 (1999), Nr. 2, S. 163–181

**Knoke und Yang 2008**

KNOKE, David ; YANG, Song: *Social network analysis*. 2. Los Angeles : SAGE Publications, 2008

**Kreps 1990**

KREPS, David M.: Corporate culture and economic theory. In: ALT, James E. (Hrsg.) ; SHEPSLE, Kenneth A. (Hrsg.): *Perspectives on Positive Political Economy*. Cambridge : Cambridge University Press, 1990

**Krohn und Weyer 1989**

KROHN, Wolfgang ; WEYER, Johannes: Gesellschaft als Labor: Die Erzeugung sozialer Risiken durch experimentelle Forschung. In: *Soziale Welt* 40 (1989), Nr. 3, S. 349–373

**Kron 2004**

KRON, Thomas: General Theory of Action? Inkonsistenzen in der Handlungstheorie von Hartmut Esser. In: *Zeitschrift für Soziologie* 33 (2004), Nr. 3, S. 186–205

**Kron 2005**

KRON, Thomas: Fuzzy-Logik für die Soziologie. In: *Österreichische Zeitschrift für Soziologie* (2005), Nr. 3, S. 51–88

**Kron 2009**

KRON, Thomas: *Zeitgenössische soziologische Theorien: Die zentralen Theorien aus Deutschland*. 1. Wiesbaden : VS Verlag für Sozialwissenschaften, 2009

**Kron und Grund 2010**

KRON, Thomas ; GRUND, Thomas: *Die Analytische Soziologie in der Diskussion*. Wiesbaden : VS Verlag für Sozialwissenschaften / Springer Fachmedien Wiesbaden GmbH Wiesbaden, 2010

**Kron und Lasarczyk 2006**

KRON, Thomas ; LASARCZYK, Christian W.: Zur sozionischen Notwendigkeit mechanistisch-soziologischer Erklärungen. In: SCHMITT, Marco (Hrsg.) ; FLORIAN, Michael (Hrsg.) ; HILLEBRANDT, Frank (Hrsg.): *Reflexive soziale Mechanismen*. Wiesbaden : VS Verlag für Sozialwissenschaften, 2006, S. 105–137

**Kron und Winter 2006**

KRON, Thomas ; WINTER, Lars: Zum bivalenten Denken bei Max Weber, Niklas Luhmann und Hartmut Esser. In: GRESHOFF, Rainer (Hrsg.) ; SCHIMANK, Uwe (Hrsg.): *Integrative Sozialtheorie? Esser — Luhmann — Weber*. Wiesbaden : VS Verl. für Sozialwiss., 2006, S. 489–514

**Kroneberg 2005**

KRONEBERG, Clemens: Die Definition der Situation und die variable Rationalität der Akteure. In: *Zeitschrift für Soziologie* 34 (2005), Nr. 5, S. 344–363

**Kroneberg 2007**

KRONEBERG, Clemens: Wertrationalität und das Modell der Frame-Selektion. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 59 (2007), Nr. 5, S. 215–239

**Kroneberg 2009**

KRONEBERG, Clemens: Methodologie statt Ontologie: Das Makro-Mikro-Makro-Modell als einheitlicher Bezugsrahmen der akteurstheoretischen Soziologie. In: GREVE, Jens (Hrsg.) ; SCHÜTZEICHEL, Rainer (Hrsg.) ; SCHNABEL, Annette (Hrsg.): *Das Mikro-Makro-Modell der soziologischen Erklärung*. Wiesbaden : VS Verlag für Sozialwissenschaften / GWV Fachverlage GmbH, Wiesbaden, 2009

**Kroneberg 2011**

KRONEBERG, Clemens: *Die Erklärung sozialen Handelns: Grundlagen und Anwendung einer integrativen Theorie*. 1. Wiesbaden : VS Verlag, 2011 (Neue Bibliothek der Sozialwissenschaften)

**Kroniger und Lücke 2010**

KRONIGER, Jens ; LÜCKE, Fabian: *Analyse und Evaluation von Steuerungsmodellen am Beispiel von Verkehrstelematiksystemen (Diplomarbeit, TU Dortmund)*. 2010

**Krummheuer 2010**

KRUMMHEUER, Antonia L.: *Interaktion mit virtuellen Agenten? Zur Aneignung eines ungewohnten Artefakts*. Stuttgart : Lucius & Lucius, 2010

**Kuls und Mohr 2010**

KULS, Norbert ; MOHR, Daniel: "Es war wie ein Torpedo": Chaos an der Wall Street. In: *Frankfurter Allgemeine Zeitung* (2010), Nr. 07.05.2010. – URL <http://www.faz.net/aktuell/finanzen/aktien/chaos-an-der-wall-street-es-war-wie-ein-torpedo-1983020.html>

**Lahno 2002**

LAHNO, Bernd: *Der Begriff des Vertrauens*. Paderborn : Mentis, 2002

**Latour 1991**

LATOUR, Bruno: Technology is society made durable. In: LAW, John (Hrsg.): *A sociology of monsters: Essays on power, technology and domination*. London : Routledge, 1991, S. 103–131

**Latour 1996a**

LATOUR, Bruno: *Der Berliner Schlüssel. Kleine Soziologie alltäglicher Gegenstände und andere Lektionen eines Liebhabers der Wissenschaften*. Berlin : Akademie Verlag, 1996

**Latour 1996b**

LATOUR, Bruno: On Actor-network Theory. A few Clarifications. In: *Soziale Welt* 47 (1996), Nr. 4, S. 369–382

**Latour 1996c**

LATOUR, Bruno: On interobjectivity. In: *Mind, Culture, and Activity* 3 (1996), Nr. 4, S. 228–245

**Latour 1998a**

LATOUR, Bruno: Aramis - oder die Liebe zur Technik. In: FRICKE, Werner (Hrsg.): *Innovationen in Technik, Wissenschaft und Gesellschaft (Forum Humane Technikgestaltung, Bd. 19)*. Bonn : Friedrich-Ebert-Stiftung, 1998, S. 147–164

**Latour 1998b**

LATOUR, Bruno: Über technische Vermittlung. Philosophie. Soziologie, Genealogie. In: RAMMERT, Werner (Hrsg.): *Technik und Sozialtheorie*. Frankfurt/M. : Campus, 1998, S. 29–81

**Latour 1999**

LATOUR, Bruno: On recalling ANT. In: LAW, John (Hrsg.) ; HASSARD, John (Hrsg.): *Actor Network Theory and After*. Oxford : Blackwell, 1999, S. 15–25

**Latour 2000**

LATOUR, Bruno: *Die Hoffnung der Pandora*. Frankfurt am Main : Suhrkamp, 2000

**Latour 2007**

LATOUR, Bruno: *Eine neue Soziologie für eine neue Gesellschaft. Einführung in die Akteur-Netzwerk-Theorie*. Frankfurt/M. : Suhrkamp, 2007

**Lee und Moray 1992**

LEE, John D. ; MORAY, Neville: Trust, control strategies and allocation of function in human- machine systems. In: *Ergonomics* 35 (1992), Nr. 10, S. 1243–1270

**Lee und Moray 1994**

LEE, John D. ; MORAY, Neville: Trust, self-confidence, and operators' adaptation to automation. In: *International Journal of Human-Computer Studies* 40 (1994), Nr. 1, S. 153–184

**Lee und See 2004**

LEE, John D. ; SEE, Katrina A.: Trust in Automation: Designing for Appropriate Reliance. In: *Human Factors* 46 (2004), Nr. 1, S. 50–80

**Lees und Lee 2007**

LEES, M. N. ; LEE, J. D.: The influence of distraction and driving context on driver response to imperfect collision warning systems. In: *Ergonomics* 50 (2007), Nr. 8, S. 1264–1286. – URL <http://dx.doi.org/10.1080/00140130701318749>

**Lindner 2012**

LINDNER, Roland: Googles selbstfahrende Autos: Das Auto wird automobil. In: *Frankfurter Allgemeine Zeitung* (2012), Nr. 129, 06.06.2012, S. T4

**Luhmann 1988**

LUHMANN, Niklas: Familiarity, Confidence, Trust: Problems and Alternatives. In: GAMBETTA, Diego (Hrsg.): *Trust: making and breaking cooperative relations*. New York and NY [u.a.] : Basil Blackwell, 1988, S. 94–107

**Luhmann 1991**

LUHMANN, Niklas: Am Ende der kritischen Soziologie. In: *Zeitschrift für Soziologie* 20 (1991), Nr. 2, S. 147–152

**Macal und North 2005**

MACAL, Charles M. (Hrsg.) ; NORTH, Michael J. (Hrsg.): *Tutorial on agent-based modeling and simulation: WSC '05: Proceedings of the 37th conference on Winter simulation*. Winter Simulation Conference, 2005

**Macal und North 2010**

MACAL, Charles M. ; NORTH, Michael J.: Tutorial on agent-based modelling and simulation. In: *Journal of Simulation* (2010), S. 151–162

**Macy und Flache 2002**

MACY, Michael W. ; FLACHE, Andreas: Learning dynamics in social dilemmas. In: *Proceedings of the National Academy of Sciences* 99 (2002), Nr. 90003, S. 7229–7236

**Malsch 1997**

MALSCH, Thomas: Die Provokation der "Artificial Societies": Warum die Soziologie sich mit den Sozialmetaphern der Verteilten Künstlichen Intelligenz beschäftigen sollte. In: *Zeitschrift für Soziologie* 26 (1997), Nr. 1, S. 3–21

**Malsch 1998**

MALSCH, Thomas (Hrsg.): *Sozionik. Soziologische Ansichten über künstliche Sozialität*. Berlin : Edition sigma, 1998

**Manzey 2012**

MANZEY, Dietrich: Systemgestaltung und Automatisierung. In: BADKE-SCHAUB, Petra (Hrsg.) ; HOFINGER, Gesine (Hrsg.) ; LAUCHE, Kristina (Hrsg.): *Human Factors*. Springer Berlin Heidelberg, 2012, S. 333–352. – URL [http://dx.doi.org/10.1007/978-3-642-19886-1\\_19](http://dx.doi.org/10.1007/978-3-642-19886-1_19)

**Manzey und Bahner 2005**

MANZEY, Dietrich ; BAHNER, Elin: Vertrauen in Automation als Aspekt der Verlässlichkeit von Mensch-Maschine-Systemen. In: KARRER, Katja (Hrsg.) ; GAUSS, Boris (Hrsg.) ; STEFFENS, Christiane (Hrsg.): *Beiträge zur Mensch-Maschine-Systemtechnik aus Forschung und Praxis*. Düsseldorf : Symposion Publishing, 2005, S. 93–109

**Manzo 2014**

MANZO, Gianluca: *Analytical Sociology: Actions and Networks*. 1. Aufl. s.l : Wiley, 2014 (Wiley Series in Computational and Quantitative Social Science). – URL [http://ebooks.ciando.com/book/index.cfm/bok\\_id/1371050](http://ebooks.ciando.com/book/index.cfm/bok_id/1371050)

**Mayntz 1967**

MAYNTZ, Renate: Modellkonstruktion - Ansatz, Typen und Zweck. In: MAYNTZ, Renate (Hrsg.): *Formalisierte Modelle in der Soziologie*. Neuwied : Luchterhand, 1967, S. 11–31

**McKnight und Chervany 2001**

MCKNIGHT, Harrison D. ; CHERVANY, Norman L.: Trust and Distrust Definitions: One Bite at a Time. In: FALCONE, Rino (Hrsg.) ; SINGH, Munindar (Hrsg.) ; TAN, Yao-Hua (Hrsg.): *Trust in Cyber-societies* Bd. 2246. Berlin : Springer, 2001, S. 27–54

**McKnight u. a. 1998**

MCKNIGHT, Harrison D. ; CUMMINGS, Larry L. ; CHERVANY, Norman L.: Initial Trust Formation in New Organizational Relationships. In: *The Academy of Management Review* 23 (1998), Nr. 3, S. 473–490

**Montgomery 2013**

MONTGOMERY, Douglas C.: *Design and analysis of experiments*. 8. ed., international student version. Hoboken and NJ : Wiley, 2013

**Moray und Inagaki 2000**

MORAY, Neville ; INAGAKI, Toshiyuki: Attention and complacency. In: *Theoretical Issues in Ergonomics Science* 1 (2000), Nr. 4, S. 354–365

**Muir und Moray 1996**

MUIR, B. M. ; MORAY, Neville: Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. In: *Ergonomics* 36 (1996), S. 429–460

**Muir 1987**

MUIR, Bonnie M.: Trust between humans and machines, and the design of decision aids. In: *International Journal of Man Machine Studies* 27 (1987), Nr. 5-6, S. 527–539

**Nikolic und Kasmire 2013**

NIKOLIC, Igor ; KASMIRE, J.: Chapter 2: Theory. In: VAN DAM, Koen H. (Hrsg.) ; NIKOLIC, Igor (Hrsg.) ; LUKSZO, Zofia (Hrsg.): *Agent-Based Modelling of Socio-Technical Systems* Bd. 9. Dordrecht : Springer Netherlands, 2013, S. 11–71

**Nooteboom 2002**

NOOTEBOOM, Bart: *Trust: Forms, Foundations, Functions, failures and Figures*. Cheltenham UK : Edward Elgar, 2002

**Norman 1983**

NORMAN, Donald A.: Some Observations on Mental Models. In: GENTNER, Dedre (Hrsg.) ; STEVENS, Albert L. (Hrsg.): *Mental Models*. Lawrence Erlbaum Associates : Erlbaum, 1983 (Cognitive science), S. 7–14

**Norman 1986**

NORMAN, Donald A.: Cognitive engineering. In: *User centered system design* (1986), S. 31–61

**Opp 1970**

OPP, Karl-Dieter: The experimental method in the social sciences. In: *Quality & Quantity* 4 (1970), Nr. 1, S. 39–54

**Opp 1998**

OPP, Karl-Dieter: Can and Should Rational Choice Theory Be Tested by Survey Research? The Example of Explaining Collective Political Action. In: BLOSSFELD, Hans-Peter (Hrsg.): *Rational choice theory and large scale data analysis*. Boulder and Colorado : Westview Press, 1998, S. 204–230

**Opp 2009**

OPP, Karl-Dieter: Das individualistische Erklärungsprogramm in der Soziologie. Entwicklung, Stand und Probleme. In: *Zeitschrift für Soziologie* 38 (2009), S. 26–47

**Osgood 1952**

OSGOOD, Charles E.: The nature and measurement of meaning. In: *Psychological*

---

*Bulletin* 49 (1952), Nr. 3, S. 197–237. – URL <http://search.ebscohost.com/login.aspx?direct=true&db=pdh&AN=1953-02510-001&site=ehost-live>

**Ostrom 1990**

OSTROM, Elinor: *Governing the commons: The evolution of institutions for collective action*. Cambridge : Cambridge University Press, 1990

**Ostrom 2005**

OSTROM, Elinor: *Russell Sage Foundation series on trust*. Bd. 6: *Trust and reciprocity: Interdisciplinary lessons from experimental research*. New York : Russell Sage Foundation, 2005

**Parasuraman u. a. 2000**

PARASURAMAN, R. ; SHERIDAN, T. B. ; WICKENS, C. D.: A model for types and levels of human interaction with automation. In: *IEEE Transactions on Systems, Man, and Cybernetics* 30 (2000), Nr. 3, S. 286–297. – URL doi: 10.1109/3468.844354

**Parasuraman und Riley 1997**

PARASURAMAN, Raja ; RILEY, Victor: Humans and Automation: Use, Misuse, Disuse, Abuse. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 39 (1997), Nr. 2, S. 230–253

**Perrow 1987**

PERROW, Charles: *Normale Katastrophen. Die unvermeidbaren Risiken der Großtechnik*. Frankfurt/M. : Campus, 1987

**Peuker und Voss 2006**

PEUKER, Birgit (Hrsg.) ; VOSS, Martin (Hrsg.): *Einleitung: Vom realen Verschwinden einer Fiktion: Verschwindet die Natur? Die Akteur-Netzwerk-Theorie in der umweltsoziologischen Diskussion*. transcript-Verlag, 2006

**Pinch und Bijker 1984**

PINCH, Trevor J. ; BIJKER, Wiebe E.: The social construction of facts and artifacts: Or how the sociology of science and the sociology of technology might benefit each other. In: *Social Studies of Science* 14 (1984), Nr. 3, S. 399–441

**Preisendörfer 1995**

PREISENDÖRFER, Peter: Vertrauen als soziologische Kategorie. Möglichkeiten und Grenzen einer entscheidungstheoretischen Fundierung des Vertrauenskonzepts. In: *Zeitschrift für Soziologie* 24 (1995), Nr. 4, S. 263–272

**Rammert 2003**

RAMMERT, Werner: Technik in Aktion: Verteiltes Handeln in soziotechnischen

Konstellationen. In: CHRISTALLER, Thomas (Hrsg.) ; WEHNER, Josef (Hrsg.): *Autonome Maschinen*. Wiesbaden : Westdeutscher Verlag, 2003, S. 289–315

**Rammert und Schubert 2006**

RAMMERT, Werner (Hrsg.) ; SCHUBERT, Cornelius (Hrsg.): *Technografie: Zur Mikrosoziologie der Technik*. Frankfurt/M. : Campus, 2006

**Rammert und Schulz-Schaeffer 2002a**

RAMMERT, Werner (Hrsg.) ; SCHULZ-SCHAEFFER, Ingo (Hrsg.): *Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik*. Frankfurt/M. : Campus, 2002

**Rammert und Schulz-Schaeffer 2002b**

RAMMERT, Werner ; SCHULZ-SCHAEFFER, Ingo: Technik und Handeln. Wenn soziales Handeln sich auf menschliches Verhalten und technische Abläufe verteilt. In: RAMMERT, Werner (Hrsg.) ; SCHULZ-SCHAEFFER, Ingo (Hrsg.): *Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik*. Frankfurt/M. : Campus, 2002, S. 11–64

**Rao und Georgeff 1995**

RAO, Anand S. ; GEORGEFF, Michael P.: BDI Agents: From Theory to Practice. In: *Proceedings of the First International Conference on Multiagent Systems (ICMAS'95)*, 1995, S. 312–319

**Raub und Buskens 2006**

RAUB, Werner ; BUSKENS, Vincent: Spieltheoretische Modellierungen und empirische Anwendungen in der Soziologie. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 44 (2006), Nr. 56, S. 560–598

**Reips und Funke 2008**

REIPS, Ulf-Dietrich ; FUNKE, Frederik: Interval-level measurement with visual analogue scales in Internet-based research: VAS Generator. In: *Behavior Research Methods* 40 (2008), Nr. 3, S. 699–704

**Robert Bosch GmbH 2013**

ROBERT BOSCH GMBH: *Driver assistance systems — How much support do German drivers want?* 2013. – URL [http://www.bosch-presse.de/presseforum/download/de/7966ks-e\\_Detailed\\_information\\_\\_driver\\_survey.pdf](http://www.bosch-presse.de/presseforum/download/de/7966ks-e_Detailed_information__driver_survey.pdf)

**Rohwer 2003**

ROHWER, Götz: Modelle ohne Akteure. Hartmut Essers Erklärung von Scheidungen. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* (2003), Nr. 2, S. 340–358

**Scheibehenne u. a. 2007**

SCHEIBEHENNE, Benjamin ; MIESLER, Linda ; TODD, Peter M.: Fast and frugal food choices: Uncovering individual decision heuristics. In: *Appetite* 49 (2007), Nr. 3, S. 578–589. – URL <http://www.sciencedirect.com/science/article/pii/S019566630700267X>

**Schelling 1971**

SHELLING, Thomas C.: Dynamic Models of Segregation. In: *Journal of Mathematical Sociology* (1971), Nr. 1, S. 143–186

**Schreckenberg und Sharma 2002**

SCHRECKENBERG, Michael (Hrsg.) ; SHARMA, Som D. (Hrsg.): *Pedestrian and evacuation dynamics*. Berlin : Springer, 2002. – ISBN 978-3-540-42690-5

**Schulte-Mecklenbeck u. a. 2013**

SCHULTE-MECKLENBECK, Michael ; SOHN, Matthias ; BELLIS, Emanuel d. ; MARTIN, Nathalie ; HERTWIG, Ralph: A lack of appetite for information and computation. Simple heuristics in food choice. In: *Appetite* 71 (2013), Nr. 0, S. 242–251. – URL <http://www.sciencedirect.com/science/article/pii/S0195666313003668>

**Schulz-Schaeffer 2000**

SCHULZ-SCHAEFFER, Ingo: *Campus Forschung*. Bd. 803: *Sozialtheorie der Technik*. Frankfurt : Campus Verlag, 2000

**Schulz-Schaeffer 2007**

SCHULZ-SCHAEFFER, Ingo: *Zugeschriebene Handlungen: Ein Beitrag zur Theorie sozialen Handelns*. Weilerswist : Velbrück, 2007

**Schulz-Schaeffer 2008**

SCHULZ-SCHAEFFER, Ingo: Die drei Logiken der Selektion. Handlungstheorie als Theorie der Situationsdefinition. In: *Zeitschrift für Soziologie* 37 (2008), Nr. 5, S. 362–379

**Schulz-Schaeffer 2011**

SCHULZ-SCHAEFFER, Ingo: Akteur-Netzwerk-Theorie. Zur Ko-Konstitution von Gesellschaft, Natur und Technik. In: WEYER, Johannes (Hrsg.): *Soziale Netzwerke. Konzepte und Methoden der sozialwissenschaftlichen Netzwerkforschung* (2. Aufl.). München : Oldenbourg, 2011, S. 264–288

**Siedenbiedel 2013**

SIEDENBIEDEL, Christian: Sklaven des Algorithmus. In: *Frankfurter Allgemeine Sonntagszeitung* (2013), Nr. 3, 20.01.2013, S. 24–25

**Sietmann 2011**

SIETMANN, Richard: Wettrüsten: Wertpapierhändler investieren in schnelle Netze. In: *c't - Magazin für Computertechnik* (2011), Nr. 12, S. 150–155

**Spiegel Online 2010**

SPIEGEL ONLINE: *Tippfehler verursachte möglicherweise Wall-Street-Chaos*. 2010. – URL <http://www.spiegel.de/wirtschaft/unternehmen/kursabsturz-tippfehler-verursachte-moeglicherweise-wall-street-chaos-a-693495.html>. – Zugriffsdatum: 07.05.2010

**Stanton und Young 2005**

STANTON, Neville A. ; YOUNG, Mark S.: Driver behaviour with Adaptive Cruise Control. In: *Ergonomics* 48 (2005), S. 1294–1313

**Stephoe u. a. 1995**

STEPHOE, Andrew ; POLLARD, Tessa M. ; WARDLE, Jane: Development of a Measure of the Motives Underlying the Selection of Food: the Food Choice Questionnaire. In: *Appetite* 25 (1995), Nr. 3, S. 267–284. – URL <http://www.sciencedirect.com/science/article/pii/S019566638570061X>

**Suchman 1987**

SUCHMAN, Lucy: *Plans and Situated Actions*. Cambridge University Press, 1987

**Thomas und Thomas 1928**

THOMAS, William I. ; THOMAS, Dorothy S.: The Methodology of Behavior Study. In: KNOPF, Alfred A. (Hrsg.): *The Child in America: Behavior Problems and Programs*. New York, 1928, S. 553–576

**Van Dam u. a. 2013**

VAN DAM, Koen H. (Hrsg.) ; NIKOLIC, Igor (Hrsg.) ; LUKSZO, Zofia (Hrsg.): *Agent-Based Social Systems*. Bd. 9: *Agent-Based Modelling of Socio-Technical Systems*. Dordrecht : Springer Netherlands, 2013. – URL <http://dx.doi.org/10.1007/978-94-007-4933-7>. – ISBN 978-94-007-4933-7

**Vanderbilt 2012**

VANDERBILT, Tom: Let the Robot Drive: The Autonomous Car of the Future is Here. In: *Wired* 20 (2012), Nr. 2, S. 86–95; 124. – URL [http://www.wired.com/magazine/2012/01/ff\\_autonomoucars/](http://www.wired.com/magazine/2012/01/ff_autonomoucars/)

**Weiser 1991**

WEISER, Marc: The Computer for the 21st Century: Specialized elements of hardware and software, connected by wires, radio waves and infrared, will be so ubiquitous that no one will notice their presence. In: *Scientific American* (1991), Nr. 1991, S. 94–104

**Weyer 1997**

WEYER, Johannes: Die Risiken der Automationsarbeit: Mensch-Maschine-Interaktion und Störfallmanagement in hochautomatisierten Verkehrsflugzeugen. In: *Zeitschrift für Soziologie* 26 (1997), Nr. 4, S. 239–257

**Weyer 2006**

WEYER, Johannes: *Die Kooperation menschlicher Akteure und nicht-menschlicher Agenten: Ansatzpunkte einer Soziologie hybrider Systeme*. 2006

**Weyer 2007**

WEYER, Johannes: Autonomie und Kontrolle: Arbeit in hybriden Systemen am Beispiel der Luftfahrt. In: *Technikfolgenabschätzung - Theorie und Praxis* 16 (2007), Nr. 2, S. 35–42

**Weyer 2008a**

WEYER, Johannes: Mixed Governance - Das Zusammenspiel von menschlichen Entscheidern und autonomer Technik im Luftverkehr der Zukunft. In: VOSS, Günter (Hrsg.) ; MATUSCHEK, Ingo (Hrsg.): *Luft-Schichten*. Berlin : Edition sigma, 2008, S. 188–208

**Weyer 2008b**

WEYER, Johannes: *Techniksoziologie: Genese, Gestaltung und Steuerung sozio-technischer Systeme*. Weinheim : Juventa, 2008

**Weyer 2011**

WEYER, Johannes (Hrsg.): *Soziale Netzwerke. Konzepte und Methoden der sozialwissenschaftlichen Netzwerkforschung (2. Aufl.)*. München : Oldenbourg, 2011

**Weyer u. a. 2014**

WEYER, Johannes ; FINK, Robin D. ; ADEL, Fabian: Human-machine cooperation in smart cars. An empirical investigation of the loss-of-control thesis: Akzeptiertes Manuskript. In: *Safety Science* (2014)

**Wilensky und Stroup 2000**

WILENSKY, Uri ; STROUP, Walter M.: Networked gridlock: Students enacting complex dynamic phenomena with the HubNet architecture. In: FISHMAN, Barry J. (Hrsg.) ; O'CONNOR-DIVELBISS, Samuel F. (Hrsg.): *Fourth International Conference of the Learning Sciences*. Mahwah and NJ : Erlbaum, 2000, S. 282–289

**Winkler 2011**

WINKLER, Fabian: *Essenswahl in der Mensa: Ein empirisch fundiertes Handlungsmodell*. Dortmund, TU Dortmund, Dissertation, 2011

**Winner u. a. 2002**

WINNER, H. ; WINTER, K. ; LUCAS, B. ; MAYER, H. ; IRION, A. ; SCHNEIDER, H.-P ; LÜDER, J. ; ZABLER, E. ; DENNER, V. ; WALTHER, M.: *Adaptive Fahrgeschwindigkeitsregelung ACC*. 2002

**Wooldridge und Jennings 1995**

WOOLDRIDGE, Michael ; JENNINGS, Nicolas R.: Intelligent agents: theory and practice. In: *The Knowledge Engineering Review* 10 (1995), Nr. 2, S. 115–152

**Yilmaz 2012**

YILMAZ, Zeynep: *Rationalität und Alltagshandeln. Eine empirische Überprüfung der Rational-Choice-Theorie (Diplomarbeit, TU Dortmund)*. 2012

# A Anhang

## A.1 Übersicht über Wikipedia-Namensräume

s. Tabelle A.1

## A.2 Wikipedia-Datenquellen

### A.2.1 Datengrundlage zu Abbildung 5.4

- Abbildung 5.4a [http://commons.wikimedia.org/w/index.php?title=User:DrTrigon/User:DrTrigonBot/logging&oldid=88934212#File:A.M.\\_Fairbairn.jpg](http://commons.wikimedia.org/w/index.php?title=User:DrTrigon/User:DrTrigonBot/logging&oldid=88934212#File:A.M._Fairbairn.jpg) (Abruf: 31. Januar 2014)
- Abbildung 5.4b: [http://commons.wikimedia.org/w/index.php?title=File:A.M.\\_Fairbairn.jpg&oldid=88929241](http://commons.wikimedia.org/w/index.php?title=File:A.M._Fairbairn.jpg&oldid=88929241) (Abruf: 31. Januar 2014)
- Abbildung 5.4c: [http://commons.wikimedia.org/w/index.php?title=File:A.M.\\_Fairbairn.jpg&action=history](http://commons.wikimedia.org/w/index.php?title=File:A.M._Fairbairn.jpg&action=history) (Abruf: 31. Januar 2014)

### A.2.2 Datengrundlage zu Abbildung 5.5

[http://de.wikipedia.org/w/index.php?title=Benutzer\\_Diskussion:The\\_weaver&oldid=113430944#Problem\\_mit\\_deiner\\_Datei\\_.2818.01.2013.29](http://de.wikipedia.org/w/index.php?title=Benutzer_Diskussion:The_weaver&oldid=113430944#Problem_mit_deiner_Datei_.2818.01.2013.29)  
(Abruf: 31. Januar 2014)

Nr.	Präfix	Beschreibung
0	<kein:>	Artikelnamensraum der Enzyklopädie. Mit diesem Namensraum kommt der größte Teil der passiven Nutzer (Leser) in Berührung
1	Diskussion:	Diskussionsseite zu Artikeln, auf der z. B. geplante Änderungen an Artikel vorbesprochen werden.
2	Benutzer:	Persönliche Nutzerseite eines angemeldeten Nutzers.
3	Benutzer_ Diskussion:	Diskussionsseite eines Nutzers, über die z. B. Nachrichten zugestellt werden.
4	Wikipedia:	Metabereich der Wikipedia.
5	Wikipedia_ Diskussion:	Diskussionsseiten zum Wikipedia-Namensraum
6	Datei:	Informationen zu einer Medien-Datei (meist Bilder)
7	Datei_ Diskussion:	Diskussionsseiten zu einer Datei
8	MediaWiki <sup>a</sup> :	Texte der MediaWiki-Software
9	MediaWiki_ Diskussion:	Diskussionsseiten zu MediaWiki
10	Vorlage:	Textbausteine und Formatierungshilfen
11	Vorlage_ Diskussion:	Diskussionsseiten zu Vorlagen
12	Hilfe:	Allgemeine Hilfeseiten
13	Hilfe_ Diskussion:	Diskussionsseiten zu allgemeinen Hilfeseiten
14	Kategorie:	Enthält Kategorien, die der Kategorisierung von Artikel dienen
15	Kategorie_ Diskussion:	Diskussionsseiten zu Kategorien
100	Portal:	Portalseiten, die Übersichten zu bestimmten Themenbereichen umfassen
101	Portal_ Diskussion:	Diskussionsseiten zu Portalen

<sup>a</sup> MediaWiki ist der Name der Web-Software, auf der die gesamte Wikipedia basiert. MediaWiki ist in der Programmiersprache PHP geschrieben und lässt sich auf nahezu jedem Webserver ausführen.

**Tabelle A.1:** Wikipedia-Namensräume besitzen eine eindeutige Nummer (0-101). Jede Editierung, die in der Wikipedia erfolgt, kann eindeutig einem Namensraum zugeordnet werden.

Quelle: <http://de.wikipedia.org/wiki/Hilfe:Namensräume>,

Abruf: 31. Januar 2014

## **A.3 Mensa-Befragung Winkler**

### **A.3.1 Durchschnittliches SEU-Modell für häufigste Wahl**

haeufigste wahl		z_satt	z_schnell	z_lecker	z_zahlen	z_gesund	wahl_satt	wahl_schnell	wahl_lecker	wahl_zahlen	wahl_gesund	veg_satt	veg_schnell	veg_lecker	veg_zahlen	veg_gesund
Vegetarisch	Mittelwert	2,44	1,22	2,48	1,59	1,96	,7533	,6189	,5348	,7778	,5130	,7352	,8637	,7504	,6107	,8193
	N	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
	Standardabweichung	,506	1,368	,802	1,366	,898	,24434	,21353	,24241	,23056	,26483	,25737	,16593	,18693	,29589	,16074
Wahlessen	Mittelwert	2,28	1,09	2,06	1,43	,91	,7532	,5528	,5785	,7247	,4434	,6368	,7653	,5981	,4766	,7426
	N	47	47	47	47	47	47	47	47	47	47	47	47	47	47	47
	Standardabweichung	,772	1,365	,919	1,118	1,139	,16436	,19089	,19535	,19604	,17142	,22110	,17792	,21198	,20002	,16664
Aktionsteller	Mittelwert	2,38	2,00	2,46	1,54	1,00	,7985	,5515	,6415	,7785	,5669	,5069	,6915	,5762	,5946	,7885
	N	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13
	Standardabweichung	,768	1,000	,660	1,127	1,414	,16732	,22086	,16856	,22349	,24622	,25401	,33682	,22017	,24367	,17860
Tagesgericht	Mittelwert	2,23	1,23	2,08	2,31	1,85	,8292	,5646	,5785	,7415	,5200	,6985	,7885	,6692	,4962	,8069
	N	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13
	Standardabweichung	,439	,832	,760	,855	1,144	,13775	,13182	,20178	,24603	,14742	,22158	,19786	,15692	,18049	,19397
Insgesamt	Mittelwert	2,33	1,26	2,23	1,60	1,33	,7690	,5720	,5749	,7482	,4882	,6545	,7853	,6456	,5307	,7776
	N	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	Standardabweichung	,667	1,284	,851	1,181	1,207	,18628	,19432	,20633	,21409	,21005	,24274	,20846	,20902	,23782	,17117

#### Bericht

haeufigste wahl		ak_satt	ak_schnell	ak_lecker	ak_zahlen	ak_gesund	tag_satt	tag_schnell	tag_lecker	tag_zahlen	tag_gesund
Vegetarisch	Mittelwert	,7544	,6426	,6137	,5744	,4396	,7711	,6722	,6141	,8356	,5478
	N	27	27	27	27	27	27	27	27	27	27
	Standardabweichung	,26258	,21504	,24201	,29849	,24791	,23492	,21800	,20539	,18383	,20209
Wahlessen	Mittelwert	,7519	,6609	,7072	,3804	,4166	,5726	,5881	,5281	,8489	,5106
	N	47	47	47	47	47	47	47	47	47	47
	Standardabweichung	,19804	,19188	,18677	,23659	,17863	,25073	,22121	,19432	,15671	,13116
Aktionsteller	Mittelwert	,7915	,6785	,7915	,5823	,4500	,5608	,6600	,5923	,8862	,5285
	N	13	13	13	13	13	13	13	13	13	13
	Standardabweichung	,25116	,22788	,18184	,27893	,27092	,26660	,15055	,22822	,11169	,18064
Tagesgericht	Mittelwert	,7462	,6177	,6615	,6154	,4069	,7831	,7138	,6115	,9315	,5500
	N	13	13	13	13	13	13	13	13	13	13
	Standardabweichung	,15777	,18895	,17024	,29413	,20894	,19311	,17003	,23727	,08640	,20265
Insgesamt	Mittelwert	,7570	,6526	,6870	,4896	,4259	,6520	,6365	,5705	,8609	,5281
	N	100	100	100	100	100	100	100	100	100	100
	Standardabweichung	,21721	,20045	,20565	,28311	,21301	,25895	,20926	,20837	,15390	,16723

## **A.4 Mensa-Befragung Yilmaz**

### **A.4.1 Durchschnittliche SEU-Modelle**

**Bericht**

haeufigste wahl		wahl_satt	wahl_schnell	wahl_lecker	wahl_zahlen	wahl_gesund	veg_satt	veg_schnell	veg_lecker	veg_zahlen	veg_gesund	ak_satt	ak_schnell	ak_lecker	ak_zahlen
Vegetarisch	Mittelwert	,7537	,6007	,4870	,7089	,3833	,7507	,7911	,7078	,6441	,6593	,7741	,5552	,5452	,5970
	N	27	27	27	27	27	27	27	27	27	27	27	27	27	27
	Standardabweichung	,21733	,24866	,22710	,24880	,21667	,21635	,19387	,18385	,24230	,23450	,20319	,24557	,21351	,32212
Wahlessen	Mittelwert	,7709	,5972	,5815	,7847	,4430	,5685	,7589	,5248	,6077	,7166	,7879	,6320	,7047	,4526
	N	92	92	92	92	92	92	92	92	92	92	92	92	92	92
	Standardabweichung	,17389	,21523	,18729	,21386	,17451	,25334	,18447	,19624	,21848	,19412	,17126	,22192	,18296	,27906
Aktionsteller	Mittelwert	,7359	,5335	,5853	,8065	,5071	,5282	,6894	,5224	,5865	,7288	,7653	,6571	,7547	,6253
	N	17	17	17	17	17	17	17	17	17	17	17	17	17	17
	Standardabweichung	,18176	,24184	,18480	,12530	,20978	,27965	,21882	,24578	,24078	,27441	,16553	,18583	,15824	,18901
Tagesgericht	Mittelwert	,7011	,5363	,6115	,7626	,4811	,4548	,7707	,5085	,5274	,7196	,7637	,6237	,6541	,4867
	N	27	27	27	27	27	27	27	27	27	27	27	27	27	27
	Standardabweichung	,21566	,22327	,16997	,19453	,21804	,23809	,13191	,25517	,24839	,18559	,15559	,19606	,15308	,28904
Insgesamt	Mittelwert	,7528	,5810	,5712	,7707	,4461	,5756	,7590	,5521	,5982	,7089	,7793	,6205	,6751	,5002
	N	163	163	163	163	163	163	163	163	163	163	163	163	163	163
	Standardabweichung	,18961	,22473	,19371	,21010	,19461	,26114	,18268	,21988	,23047	,20848	,17263	,21875	,19079	,28611

**Bericht**

haeufigste wahl		ak_gesund	tag_satt	tag_schnell	tag_lecker	tag_zahlen	tag_gesund	z_satt	z_schnell	z_lecker	z_zahlen	z_gesund
Vegetarisch	Mittelwert	,303	,7033	,5141	,486	,7404	,422	2,222	,500	2,389	1,000	1,778
	N	27	27	27	27	27	27	27	27	27	27	27
	Standardabweichung	,2045	,24726	,21286	,2112	,24154	,2031	,8697	1,3156	,7511	1,2481	1,1794
Wahlessen	Mittelwert	,404	,6565	,6405	,553	,7659	,496	2,152	,962	2,103	,880	,489
	N	92	92	92	92	92	92	92	92	92	92	92
	Standardabweichung	,1829	,20938	,19080	,1866	,22689	,1526	,8979	1,2497	,9960	1,4265	1,5583
Aktionsteller	Mittelwert	,501	,6671	,6206	,535	,7682	,493	2,294	,882	2,118	,882	,794
	N	17	17	17	17	17	17	17	17	17	17	17
	Standardabweichung	,1443	,17832	,15164	,1948	,16712	,1438	,7717	1,4090	,9275	1,4090	1,5110
Tagesgericht	Mittelwert	,441	,6704	,6178	,642	,8089	,506	2,222	1,111	1,944	1,722	,944
	N	27	27	27	27	27	27	27	27	27	27	27
	Standardabweichung	,1902	,18035	,19407	,1793	,16823	,1743	,8697	1,2195	1,1630	1,2274	1,5648
Insgesamt	Mittelwert	,403	,6677	,6137	,555	,7690	,485	2,190	,902	2,126	1,040	,810
	N	163	163	163	163	163	163	163	163	163	163	163
	Standardabweichung	,1905	,20757	,19521	,1942	,21445	,1656	,8700	1,2752	,9834	1,3878	1,5568

## A.5 MensaAssist

### A.5.1 Fragebogen zu MensaAssist

**Umfrage Mensa-Assist**





Liebe Teilnehmerin, lieber Teilnehmer,

im Rahmen eines Projektseminars am Fachgebiet Techniksoziologie der Technischen Universität Dortmund (Fakultät WiSo) untersuchen wir als studentische Forschungsgruppe den Umgang mit moderner Technik. Wir bitten Sie die folgenden Fragen möglichst ehrlich zu beantworten. Wichtig ist dabei Ihre individuelle Einschätzung – es gibt daher keine „richtigen“ oder „falschen“ Antworten!

Bitte wählen Sie für Ihre Antworten die entsprechenden Kästchen aus. Alle Angaben werden selbstverständlich anonymisiert und werden nicht mit Ihrer Person in Verbindung gebracht. Die Ergebnisse werden ausschließlich für Forschungszwecke verwendet.

Vielen Dank für Ihre Bereitschaft zur Teilnahme!

Zwischengespeicherte Umfrage laden
Weiter >>
Umfrage verlassen und löschen

0%  100%

### Moderne Technik

**\* Wie denken Sie über neue technische Geräte und neue Technik im Allgemeinen?**

	Trifft überhaupt nicht zu 1	2	3	4	Trifft voll zu 5	keine Angabe
Wenn ein neues technisches Gerät auf den Markt kommt, bin ich einer der ersten, der es kauft.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wenn ich ein neues technisches Gerät erworben habe, bin ich schnell mit allen Funktionen vertraut.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Neue Technik enthält oft überflüssige Funktionen, die ich nicht benötige.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich zögere, neue Technik zu nutzen, weil ich Angst habe, bei der Bedienung Fehler zu machen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mich stört es, dass ständig neue technische Geräte auf den Markt kommen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
In Sachen Technik verlasse ich mich lieber auf Altbewährtes.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Neue Technik macht das Leben leichter.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Später Fortfahren
Weiter >>
Umfrage verlassen und löschen

0%  100%

## Einleitung

**Wichtig: Bitte unbedingt aufmerksam lesen!**

*Es folgen zwei Seiten erklärender Text; danach beginnt die eigentliche Umfrage.*

**Bitte versetzen Sie sich für die Bearbeitung des nächsten Fragenabschnittes in die folgende Situation:**

Sie gehen in die Mensa und haben drei verschiedene Wahlmöglichkeiten für ein Essen. Die angebotenen Gerichte unterscheiden sich anhand von fünf Kriterien:

- Geschmack
- Menge
- Gesunde Ernährung
- Wartezeit
- Preis

Bei der Bewertung der einzelnen Eigenschaften bedeutet 1 Punkt (●) "ist nicht erfüllt" bis 5 Punkte (●●●●●) "ist voll erfüllt".

Das folgende Beispiellesen zeichnet sich somit durch eine **große Menge** (●●●●) aus. Weiterhin ist das Kriterium des **geringen Preises voll erfüllt** (●●●●●), das Essen ist also sehr günstig. Das Kriterium der **kurzen Wartezeit** ist jedoch **nicht erfüllt** (●), somit muss man lange auf sein Essen warten.

	Essen 1
Guter Geschmack	●●●
Große Menge	●●●●
Gesunde Ernährung	●●
Kurze Wartezeit	●
Geringer Preis	●●●●●

**MensaAssist**

Wie eingangs erwähnt, geht es in der Befragung um den Umgang mit moderner Technik. Nachdem Sie also Ihre Entscheidung für ein Essen getroffen haben (unter Berücksichtigung Ihrer Vorlieben) kommt ein technisches Hilfssystem ins Spiel.

Bei dem hier betrachteten technischen Gerät handelt es sich um "MensaAssist", einem noch in der Entwicklung befindlichen Assistenzsystem, das Sie in Zukunft bei der für Sie passenden Essenswahl in überfüllten Mensen und Cafeterien unterstützen könnte. In diesem Fragebogen geht es erst einmal darum, Ihre Entscheidungen mit denen von MensaAssist zu vergleichen und zu bewerten.

0%  100%

### Typenauswahl

Um Ihnen in den folgenden Entscheidungssituationen die Möglichkeit zu geben, sich selbst und Ihre eigenen Präferenzen wiederzufinden, stehen Ihnen drei Typen zur Auswahl.

Diese drei Typen unterscheiden sich in ihrer Entscheidungsfindung in Bezug auf den Preis, die Wartezeit und die gesundheitsfördernde Wirkung der Gerichte.

Im Einzelnen lassen sich die Typen wie folgt beschreiben:

**Der Sparfuchs:**  
"Sehr wichtig ist mir, dass ich Geld spare. Danach achte ich zuerst auf die Menge und dann auf den Geschmack. Ich habe im Prinzip keinen Stress (Wartezeit) und brauche absolut kein gesundes Essen (Gesundheit)."

**Der Gesundheitsbewusste:**  
"Sehr wichtig ist mir, dass ich satt werde UND es gut schmeckt. Danach achte ich zuerst darauf Gesundes zu wählen und dann auf die Wartezeit. Ich habe im Prinzip keine Geldsorgen (Preis)."

**Der Eilige:**  
"Sehr wichtig ist mir, dass es gut schmeckt und dann, dass ich satt werde. Dann achte ich zuerst darauf, dass die Wartezeit gering ist und ich Geld spare. Ich verschwende keine Gedanken an meine Gesundheit."

**\* Welcher der oben beschriebenen Typen entspricht am ehesten Ihren eigenen Neigungen?**

Der Sparfuchs  
 Der Gesundheitsbewusste  
 Der Eilige

0%  100%

### Tag 1

Zur Erinnerung: Sie haben sich am ehesten dem Typ **'Der Sparfuchs'** zugeordnet, der seine Essenswahl wie folgt trifft:

"Sehr wichtig ist mir, dass ich Geld spare. Danach achte ich zuerst auf die Menge und dann auf den Geschmack. Ich habe im Prinzip keinen Stress (Wartezeit) und brauche absolut kein gesundes Essen (Gesundheit)."

**Tag 1/6:**

Die Eigenschaften der Gerichte des heutigen Tages gestalten sich wie folgt:  
[Hinweis: 1 Punkt (●) = trifft nicht zu bis 5 Punkte (●●●●●) = trifft zu]

	Essen 1	Essen 2	Essen 3
Guter Geschmack	●●●●●	●●●●●	●●●
Große Menge	●●●●●	●	●●●●
Gesunde Ernährung	●●	●●●	●
Kurze Wartezeit	●●	●●●●	●●●●●
Geringer Preis	●	●●●●●	●●

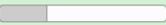
**\* Für welches dieser Gerichte würden Sie sich entscheiden?**

Essen 1  
 Essen 2  
 Essen 3

**\* MensaAssist rät Ihnen, Essen 2 zu wählen.**

Würden Sie gerne dem Vorschlag von MensaAssist folgen?

Ja     Nein

0%  100%**Tag 2**

Zur Erinnerung: Sie haben sich am ehesten dem Typ **'Der Sparfuchs'** zugeordnet, der seine Essenswahl wie folgt trifft:

"Sehr wichtig ist mir, dass ich Geld spare. Danach achte ich zuerst auf die Menge und dann auf den Geschmack. Ich habe im Prinzip keinen Stress (Wartezeit) und brauche absolut kein gesundes Essen (Gesundheit)."

**Tag 2/6:**

Die Eigenschaften der Gerichte des heutigen Tages gestalten sich wie folgt:  
[Hinweis: 1 Punkt (●) = trifft nicht zu bis 5 Punkte (●●●●●) = trifft zu]

	Essen 1	Essen 2	Essen 3
Guter Geschmack	●●●●●	●●	●●
Große Menge	●●●	●	●●●●●
Gesunde Ernährung	●●	●●●●●	●●
Kurze Wartezeit	●●●	●●●	●●●
Geringer Preis	●	●●●●●	●

\* Für welches dieser Gerichte würden Sie sich entscheiden?

- Essen 1  
 Essen 2  
 Essen 3

MensaAssist schlägt Ihnen **Essen 1** vor.

[Später Fortfahren](#)
[Weiter >>](#)
[Umfrage verlassen und löschen](#)

0%  100%

### Tag 3

Zur Erinnerung: Sie haben sich am ehesten dem Typ **'Der Sparfuchs'** zugeordnet, der seine Essenswahl wie folgt trifft:

"Sehr wichtig ist mir, dass ich Geld spare. Danach achte ich zuerst auf die Menge und dann auf den Geschmack. Ich habe im Prinzip keinen Stress (Wartezeit) und brauche absolut kein gesundes Essen (Gesundheit)."

---

**Tag 3/6:**

Die Eigenschaften der Gerichte des heutigen Tages gestalten sich wie folgt:  
[Hinweis: 1 Punkt (●) = trifft nicht zu bis 5 Punkte (●●●●●) = trifft zu]

	Essen 1	Essen 2	Essen 3
Guter Geschmack	●●●●	●	●●●●●
Große Menge	●●●●●	●●●●	●●
Gesunde Ernährung	●	●●●●	●●
Kurze Wartezeit	●●●	●●	●●●●●
Geringer Preis	●	●●●●●	●

---

\* Für welches dieser Gerichte würden Sie sich entscheiden?

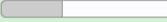
Essen 1  
 Essen 2  
 Essen 3

---

\* MensaAssist schlägt Ihnen **Essen 3** vor.

Würden Sie gerne dem Vorschlag von MensaAssist folgen?

Ja     Nein

0%  100%

### Tag 3

Zur Erinnerung: Sie haben sich am ehesten dem Typ **'Der Sparfuchs'** zugeordnet, der seine Essenswahl wie folgt trifft:

"Sehr wichtig ist mir, dass ich Geld spare. Danach achte ich zuerst auf die Menge und dann auf den Geschmack. Ich habe im Prinzip keinen Stress (Wartezeit) und brauche absolut kein gesundes Essen (Gesundheit)."

---

**Tag 3/6:**

Die Eigenschaften der Gerichte des heutigen Tages gestalten sich wie folgt:  
[Hinweis: 1 Punkt (●) = trifft nicht zu bis 5 Punkte (●●●●●) = trifft zu]

	Essen 1	Essen 2	Essen 3
Guter Geschmack	●●●●	●	●●●●●
Große Menge	●●●●●	●●●●	●●
Gesunde Ernährung	●	●●●●	●●
Kurze Wartezeit	●●●	●●	●●●●●
Geringer Preis	●	●●●●●	●

---

\* Für welches dieser Gerichte würden Sie sich entscheiden?

Essen 1  
 Essen 2  
 Essen 3

---

\* MensaAssist schlägt Ihnen **Essen 3** vor.

Würden Sie gerne dem Vorschlag von MensaAssist folgen?

Ja     Nein

0%  100%

### Tag 3

Zur Erinnerung: Sie haben sich am ehesten dem Typ **'Der Sparfuchs'** zugeordnet, der seine Essenswahl wie folgt trifft:

"Sehr wichtig ist mir, dass ich Geld spare. Danach achte ich zuerst auf die Menge und dann auf den Geschmack. Ich habe im Prinzip keinen Stress (Wartezeit) und brauche absolut kein gesundes Essen (Gesundheit)."

---

**Tag 3/6:**

Die Eigenschaften der Gerichte des heutigen Tages gestalten sich wie folgt:  
[Hinweis: 1 Punkt (●) = trifft nicht zu bis 5 Punkte (●●●●●) = trifft zu]

	Essen 1	Essen 2	Essen 3
Guter Geschmack	●●●●	●	●●●●●
Große Menge	●●●●●	●●●●	●●
Gesunde Ernährung	●	●●●●	●●
Kurze Wartezeit	●●●	●●	●●●●●
Geringer Preis	●	●●●●●	●

---

\* Für welches dieser Gerichte würden Sie sich entscheiden?

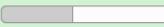
Essen 1  
 Essen 2  
 Essen 3

---

\* MensaAssist schlägt Ihnen **Essen 3** vor.

Würden Sie gerne dem Vorschlag von MensaAssist folgen?

Ja     Nein

0%  100%**Tag 4**

Zur Erinnerung: Sie haben sich am ehesten dem Typ **'Der Sparfuchs'** zugeordnet, der seine Essenswahl wie folgt trifft:

"Sehr wichtig ist mir, dass ich Geld spare. Danach achte ich zuerst auf die Menge und dann auf den Geschmack. Ich habe im Prinzip keinen Stress (Wartezeit) und brauche absolut kein gesundes Essen (Gesundheit)."

**Tag 4/6:**

Die Eigenschaften der Gerichte des heutigen Tages gestalten sich wie folgt:  
[Hinweis: 1 Punkt (●) = trifft nicht zu bis 5 Punkte (●●●●●) = trifft zu]

	Essen 1	Essen 2	Essen 3
Guter Geschmack	●●	●●●	●
Große Menge	●●●●●	●	●●
Gesunde Ernährung	●	●●●●	●●
Kurze Wartezeit	●●●	●●●	●●●●●
Geringer Preis	●	●●●	●●●

\* Für welches dieser Gerichte würden Sie sich entscheiden?

- Essen 1  
 Essen 2  
 Essen 3

MensaAssist schlägt Ihnen **Essen 2** vor.

[Später Fortfahren](#)

[Weiter >>](#)

[Umfrage verlassen und löschen](#)

## Tag 5

Zur Erinnerung: Sie haben sich am ehesten dem Typ **'Der Sparfuchs'** zugeordnet, der seine Essenswahl wie folgt trifft:

"Sehr wichtig ist mir, dass ich Geld spare. Danach achte ich zuerst auf die Menge und dann auf den Geschmack. Ich habe im Prinzip keinen Stress (Wartezeit) und brauche absolut kein gesundes Essen (Gesundheit)."

Zur Erinnerung: Sie haben sich am ehesten dem Typ **'Der Gesundheitsbewusste'** zugeordnet, der seine Essenswahl wie folgt trifft:

"Sehr wichtig ist mir, dass ich satt werde UND es gut schmeckt. Danach achte ich zuerst darauf Gesundes zu wählen und dann auf die Wartezeit. Ich habe im Prinzip keine Geldsorgen (Preis)."

Zur Erinnerung: Sie haben sich am ehesten dem Typ **'Der Eilige'** zugeordnet, der seine Essenswahl wie folgt trifft:

"Sehr wichtig ist mir, dass es gut schmeckt und dann, dass ich satt werde. Dann achte ich zuerst darauf, dass die Wartezeit gering ist und ich Geld spare. Ich verschwende keine Gedanken an meine Gesundheit."

### Tag 5/6:

Die Eigenschaften der Gerichte des heutigen Tages gestalten sich wie folgt:  
[Hinweis: 1 Punkt (●) = trifft nicht zu bis 5 Punkte (●●●●●) = trifft zu]

	Essen 1	Essen 2	Essen 3
Guter Geschmack	●●●●	●●●●●	●●
Große Menge	●●	●●●	●●●●●
Gesunde Ernährung	●●●●●	●●●●	●●
Kurze Wartezeit	●	●●●	●●●
Geringer Preis	●●●●●	●	●

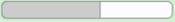
• Für welches dieser Gerichte würden Sie sich entscheiden?

- Essen 1  
 Essen 2  
 Essen 3

• MensaAssist schlägt Ihnen **Essen 2** vor.

Würden Sie gerne dem Vorschlag von MensaAssist folgen?

- Ja     Nein

0%  100%

## Tag 6

Zur Erinnerung: Sie haben sich am ehesten dem Typ **'Der Sparfuchs'** zugeordnet, der seine Essenswahl wie folgt trifft:

"Sehr wichtig ist mir, dass ich Geld spare. Danach achte ich zuerst auf die Menge und dann auf den Geschmack. Ich habe im Prinzip keinen Stress (Wartezeit) und brauche absolut kein gesundes Essen (Gesundheit)."

**Tag 6/6:**

Die Eigenschaften der Gerichte des heutigen Tages gestalten sich wie folgt:  
[Hinweis: 1 Punkt (●) = trifft nicht zu bis 5 Punkte (●●●●●) = trifft zu]

	EsSEN 1	EsSEN 2	EsSEN 3
Guter Geschmack	●●●	●●●	●●●●
Große Menge	●●●●●	●●	●●●
Gesunde Ernährung	●	●●●●	●
Kurze Wartezeit	●	●	●●●●●
Geringer Preis	●●●●	●●●●●	●

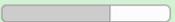
\* Für welches dieser Gerichte würden Sie sich entscheiden?

Essen 1  
 Essen 2  
 Essen 3

\* MensaAssist schlägt Ihnen **EsSEN 3** vor.

Würden Sie gerne dem Vorschlag von MensaAssist folgen?

Ja  Nein

0%  100%

## Ihre Entscheidungsgrundlage

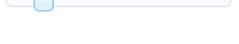
\* Wie wichtig sind Ihnen die untenstehenden Kriterien bei Ihrer Entscheidung für eines der Essen gewesen?

Guter Geschmack    unwichtig  wichtig    40

Große Menge    unwichtig  wichtig    70

Gesunde Ernährung    unwichtig  wichtig    50

Kurze Wartezeit    unwichtig  wichtig    77

Geringer Preis    unwichtig  wichtig    17

0%  100%

### Zufriedenheit

\* Wie zufrieden sind Sie mit den Vorschlägen von MensaAssist? (0 - gar nicht zufrieden , 50 - unentschieden, 100 - sehr zufrieden)?

sehr unzufrieden  sehr zufrieden

Später Fortfahren
Weiter >>
Umfrage verlassen und löschen

0%  100%

### Ihre Bewertung von MensaAssist

\*  
Nachdem Sie MensaAssist bei der Essenswahl unterstützt hat, haben Sie nun die Möglichkeit das Assistenzsystem im Detail zu bewerten.

	Trifft überhaupt nicht zu 1	2	3	4	Trifft voll zu 5	<i>keine Angabe</i>
MensaAssist leitet mich fehl.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin misstrauisch, was die Empfehlungen von MensaAssist angeht.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich stehe MensaAssist insgesamt skeptisch gegenüber.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Insgesamt bin ich überzeugt von MensaAssist.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
MensaAssist gibt mir Sicherheit (bzgl. meiner Essenswahl).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
MensaAssist ist zuverlässig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich kann MensaAssist vertrauen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Funktionsweise von MensaAssist ist mir vertraut.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Später Fortfahren
Weiter >>
Umfrage verlassen und löschen

0%  100%

### Ihre Bewertung von MensaAssist

\*

Nachdem Sie MensaAssist bei der Essenswahl unterstützt hat, haben Sie nun die Möglichkeit das Assistenzsystem im Detail zu bewerten.

	Trifft überhaupt nicht zu 1	2	3	4	Trifft voll zu 5	<i>keine Angabe</i>
MensaAssist leitet mich fehl.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin misstrauisch, was die Empfehlungen von MensaAssist angeht.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich stehe MensaAssist insgesamt skeptisch gegenüber.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Insgesamt bin ich überzeugt von MensaAssist.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
MensaAssist gibt mir Sicherheit (bzgl. meiner Essenswahl).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
MensaAssist ist zuverlässig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich kann MensaAssist vertrauen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Funktionsweise von MensaAssist ist mir vertraut.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

0%  100%

### Angaben zu Ihrer Person

**Wie alt sind Sie?**

*In dieses Feld dürfen nur Ziffern eingetragen werden.*

**Welches Geschlecht haben Sie?**

weiblich  männlich  keine Antwort

**Welcher ist Ihr höchster Bildungsabschluss?**

Haupt-/Realschulabschluss  
 Abitur/Fachabitur  
 Berufsausbildung/Lehre  
 FH/Universität  
 Sonstiges:  
  
 keine Antwort

**Wie sieht Ihre derzeitige berufliche Beschäftigung aus?**

Professor/Professorin  
 Wissenschaftlicher Mitarbeiter/Mitarbeiterin  
 Student/Studentin  
 Sonstige Mitarbeiter (Verwaltung etc.)  
 Andere  
 keine Antwort

**Umfrage Mensa-Assist**

0%  100%

### Angaben zu Ihrer Person

**Wie alt sind Sie?**

*In dieses Feld dürfen nur Ziffern eingetragen werden.*

**Welches Geschlecht haben Sie?**

weiblich     männlich     keine Antwort

**Welcher ist Ihr höchster Bildungsabschluss?**

Haupt-/Realschulabschluss  
 Abitur/Fachabitur  
 Berufsausbildung/Lehre  
 FH/Universität  
 Sonstiges:

keine Antwort

**Wie sieht Ihre derzeitige berufliche Beschäftigung aus?**

Professor/Professorin  
 Wissenschaftlicher Mitarbeiter/Mitarbeiterin  
 Student/Studentin  
 Sonstige Mitarbeiter (Verwaltung etc.)  
 Andere  
 keine Antwort

**In welchem Studiengang studieren Sie?**

0%  100%

### Abschluss

Falls Sie Anmerkungen oder Verbesserungsvorschläge haben, können Sie uns diese hier mitteilen.

Bei weiteren Frage zu dieser Umfrage wenden Sie sich bitte an folgende Adresse: [robin.fink@tu-dortmund.de](mailto:robin.fink@tu-dortmund.de)

Vielen Dank für Ihre Teilnahme!

## A.5.2 Entscheidungsgrundlage MensaAssist / Übereinstimmungen

	satt	schnell	lecker	zahlen	gesund	Sparf	Gesund	Eiliger (*)
<b>Essen 1</b>	1	0.25	1	0	0.25	7.5	11.5	10
<b>Essen 2</b>	0	0.75	1	1	0.5	9.25	9.75	9.75
<b>Essen 3</b>	0.75	0.75	0.5	0.25	0	6.5	8.75	8.25

	satt	schnell	lecker	zahlen	gesund			
<b>Essen 1</b>	0.5	0.5	1	0	0.25	5.75	9.75	8.75
<b>Essen 2</b>	0	0.5	0.25	1	1	7.25	6.75	5.75
<b>Essen 3</b>	1	0.5	0.25	0	0.25	5.5	8.5	7

<b>Essen 1</b>	1	0.5	0.75	0	0	6.75	10.25	9.25
<b>Essen 2</b>	0.75	0.25	0	1	0.75	9	7.75	6.5
<b>Essen 3</b>	0.25	1	1	0	0.25	5.25	10	9.25

<b>Essen 1</b>	1	0.5	0.25	0	0	5.25	7.75	6.75
<b>Essen 2</b>	0	0.5	0.5	0.5	0.75	5.25	6.75	5.75
<b>Essen 3</b>	0.25	1	0	0.5	0.25	4.75	5.5	5.25

<b>Essen 1</b>	0.25	0	0.75	1	1	9.25	9	7.75
<b>Essen 2</b>	0.5	0.5	1	0	0.75	6.25	11.25	9.25
<b>Essen 3</b>	1	0.5	0.25	0	0.25	5.5	8.5	7

<b>Essen 1</b>	1	0	0.5	0.75	0	9.25	8.25	8
<b>Essen 2</b>	0.25	0	0.5	1	0.75	8.25	7	6.25
<b>Essen 3</b>	0.5	1	0.75	0	0	5.25	9.25	8.75

Operationalisierung strikt nach Formulierungen aus Fragebogen mit 1 und 5 als Extreme

	satt	schnell	lecker	zahlen	gesund
Sparfuchs	4	1	3	5	1
Gesundheitsbewusster	5	3	5	1	3
Eiliger	4	3	5	2	1

14  
17  
15

Mapping Punkte --> p - Werte

1	0
2	0.25
3	0.5
4	0.75
5	1

1	0
2	0.25
3	0.5
4	0.75
5	1

	satt	schnell	lecker	zahlen	gesund
Sparfuchs	4	2	3	5	1
Gesundheitsbewusster	5	2	5	1	4
Eiliger	4	3	5	3	1

MensaAssist ist auf Typ

Übereinstimmungen

2	3	4
---	---	---

satt      lecker      besonders wichtig

Codierung im Fragebogen

	true?	lecker	satt	gesund	schnell	zahlen
<b>Essen 1</b>		1	1	0.25	0.25	0
<b>Essen 2</b>	1	1	0	0.5	0.75	1
<b>Essen 3</b>		0.5	0.75	0	0.75	0.25

	lecker	satt	gesund	schnell	zahlen	
<u>Essen 1</u>	<u>1</u>	1	0.5	0.25	0.5	0
<u>Essen 2</u>		0.25	0	1	0.5	1
<u>Essen 3</u>		0.25	1	0.25	0.5	0
<u>Essen 1</u>		0.75	1	0	0.5	0
<u>Essen 2</u>		0	0.75	0.75	0.25	1
<u>Essen 3</u>	<u>1</u>	1	0.25	0.25	1	0
<u>Essen 1</u>		0.25	1	0	0.5	0
<u>Essen 2</u>	<u>1</u>	0.5	0	0.75	0.5	0.5
<u>Essen 3</u>		0	0.25	0.25	1	0.5
<u>Essen 1</u>		0.75	0.25	1	0	1
<u>Essen 2</u>	<u>1</u>	1	0.5	0.75	0.5	0
<u>Essen 3</u>		0.25	1	0.25	0.5	0
<u>Essen 1</u>		0.5	1	0	0	0.75
<u>Essen 2</u>		0.5	0.25	0.75	0	1
<u>Essen 3</u>	<u>1</u>	0.75	0.5	0	1	0

### A.5.3 Auswertung MensaAssist

#### A.5.3.1 Faktor Vertrauen MensaAssist

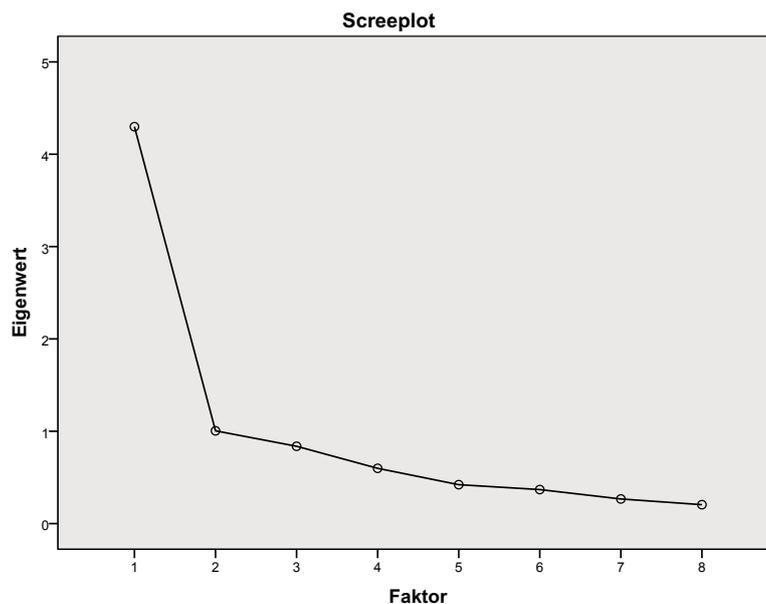
Kaiser-Kriterium zur optimalen Faktorenzahl liefert zunächst zwei Faktoren, da das Kriterium für die anfänglichen Eigenwerte verlangt, dass diese größer gleich eins sein müssen:

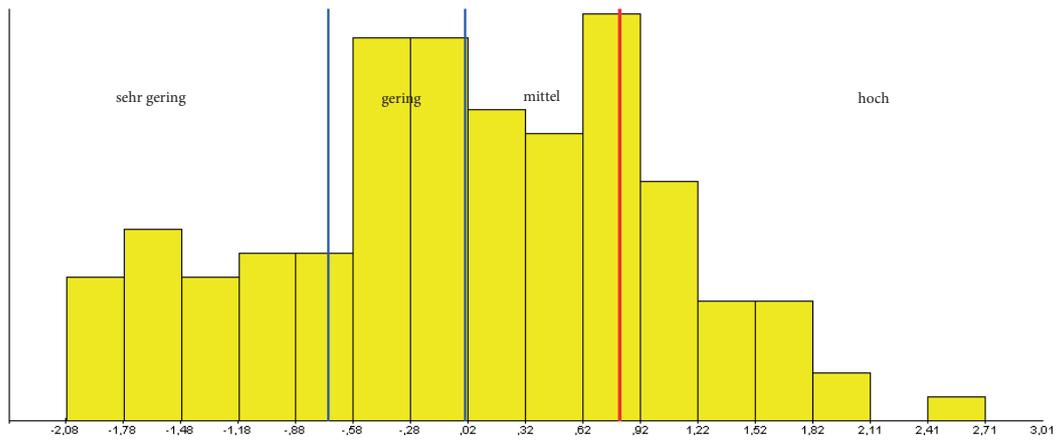
**Erklärte Gesamtvarianz / Kaiser-Kriterium zur optimalen Faktorenzahl**

Komponente	Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion			Rotierte Summe der quadrierten Ladungen		
	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	4,298	53,725	53,725	4,298	53,725	53,725	4,235	52,932	52,932
2	1,005	12,559	66,284	1,005	12,559	66,284	1,068	13,352	66,284
3	,837	10,469	76,753						
4	,599	7,484	84,236						
5	,421	5,267	89,504						
6	,368	4,606	94,110						
7	,267	3,332	97,442						
8	,205	2,558	100,000						

Extraktionsmethode: Hauptkomponentenanalyse.

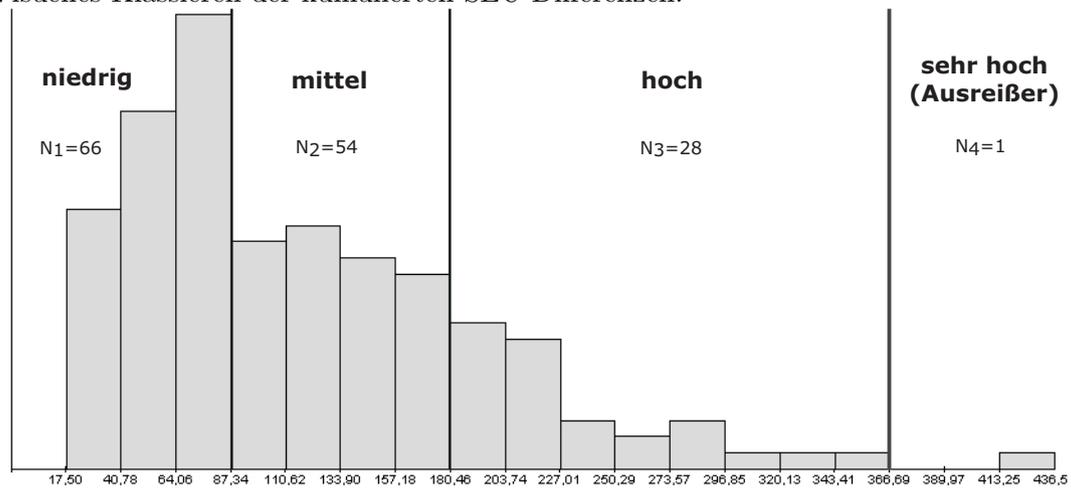
Bestätigung durch Screeplot, da Knick bei 2:





### A.5.3.2 Kumulierte SEU-Differenzen

Visuelles Klassieren der kumulierten SEU-Differenzen:



Klassierung der Faktorwerte „Vertrauen“ in vier gleich große Gruppen:

### A.5.3.3 Inhaltliche Interpretation von mangelndem Vertrauen

**Vertrauen: sehr gering** Clusterverfahren: Ward, Distanzmaß: Euklidischer Abstand  
Zuordnungsübersicht, die zeigt, dass der größte Koeffizientensprung zwischen Schritt 29 und Schritt 30 (s. Markierung) liegt und von daher eine Zwei-Clusterlösung anzustreben ist.

Zuordnungsübersicht						
Schritt	Zusammengeführte Cluster		Koeffizienten	Erstes Vorkommen des Clusters		Nächster Schritt
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	2	29	,184	0	0	7
2	7	22	,420	0	0	26
3	11	16	,925	0	0	16
4	18	23	1,473	0	0	12
5	14	20	2,056	0	0	15
6	12	30	2,645	0	0	21
7	2	3	3,242	1	0	19
8	15	25	3,875	0	0	17
9	10	21	4,528	0	0	14
10	8	24	5,220	0	0	15
11	6	26	5,918	0	0	20
12	18	31	6,617	4	0	16
13	17	27	7,323	0	0	19
14	10	28	8,213	9	0	21
15	8	14	9,194	10	5	22
16	11	18	10,201	3	12	25
17	1	15	11,363	0	8	22
18	4	19	12,578	0	0	25
19	2	17	13,823	7	13	24
20	6	9	15,113	11	0	24
21	10	12	16,596	14	6	23
22	1	8	18,163	17	15	26
23	10	13	19,940	21	0	27
24	2	6	22,387	19	20	28
25	4	11	24,863	18	16	29
26	1	7	27,502	22	2	28
27	5	10	30,933	0	23	29
28	1	2	34,496	26	24	30
29	4	5	38,227	25	27	30
30	1	4	44,562	28	29	0

**Vertrauen: hoch** Clusterverfahren: Ward, Distanzmaß: Euklidischer Abstand  
Größter Sprung auch hier beim letzten Fusionsschritt. Aber bereits vorher sind größere Sprünge zu verzeichnen. Deswegen ist die Anzahl der Cluster hier weniger eindeutig als bei der Fallauswahl nach sehr gering Vertrauenden.

Zuordnungsübersicht

Schritt	Zusammengeführte Cluster		Koeffizienten	Erstes Vorkommen des Clusters		Nächster Schritt
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	7	8	,288	0	0	15
2	18	32	,595	0	0	7
3	3	10	,983	0	0	10
4	21	22	1,417	0	0	24
5	1	16	1,925	0	0	20
6	9	24	2,496	0	0	24
7	12	18	3,081	0	2	12
8	14	31	3,721	0	0	13
9	4	23	4,370	0	0	18
10	3	17	5,064	3	0	17
11	13	33	5,806	0	0	25
12	6	12	6,553	0	7	20
13	14	15	7,304	8	0	21
14	5	19	8,130	0	0	19
15	7	26	8,961	1	0	25
16	28	30	9,816	0	0	21
17	3	20	10,699	10	0	31
18	4	29	11,612	9	0	22
19	5	27	12,848	14	0	28
20	1	6	14,149	5	12	26
21	14	28	15,459	13	16	29
22	4	11	16,943	18	0	27
23	2	25	18,536	0	0	27
24	9	21	20,143	6	4	26
25	7	13	21,892	15	11	30
26	1	9	24,078	20	24	28
27	2	4	26,630	23	22	29
28	1	5	29,519	26	19	30
29	2	14	32,999	27	21	31
30	1	7	37,230	28	25	32
31	2	3	42,435	29	17	32
32	1	2	48,377	30	31	0

## A.6 Simhybs

### A.6.1 Fragebogen



#### Bewertung von Fahrerassistenzsystemen

Wir sind Studenten, die ein Forschungsprojekt am Fachbereich Techniksoziologie der Technischen Universität Dortmund durchführen. Im Rahmen der folgenden Untersuchung erhalten Sie die Möglichkeit, mehrere Einstellungen eines Fahrerassistenzsystems zu testen und zu bewerten.

Ihre Teilnahme ist wesentlich für das Gelingen des Forschungsprojekts. Ihre persönlichen Erfahrungen und Meinungen werden im Laufe der Untersuchung anhand von kurzen Fragebögen erhoben.

Für die Beantwortung benötigen Sie nur wenige Minuten. Bitte beantworten Sie die Fragen vollständig, spontan und ehrlich. Die Antworten verlangen kein technisches Wissen, somit gibt es keine richtigen oder falschen Antworten.

Bitte beantworten Sie die Fragen, indem sie jeweils eines der Kästchen deutlich ankreuzen. Falls sie ein Kästchen versehentlich angekreuzt haben, schwärzen Sie bitte diese Markierung und markieren Sie das gewünschte Kästchen. Ihre Angaben werden anonym und streng vertraulich behandelt. Sie werden ausschließlich für wissenschaftliche Zwecke genutzt und nicht an Dritte weitergereicht.

Vielen Dank für Ihre Teilnahme an dieser Untersuchung!

Nachdem Sie einen kurzen Einführungsfragebogen ausgefüllt haben, folgt eine kurze Probephase im Modus **MANUELL**, damit Sie die Simulation kennenlernen. Danach werden Sie in 5 Durchläufen die verschiedenen Modi testen. Diese Modi heißen **MANUELL**, **GESCHWINDIGKEIT-AUTO**, **LENK-AUTO** und **VOLL-AUTO**.

- **MANUELL**: Das Assistenzsystem greift nicht ein, sondern warnt lediglich bei einer Geschwindigkeitsüberschreitung.
- **GESCHWINDIGKEIT-AUTO**: Das Assistenzsystem übernimmt die Erhöhung und Verringerung der Geschwindigkeit.
- **LENK-AUTO**: Das Assistenzsystem übernimmt die Aktionen Links und Rechts.
- **VOLL-AUTO**: Das Assistenzsystem übernimmt alle Aktionen.

Ihre Aufgaben bei der Simulation bestehen darin, eine möglichst weite **Strecke zurückzulegen**, die **Geschwindigkeitsbegrenzung nicht zu überschreiten** sowie **Unfälle** zu vermeiden.

### Informationen zur Simulation



Gesamtoberfläche: Am rechten unteren Rand ist der aktuelle Modus zu sehen



Geahndete Geschwindigkeitsüberschreitungen: Die Zahl gibt die Anzahl der Überschreitung je Durchlauf an



Unfälle: Die Zahl gibt die Anzahl der Unfälle je Durchlauf an

## Einführungsfragebogen

<b>1. Bitte beurteilen Sie die folgenden Aussagen</b> Bitte kreuzen Sie dazu Werte von „1“ = „trifft gar nicht zu“ bis „5“ = „trifft voll zu“ an.		☹	↔			☺
Folgende Aussage ...		1	2	3	4	5
1.1	Wenn ein neues technisches Gerät auf den Markt kommt, bin ich einer der ersten, der es kauft.	<input type="checkbox"/>				
1.2	Wenn ich ein neues technisches Gerät erworben habe, bin ich schnell mit allen Funktionen vertraut.	<input type="checkbox"/>				
1.3	Neue Technik enthält oft überflüssige Funktionen, die ich nicht benötige.	<input type="checkbox"/>				
1.4	Ich zögere, neue Technik zu nutzen, weil ich Angst habe, bei der Bedienung Fehler zu machen.	<input type="checkbox"/>				
1.5	Mich stört es, dass ständig neue technische Geräte auf den Markt kommen.	<input type="checkbox"/>				
1.6	In Sachen Technik verlasse ich mich lieber auf Altbewährtes.	<input type="checkbox"/>				
1.7	Neue Technik macht das Leben leichter.	<input type="checkbox"/>				
1.8	Eine Welt ohne Technik kann ich mir nicht vorstellen.	<input type="checkbox"/>				
1.9	Trotz negativer Erfahrungen würde ich meinen technischen Geräten weiterhin trauen.	<input type="checkbox"/>				
1.10	Ich überprüfe die Aktionen der Technik, die ich nutze.	<input type="checkbox"/>				



### Fragebogen zum Modus

2. Bitte bewerten Sie die folgenden Fragen in Bezug auf den gerade durchgeführten Modus <small>Bitte kreuze dazu Werte von „1“ = „trifft gar nicht zu“ bis „5“ = „trifft voll zu“ an.</small>		trifft gar nicht zu $\longleftrightarrow$ trifft voll zu				
Folgende Aussage ...		1	2	3	4	5
2.1	Das System leitet mich fehl.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.2	Ich bin misstrauisch, was die Ergebnisse des Systems angeht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.3	Ich stehe dem System insgesamt skeptisch gegenüber.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.4	Insgesamt bin ich überzeugt von dem Fahrassistenzmodus.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.5	Das System gibt mir Sicherheit (bzgl. meinen verfolgten Zielen).	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.6	Das System ist zuverlässig.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.7	Ich kann dem System vertrauen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.8	Die Funktionsweise des Systems ist mir vertraut.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.9	Das System verfolgt das Ziel der Geschwindigkeitseinhaltung.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.10	Das System verfolgt das Ziel der Unfallvermeidung.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.11	Das System verfolgt das Ziel einer möglichst langen zurückgelegten Strecke.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3. Bitte beurteilen Sie den soeben verwendeten Modus mit seinen spezifischen Aktionen nach den folgenden Kriterien. <small>(Bitte machen Sie nur ein Kreuz pro Zeile)</small>							
		1	2	3	4	5	
3.1	positiv	<input type="checkbox"/>	negativ				
3.2	interessant	<input type="checkbox"/>	uninteressant				
3.3	leicht bedienbar	<input type="checkbox"/>	schwer bedienbar				
3.4	vertrauenswürdig	<input type="checkbox"/>	nicht vertrauenswürdig				
3.5	hilfreich	<input type="checkbox"/>	nicht hilfreich				
3.6	hohe Qualität	<input type="checkbox"/>	niedrige Qualität				
3.7	nachvollziehbar	<input type="checkbox"/>	undurchschaubar				
3.8	zuverlässig	<input type="checkbox"/>	unzuverlässig				
3.9	sicher	<input type="checkbox"/>	unsicher				

Probandennr.: \_\_\_\_

Fragebogennr.: \_\_\_\_

Modusname: \_\_\_\_\_

## Fragebogen zum Modus im Wahlteil

4.1 Warum haben Sie diesen Modus gewählt?

4.2 Was war der Grund für den/die Moduswechsel/-beibehaltung?

4.3 Ist Ihnen noch etwas sehr wichtig, was hier nicht erwähnt wurde?



## Abschlussfragebogen

### 5. Angaben zu Ihrer Person

<b>Geschlecht:</b>	<input type="checkbox"/> weiblich <input type="checkbox"/> männlich	<b>Alter:</b>	_____ Jahre
<b>Abgeschlossene Ausbildung:</b>	<input type="checkbox"/> Berufsausbildung/Lehre	<input type="checkbox"/> Fachabitur/Abitur	
	<input type="checkbox"/> Haupt-/Realschulabschluss	<input type="checkbox"/> Fachhochschule/Universität	
	<input type="checkbox"/> andere: _____		
<b>Welchen Beruf führen Sie zurzeit aus?</b>		_____	

**Vielen Dank für Ihre Teilnahme!**

### A.6.2 Faktoranalyse Vertrauen

Die Variable `vertrauen_8` weist eine sehr geringe Faktorladung auf und kann nach Backhaus u. a. (2006: 299) nicht zur Interpretation des Faktors herangezogen werden. Daher wird diese aus dem Faktor entfernt.

**Komponentenmatrix<sup>a</sup>**

	Komponente
	1
<code>vertrauen_01</code>	-,741
<code>vertrauen_02</code>	-,800
<code>vertrauen_03</code>	-,808
<code>vertrauen_04</code>	,832
<code>vertrauen_05</code>	,857
<code>vertrauen_06</code>	,859
<code>vertrauen_07</code>	,894
<code>vertrauen_08</code>	,381

Extraktionsmethode:  
Hauptkomponentenanalyse.

a. 1 Komponenten extrahiert

## Faktoranalyse für Variablen vertrauen\_1...vertrauen\_7

Korrelationsmatrix<sup>a</sup>

		vertrauen_01	vertrauen_02	vertrauen_03	vertrauen_04
Korrelation	vertrauen_01	1,000	,625	,571	-,480
	vertrauen_02	,625	1,000	,745	-,556
	vertrauen_03	,571	,745	1,000	-,642
	vertrauen_04	-,480	-,556	-,642	1,000
	vertrauen_05	-,530	-,568	-,574	,757
	vertrauen_06	-,602	-,603	-,545	,664
	vertrauen_07	-,572	-,617	-,662	,735
Signifikanz (1-seitig)	vertrauen_01		,000	,000	,000
	vertrauen_02	,000		,000	,000
	vertrauen_03	,000	,000		,000
	vertrauen_04	,000	,000	,000	
	vertrauen_05	,000	,000	,000	,000
	vertrauen_06	,000	,000	,000	,000
	vertrauen_07	,000	,000	,000	,000

Korrelationsmatrix<sup>a</sup>

		vertrauen_05	vertrauen_06	vertrauen_07
Korrelation	vertrauen_01	-,530	-,602	-,572
	vertrauen_02	-,568	-,603	-,617
	vertrauen_03	-,574	-,545	-,662
	vertrauen_04	,757	,664	,735
	vertrauen_05	1,000	,741	,783
	vertrauen_06	,741	1,000	,796
	vertrauen_07	,783	,796	1,000
Signifikanz (1-seitig)	vertrauen_01	,000	,000	,000
	vertrauen_02	,000	,000	,000
	vertrauen_03	,000	,000	,000
	vertrauen_04	,000	,000	,000
	vertrauen_05		,000	,000
	vertrauen_06	,000		,000
	vertrauen_07	,000	,000	

a. Determinante = ,005

## KMO- und Bartlett-Test

Maß der Stichprobeneignung nach Kaiser-Meyer-Olkin.		,888
Bartlett-Test auf Sphärizität	Ungefähres Chi-Quadrat	2131,543
	df	21
	Signifikanz nach Bartlett	,000

**Komponentenmatrix<sup>a</sup>**

	Komponente
	1
vertrauen_01	-,744
vertrauen_02	-,806
vertrauen_03	-,812
vertrauen_04	,836
vertrauen_05	,858
vertrauen_06	,856
vertrauen_07	,895

Extraktionsmethode:  
Hauptkomponentenanalyse.

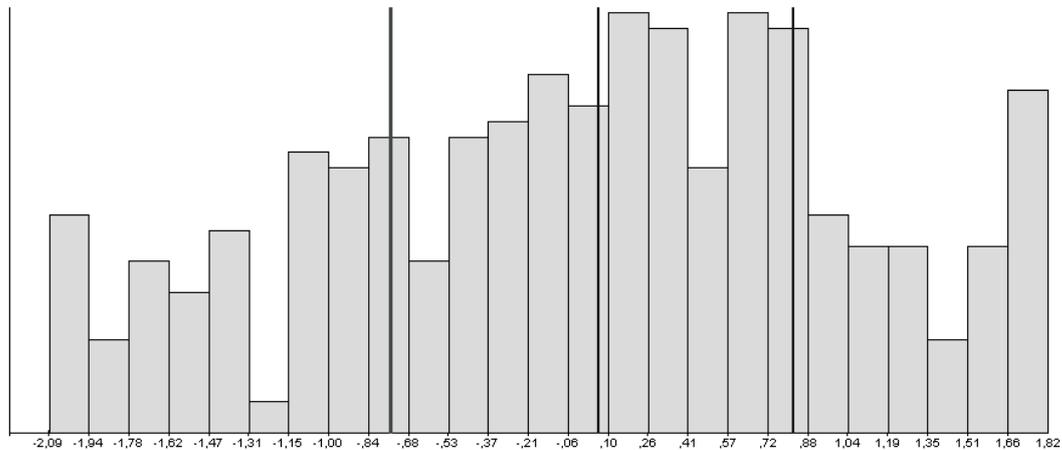
a. 1 Komponenten extrahiert

**Erklärte Gesamtvarianz**

Komponente	Summen von quadrierten Faktorladungen für Extraktion		
	Gesamt	% der Varianz	Kumulierte %
1	4,832	69,028	69,028

Extraktionsmethode: Hauptkomponentenanalyse.

Der gebildete Faktor FAC\_Vertrauen wird gemäß seiner Perzentile in vier gleich große Gruppen unterteilt:



### A.6.2.1 Korrelationen zwischen Vertrauensfaktorwerten je Proband

Vertrauensfaktorwerte sind in hohem Maße korreliert:

Korrelationen

		FAC_Vertraue n.2	FAC_Vertraue n.3	FAC_Vertraue n.4	FAC_Vertraue n.5	FAC_Vertraue n.6
FAC_Vertrauen.2	Korrelation nach Pearson	1	,340**	,360**	,480**	,295*
	Signifikanz (2-seitig)		,001	,000	,001	,020
	N	100	100	100	42	62
FAC_Vertrauen.3	Korrelation nach Pearson	,340**	1	,281**	,254	,167
	Signifikanz (2-seitig)	,001		,005	,104	,194
	N	100	100	100	42	62
FAC_Vertrauen.4	Korrelation nach Pearson	,360**	,281**	1	,577**	,250
	Signifikanz (2-seitig)	,000	,005		,000	,050
	N	100	100	100	42	62
FAC_Vertrauen.5	Korrelation nach Pearson	,480**	,254	,577**	1	,171
	Signifikanz (2-seitig)	,001	,104	,000		,404
	N	42	42	42	42	26
FAC_Vertrauen.6	Korrelation nach Pearson	,295*	,167	,250	,171	1
	Signifikanz (2-seitig)	,020	,194	,050	,404	
	N	62	62	62	26	62

\*\* Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

\* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

Zu den jeweiligen Vertrauensmessungen in den Durchgängen können entsprechende Regressionsrechnungen angestellt werden:

#### Aufgenommene/Entfernte Variablen<sup>a</sup>

Modell	Aufgenommene Variablen	Entfernte Variablen	Methode
1	FAC_Vertrauen.2 <sup>b</sup>	.	Einschluß

a. Abhängige Variable: FAC\_Vertrauen.3

b. Alle gewünschten Variablen wurden eingegeben.

#### Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,340 <sup>a</sup>	,116	,107	,99594107

a. Einflußvariablen : (Konstante), FAC\_Vertrauen.2

#### ANOVA<sup>a</sup>

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	12,724	1	12,724	12,828	,001 <sup>b</sup>
	Nicht standardisierte Residuen	97,206	98	,992		
	Gesamt	109,930	99			

a. Abhängige Variable: FAC\_Vertrauen.3

b. Einflußvariablen : (Konstante), FAC\_Vertrauen.2

#### Koeffizienten<sup>a</sup>

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.
		Regressionskoeffizient B	Standardfehler	Beta		
1	(Konstante)	-,677	,102		-6,645	,000
	FAC_Vertrauen.2	,404	,113	,340	3,582	,001

a. Abhängige Variable: FAC\_Vertrauen.3

**Aufgenommene/Entfernte Variablen<sup>a</sup>**

Modell	Aufgenommene Variablen	Entfernte Variablen	Methode
1	FAC_Vertrauen.3, FAC_Vertrauen.2 <sup>b</sup>	.	Einschluß

a. Abhängige Variable: FAC\_Vertrauen.4

b. Alle gewünschten Variablen wurden eingegeben.

**Modellzusammenfassung**

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,398 <sup>a</sup>	,158	,141	,83557674

a. Einflußvariablen : (Konstante), FAC\_Vertrauen.3, FAC\_Vertrauen.2

**ANOVA<sup>a</sup>**

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	12,727	2	6,364	9,115	,000 <sup>b</sup>
	Nicht standardisierte Residuen	67,724	97	,698		
	Gesamt	80,452	99			

a. Abhängige Variable: FAC\_Vertrauen.4

b. Einflußvariablen : (Konstante), FAC\_Vertrauen.3, FAC\_Vertrauen.2

**Koeffizienten<sup>a</sup>**

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.
		Regressionskoeffizient B	Standardfehler	Beta		
1	(Konstante)	,266	,103		2,587	,011
	FAC_Vertrauen.2	,304	,101	,299	3,022	,003
	FAC_Vertrauen.3	,153	,085	,179	1,808	,074

a. Abhängige Variable: FAC\_Vertrauen.4

**Aufgenommene/Entfernte Variablen<sup>a</sup>**

Modell	Aufgenommene Variablen	Entfernte Variablen	Methode
1	FAC_Vertrauen.4, FAC_Vertrauen.2, FAC_Vertrauen.3 <sup>b</sup>		Einschluß

- a. Abhängige Variable: FAC\_Vertrauen.5  
b. Alle gewünschten Variablen wurden eingegeben.

**Modellzusammenfassung**

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,642 <sup>a</sup>	,413	,366	,70232269

- a. Einflussvariablen : (Konstante), FAC\_Vertrauen.4, FAC\_Vertrauen.2, FAC\_Vertrauen.3

**ANOVA<sup>a</sup>**

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	13,162	3	4,387	8,895	,000 <sup>b</sup>
	Nicht standardisierte Residuen	18,744	38	,493		
	Gesamt	31,906	41			

- a. Abhängige Variable: FAC\_Vertrauen.5  
b. Einflussvariablen : (Konstante), FAC\_Vertrauen.4, FAC\_Vertrauen.2, FAC\_Vertrauen.3

**Koeffizienten<sup>a</sup>**

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.
		Regressionskoeffizient B	Standardfehler	Beta		
1	(Konstante)	,197	,144		1,366	,180
	FAC_Vertrauen.2	,304	,134	,306	2,260	,030
	FAC_Vertrauen.3	-,007	,122	-,008	-,055	,956
	FAC_Vertrauen.4	,487	,150	,464	3,246	,002

- a. Abhängige Variable: FAC\_Vertrauen.5

**Aufgenommene/Entfernte Variablen<sup>a</sup>**

Modell	Aufgenommene Variablen	Entfernte Variablen	Methode
1	FAC_Vertrauen.5, FAC_Vertrauen.3, FAC_Vertrauen.2, FAC_Vertrauen.4 <sup>b</sup>		Einschluß

a. Abhängige Variable: FAC\_Vertrauen.6

b. Alle gewünschten Variablen wurden eingegeben.

**Modellzusammenfassung**

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,326 <sup>a</sup>	,107	-,064	,84978405

a. Einflußvariablen : (Konstante), FAC\_Vertrauen.5, FAC\_Vertrauen.3, FAC\_Vertrauen.2, FAC\_Vertrauen.4

**ANOVA<sup>a</sup>**

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	1,809	4	,452	,626	,649 <sup>b</sup>
	Nicht standardisierte Residuen	15,165	21	,722		
	Gesamt	16,973	25			

a. Abhängige Variable: FAC\_Vertrauen.6

b. Einflußvariablen : (Konstante), FAC\_Vertrauen.5, FAC\_Vertrauen.3, FAC\_Vertrauen.2, FAC\_Vertrauen.4

**Koeffizienten<sup>a</sup>**

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.
		Regressionskoeffizient B	Standardfehler	Beta		
1	(Konstante)	,241	,235		1,022	,318
	FAC_Vertrauen.2	-,182	,270	-,195	-,673	,508
	FAC_Vertrauen.3	,303	,250	,331	1,214	,238
	FAC_Vertrauen.4	-,035	,307	-,034	-,115	,910
	FAC_Vertrauen.5	,170	,274	,181	,621	,541

a. Abhängige Variable: FAC\_Vertrauen.6

Schränkt man die Überprüfung der Korrelation auf gleiche Modi ein, sind die gefundenen Zusammenhänge noch stärker:

**Korrelationen | modus.5=Geschwindigkeit-Auto**

		FAC_Vertraue n.2	FAC_Vertraue n.5
FAC_Vertrauen.2	Korrelation nach Pearson	1	,876**
	Signifikanz (2-seitig)		,000
	N	21	21
FAC_Vertrauen.5	Korrelation nach Pearson	,876**	1
	Signifikanz (2-seitig)	,000	
	N	21	21

\*\* . Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

**Korrelationen | modus.6=Geschwindigkeit-Auto**

		FAC_Vertraue n.2	FAC_Vertraue n.6
FAC_Vertrauen.2	Korrelation nach Pearson	1	,616**
	Signifikanz (2-seitig)		,000
	N	28	28
FAC_Vertrauen.6	Korrelation nach Pearson	,616**	1
	Signifikanz (2-seitig)	,000	
	N	28	28

\*\* . Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

**Korrelationen | modus.6=Voll-Auto**

		FAC_Vertraue n.4	FAC_Vertraue n.6
FAC_Vertrauen.4	Korrelation nach Pearson	1	,847**
	Signifikanz (2-seitig)		,000
	N	21	21
FAC_Vertrauen.6	Korrelation nach Pearson	,847**	1
	Signifikanz (2-seitig)	,000	
	N	21	21

\*\* . Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

### A.6.3 Faktorenanalyse Semantisches Differenzial

Ggf. Screenshot der Fragebatterie aus Fragebogen.

Die Variablen zu Frage 3 aus dem Fragebogen je Modus (Variablen Frag3\_1...Frag3\_9) werden so umkodiert, dass geringe Werte eine negative Beurteilung bedeuten, hohe Werte eine positive Bedeutung. Die Skala umfasst die Werte 1, 2, 3, 4 und 5.

Faktoranalyse für Variablen Frag3\_1\_recoded...Frag3\_9\_recoded. Ein KMO von 0.918 zeigt, dass die Eingangsvariablen sehr gut für eine Faktorenanalyse geeignet sind. Die automatische Faktorextraktion extrahiert genau einen Faktor:

**KMO- und Bartlett-Test**

Maß der Stichprobeneignung nach Kaiser-Meyer-Olkin.		,918
Bartlett-Test auf Sphärizität	Ungefähres Chi-Quadrat	2496,012
	df	36
	Signifikanz nach Bartlett	,000

**Kommunalitäten**

	Anfänglich	Extraktion
Frag3_1_recoded	1,000	,767
Frag3_2_recoded	1,000	,373
Frag3_3_recoded	1,000	,282
Frag3_4_recoded	1,000	,768
Frag3_5_recoded	1,000	,722
Frag3_6_recoded	1,000	,694
Frag3_7_recoded	1,000	,308
Frag3_8_recoded	1,000	,750
Frag3_9_recoded	1,000	,759

Extraktionsmethode:  
Hauptkomponentenanalyse.

**Erklärte Gesamtvarianz**

Komponente	Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion		
	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	5,423	60,258	60,258	5,423	60,258	60,258
2	,978	10,868	71,126			
3	,771	8,570	79,696			
4	,571	6,343	86,039			
5	,381	4,237	90,276			
6	,282	3,137	93,413			
7	,215	2,390	95,803			
8	,198	2,204	98,007			
9	,179	1,993	100,000			

Extraktionsmethode: Hauptkomponentenanalyse.

**Komponentenmatrix<sup>a</sup>**

	Komponente
	1
Frag3_1_recoded	,876
Frag3_2_recoded	,610
Frag3_3_recoded	,531
Frag3_4_recoded	,876
Frag3_5_recoded	,849
Frag3_6_recoded	,833
Frag3_7_recoded	,555
Frag3_8_recoded	,866
Frag3_9_recoded	,871

Extraktionsmethode:  
Hauptkomponentenanalyse.

a. 1 Komponenten extrahiert

Koeffizientenmatrix der  
Komponentenwerte

	Komponente
	1
Frag3_1_recoded	,162
Frag3_2_recoded	,113
Frag3_3_recoded	,098
Frag3_4_recoded	,162
Frag3_5_recoded	,157
Frag3_6_recoded	,154
Frag3_7_recoded	,102
Frag3_8_recoded	,160
Frag3_9_recoded	,161

Extraktionsmethode:  
Hauptkomponentenanalyse.  
Rotationsmethode: Varimax mit  
Kaiser-Normalisierung.  
Komponentenwerte.

#### A.6.4 Mittelwertdifferenzen und Konfidenzintervalle

Für eine gegebene Variable ist man hinsichtlich eines Punktschätzers, z. B. einem Mittelwert, meist daran interessiert, ob sich die Mittelwerte eines Merkmals zwischen zwei Gruppen signifikant voneinander unterscheiden. Um dies zu entscheiden, kann auf zwei Verfahren zurückgegriffen werden: Konfidenzintervallberechnung mit Überschneidungsüberprüfung oder T-Test.

Konfidenzintervalle für den Mittelwert  $\bar{X}$  einer unbekanntem empirischen Verteilung mit Standardabweichung  $\sigma$  und  $N$  Fällen werden wie folgt berechnet (vgl. Bortz und Schuster 2010: 119)

$$CI_{low} = \bar{X} - t_{\alpha/2, N-1} \frac{\sigma}{\sqrt{N}} \quad CI_{high} = \bar{X} + t_{\alpha/2, N-1} \frac{\sigma}{\sqrt{N}}$$

wobei  $\frac{\sigma}{\sqrt{N}}$  der Standardfehler  $SE$  ist.

Das zugehörige Konfidenzintervall lautet dann  $(CI_{low}, CI_{high})$

Die Berechnung von Konfidenzintervallen ist vor allem zur besseren Interpretation von Mittelwerten erforderlich. Überschneiden sich Konfidenzintervalle nicht, kann mit einer Irrtumswahrscheinlichkeit von z. B. 5% davon ausgegangen werden, dass sich die Mittelwerte zweier Variablen tatsächlich voneinander unterscheiden.

Zu unterscheiden ist die Berechnung der Konfidenzintervalle für Mittelwerte von der Berechnung von Konfidenzintervallen für den T-Test, der die Signifikanz von Mittelwertdifferenzen untersucht. Es gilt: Sind die Konfidenzintervalle der Mittelwerte überschneidungsfrei, so enthält das Konfidenzintervall für die Mittelwertdifferenzen sicher nicht die 0. Sind die Konfidenzintervalle nicht überschneidungsfrei, kann es dennoch vorkommen, dass das Konfidenzintervall für die Mittelwertdifferenzen gemäß T-Test

die 0 *nicht* enthält. Für Details s. <http://www.cscu.cornell.edu/news/statnews/stnews73.pdf>.

### A.6.5 Clusteranalyse Vertrauensentwicklungstypen

Zunächst wurde eine Anomalieerkennung über die Faktoren FAC\_Vertrauen.2... FAC\_Vertrauen.4 durchgeführt. Anomalien stören in der Regel die Cluster-Analyse, da sie meist einzelne Cluster bilden, oder bei Verfahren, die gleichgroße Cluster anstreben, unproportional stark in die Cluster-Bildung eingehen.

#### Anomalie erkennen

Liste der Indizes  
anomaler Fälle

Fall	Anomalie-Index
78	2,993
29	2,641
24	2,254
27	2,075

Liste der Gruppen-IDs anomaler Fälle

Fall	Gruppen-ID	Gruppengröße	Gruppengröße in Prozent
78	4	29	29,0%
29	1	23	23,0%
24	5	23	23,0%
27	4	29	29,0%

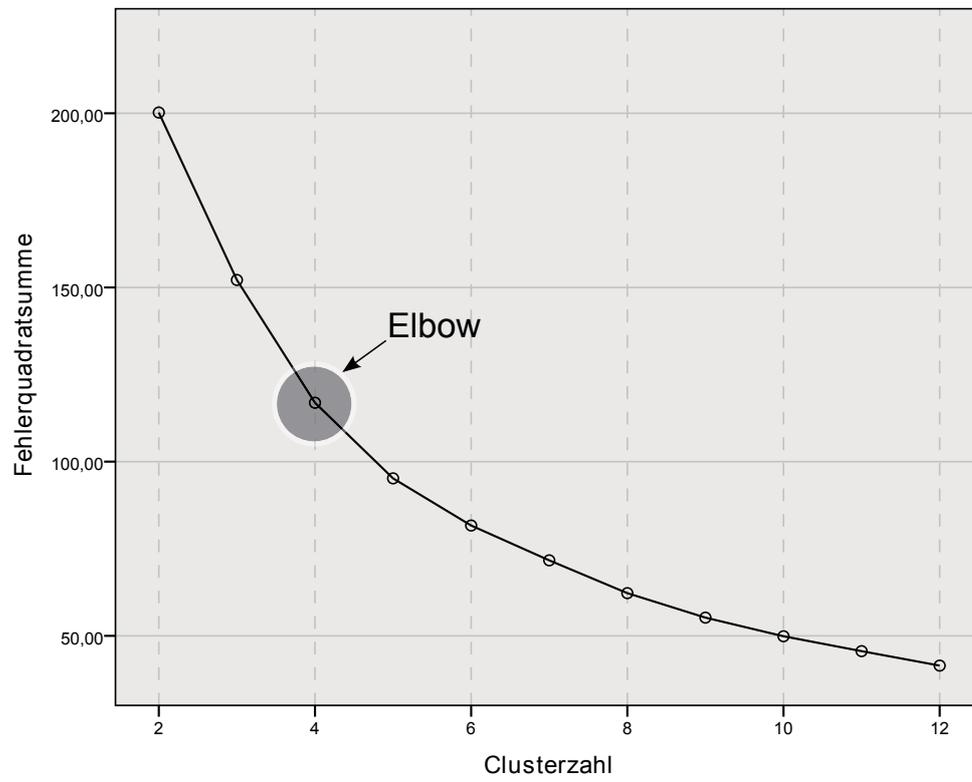
Liste der Gründe anomaler Fälle

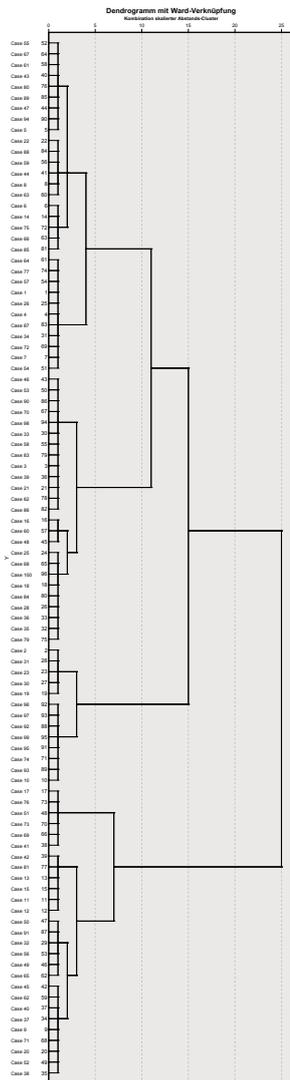
Grund: 1

Fall	Grundvariable	Variablenbeeinflussung	Variablenwert	Normwert der Variablen
78	FAC_Vertrauen.2	,499	1,66498	,3099610
29	FAC_Vertrauen.4	,470	1,66498	,6312326
24	FAC_Vertrauen.3	,605	,28816	-1,3394503
27	FAC_Vertrauen.2	,720	1,66498	,3099610

#### Die optimale Clusterzahl

Die optimale Zahl der Cluster lässt sich grafisch anhand des Elbow-Kriteriums abschätzen. Trägt man die Fehlerquadratsummen gegen die Clusterzahl auf, so gibt die Stelle des ersten Sprungs einen guten Anhaltspunkt für die optimale Anzahl der Cluster.





Basierend auf den gefilterten Daten und eine vorgegebenen fixen Clusterzahl wurde dann eine Analyse mit folgenden Einstellungen durchgeführt:

- Typ: Hierarchisch
- Methode: Ward
- Unterschiedlichkeitsmaß: Quadrierte Euklidische Distanz
- Standardisierung: Z-Scores
- Rotation: Varimax

**Vier-Cluster-Lösung für  
unterschiedliche Vertrauensentwicklungstypen**

CLU_Trust_Development		FAC_Vertrauen .2	FAC_Vertrauen .3	FAC_Vertrauen .4
1	Mittelwert	,1002	,1416	,1353
	N	31	31	31
	Standardabweichung	,4210	,5133	,6875
2	Mittelwert	1,3807	,7644	1,3484
	N	13	13	13
	Standardabweichung	,2461	,8555	,3279
3	Mittelwert	,6375	-1,3595	,3896
	N	25	25	25
	Standardabweichung	,3696	,3963	,7256
4	Mittelwert	-,7752	-1,4615	-,2950
	N	27	27	27
	Standardabweichung	,6244	,5416	,9395
Insgesamt	Mittelwert	,1673	-,6159	,2448
	N	96	96	96
	Standardabweichung	,8476	1,0451	,8924



### A.6.6 Korrelationen Vertrauensentwicklungstypen

		Korrelationen												
CLU_Trust_Development		FAC_Vertrauen.2	FAC_Vertrauen.3	FAC_Vertrauen.4	u_gw.2	u_gw.3	u_gw.4	u_cr.2	u_cr.3	u_cr.4	u_md.2	u_md.3	u_md.4	
1	FAC_Vertrauen.2	Pearson-Korrelation	1	-.061	-.057	.156	-.112	.054	.051	.028	.054	-.335	.042	.041
		Sig. (2-seitig)		.745	.761	.401	.549	.774	.785	.881	.772	.066	.822	.826
		N	31	31	31	31	31	31	31	31	31	31	31	31
	FAC_Vertrauen.3	Pearson-Korrelation	-.061	1	.240	-.357	.183	.173	-.245	.185	.196	-.245	-.197	-.129
		Sig. (2-seitig)	.745		.049	.025	.325	.351	.184	.320	.291	.185	.289	.489
		N	31	31	31	31	31	31	31	31	31	31	31	31
	FAC_Vertrauen.4	Pearson-Korrelation	-.057	.240	1	-.160	.358	.470	.019	.263	.564	-.176	.092	.044
		Sig. (2-seitig)	.761	.194		.391	.048	.008	.920	.153	.001	.345	.621	.812
		N	31	31	31	31	31	31	31	31	31	31	31	31
	u_gw.2	Pearson-Korrelation	.156	-.357	-.160	1	-.391	.177	.094	.249	.157	.074	-.365	-.298
		Sig. (2-seitig)	.401	.025	.391		.030	.342	.614	.177	.399	.692	.043	.104
		N	31	31	31	31	31	31	31	31	31	31	31	31
u_gw.3	Pearson-Korrelation	-.112	.183	.358	-.391	1	.098	.329	.107	.187	.166	.477	.338	
	Sig. (2-seitig)	.549	.325	.048	.030		.600	.071	.565	.314	.372	.007	.063	
	N	31	31	31	31	31	31	31	31	31	31	31	31	
u_gw.4	Pearson-Korrelation	.054	.173	.470	.177	.098	1	-.012	.482	.832	-.005	.235	.115	
	Sig. (2-seitig)	.774	.351	.008	.342	.600		.951	.006	.000	.979	.202	.537	
	N	31	31	31	31	31	31	31	31	31	31	31	31	
u_cr.2	Pearson-Korrelation	.051	-.245	.019	.094	.329	-.012	1	.139	.055	.447	.033	.243	
	Sig. (2-seitig)	.785	.184	.920	.614	.071	.951		.456	.770	.012	.859	.189	
	N	31	31	31	31	31	31	31	31	31	31	31	31	
u_cr.3	Pearson-Korrelation	.028	.185	.263	.249	.107	.482	.139	1	.433	-.089	.006	-.266	
	Sig. (2-seitig)	.881	.320	.153	.177	.565	.006	.456		.015	.632	.975	.147	
	N	31	31	31	31	31	31	31	31	31	31	31	31	
u_cr.4	Pearson-Korrelation	.054	.196	.564	.157	.187	.832	.055	.433	1	-.029	.178	.155	
	Sig. (2-seitig)	.772	.291	.001	.399	.314	.000	.770	.015		.875	.337	.404	
	N	31	31	31	31	31	31	31	31	31	31	31	31	
u_md.2	Pearson-Korrelation	.335	-.245	-.176	.074	.166	-.005	.447	-.089	-.029	1	.401	.446	
	Sig. (2-seitig)	.066	.185	.345	.692	.372	.979	.012	.632	.875		.025	.012	
	N	31	31	31	31	31	31	31	31	31	31	31	31	
u_md.3	Pearson-Korrelation	.822	.289	.092	-.365	.477	.235	.033	.006	.178	.401	1	.436	
	Sig. (2-seitig)	.822	.289	.621	.043	.007	.202	.859	.975	.337	.025		.014	
	N	31	31	31	31	31	31	31	31	31	31	31	31	
u_md.4	Pearson-Korrelation	.041	.129	.044	-.298	.338	.115	.243	-.266	.155	.446	.436	1	
	Sig. (2-seitig)	.826	.489	.812	.104	.063	.537	.189	.147	.404	.012	.014		
	N	31	31	31	31	31	31	31	31	31	31	31	31	
2	FAC_Vertrauen.2	Pearson-Korrelation	1	-.260	.637	.5	.523	.428	.177	-.253	.503	.388	.086	.479
		Sig. (2-seitig)		.390	.019	.5	.067	.145	.564	.405	.079	.190	.781	.097
		N	13	13	13	13	13	13	13	13	13	13	13	13
	FAC_Vertrauen.3	Pearson-Korrelation	-.260	1	.173	.5	-.656	.015	-.620	.730	.129	-.459	-.438	-.752
		Sig. (2-seitig)	.390		.571	.5	.015	.961	.024	.005	.674	.115	.135	.003
		N	13	13	13	13	13	13	13	13	13	13	13	13
	FAC_Vertrauen.4	Pearson-Korrelation	.637	.173	1	.5	-.054	.646	-.380	.219	.714	.524	-.365	.285
		Sig. (2-seitig)	.019	.571		.5	.862	.017	.201	.473	.006	.066	.220	.346
		N	13	13	13	13	13	13	13	13	13	13	13	13
	u_gw.2	Pearson-Korrelation	.5	.5	.5	1	.5	.5	.5	.5	.5	.5	.5	.5
		Sig. (2-seitig)												
		N	13	13	13	13	13	13	13	13	13	13	13	13
u_gw.3	Pearson-Korrelation	.523	-.656	-.054	.5	1	-.090	.852	-.618	-.037	.224	.635	.543	
	Sig. (2-seitig)	.067	.015	.862			.771	.000	.025	.905	.463	.020	.056	
	N	13	13	13	13	13	13	13	13	13	13	13	13	
u_gw.4	Pearson-Korrelation	.428	.015	.646	.5	-.090	1	-.490	.249	.927	.246	-.023	.233	
	Sig. (2-seitig)	.145	.961	.017		.771		.089	.411	.000	.418	.940	.444	
	N	13	13	13	13	13	13	13	13	13	13	13	13	
u_cr.2	Pearson-Korrelation	.177	-.620	-.380	.5	.852	-.490	1	-.702	-.494	.090	.699	.418	
	Sig. (2-seitig)	.564	.024	.201		.000	.089		.008	.086	.770	.008	.156	
	N	13	13	13	13	13	13	13	13	13	13	13	13	
u_cr.3	Pearson-Korrelation	-.253	.730	.219	.5	-.618	.249	-.702	1	.309	-.351	-.470	-.318	
	Sig. (2-seitig)	.405	.005	.473		.025	.411	.008		.304	.239	.105	.290	
	N	13	13	13	13	13	13	13	13	13	13	13	13	
u_cr.4	Pearson-Korrelation	.503	.129	.714	.5	-.037	.927	-.494	.309	1	.228	-.133	.120	
	Sig. (2-seitig)	.079	.674	.006		.905	.000	.086	.304		.453	.664	.695	
	N	13	13	13	13	13	13	13	13	13	13	13	13	
u_md.2	Pearson-Korrelation	.388	-.459	.524	.5	.224	.246	.090	-.351	.228	1	.068	.504	
	Sig. (2-seitig)	.190	.115	.066		.463	.418	.770	.239	.453		.824	.079	
	N	13	13	13	13	13	13	13	13	13	13	13	13	
u_md.3	Pearson-Korrelation	.086	-.438	-.365	.5	.635	-.023	.699	-.470	-.133	.068	1	.228	
	Sig. (2-seitig)	.781	.135	.220		.020	.940	.008	.105	.664	.824		.455	
	N	13	13	13	13	13	13	13	13	13	13	13	13	
u_md.4	Pearson-Korrelation	.479	-.752	.285	.5	.543	.233	.418	-.318	.120	.504	.228	1	
	Sig. (2-seitig)	.097	.003	.346		.055	.444	.156	.290	.695	.079	.455		
	N	13	13	13	13	13	13	13	13	13	13	13	13	
3	FAC_Vertrauen.2	Pearson-Korrelation	1	.099	-.353	.439	-.118	-.224	-.122	.012	-.226	-.174	-.200	.278
		Sig. (2-seitig)		.638	.083	.028	.573	.281	.562	.954	.277	.405	.339	.178
		N	25	25	25	25	25	25	25	25	25	25	25	25
	FAC_Vertrauen.3	Pearson-Korrelation	.099	1	-.101	-.075	.417	-.193	-.111	.221	.095	.224	-.202	.263
		Sig. (2-seitig)	.638		.632	.721	.038	.355	.598	.289	.652	.281	.333	.205
		N	25	25	25	25	25	25	25	25	25	25	25	25
	FAC_Vertrauen.4	Pearson-Korrelation	-.353	-.101	1	-.122	.057	.076	.098	.311	.323	.375	.151	.437
		Sig. (2-seitig)	.083	.632		.561	.788	.720	.643	.130	.116	.065	.471	.029
		N	25	25	25	25	25	25	25	25	25	25	25	25
	u_gw.2	Pearson-Korrelation	.439	-.075	-.122	1	-.676	-.122	-.337	.315	-.048	-.424	-.522	-.202
		Sig. (2-seitig)	.028	.721	.561		.000	.561	.099	.125	.819	.035	.007	.333
		N	25	25	25	25	25	25	25	25	25	25	25	25
u_gw.3	Pearson-Korrelation	-.118	.417	.057	-.676	1	-.297	.470	-.132	-.122	.428	.428	.195	
	Sig. (2-seitig)	.573	.038	.788	.000		.149	.018	.530	.562	.033	.033	.350	
	N	25	25	25	25	25	25	25	25	25	25	25	25	
u_gw.4	Pearson-Korrelation	-.224	-.193	.076	-.122	-.297	1	-.047	-.127	.676	.037	.091	.060	
	Sig. (2-seitig)	.281	.355	.720	.561	.149		.823	.546	.000	.860	.665	.774	
	N	25	25	25	25	25	25	25	25	25	25	25	25	
u_cr.2	Pearson-Korrelation	-.122	-.111	.098	-.337	.470	-.047	1	.044	.097	-.024	.352	.156	
	Sig. (2-seitig)	.562	.598	.643	.099	.018	.823		.834	.643	.910	.085	.457	
	N	25	25	25	25	25	25	25	25	25	25	25	25	
u_cr.3	Pearson-Korrelation	.012	.221	.311	.315	-.132	-.127	.044	1	.231	.097	-.058	.227	
	Sig. (2-seitig)	.954	.269	.130	.125	.530	.546	.834		.267	.646	.784	.276	

Korrelationen

CLU_Trust_Development		FAC_Vertrauen.2	FAC_Vertrauen.3	FAC_Vertrauen.4	u_gw.2	u_gw.3	u_gw.4	u_cr.2	u_cr.3	u_cr.4	u_rnd.2	u_rnd.3	u_rnd.4
u_rnd.3	Pearson-Korrelation	-.200	-.202	.151	-.522*	.428	.091	.352	-.058	.036	.832**	1	.495
	Sig. (2-seitig)	.339	.333	.471	.007	.033	.665	.085	.784	.865	.001		.012
	N	25	25	25	25	25	25	25	25	25	25	25	25
u_rnd.4	Pearson-Korrelation	-.278	.263	.437	-.202	-.195	.060	.156	.227	.269	.572**	.495*	1
	Sig. (2-seitig)	.178	.205	.029	.333	.350	.774	.457	.276	.194	.003	.012	
	N	25	25	25	25	25	25	25	25	25	25	25	25
4 FAC_Vertrauen.2	Pearson-Korrelation	1	-.055	-.234	.116	-.213	.003	-.104	-.187	-.258	.240	.129	.274
	Sig. (2-seitig)		.787	.240	.566	.286	.989	.606	.351	.195	.227	.521	.167
	N	27	27	27	27	27	27	27	27	27	27	27	27
FAC_Vertrauen.3	Pearson-Korrelation	-.055	1	.077	-.212	-.193	-.104	-.093	.131	.243	.196	.188	.327
	Sig. (2-seitig)	.787		.703	.288	.334	.607	.644	.516	.222	.327	.348	.095
	N	27	27	27	27	27	27	27	27	27	27	27	27
FAC_Vertrauen.4	Pearson-Korrelation	-.234	.077	1	-.210	-.149	.094	.042	-.446*	.096	-.313	-.067	.242
	Sig. (2-seitig)	.240	.703		.293	.459	.640	.836	.020	.632	.111	.741	.223
	N	27	27	27	27	27	27	27	27	27	27	27	27
u_gw.2	Pearson-Korrelation	.116	-.212	-.210	1	-.003	.382	.146	.231	.157	-.193	-.187	-.275
	Sig. (2-seitig)	.566	.288	.293		.988	.049	.468	.245	.435	.336	.350	.165
	N	27	27	27	27	27	27	27	27	27	27	27	27
u_gw.3	Pearson-Korrelation	-.213	-.193	-.149	-.003	1	.043	-.091	-.054	-.003	-.057	.088	-.032
	Sig. (2-seitig)	.286	.334	.459	.988		.832	.651	.790	.987	.777	.664	.872
	N	27	27	27	27	27	27	27	27	27	27	27	27
u_gw.4	Pearson-Korrelation	.003	-.104	.094	.382	.043	1	.104	.289	.555**	-.204	.044	.008
	Sig. (2-seitig)	.989	.607	.640	.049	.832		.606	.143	.003	.307	.826	.968
	N	27	27	27	27	27	27	27	27	27	27	27	27
u_cr.2	Pearson-Korrelation	-.104	-.093	.042	.146	-.091	.104	1	-.025	-.019	-.016	-.183	-.247
	Sig. (2-seitig)	.606	.644	.836	.468	.651	.606		.900	.926	.938	.362	.215
	N	27	27	27	27	27	27	27	27	27	27	27	27
u_cr.3	Pearson-Korrelation	-.187	.131	-.446*	.231	-.054	.289	-.025	1	.426	.236	.157	-.267
	Sig. (2-seitig)	.351	.516	.020	.245	.790	.143	.900		.027	.236	.435	.179
	N	27	27	27	27	27	27	27	27	27	27	27	27
u_cr.4	Pearson-Korrelation	-.258	.243	.096	.157	-.003	.555**	-.019	.426*	1	-.001	.122	-.084
	Sig. (2-seitig)	.195	.222	.632	.435	.987	.003	.926	.027		.544	.678	.678
	N	27	27	27	27	27	27	27	27	27	27	27	27
u_rnd.2	Pearson-Korrelation	.240	.196	-.313	-.193	-.057	-.204	-.016	.236	-.001	1	.363	.231
	Sig. (2-seitig)	.227	.327	.111	.336	.777	.307	.938	.236	.995		.063	.247
	N	27	27	27	27	27	27	27	27	27	27	27	27
u_rnd.3	Pearson-Korrelation	.129	.188	-.067	-.187	.088	.044	-.183	.157	.122	.363	1	.496**
	Sig. (2-seitig)	.521	.348	.741	.350	.664	.826	.362	.435	.544	.063		.009
	N	27	27	27	27	27	27	27	27	27	27	27	27
u_rnd.4	Pearson-Korrelation	.274	.327	.242	-.275	-.032	.008	-.247	-.267	-.084	.231	.496**	1
	Sig. (2-seitig)	.167	.095	.223	.165	.872	.968	.215	.179	.678	.247	.009	
	N	27	27	27	27	27	27	27	27	27	27	27	27

\*. Korrelation ist bei Niveau 0,05 signifikant (zweiseitig).

\*\* . Korrelation ist bei Niveau 0,01 signifikant (zweiseitig).

c. Kann nicht berechnet werden, da mindestens eine der Variablen konstant ist.

A.6.7 Performancedaten Vertrauensentwicklungstypen

	distanceTravelled <sub>2</sub>	distanceTravelled <sub>3</sub>	distanceTravelled <sub>4</sub>	crashes <sub>2</sub>	crashes <sub>3</sub>	crashes <sub>4</sub>	gwueberschreitungen <sub>2</sub>	gwueberschreitungen <sub>3</sub>	gwueberschreitungen <sub>4</sub>	lastaturschlaege <sub>2</sub>	lastaturschlaege <sub>3</sub>	lastaturschlaege <sub>4</sub>
1 $\bar{X}$	5,3855	4,9733	5,6184	,48	,61	,00	,03	,84	,00	72,29	68,10	14,58
$\sigma$	,0979	,5282	,0000	,769	,919	,000	,180	1,319	,000	143,376	232,361	31,280
2 $\bar{X}$	5,4702	4,6925	5,6184	,62	,54	,00	,31	,15	,00	18,23	31,38	5,69
$\sigma$	,1073	,3416	,0000	,650	1,127	,000	,630	,376	,000	17,532	13,320	8,693
3 $\bar{X}$	5,3802	4,9251	5,6184	,32	1,36	,00	,08	,88	,00	60,92	46,48	14,72
$\sigma$	,0782	,3842	,0000	,627	,860	,000	,277	1,333	,000	96,512	49,696	23,767
4 $\bar{X}$	5,4460	4,8906	5,5953	,59	1,59	,00	,22	,59	,00	43,89	51,41	12,70
$\sigma$	,2105	,2583	,1201	,888	1,279	,000	,424	,991	,000	51,643	48,632	21,069
$\Sigma$	5,4126	4,8994	5,6119	,49	1,07	,00	,14	,69	,00	54,02	59,26	12,89
	,1394	,4077	,0637	,754	1,126	,000	,373	1,145	,000	99,696	137,047	24,324

Cluster



# Eidesstattliche Versicherung

Hiermit erkläre ich an Eides statt, dass ich die vorgelegte Dissertation selbständig und ohne unzulässige fremde Hilfe angefertigt und verfasst habe, dass alle Hilfsmittel und sonstigen Hilfen angegeben und dass alle Stellen, die ich wörtlich oder dem Sinne nach aus anderen Veröffentlichungen entnommen habe, kenntlich gemacht worden sind.

Dortmund, den

Robin D. Fink