

# Vertrauen in Online-Bewertungen

Integrative Literaturübersicht, theoretisches Modell und  
empirische Exploration am Beispiel von Arztbewertungen

Inaugural-Dissertation  
zur Erlangung des akademischen Grades  
eines Doktors der Philosophie  
der Universität Mannheim

vorgelegt von Sevda Can Arslan

Mannheim, November 2018

Erstgutachter: Prof. Dr. Matthias Kohring, Universität Mannheim

Zweitgutachter: Prof. Dr. Eugène Loos, Universiteit van Amsterdam

Dekan: Prof. Dr. Philipp Gassert

## INHALTSVERZEICHNIS

Inhaltsverzeichnis .....	iii
Abbildungsverzeichnis .....	iv
Tabellenverzeichnis .....	v
<b>1 Einleitung .....</b>	<b>1</b>
<b>2 Integrative Literaturübersicht: Faktoren des Vertrauens .....</b>	<b>5</b>
2.1 Einführung .....	5
2.2 Definition von Online-Bewertung.....	6
2.3 Notwendigkeit einer Literaturübersicht .....	7
2.4 Literaturrecherche und Aufbereitung.....	9
2.5 Geschichte und gegenwärtige Forschungslandschaft.....	13
2.6 Verständnis von Vertrauen .....	20
2.7 Methodisches Vorgehen .....	29
2.8 Ergebnisse zu den Faktoren.....	34
2.9 Fazit.....	59
<b>3 Theoretisches Modell: Vertrauen in Online-Bewertungen .....</b>	<b>63</b>
3.1 Einführung .....	63
3.2 Definition von Vertrauen .....	63
3.3 Voraussetzungen für Vertrauen .....	65
3.4 Objekt und Dimensionen des Vertrauens .....	69
3.5 Gründe für Vertrauen .....	72
3.6 Resultat und Konsequenz des Vertrauens .....	74
3.7 Literaturübersicht im Licht des theoretischen Modells .....	75
3.8 Fazit.....	84
<b>4 Empirische Exploration: Vertrauen in Arztbewertungen .....</b>	<b>87</b>
4.1 Einführung .....	87
4.2 Relevanz des Fallbeispiels Arztbewertung .....	87
4.3 Methodisches Vorgehen: Think-Aloud, PZI und Qual. IA .....	92

---

4.4	Auswahl der Bewertungen und der Teilnehmer.....	97
4.5	Voraussetzungen für Vertrauen in Arztbewertungen .....	102
4.6	Dimensionen von Vertrauen in Arztbewertungen .....	106
4.7	Gründe für Vertrauen in Arztbewertungen .....	115
4.8	Weitere Faktoren und Zusammenhänge zwischen den Faktoren.....	129
4.9	Konsequenzen des Vertrauen in Arztbewertungen .....	133
4.10	Fazit.....	134
<b>5</b>	<b>Schluss .....</b>	<b>141</b>
<b>6</b>	<b>Literaturverzeichnis .....</b>	<b>145</b>
<b>7</b>	<b>Anhang.....</b>	<b>159</b>
7.1	Einführung .....	159
7.2	Aufgabenstellung für Think-Aloud.....	159
7.3	Arztbewertungen.....	161
7.4	Leitfaden für Problemzentriertes Interview .....	171
7.5	Codebuch für Qualitative Inhaltsanalyse .....	172
7.6	Transkripte.....	177
<b>8</b>	<b>Eidesstattliche Erklärung.....</b>	<b>178</b>

## ABBILDUNGSVERZEICHNIS

Abbildung 1:	Forschungsgeschichte .....	15
Abbildung 2:	Artikel pro Jahr.....	17
Abbildung 3:	Artikel pro Fachgebiet .....	18
Abbildung 4:	Studien pro Methode .....	30
Abbildung 5:	Online-Bewertungs-Prozess .....	40
Abbildung 6:	Vertrauensmodell.....	65

---

## TABELLENVERZEICHNIS

Tabelle 1: Suche nach Studien zu Vertrauen in Online-Bewertungen.....	10
Tabelle 2: Zitierhäufigkeit nach Studie .....	19
Tabelle 3: Skalen und Konzepte von Vertrauen .....	22
<i>Tabelle 4: Items für Vertrauen .....</i>	<i>26</i>
Tabelle 5: Theoretische Grundlagen der Studien (sortiert nach Häufigkeit) .....	35
Tabelle 6: Anbieter-bezogene Faktoren .....	41
Tabelle 7: Bewertetes-bezogene Faktoren .....	42
Tabelle 8: Bewerter-bezogene Faktoren .....	43
Tabelle 9: Bewertungs-bezogene Faktoren .....	46
Tabelle 10: Community-bezogene Faktoren .....	50
Tabelle 11: Plattform-bezogene Faktoren .....	52
Tabelle 12: Leser-bezogene Faktoren .....	53
Tabelle 13: Externer Akteur-bezogene Faktoren .....	57
Tabelle 14: Gründe für Vertrauen .....	73
Tabelle 15: Items für Vertrauen.....	78
Tabelle 16: Überblick Arztbewertungen .....	98



# 1 EINLEITUNG

Ein Autor der ZEIT Online schimpfte vor Kurzem über „blinde[s] Vertrauen in die Bewertungsökonomie des Internets“ (Kühl, 2018). Er beschreibt eine „im Internet weit verbreitete [...] Benoteritis“ (Kühl, 2018). Damit meint er „dass inzwischen alles und jeder bewertet wird“. Kühl (2018) befürchtet, dass „im schlimmsten Fall die Menschen den Noten blind vertrauen.“ Anlass für seinen Kommentar bot der Abschluss eines längeren Gerichtsverfahrens, in dem eine Ärztin gegen ein Arztbewertungsportal geklagt hatte.

Tatsächlich lässt sich im Internet eine wachsende Zahl von Rankings und Ratings beobachten. Auf Online-Shopping-Seiten können Händler, Käufer und Produkte bewertet werden, auf allgemeinen oder spezifischen Bewertungsportalen werden Restaurants, Hotels, Lehrer und Ärzte bewertet. Selbst die Bewerter werden mittlerweile bewertet. Die Relevanz von Bewertungen wird immer größer (Friedl & Wietlisbach, 2008, S. 29). Borchers (2016) erklärt die wachsende Nutzung von Online-Bewertungen mit ihrer gesellschaftlichen Funktion. Er identifiziert Optionalisierung und Individualisierung als zwei stark miteinander verwobene gesellschaftliche Entwicklungen, die das Bedürfnis nach Orientierung erzeugen. Laut Borchers (2016) erfüllen Online-Bewertungen genau dieses Bedürfnis. Unter „Optionalisierung“ versteht Borchers (2016), dass die Handlungsmöglichkeiten der Menschen zunehmen. Mit „Individualisierung“ meint er, dass die Wahl der eigenen Handlung aus dem Meer an Möglichkeiten nun am einzelnen Individuum hängt. Beide Trends führen dazu, dass der Mensch heute mehr Entscheidungen treffen muss als früher. Borchers (2016) argumentiert, dass der Mensch dazu mehr Orientierung braucht und daher vermehrt auf die Erfahrungen anderer zurückgreift. Online-Bewertungen erfüllen genau dieses Bedürfnis, da sie jedem Menschen überall eine nahezu unendliche Anzahl von Erfahrungen anderer zur Verfügung stellen.

Bewertungen werden aber nur dann gesellschaftlich wichtig, wenn sie Folgen haben. Die Folgenhaftigkeit ist wiederum davon abhängig, ob der Leser dem Bewerter vertraut. Das Vertrauen in Online-Bewertungen ist also besonders wichtig, um die Folgen von Bewertungen zu erklären.

Einem Bewerter und seinen Bewertungen zu vertrauen, ist ein hochriskantes Unterfangen. Zum einen hat der Leser einer Bewertung keine Informationen zu den Motivationen und Kompetenzen des Bewerter. Zweitens profitiert der Bereitsteller des Bewerteten von guten Bewertungen und ist daher sehr daran interessiert, die Bewertungen in seinem Sinne zu beeinflussen. So begründet der eingangs zitierte Journalist Kühl (2018) seine kritische Haltung gegenüber Online-Bewertungen damit, dass sie „gerne gefälscht oder eingekauft werden“. Außerdem schreibt er:

„Reine Bewertungen in Form von Noten oder Sternchen entstehen aus zu vielen Faktoren und sind häufig irreführend. Sie sind im besten Fall eine grobe Orientierung und im schlimmsten Fall absolut nichtssagend.“  
(Kühl, 2018)

In der darunter geschalteten Leserdiskussion zum Thema „Vertrauen Sie Bewertungsportalen?“ (Leserdiskussion, 2018) sehen das viele Leser von ZEIT Online ähnlich. Ein Nutzer vertraut eher negativen Bewertungen, weil er die positiven Bewertungen für gefälscht hält (Leserdiskussion, 2018). Ein anderer meint dazu: „Das Lob ist ebenso frei erfunden wie der Tadel oder Verriß. Nichts kann als ehrlich betrachtet werden, denn Ehrlichkeit ist unidentifizierbar.“. Ein dritter Online-Diskutant widerspricht dem Journalisten und den anderen Nutzern:

„Ich denke aber, hier wird der gesunde Menschenverstand und das Urteilsvermögen der schweigenden Mehrheit an Durchschnittsusern unterschätzt. Die meisten unter uns dürften nach vielen Jahren im Internet sehr wohl in der Lage sein, Bewertungen auf Onlineportalen differenziert und mit Distanz zu betrachten. [...] Für mich jedenfalls hat diese Art von Fortschritt unterm Strich einen Mehrwert, wenngleich man natürlich kritisch distanziert bleiben und auch auf andere Infoquellen zurückgreifen sollte - sofern man überhaupt welche hat.“

Schon dieser kurze Einblick in eine aktuelle Debatte zeigt die Heterogenität an Positionen zum Thema Vertrauen in Online-Bewertungen. Angesichts der gesellschaftlichen Relevanz, die dem Phänomen der Bewertungsportale von den Massenmedien zugeschrieben wird, ist es höchste Zeit, sich dem Phänomen aus wissenschaftlicher Perspektive zu nähern. Ausgehend von der Beobachtung der zunehmenden Zahl an Online-Bewertungen sowie vor dem gesellschaftlichen Hintergrund, mit dem das Aufkommen dieses Phänomens erklärt werden kann, stellt sich die Frage danach, wie die Nutzung von Bewertungen überhaupt funktionieren

---

*kann*: Wie kann der Leser entscheiden, ob eine Bewertung ihm hilft, sich zu orientieren?

Wie bei vielen Informationen aus dem World Wide Web bleibt die Quelle auch bei Bewertungen häufig persönlich unbekannt. Was, wenn es sich beim Bewerter um den Bewerteten selbst handelt, der so für sich Werbung machen möchte? Oder um einen Mitbewerber, der seinen Konkurrenten schlecht machen möchte? Von der gezielten Manipulation einer Online-Bewertung abgesehen: Wie kann der Leser sich für eine Handlungsoption entscheiden, wenn er zwei einander widersprechende Bewertungen vorfindet?

Ich möchte dem Rätsel, wie der Leser die Entscheidung darüber trifft, welche Bewertung er als Orientierung heranzieht, aus kommunikations- und sprachwissenschaftlicher Perspektive auf den Grund gehen: Wie gestaltet sich Vertrauen in Online-Bewertungen?

Um diese Frage zu beantworten, führe ich zunächst eine umfangreiche Literaturstudie zu den bisherigen empirischen Befunden zu Vertrauen in Online-Bewertungen durch. Mithilfe einer disziplinäre und methodologische Grenzen überschreitenden integrativen Literaturübersicht identifiziere ich zunächst unzählige Faktoren, die mit dem Vertrauen in Online-Bewertungen zusammenhängen. Dabei wird deutlich, dass diese Faktoren theoretisch wenig integriert sind und sich Befunde häufig widersprechen. Auf Basis einer Theorie zu Vertrauen in Journalismus (Kohring, 2004) entwickle ich daher anschließend ein theoretisches Modell zu Vertrauen in Online-Bewertungen. Mit diesem Modell können die theoretischen Lücken gefüllt und manche Befunde der bisherigen Forschung erklärt werden. Um die weiter bestehenden empirischen Unklarheiten aufzulösen und die umfassende Belastbarkeit des theoretischen Modells zu demonstrieren, führe ich abschließend eine empirische Studie am Beispiel von Vertrauen in Arztbewertungen durch. Diese drei Schritte vom aktuellen Forschungsstand über einen Vorschlag für ein theoretisches Modell bis zur empirischen Exploration sollen mehr Aufschluss darüber geben, welche Faktoren mit Vertrauen in Online-Bewertungen zusammenhängen.



## 2 INTEGRATIVE LITERATURÜBERSICHT: FAKTOREN DES VERTRAUENS

### 2.1 EINFÜHRUNG

Zu Beginn der Arbeit setze ich mich mit der bestehenden Forschung zu Vertrauen in Online-Bewertungen auseinander. Zunächst bestimme ich Online-Bewertungen als zentralen Gegenstand dieser Arbeit genauer (ab S. 6). Dann begründe ich die Notwendigkeit einer Literaturübersicht (S. 7). Anschließend folgt ein ausführlicher Überblick über den aktuellen Forschungsstand. Das Vorgehen bei der Literaturrecherche, d. h. den Suchterm, die verwendeten Datenbanken und die Aufgreifkriterien sowie die der Literaturübersicht zugrunde liegende Methode erläutere ich im nächsten Unterkapitel (ab S. 6). Es folgt ein kurzer Abriss über die allmähliche Ausbildung des Forschungsfeldes „Vertrauen in Online-Bewertungen“ als eigenes Thema innerhalb der allgemeineren Forschung zu Online-Bewertungen. Dieses eigene Forschungsfeld wird daraufhin anhand der publizierenden Autoren und Journals der recherchierten Studien breiter aufgefächert (ab S. 13).

Dann beginnt die strukturierte Darstellung und kritische Betrachtung der zentralen Studieninhalte. Zunächst stelle ich die in den Studien verwendeten Konzepte des im Mittelpunkt meiner Arbeit stehenden Begriffs „Vertrauen“ (bzw. „Glaubwürdigkeit“) vor. Ich erläutere die Verständnisse von Vertrauen und ordne die Studien nach ihrer empirischen Messung dieses Konzepts (ab S. 20). Im nächsten Unterkapitel beschreibe ich das methodische Vorgehen in der bisherigen Forschung anhand der Art der Online-Bewertung, anhand der Teilnehmer der Studien sowie anhand der Methoden der Datenerhebung (ab S. 29).

Der Fokus dieses Kapitels liegt auf der dann folgenden ausführlichen Darstellung der bisherigen Ergebnisse der Studien (ab S. 34). Zunächst gebe ich die theoretischen Ableitungen und Erklärungen der Faktoren für Vertrauen in Online-Bewertungen wieder. Anschließend erläutere ich meinen Ansatz zur Systematisierung aller bisher untersuchten Faktoren. Die Gruppen der zusammengefassten Faktoren werden in sortierten Tabellen beschrieben. Daneben werden die Ergebnisse der jeweiligen Studien übersichtlich aufgeführt. Die so geordnete Darstellung der Ergebnisse ergänze ich mit den daraus folgenden

Schlüssen und Fragen. Am Ende fasse ich die zentralen Erkenntnisse noch einmal zusammen.

Das zweite Kapitel schließt mit einem Fazit des gesamten Kapitels (ab S. 59). Darin reflektiere ich kurz meine Vorgehensweise der integrativen Literaturübersicht. Aus der Zusammenfassung und Diskussion der bisherigen Forschung folgere ich dann die noch offenen Fragen, die meine weitere Arbeit anleiten.

## 2.2 DEFINITION VON ONLINE-BEWERTUNG

Nun zunächst zum Gegenstand der Online-Bewertung. Online-Bewertungen werden normalerweise auf Bewertungsplattformen gepostet. Diese können in größere Verkaufsplattformen eingebunden sein (z. B. *Amazon* oder *Bookings*), in Fan Communitys (z. B. *Metalstorm*) oder auf Sozialen Medien (wie *Facebook* und *Google+*), aber auch unabhängig gegründet werden (z. B. *Yelp* oder *HealthGrades*) (Borchers, 2016). Die meisten Studien zu Online-Bewertungen beziehen sich bei der Definition ihres Gegenstands auf „electronic Word of Mouth“ („eWOM“) als „any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet“ (Hennig-Thurau et al., 2004, S. 39). Genau genommen deckt diese sehr breite Definition allerdings noch viel mehr Formen computervermittelter Kommunikation von Konsumenten ab. Eine spezifischere Definition von Online-Bewertungen hat sich bisher aber noch nicht etabliert.

In meiner Arbeit zu Vertrauen in Online-Bewertungen grenze ich den Gegenstand explizit näher ein. Ich konzentriere mich auf Online-Bewertungen, die auf Bewertungsplattformen publiziert werden. Damit werden, anders als bei der obigen Definition von Hennig-Thurau et al. (2004), Online-Bewertungen z. B. auf individuellen Blogs ausgeschlossen. Dieser Ausschluss liegt darin begründet, dass die Bedingungen, unter denen Vertrauen auf Bewertungsplattformen stattfindet, bedeutend von Blogs abweichen. Das liegt an den unterschiedlichen Organisationsformen von Blogs und Bewertungsplattformen. Während Blogs meist um die Person des Bloggers, d. h. des Bewerter, organisiert sind, folgen Bewertungsplattformen in ihrem Aufbau dem Bewerteten. Das führt z. B. dazu, dass es auf Blogs zu einem Bewerteten meist nur eine Bewertung, nämlich die des Bloggers, gibt. Auf Bewertungsplattformen steht dem Leser dagegen eine Vielzahl

von Bewertungen verschiedener Bewerter zum selben Bewerteten zur Verfügung. Des Weiteren besteht auf Bewertungsplattformen die Möglichkeit des Lesers, sich in eine Community einzubringen, während beim Blog v.a. der Blogger im Fokus steht und die Leser meist passiv bleiben. Auf Bewertungsplattformen werden die Bewerter zudem hinsichtlich ihrer Bewertung angeleitet, sie erhalten z. B. Vorgaben zu den Bewertungskategorien, der maximalen Länge und der Benotungsskala. Auf Blogs sind Bewertungen dagegen nicht beschränkt.

Ich fasse den hier behandelten Gegenstand mit der Eingrenzung auf Online-Bewertungen von Bewertungsplattformen daher gezielt spezifischer. Damit möchte ich die Vergleichbarkeit der im Folgenden vorgestellten Forschung hinsichtlich der Faktoren des Vertrauens gewährleisten. Die Festlegung auf Online-Bewertungen von Bewertungsplattformen bildet die Grundlage für die gesamte weitere Arbeit. Ab jetzt sind bei meiner Verwendung des Begriffs „Online-Bewertung“ ausschließlich solche von Bewertungsplattformen gemeint. Nach dieser Definition des konkreten Gegenstandes dieser Arbeit argumentiere ich im nächsten Abschnitt dafür, eine ausführliche Literaturübersicht zum Thema dieser Studie zu erarbeiten.

### **2.3 NOTWENDIGKEIT EINER LITERATURÜBERSICHT**

Die Recherche zum Vertrauen in Online-Bewertungen (auf Bewertungsplattformen, s.o.) fördert zahlreiche Studien aus verschiedenen Disziplinen zutage. Vor allem in den letzten Jahren wurde einige Forschung zu den Faktoren, die mit der wahrgenommenen Glaub- bzw. Vertrauenswürdigkeit von Online-Bewertungen zusammenhängen, publiziert. Das macht die steigende wissenschaftliche Relevanz der Vertrauensfrage auch für den Gegenstand der Online-Bewertungen deutlich.

Der große Output birgt aber eine große Herausforderung. Die Autoren der verschiedenen Studien beziehen sich kaum aufeinander, d.h. sie benutzen verschiedene Konzepte von Vertrauen, untersuchen unterschiedliche Faktoren und binden ihre Ergebnisse nicht an die bisherige Forschung rück. So ist es sehr schwierig, zu erkennen, welche Fragen zum Vertrauen in Online-Bewertungen bereits beantwortet wurden und welche noch offen sind. Die Diversität der vielen beteiligten Disziplinen, die Fragmentierung und das Fehlen eines gemeinsamen konzeptionellen Rahmens haben eine verwirrende Situation in diesem Forschungsfeld hinterlassen.

Es gibt bisher nur eine Übersicht zum Forschungsstand zu Online-Bewertungen von Cheung und Thadani (2012). Sie fassen 22 Studien zusammen, die zwischen 2000 und 2010 erschienen sind. Allerdings geht es nur in wenigen der zitierten Studien um Vertrauen bzw. Glaubwürdigkeit im Zusammenhang mit Online-Bewertungen. Deswegen und auch wegen der großen Anzahl der nach 2010 publizierten Studien ist diese Literaturübersicht bei weitem nicht erschöpfend. Ein jüngeres Review von Choi und Stvilia (2015) beschäftigt sich mit „Web Credibility“. Die Autoren handeln das spezifische Thema des Vertrauens in Online-Bewertungen allerdings mit lediglich einer Studie von Flanagin und Metzger (2008) ab. Es fehlt demnach ein aktueller Überblick, der sich eigens auf die Forschung zu Vertrauen in Online-Bewertungen konzentriert.

Das Ziel dieses Kapitels ist es daher, einen solchen Überblick zur bisherigen Forschung zu liefern. Dazu sollen die einschlägigen Studien recherchiert, sortiert und integriert werden. Nur eine systematische Synthese der bestehenden Ergebnisse ermöglicht es mir, die bisherigen Erkenntnisse zusammenzufassen und mögliche Leerstellen bisheriger Forschung zu identifizieren. Für eine gründliche und sorgfältige Auseinandersetzung mit der bisherigen Forschung bietet sich die Formulierung konkreter Fragen an, die die Literaturübersicht strukturieren. Folgende Fragen sollen dazu dienen, das Forschungsfeld zu Vertrauen in Online-Bewertungen systematisch zu erschließen:

- Zum Forschungsfeld: Wann wurden Studien zum Vertrauen in Online-Bewertungen veröffentlicht? Welche Autoren forschen dazu? Welche Fachzeitschriften publizieren diese Forschung?
- Zur Theorie: Was verstehen die Autoren bisheriger Studien unter „Vertrauen“? Wie wird „Vertrauen“ gemessen?
- Zur Methode: Welche Methoden benutzen die Autoren? Welche Menschen nehmen an den Studien teil? Um welche Gegenstände geht es bei den untersuchten Online-Bewertungen?
- Zu den Ergebnissen: Welche Faktoren, die mit Vertrauen in Online-Bewertungen zusammenhängen, wurden bisher untersucht? Wie werden die Faktoren erklärt? Mit welchen Ergebnissen werden die Faktoren untersucht?

---

Nach dieser kurzen Einführung in die bisher konfuse Lage der Forschung und dem daraus hervorgehenden Plädoyer für eine klärende Literaturübersicht erfolgt im nächsten Unterkapitel die Vorstellung der Recherchestrategie und der hier angewandten Methode mit dem Ziel eines solchen systematisierenden Forschungsüberblicks.

## 2.4 LITERATURRECHERCHE UND AUFBEREITUNG

Eine angemessene und verständliche Strategie zur Literaturrecherche bildet die Basis jedes starken Forschungsstands. Mit dieser Strategie sollen alle relevanten Artikel identifiziert werden. Nur wenn die Literaturübersicht auf einer guten Datenbasis fußt, kann sie genaue Ergebnisse liefern (Whittemore, 2005). Im Folgenden beschreibe und begründe ich meine Literaturrecherche anhand des Suchterms, der Datenbanken sowie der formalen und inhaltlichen Aufgreifkriterien.

Für die Entwicklung eines geeigneten Suchterms ist es wichtig, daran zu erinnern, dass ich für diese Arbeit zwar den engen Begriff von Online-Bewertungen verwende, ein großer Teil der Forschung zum selben Phänomen aber die Begriffe „ewom“ bzw. „electronic word-of-mouth“ benutzt. Daher müssen diese Begriffe in all ihren Varianten (z. B. auch als „word of mouse“) in die Suche einbezogen werden. Zusätzlich dazu habe ich alle englischen Synonyme zu „Online“ bzw. „Konsument“ und zu „Bewertung“ in den Suchterm integriert. So sollen alle Studien aufgegriffen werden, die sich mit Online-Bewertungen beschäftigen, diese aber unter unterschiedlichen Namen verhandeln.

Neben dem konkreten Gegenstand der Online-Bewertungen geht es in dieser Arbeit um den abstrakten Gegenstand des Vertrauens. Vertrauen hängt zusammen mit Vertrauenswürdigkeit („trustworthiness“) und in der englischsprachigen Forschung häufig auch mit Glaubwürdigkeit („credibility“). Daher werden beide Begriffe in den Suchterm aufgenommen.

Neben dem Suchterm spielen die verwendeten Datenbanken eine große Rolle für die Qualität der Ergebnisse der integrativen Literaturübersicht. Als Phänomen der öffentlichen Kommunikation gehören Online-Bewertungen zum Gegenstand der Kommunikationswissenschaft. Online-Bewertungen werden aber auch in anderen Disziplinen wie im Marketing, der Psychologie, der Wirtschaftsinformatik und der Tourismusforschung untersucht. Daher erfolgte die Literaturrecherche für das

Korpus meiner Literaturübersicht nicht nur in kommunikationswissenschaftlichen, sondern auch in den etablierten internationalen wissenschaftlichen Datenbanken der genannten Fächer.

Mit der Auswahl des Suchterms und der Datenbanken stelle ich sicher, dass das Korpus für diesen Forschungsstand erschöpfend und bedeutsam ist. Aus allen Überlegungen zum Suchterm und zu den Datenbanken folgt die in Tabelle 1 dargestellte Suchstrategie.

*Tabelle 1: Suche nach Studien zu Vertrauen in Online-Bewertungen*

<b>Gegenstand</b>	<b>Suchterm</b>
<b>Vertrauen</b>	(trust* OR cred*)
<b>Online-Bewertungen</b>	AND ("ewom" OR "electronic word-of-mouth" OR "electronic word of mouth" OR "word-of-mouth" OR "word of mouse" OR "online review" OR "online rating" OR "online recommendation" OR "internet review" OR "internet rating" OR "internet recommendation" OR "digital review" OR "digital rating" OR "digital recommendation" OR "consumer review" OR "consumer rating" OR "consumer recommendation")

*Datenbanken: Business Source Premier, Communication & Mass Media Complete, PsycARTICLES, PsycINFO, PSYINDEX: Literature and Audiovisual Media with PSYINDEX Tests, Library Information Science and Technology Abstracts, Web of Science.*

*Letzte Suche am 09.08.2017*

Die Recherche auf Grundlage dieser Suchstrategie führt zu 703 Treffern. Wie in der Wissenschaft weitestgehend üblich, müssen die für eine Literaturübersicht recherchierten Studien aber noch formale Auswahlkriterien erfüllen. Um die Qualität der Studien zu gewährleisten, müssen die Studien in einem wissenschaftlichen Journal mit Peer-Review erschienen sein. Damit ich die Studien verstehen kann, müssen diese außerdem in Englisch als aktueller Lingua franca der Wissenschaft

verfasst sein. Von den 703 im ersten Durchlauf gefundenen Studien entsprachen 533 Studien diesen formalen Kriterien. Die Entfernung der Duplikate reduzierte die Anzahl der Studien auf 311.

Neben den formalen Kriterien müssen die Studien im letzten Schritt auch inhaltliche Kriterien erfüllen, um in das Korpus für diese Literaturübersicht integriert zu werden. Eine Studie muss (1) den Zusammenhang mindestens eines Faktors mit (2) Vertrauen in eine bzw. Glaubwürdigkeit von einer (3) Online-Bewertung auf einer Bewertungsplattform (4) theoretisch und/ oder empirisch untersuchen.

Das Erkenntnisinteresse dieser Arbeit liegt also auf Faktoren, die mit Vertrauen in und wahrgenommener Glaubwürdigkeit von Online-Bewertungen zusammenhängen. Studien zu anderen Konzepten wurden allerdings teilweise mit Begriffen aus dem Suchterm verschlagwortet und werden daher durch meine Suchstrategie zunächst mit aufgegriffen. Darum gilt es, die Studien mit anderen Konzepten zu identifizieren und auszuschließen. So werden z. B. Studien, die untersuchen, welche Faktoren mit der Einstellung gegenüber („attitude“, z. B. Willemsen, 2009) bzw. Übernahme („adoption“), Überzeugungskraft („persuasiveness“, z. B. Zhang et al., 2010) oder Annahme („acceptance“, z. B. Gershoff et al., 2003) von Online-Bewertungen zusammenhängen, nicht berücksichtigt. Denn bei diesen Konzepten spielt die Glaub- bzw. Vertrauenswürdigkeit nur eine nachgelagerte Rolle. Auch die Studien zur Nützlichkeit bzw. Brauchbarkeit („helpfulness“ bzw. „usefulness“) von Online-Bewertungen werden ausgeschlossen. Diese hauptsächlich aus der Informatik stammenden Studien untersuchen welche Eigenschaften einer Bewertung oder eines Bewerbers damit zusammenhängen, dass Online-Bewertungen von Lesern mithilfe eines Buttons als „hilfreich“ markiert wird. Es findet keine theoretische Auseinandersetzung mit dem Konzept der Nützlichkeit bzw. Brauchbarkeit statt, der Zusammenhang mit Vertrauen bleibt unklar. Friedlandt (2013) erhebt z. B. „Glaubwürdigkeit“ und „Usefulness“ als getrennte Variablen. Zum Verhältnis von Glaubwürdigkeit und Nützlichkeit siehe auch Yoon (2008) oder Pavlou und Gefen (2004).

Um zu überprüfen, welche der formal passenden 311 Studien den inhaltlichen Kriterien genügen, las ich deren Titel, Schlüsselworte und Abstracts. Falls

die Informationen daraus nicht ausreichen, um über die endgültige Aufnahme in das Korpus des Reviews zu entscheiden, las ich den gesamten Artikel. Außerdem wertete ich die Literaturverzeichnisse der als relevant identifizierten Studien aus und stieß so auf weitere Artikel. Unter Berücksichtigung aller formalen und inhaltlichen Aufgreifkriterien wurden so 53 Studien identifiziert. Diese bilden nun das Korpus der folgenden Literaturübersicht.

Bei der Interpretation der Ergebnisse dieser Literaturübersicht muss beachtet werden, dass auch dieses Korpus Beschränkungen unterliegt. Zum einen muss der Publikationsbias berücksichtigt werden. Es existieren sicher noch weitere Studien zu diesem Thema, die wegen statistisch nicht signifikanter Ergebnisse nicht veröffentlicht werden. Studien, die nach der letzten Aktualisierung der Recherche im August 2017 publiziert wurden, sind auch nicht Teil des Korpus. Außerdem besteht das Risiko, dass einzelne Studien nicht mit meinem Suchterm oder in den verwendeten Datenbanken erfasst wurden. Diese Einschränkungen des Korpus gilt es in der weiteren Arbeit miteinzubeziehen.

Um das Korpus der 53 recherchierten Studien angemessen aufbereiten zu können, verwende ich die Methode der integrativen Literaturübersicht („Integrative Literature Review“, siehe Torraco, 2005, und Whittemore, 2005). Ein mit dieser als methodischer Basis angefertigter Forschungsstand „reviews, critiques, and synthesizes representative literature on a topic in an integrated way“ (Torraco, 2005, S. 356). Auf diese Weise sollen die Forschungstrends zusammengefasst sowie neue Perspektiven und ein Rahmen für den untersuchten Gegenstand gefunden werden (Torraco, 2005; Khoo, 2011).

Für die Analyse der Studien folge ich daher den Vorschlägen von Whittemore und Knafl (2005). Den ersten Schritt bildet die Reduzierung der Studien. Dafür codiere ich die Studien entsprechend der Fragen, die ich am Ende des letzten Unterkapitels (ab S. 7) bereits formulierte. Ich notiere die Autoren, Erscheinungsjahre und Journals und durchsuche die Studien nach Antworten auf die weiteren Fragen zu den von ihnen verwendeten Theorien und Methoden sowie zu den Ergebnissen der Faktoren. Nach der Reduktion der Studien auf die für diesen Forschungsstand relevanten Daten erfolgt im nächsten Schritt die Darstellung der Ergebnisse dieses Prozesses. Die Darstellung soll die Ergebnisse übersichtlich

---

zusammenfassen. Um die codierten Ergebnisse zusammenfassen zu können, werden sie anhand theoretisch bestimmter Merkmale passend gruppiert. Dafür schlage ich jeweils ein Modell vor, in dessen Rahmen die Ergebnisse einsortiert werden können. So können die Ergebnisse im dritten Schritt miteinander verglichen werden. Beim Vergleich wird deutlich, welche Studien zu ähnlichen Ergebnissen kommen und wo sich die Studien unterscheiden oder widersprechen. Im Anschluss werden diese Ergebnisse kritisch analysiert, wie von Torraco (2005) beschrieben.

Dieses systematische Vorgehen ermöglicht es, die Ergebnisse der einzelnen Studien mithilfe eines konzeptionellen Frameworks zu integrieren. So können die Ergebnisse zusammengefasst und verglichen, sowie daraus Fragen an die zukünftige Forschung entwickelt werden. Das stellt auch den größten Vorteil der integrativen Literaturübersicht gegenüber anderen Reviewtypen dar. Es bleibt nicht bei einer bloßen Beschreibung der bestehenden Forschung. Das Ziel der integrativen Literaturübersicht besteht vielmehr darin, ein Konzept anzuwenden oder zu entwickeln, mit dem alle relevanten Studienergebnisse integriert werden können. So biete ich in einem Unterkapitel (ab S. 34) z. B. ein alternatives konzeptuelles Framework an, anhand dessen ich alle bisher untersuchten Faktoren zu Vertrauen in Online-Bewertungen erfassen und einordnen kann.

Ein entscheidender Vorteil gegenüber anderen Reviewtypen besteht bei der integrativen Literaturübersicht außerdem darin, dass hier Studien mit verschiedenen Untersuchungsmethoden berücksichtigt werden können (Whittemore, 2005, S. 547). Denn obwohl die Mehrzahl der Studien zu Vertrauen in Online-Bewertungen auf Experimenten beruht, gibt es dennoch relevante Studien, die ein anderes methodisches Design einsetzen. Alle können hier integriert und miteinander verglichen werden.

Mit der ausführlichen Darstellung der Literaturrecherche und dem methodischen Vorgehen wurde in diesem Kapitel das Fundament gegossen für die Darstellung der Ergebnisse in den folgenden Unterkapiteln. Zunächst geht es um die Geschichte des Forschungsfeldes und die gegenwärtige Forschungslandschaft.

## **2.5 GESCHICHTE UND GEGENWÄRTIGE FORSCHUNGSLANDSCHAFT**

In diesem Kapitel werden die Fragen rund um das Forschungsfeld beantwortet: Wann wurden Studien zum Vertrauen in Online-Bewertungen

veröffentlicht? Welche Autoren forschten dazu? Welche Fachzeitschriften publizierten diese Forschung?

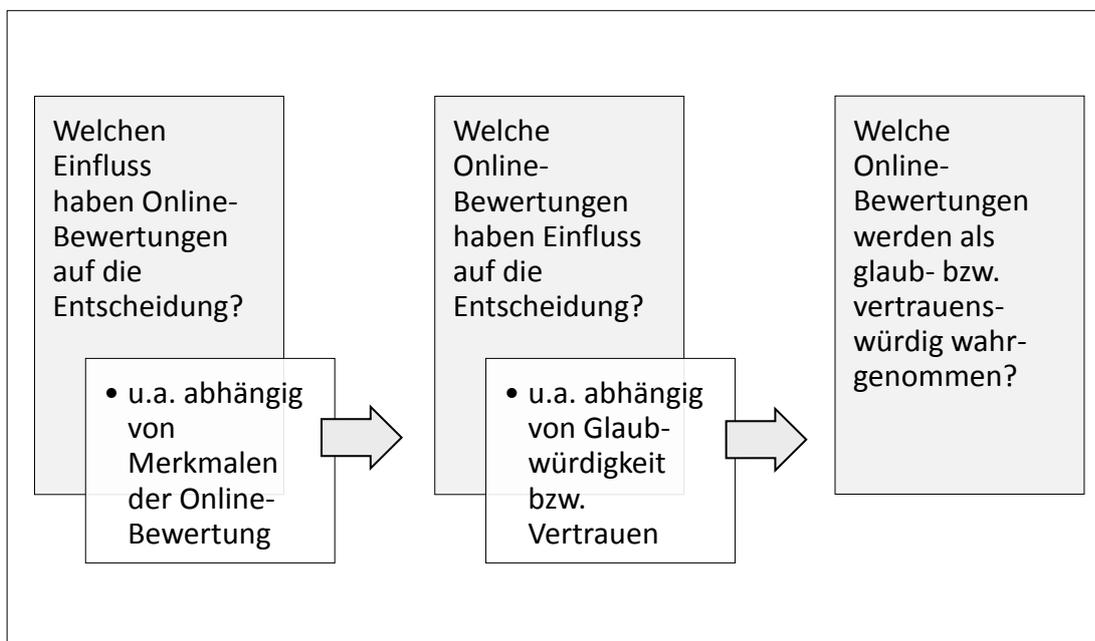
Ich beginne mit einer Einführung in die Entwicklung des Forschungsinteresses an Vertrauen in Online-Bewertungen. Bei diesem Forschungsbereich handelt es sich um eine Subdisziplin der allgemeinen Forschung zu „electronic word of mouth“ („ewom“, für einen Überblick über die eWOM-Forschung siehe Trenz & Berger, 2013, und King et al., 2014). Die bisherige Forschung lässt sich also hauptsächlich im Umfeld des konkreten Gegenstands der „Online-Bewertung“ verorten. Die Forschung zu dem abstrakten Gegenstand des „Vertrauens“, d. h. die Glaubwürdigkeits- bzw. Vertrauensforschung, hat Online-Kommunikation zwar untersucht, aber den konkreten Gegenstand der Online-Bewertungen bisher vernachlässigt. Daher konzentriere ich mich bei hiesiger Darstellung auf die Ausbildung des Forschungsinteresses innerhalb der eWOM-Forschung. Diese hat, als jüngster Spross der Forschung über sog. Mund-zu-Mund-Propaganda („Word of Mouth“), ihren Ursprung in der Marketing- und Managementforschung. Aus diesen Disziplinen stammt auch jetzt noch der Großteil der Studien. Cheng und Ho (2015) bestätigen meinen Eindruck, dass sich jene Studien zunächst dem Einfluss von Online-Bewertungen oder bestimmter Eigenschaften von Online-Bewertungen (wie der Valenz oder der Anzahl der Bewertungen) auf die Wahrnehmung des Händlers bzw. des Produkts, auf die Kaufabsicht oder auf das tatsächliche Kaufverhalten widmeten. Dieser Fokus ist sicher der Anwendungsbezogenheit der Disziplinen geschuldet.

Im weiteren Forschungsfeld zum Gegenstand der Online-Bewertungen beschäftigen sich Studien z. B. mit der Analyse von eWOM selbst. Dabei steht die Frage im Mittelpunkt, ob Konsumenten eine Ware gut oder schlecht bewerten und anhand welcher Kriterien sie das tun. Außerdem gibt es Forschung zur Nutzung von Online-Bewertungen bei der Informationssuche (z. B. Reichelt et al., 2013). Die Verfasser von Online-Bewertungen und ihre Motivation werden seltener untersucht. Neben den vielen empirischen Studien ist mir nur eine Studie bekannt, die sich mit dem theoretischen Verständnis von eWOM und seiner gesellschaftlichen Funktion auseinandersetzt (Borchers, 2016).

Aus der oben genannten Frage nach dem Einfluss von Online-Bewertungen auf die Entscheidung des Lesers entwickelt sich mit der Zeit die Frage nach dem

Vertrauen in Online-Bewertungen, die auch die vorliegende Arbeit beschäftigt (siehe Abbildung 1). Die Frage differenzierte sich mit der Zeit aus. Mit der Erkenntnis darüber, dass der Einfluss u.a. abhängig von bestimmten Merkmalen der Online-Bewertungen ist, entwickelte sich zunächst die Frage, welche der Online-Bewertungen einen Einfluss haben. Hier taucht Vertrauens- bzw. Glaubwürdigkeit als einer der möglichen Faktoren auf. Das führt zu der Frage, welche Online-Bewertungen als glaub- bzw. vertrauenswürdig wahrgenommen werden. Diese Frage etablierte sich innerhalb des letzten Jahrzehnts langsam.

Abbildung 1: Forschungsgeschichte



Hennig-Thurau et al. veröffentlichten 2002 die erste Studie, die sich damit beschäftigt, welche Faktoren die wahrgenommene Glaubwürdigkeit von Online-Bewertungen beeinflussen. Die erste englischsprachige Studie erschien drei Jahre darauf (Smith et al., 2005). Bis 2008 wurden nur zwei weitere Studien dazu publiziert, sodass Lee et al. (2008) zum Ende ihres Artikels über die Effekte von eWOM feststellten: „We have not yet considered the credibility issue.“ (S. 351).

Daneben entwickelte sich allerdings ein Forschungsstrang, der das Vertrauen in Online-Bewertungen im Vergleich zu anderen Informationsquellen misst, d. h. die relative Glaubwürdigkeit (z. B. Senecal & Nantel, 2004). Für den folgenden Forschungsstand wurden darunter diejenigen Studien aufgegriffen, in denen die

Ergebnisse zu Faktoren von Vertrauen in Online-Bewertungen einzeln besprochen werden (z. B. Seo & Lee, 2014).

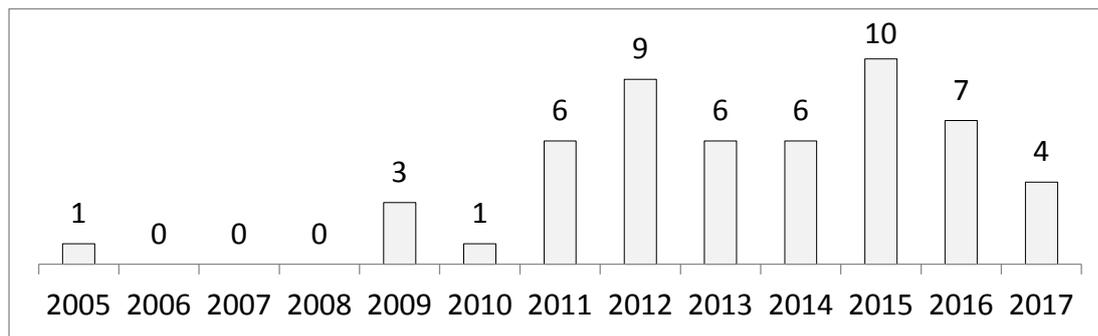
In den letzten Jahren griffen immer mehr Autoren in ihren Modellen die wahrgenommene Glaubwürdigkeit von Online-Bewertungen zumindest als moderierenden Faktor mit auf. Trotzdem stellten verschiedene Autoren fest, dass das Wissen über die Wahrnehmung der Online-Bewertungen (Ong, 2011) bzw. Einflussfaktoren auf die Glaubwürdigkeit (Kusumasondjaja et al., 2012) und das Vertrauen in Online-Bewertungen (Hsiao et al., 2010) noch sehr begrenzt ist. Jensen et al. (2013) führten die teilweise widersprüchlichen Ergebnisse bisheriger Studien über Effekte von eWOM darauf zurück, dass bei vielen Studien die Wahrnehmung der Bewertungen durch den Rezipienten als entscheidender Faktor übersehen wurde und schlagen vor

„we adopt a more nuanced approach for understanding the effects of review characteristics on purchase decisions that considers the level of credibility attributed to the reviewer. Such a conceptualization may assist in moving past previous, sometimes conflicting research findings. It is the additional step of attributing credibility to a reviewer that affects how a product review will influence perceptions and intentions about a product. Characteristics of the reviewer and the review may appear to directly influence potential buyer intentions and perceptions, but without considering the effect of credibility, these apparent relationships may be unstable" (S. 301)

Ebenfalls 2013 findet Lis, dass Glaubwürdigkeit zwar mittlerweile von einigen Studien als der wichtigste Einflussfaktor bei der Übernahme von eWOM erkannt wurde, es aber nur wenige Erkenntnisse über das Entstehen dieser Glaubwürdigkeit gibt.

Doch mittlerweile ist das Thema in der Forschung angekommen. Sortiert nach Erscheinungsjahr wird mit den Studien aus dem vorliegenden Korpus deutlich, dass das Interesse an Vertrauen in Online-Bewertungen ab 2011 wächst (siehe Abbildung 2). Seit sieben Jahren werden jährlich mindestens vier Studien zu den Einflussfaktoren von Vertrauen in Online-Bewertungen publiziert.

Abbildung 2: Artikel pro Jahr

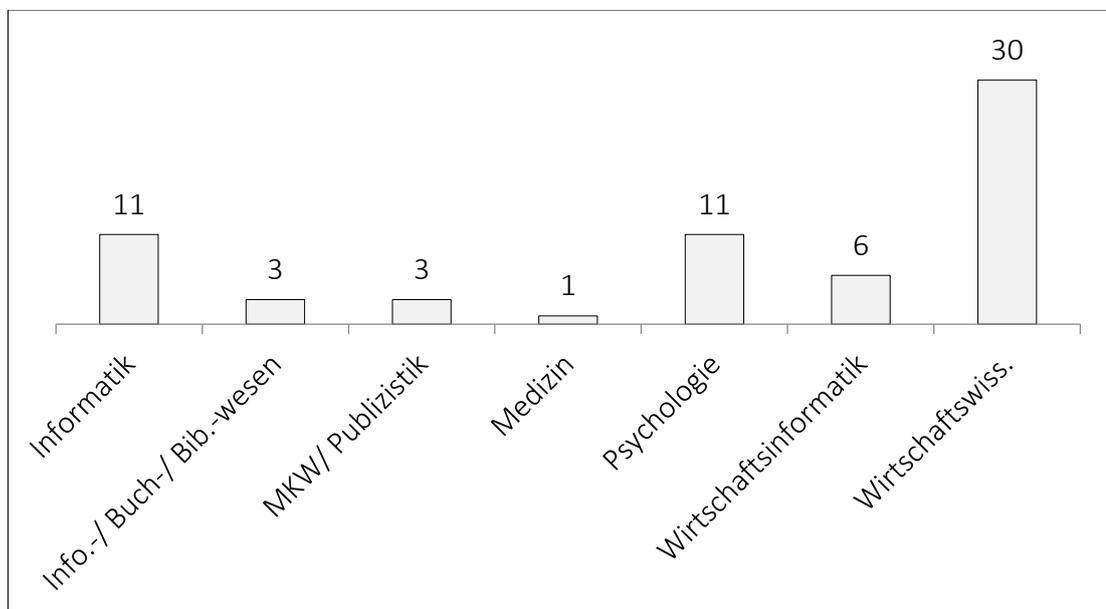


Ein Blick auf die publizierenden Journals zeigt, dass die meisten Studien bisher in *Computers in Human Behavior* veröffentlicht wurden. Dort scheint der Schwerpunkt der bisherigen Forschung zu liegen. Jeweils zwei Studien wurden in *Decision Support Systems*, *International Journal of Hospitality Management*, *Internet Research*, *Journal of Business Research*, *Journal of Computer-Mediated Communication*, *Journal of Interactive Advertising*, *Journal of Interactive Marketing*, *Journal of Travel Research* und *Online Information Review* veröffentlicht. 28 weitere Journals haben je eine Studie publiziert: *Annals of Tourism Research*, *Business & Information Systems Engineering*, *Communication Research*, *Cyberpsychology & Behavior*, *Cyberpsychology, Behaviour, and Social Networking*, *Electronic Commerce Research*, *Electronic Markets*, *European Journal of Marketing*, *Industrial Management & Data Systems*, *Information & Management*, *International Journal of Contemporary Hospitality Management*, *International Journal of Electronic Commerce*, *International Journal of Information Management*, *International Journal of Integrated Marketing Communications*, *International Review of Retail, Distribution & Consumer Research*, *Journal of Consumer Behavior*, *Journal of Electronic Commerce Research*, *Journal of Internet Commerce*, *Journal of Management Information Systems*, *Journal of Medical Internet research*, *Journal of Service Theory and Practice*, *Journal of the Association for Information Systems*, *Journal of Vacation Marketing*, *Management Information Systems Quarterly*, *PsychoNology Journal*, *Qualitative Market Research: An International Journal*, *Recherche et Applications en Marketing (English Edition)* und *Tourism Management*.

Schon an den Titeln der Fachzeitschriften wird deutlich, dass Vertrauen in Online-Bewertungen in vielen verschiedenen Disziplinen erforscht wird. Um einen

Überblick darüber zu bekommen, welche Disziplinen wie stark involviert sind, ordne ich die Studien nach den Fachgebieten der Journals, in denen sie erschienen sind (siehe Abbildung 3). Dabei orientiere ich mich an der Einteilung der Elektronischen Zeitschriftenbibliothek, hier sind Mehrfachzuteilungen möglich. *Computers in Human Behavior*, die Fachzeitschrift, die die meisten Studien aus dem Korpus veröffentlicht hat, wird hier z. B. den Fachgebieten Informatik, Medizin und Psychologie zugeordnet. Die meisten publizierenden Fachzeitschriften lassen sich den Wirtschaftswissenschaften zuordnen. Dort sind die häufigsten Schlagworte „Marketing“ (bei neun Fachzeitschriften) und „E Commerce“ (bei sechs Fachzeitschriften). Nur drei der Studien zu Vertrauen in Online-Bewertungen wurden in Fachzeitschriften veröffentlicht, die dem Fachbereich der Medien- und Kommunikationswissenschaft bzw. Publizistik zugeordnet werden können. Zwei der Studien erschienen im *Journal of Computer Mediated Communication*, eine in der Zeitschrift *New Media & Society*. Das zeigt, dass es bisher kaum Arbeiten zu diesem Thema aus kommunikationswissenschaftlicher Perspektive gibt.

Abbildung 3: Artikel pro Fachgebiet



Die Zitierhäufigkeit kann als Indikator für die Relevanz einer Studie herangezogen werden. Für die Angaben dazu in Tabelle 2 benutze ich die von *Web of Science* zur Verfügung gestellten Meta-Daten („k. A.“ steht für „keine Angabe“). Die drei am häufigsten zitierten Studien zum Thema sind Cheung et al. (2009) mit 208, Doh und

Hwang (2009) mit 108 und Ayeh et al. (2013) mit 93 Zitationen. Alle weiteren Studien zum Thema wurden bisher gar nicht bis 78 Mal zitiert. Deutlich wird auch, dass es viele verschiedene Autoren gibt, die zu diesem Thema publizieren. Es lassen sich bisher also kaum kanonische Studien oder zentrale Autoren ausmachen, die den Schwerpunkt der Forschung bilden. Das lässt sich dadurch erklären, dass es sich um ein junges und fachlich breites Forschungsfeld handelt.

*Tabelle 2: Zitierhäufigkeit nach Studie*

Studie	Zitierhäufigkeit <sup>1</sup>
Ayeh et al. (2013)	93
Bae & Lee (2011)	24
Casaló et al. (2015)	19
Cheung et al. (2012)	23
Cheung et al. (2009)	208
Chih et al. (2013)	8
Cox et al. (2017)	0
Dickinger (2011)	57
Doh & Hwang (2009)	108
Dou et al. (2012)	34
Filieri (2016)	5
Filieri (2015)	31
Filieri et al. (2015)	34
Flanagin & Metzger (2013)	43
Furner et al. (2016)	0
Grabner-Kräuter & Waiguny (2015)	7
Hong & Park (2012)	22
Hsiao et al. (2010)	57
Hsu et al. (2016)	3
Jensen et al. (2013)	40
Jensen & Yetgin (2017)	0
Kusumasondjaja et al. (2012)	k. A.
Lee et al. (2011)	53
Lee & Koo (2012)	34
Lee (2014)	7
Lim & van der Heide (2015)	16
Lin & Xu (2017)	3
Lis (2013)	10
Llamero (2014)	5
López & Sicilia (2014)	k. A.
Luo et al. (2013)	17

1 bis 02.02.2018 laut *Web of Science* (alle Datenbanken)

Luo et al. (2015)	5
Luo et al. (2014)	9
McGloin et al. (2014)	k. A.
Munzel (2015)	k. A.
Nekmat & Gower (2012)	k. A.
O'Reilly et al. (2016)	k. A.
O'Reilly & Marx (2011)	k. A.
Pan & Chiou (2011)	64
Qiu et al. (2012)	35
Racherla et al. (2012)	23
Reimer & Benkenstein (2016)	0
Shamhuyenzva et al. (2016)	k. A.
Smith et al. (2005)	k. A.
Tsang & Prendergast (2009)	32
Tsao & Hsieh (2015)	4
van der Heide & Lim (2016)	5
Wang, S. et al. (2015)	k. A.
Wang, X. et al. (2015)	2
Willemsen et al. (2012)	30
Wu et al. (2017)	0
Xie et al. (2011)	78
Xu (2014)	25

Alles in allem wird deutlich, dass Vertrauen in Online-Bewertungen sich langsam als eigenes Forschungsthema etabliert. Forscher aus vielen verschiedenen Fächern beschäftigen sich mit diesem Phänomen. Bisher wurde das Thema aber vor allem aus der Perspektive der Wirtschaftswissenschaften untersucht. Aus der Perspektive der Medien- und Kommunikationswissenschaft gibt es dagegen nur vereinzelte Studien. Mit diesem Wissen über die Entwicklung und aktuelle Gestaltung der Forschungslandschaft ist jetzt eine Auseinandersetzung mit den Inhalten des Studienkorpus möglich.

## 2.6 VERSTÄNDNIS VON VERTRAUEN

Den konkreten Gegenstand der Online-Bewertungen habe ich am Anfang des Kapitels definiert und die Literaturrecherche dementsprechend durchgeführt. Alle ausgewählten Studien behandeln daher einen ähnlichen konkreten Gegenstand.

Doch der abstrakte Gegenstand des Vertrauens ist bisher ungeklärt. Für die Literaturrecherche wird im Suchterm nicht zwischen Vertrauen bzw. Vertrauenswürdigkeit und Glaubwürdigkeit unterschieden. Denn manche Autoren nennen den einen Begriff, manche benutzen den anderen, fast alle verwenden sogar

beide Begriffe im selben Text für dasselbe Konstrukt. Um die Ergebnisse der Studien vergleichen zu können, muss aber klar sein, ob sie tatsächlich dasselbe Phänomen untersuchen. Daher geht es in diesem Unterkapitel darum, was die Autoren bisheriger Studien unter „Vertrauen“ bzw. „Glaubwürdigkeit“ verstehen und wie sie das Konstrukt messen.

In der oben aufgeführten flexiblen Handhabung der unterschiedlichen Begriffe zeigt sich schon die konzeptionelle Unschärfe vieler Studien. Für die Forschung zu Vertrauen in Online-Bewertungen lässt sich daher dieselbe Kritik anführen, die Kohring (2004) schon an Hovland et al. (1953) übte an der „unzureichende[n] Reflexion der Begriffe Glaubwürdigkeit und Vertrauen(swürdigkeit), die oft sogar synonym verwendet werden“. Das Verhältnis von Vertrauen und wahrgenommener Glaubwürdigkeit wird, wie in älteren Studien zu diesem Phänomen (Kohring, 2004), häufig nicht geklärt.

Diese unzureichende Reflexion der Begriffe lässt sich als Indikator lesen für die mangelnde theoretische Auseinandersetzung mit dem für das Forschungsthema eigentlich zentralen Konzept des Vertrauens. Die überwiegende Mehrheit der Studien zu Vertrauen in Online-Bewertungen hat einen empirischen Schwerpunkt mit zum Teil hohem methodischen Aufwand und großen Datensätzen. Das theoretische Verständnis von Vertrauen wird dagegen nicht expliziert. Dasselbe gilt aber auch für die wenigen qualitativen Studien. Meist interpretieren die Autoren die Aussagen ihrer Teilnehmer zu Vertrauen entsprechend der Messitems einer Skala aus der quantitativen Forschung. So findet in Studien dieses Korpus fast nie eine Konzeptualisierung von Vertrauen statt. Die Beschäftigung mit Vertrauen wird stattdessen von der theoretischen auf die methodische Ebene verlagert. Erst dort wird das vorher nicht genau erklärte, wie als bekannt vorausgesetzte Konstrukt des Vertrauens nun in verschiedene Items zur Messung zerlegt. So geschieht die Definition von Vertrauen gewissermaßen verdeckt, auf dem Weg (siehe dazu auch Kohring, 2004, S. 37). Doch auch diese Operationalisierung von Vertrauens- bzw. Glaubwürdigkeit erfolgt meist unbegründet. Das hat zur Folge, dass unklar ist, was mit dem Instrument genau gemessen werden soll.

Um die Ergebnisse der Studien deuten und miteinander in Bezug setzen zu können, ist es aber notwendig, nachzuvollziehen, was in den Studien als Vertrauen

bzw. Glaubwürdigkeit verhandelt wird. Daher schaue ich mir die konkrete Manifestation der vernachlässigten theoretischen Konzeptualisierung auf empirischer Ebene an: die Skalen zur Messung von Vertrauen (siehe Tabelle 3).

Die Studien messen Vertrauen auf unterschiedliche Weise. Der Großteil der Autoren greift dabei vollständig oder teilweise auf bestehende Skalen zurück. Am häufigsten wird die Skala von Ohanian (1990 bzw. 1991) benutzt, 13 Studien nutzen sie. Sieben Studien verwenden die Skala von Cheung et al. (2008 bzw. 2009), welche selbst wiederum auf Block & Keller (1995) und Smith & Vogt (1995) basiert (auf letzterer fußt auch Luo et al., 2015). Außerdem werden die Skalen von Flanagin und Metzger (2000 bzw. 2007) vier Mal übernommen. Seltener, aber mehrfach kommen die Skalen folgender Autoren(teams) vor: Gefen (2000) bzw. Gefen & Straub (2004), Hovland et al. (1953) bzw. Hovland & Weiss (1951), Johnson & Kaye (2004 bzw. 2009) bzw. Johnson et al. (2008) bzw. Johnson & Grayson (2005), sowie McCroskey & Teven (1999).

*Tabelle 3: Skalen und Konzepte von Vertrauen*

<b>Vertrauen nach</b>	<b>Studie</b>
Andrews et al. (1998)	Munzel (2015)
Berlo et al. (1969)	Cheung et al. (2009)
Bettman & Park 1980	Doh & Hwang (2009)
Block & Keller (1995)	Cheung et al. (2009) Cheung et al. (2012)
Chang (2009)	Chang (2014)
Cheung et al. (2008)	Hsu et al. (2016)
Cheung et al. (2009)	Lee & Koo (2012) Lis (2013) O'Reilley & Marx (2011) O'Reilley et al. (2016) Qui et al. (2013) Shamhuyenzva et al. (2016)
Fan & Miao (2012)	Shamhuyenzva et al. (2016)
Fisher et al. (1979)	Lopéz & Sicilia (2014)
Flanagin & Metzger (2000)	Hong & Park (2012)
Flanagin & Metzger (2007)	Jensen & Averbeck (2013) Kusumasondjaja et al. (2012) Nekmat & Gower (2012)
Furner et al. (2012)	Furner et al. (2016)
Gefen (2000)	Dickinger (2011)
Gefen & Straub (2004)	Hsiao et al. (2010)
Hilligoss & Rieh (2008)	Llamero (2014)

Hovland & Weiss (1951)	Shamhuyenhazva et al. (2016)
Hovland et al. (1953)	Xie et al. (2011)
Huang & Chen (2006)	O'Reilley et al. (2016)
Johnson & Grayson (2005)	Bae & Lee (2011)
Johnson & Kaye (2004)	Xu (2014)
Johnson & Kaye (2009)	
Johnson et al. (2008)	
Kiecker & Cowles (2001)	O'Reilley et al. (2016)
Lis (2013)	Shamhuyenhazva et al. (2016)
McAllister (1995)	Xu (2014)
McCroskey & Teven (1999)	Lim & van der Heide (2015) van der Heide & Lim (2016)
McKnight & Chervany (2003)	Racherla et al. (2012)
Ohanian (1990)	Ayeh et al. (2013)
Ohanian (1991)	Cox et al. (2017) Dou et al. (2012) Filieri (2015) Filieri et al. (2015) Grabner-Kräuter & Waiguny (2015) Lin & Xu (2017) Lis (2013) Reimer & Benkenstein (2016) Tsao & Hsieh (2015) Wang, S. et al. (2015) Wang, X. et al. (2015) Willemssen et al. (2012)
Park & Kim (2008)	O'Reilley et al. (2016)
Smith & Vogt (1995)	Cheung et al. (2009) Cheung et al. (2012) Luo et al. (2015)
Wathen & Burkell (2002)	Filieri (2016)
Willemssen et al. (2012)	Shamhuyenhazva et al. (2016)
Zhang & Watts (2003)	Cheung et al. (2009)
Zhang (1996)	Cheung et al. (2009)
<i>eigene Skala</i>	Flanagin & Metzger (2013) McGloin et al. (2014) Pan & Chiou (2011) Smith et al. (2005) Tsang & Prendergast (2009) Xu (2014)

Diese Vielfalt an Skalen verdeutlicht, dass sich bisher kein Messinstrument für das Konstrukt Vertrauen bzw. Glaubwürdigkeit im Kontext von Online-Bewertungen durchsetzen konnte. Diese sehr unterschiedliche Operationalisierung erschwert zunächst die Zusammenführung der Ergebnisse aller Studien.

Ein Blick auf die einzelnen Items der verwendeten Skalen verspricht aber einen Ausweg aus diesem Problem. Die vielen Skalen überschneiden sich häufig hinsichtlich der verwendeten Items. Das lässt eine gemeinsame Interpretation möglich erscheinen. Dafür müssen die Items aber zunächst genauer bestimmt werden. Daher folgt nun eine intensive Auseinandersetzung mit den einzelnen Items.

In den meisten Studien werden drei bis fünf Items benutzt. Bei der näheren theoretischen Betrachtung der Items zeigt sich, dass diese verschiedenen Gruppen zugeordnet werden können (siehe Tabelle 4). Die Messung erfolgt anhand einer Abfrage (1) des Abstraktums „Vertrauen“, (2) der Konsequenzen des Vertrauens oder (3) der Definition von Vertrauen. Dazu existiert eine große Gruppe mit allen restlichen Items, deren theoretische Bedeutung an dieser Stelle noch nicht geklärt werden kann. Die Items werden häufig über die theoretisch unterscheidbaren Gruppen hinweg gemischt verwendet. In den meisten Studien werden zwei Aspekte von Vertrauen mit demselben Instrument gemessen. Es folgt eine genauere Darstellung und Kritik dieser vier Gruppen von Items.

Mit der Messung anhand des Abstraktums „Vertrauen“ ist die direkte Abfrage von „Vertrauen“ oder abstrakter Synonyme von „Vertrauen“ gemeint. Es fällt auf, dass die am häufigsten genutzten Items in den Studien das zu messende Konstrukt einfach mit denselben abstrakten Begriffen messen, mit denen sie es bezeichnen, oder mit ähnlich abstrakten Synonymen. Darunter fallen „trustworthy“, „reliable“, „credible“, „believable“, „dependable“ und „confidence“. Auch in den qualitativen Studien wird Vertrauen u.a. als Abstraktum verstanden, so z. B. bei O’Reilly et al. (2016): „The authors posit that online source credibility is directly linked to how likely a consumer is to believe and use the information from a particular source. In this way, credibility is viewed as synonymous with believability.“ Bei O’Reilly & Marx (2011) steht „believable“ für Vertrauen, bei Llamero (2014) „trustworthiness“.

Bei der Verwendung dieser Items bleibt allerdings unklar, was genau gemessen wird. Der Autor der Studien kann nicht wissen, was der Studienteilnehmer unter dem abstrakten Begriff versteht und ob alle Studienteilnehmer dasselbe darunter verstehen. Kohring kritisiert an dieser „Gleichsetzung von Begriff und Konstrukt“ (2004, S. 39): „als verstehe es sich von selbst, dass das relevante *Konstrukt* ‚Glaubwürdigkeit‘ auch durch den *Begriff* ‚Glaubwürdigkeit‘ repräsentiert

werde – würde es vorläufig einfach X genannt werden, könnte man nicht einfach so schlicht danach fragen.“ (2004, S. 39). Er schlägt stattdessen vor: „Ein theoretisch abgeleitetes Konstrukt explorativ zu erkunden heißt, es in die Sprache von Rezipienten und Befragten zu transformieren, z. B. durch die Suche nach semantisch geeigneten Indikatoren.“ (Kohring, 2004, S. 38f). Wie man es anders besser machen kann, zeigt Llameros (2014) Studie. Die Autorin überlässt die Definition von Vertrauen zwar auch den Teilnehmern, aber erhebt diese mit, um zu erfahren, was die Teilnehmer darunter verstehen. Am wenigsten valide und reliabel sind dagegen die Studien, in denen das Konstrukt nur mit einem Item gemessen wird, z. B. „trust/trustworthy“ (Dickinger, 2011; Seo & Lee, 2014).

Die folgenden zwei Gruppen von Items kommen nur selten vor. Dazu gehört die Messung von Vertrauen anhand der Konsequenzen des Vertrauens. In drei Studien gibt es Items, die nicht direkt das Vertrauen erfassen, sondern das, was aus dem Vertrauen folgt. Das Vertrauen des Lesers wird darüber gemessen werden, ob der Leser aus dem Vertrauen eine bestimmte Handlung ableitet. In drei Fällen besteht die als Vertrauen in die Bewertung abgefragte Konsequenz darin, das empfohlene Produkt zu nutzen bzw. zu kaufen (Racherla et al., 2012; Tsang & Prendergast, 2009; Lee, 2014). In einem Fall geht es darum, die Bewertungen an Verwandte und Freunde weiterzuempfehlen (Lee, 2014).

Solche eine Messung ist zwar prinzipiell möglich und gültig. Dennoch bleibt so unklar, worin das Vertrauen des Lesers genau besteht. Außerdem kann es neben Vertrauen andere Faktoren geben, die beeinflussen, warum ein Leser eine bestimmte Handlung vollzieht (z. B. Bequemlichkeit).

Die Messung anhand der Definition von Vertrauen ist eine Besonderheit, die sich nur in zwei Studien zeigt. Hier wird in der Operationalisierung von Vertrauen mithilfe bestimmter Items definiert, was die Autoren unter Vertrauen verstehen: „I would trust the reviewer to make a decision for me.“ (McGloin et al. 2014) und „I trust the peer recommender to the extent that if I were unable to make this decision on my own, I would allow him to make it for me.“ (Smith et al., 2015). Dem Bewerter zu vertrauen bedeutet also hier, ihn für sich entscheiden zu lassen.

Die Messung anhand der Definition von Vertrauen ist grundsätzlich möglich und gültig. Aber die bisher verwendeten Definitionen von Vertrauen sind zu allgemein für

die Messung von spezifischem Vertrauen. Denn dabei bleibt unklar, worin genau das Vertrauen in Online-Bewertungen besteht. Auf das obige Beispiel bezogen stellt sich nämlich die Frage, welche der vielfältigen Entscheidungen in seinem Leben der Leser dem Bewerter überlassen würde. Sicher wollen die Autoren der Studie nicht wissen, ob das Vertrauen des Lesers darin besteht, dem Bewerter die Entscheidung bezogen auf *jegliche* Handlungsbereiche zu übertragen. Das Spezifische des Vertrauens in Online-Bewertungen bleibt unklar. Das Vertrauen in den Bewerter wird sich wohl nur auf bestimmte Fähigkeiten des Bewerter und damit nur auf bestimmte Handlungsbereiche beziehen. Diese bleiben in den verwendeten Definitionen aber unbestimmt.

Die vierte Gruppe enthält alle Items, die in keine der anderen drei Gruppen passten. Die theoretische Bedeutung dieser Gruppe von Items ist noch nicht geklärt. Die Items betreffen zum einen die Bewertung selbst („accurate“, „authentic“, „complete“, „convincing“, „depth“, „factual“, „fair“, „honest“, „no false information included“, „realistic“, „reasonable“, „sincere“, „unbiased“, „well written“, „easy to read“, „communicated clearly“, „logic of this review is easy to follow“) oder die Items beziehen sich auf Eigenschaften des Bewerter („bright“, „competent“, „informed“, „trained“, „intelligent“, „experienced“, „knowledgable“, „expert“, „qualified“, „skilled“, „good“, „attractive“, „open-minded“, „ethical“, „moral“, „caring“, „concerned with me“, „warm“, „sensitive“, „not self-centered“, „understanding“, „constructive“, „very open“, „honorable“, „genuine“, „has my interests at heart“ und „I can talk freely with him“). Llamero (2014) findet heraus, dass ihre Teilnehmer unter Vertrauen verstehen, dass der Bewerter kompetent und die Bewertung plausibel ist. Hier ist wiederum die Verwendung der abstrakten Begriffe „good“ und „quality“ zu kritisieren. Denn auch hier bleibt wieder unklar, was der Rezipient darunter versteht.

*Tabelle 4: Items für Vertrauen*

Studie	Abstrakt um	Konse- quenz	Defini- tion	Sonstige
Ayeh et al. (2013)		3		7
Bae & Lee (2011)		1		3
Casaló et al. (2015)		1		2
Chang (2014)		1		3

Studie	Abstrakt um	Konse- quenz	Defini- tion	Sonstige
Cheung et al. (2009)	1			2
Cheung et al. (2012)	2			2
Chih et al. (2013)	2			2
Cox et al. (2017)	2			3
Dickinger (2011)	1			
Doh & Hwang (2009) <sup>2</sup>				
Dou et al. (2012)	3			7
Filieri (2015)	3			1
Filieri (2016)	1			1
Filieri et al. (2015)	3			1
Flanagin & Metzger (2013)	2			
Furner et al. (2016)	2	1		
Grabner-Kräuter & Waiguny (2015)	3			7
Grabner-Kräuter & Waiguny (2015)	1			1
Hong & Park (2012)	2			3
Hsiao et al. (2010)	3			
Hsu et al. (2016)	1			2
Jensen & Averbeck (2013)	3			1
Jensen & Yetgin (2017)	3			1
Kusumasondjaja et al. (2012)	2			3
Lee & Koo (2012)	1			1
Lee (2014)	4	2		
Lee et al. (2011) <sup>3</sup>				
Lim & van der Heide (2015)	1			17
Lin & Xu (2017)	2			2
Lis (2013)	1			2
Lis (2013)	3			7
Llamero (2014)	1			2
Lopéz & Sicilia (2014)	1			1
Luo et al. (2013)	3			1
Luo et al. (2014)	3			1
Luo et al. (2015)	3			1
McGloin et al. (2014) <sup>4</sup>		1		2
Munzel (2015)	3			
Nekmat & Gower (2012)	3			6
O'Reilley & Marx (2011)	1			2
O'Reilley et al. (2016)	1			1
Pan & Chiou (2011)	3			
Qui et al. (2013)	3			

---

2 Anzahl und Items nicht genannt

3 Items nicht genannt

4 6 Items nicht genannt

Studie	Abstrakt um	Konse- quenz	Defini- tion	Sonstige
Racherla et al. (2012) <sup>5</sup>				
Reimer & Benkenstein (2016)	2			2
Shamhuyenhazva et al. (2016)	2			2
Shamhuyenhazva et al. (2016)	5			
Smith et al. (2005)	1		1	
Tsang & Prendergast (2009)	1	1		1
Tsao & Hsieh (2015)	3			7
van der Heide & Lim (2016)	1			11
Wang, S. et al. (2015)	3			4
Wang, X. et al. (2015)	2			4
Willemsen et al. (2012) <sup>6</sup>				4
Wu et al. (2017)	1			1
Xie et al. (2011)	1			1
Xu (2014)	3			

Schon allein die Nutzung verschiedener Skalen aber auch die zahlreichen einzelnen Items der Skalen machen deutlich, dass es bisher keine Einigung über die Messung von „Vertrauen“ bzw. „Glaubwürdigkeit“ gibt. Alle Studien des Korpus messen verschiedene Aspekte des Konstrukts, die aber miteinander zusammenzuhängen scheinen. Das ermöglicht auch über die verschiedenen Messinstrumente hinweg eine Vergleichbarkeit der Studien. Dennoch muss bei der Interpretation der zusammengeführten Ergebnisse der Studien immer mitbedacht werden, dass sie sich vielleicht durch unterschiedliche Messungen von Vertrauen erklären lassen.

Neben den unterschiedlichen Skalen fällt bei der genauen Betrachtung der Messung auch noch auf, dass sich die Referenz der Messung unterscheidet. Damit meinte ich, dass sich die Items auf verschiedenen Einheiten beziehen. Bei etwas mehr als der Hälfte der Studien wird das Vertrauen in die bzw. die Glaubwürdigkeit von der Bewertung erhoben. Bei den anderen Studien geht es um das Vertrauen in den Bewerter oder auch in die Bewertung und in den Bewerter gleichzeitig. Zwischen den Studien herrscht also auch Uneinigkeit darüber, was die relevante Einheit ist. So wenden manche Autoren, die dieselbe Skala benutzen, sie auf verschiedene Vertrauensobjekte an. Tsao und Hsieh (2015) beziehen z. B. die am meisten genutzte

---

5 Anzahl und Items nicht genannt

6 2 Items nicht genannt

Skala von Ohanian (1990) auf die Bewertung selbst, alle anderen Autoren beziehen sie auf den Bewerter. Die Skala von Flanagin und Metzger (2000, 2007) wird von Hong und Park (2012) sowie Kusumasondjaja et al. (2012) auf die Bewertung bezogen, bei Jensen et al. (2013) und Nekmat & Gower (2012) dagegen auf den Bewerter. In einigen Fällen beziehen sich einige Items für dasselbe Konstrukt auf die Bewertung, in anderen auf den Bewerter (Dickinger, 2011; Huang, 2013). Bei der explorativen Studie von O'Reilley und Marx (2011) steht ebenfalls „Factors affecting online review(ers) credibility“. Das zeigt deutlich, dass das Vertrauensobjekt unklar ist, „review“ und „reviewer“ werden hier einfach als eins gefasst.

Hennig-Thurau et al. (2002) meinen, dass „es eine Frage der Perspektive [ist], ob von Vertrauen in den Verfasser von Meinungsbeiträgen oder in den Beitrag selbst die Rede ist“ (S. 9). Die gewählte Perspektive wird aber in den bisherigen Untersuchungen nicht expliziert oder begründet. Die mangelnde Reflexion bei der Wahl der Bezugseinheit sowie die flexible Handhabung der Bezugseinheit bei derselben Skala über verschiedene Studien hinweg, weisen auf eine theoretische Unklarheit hin. Es bleibt die Frage offen, ob sich das Vertrauen des Lesers auf die Bewertung oder den Bewerter bezieht sowie welcher Zusammenhang dazwischen besteht.

In diesem Unterkapitel wird die theoretische Schwäche der bisherigen Forschung deutlich. Es fehlt ein theoretisches Konzept von Vertrauen. Dem Problem der vielen Skalen konnte dagegen mit einem näheren Blick auf die Items begegnet werden. Die Eingruppierung der Items machte klar, dass diese miteinander zusammenhängen. Die theoretische Bedeutung der vierten Gruppe an Items ist aber noch unklar, genauso wie die Bestimmung von Bewertung oder Bewerter als Referenz der Messung. Diese Herausforderungen rufen nach einer theoretischen Reflexion des untersuchten Phänomens Vertrauen in Online-Bewertungen. Im nächsten Unterkapitel soll es aber zunächst darum gehen, das methodische Vorgehen der Studien kritisch zu beschreiben.

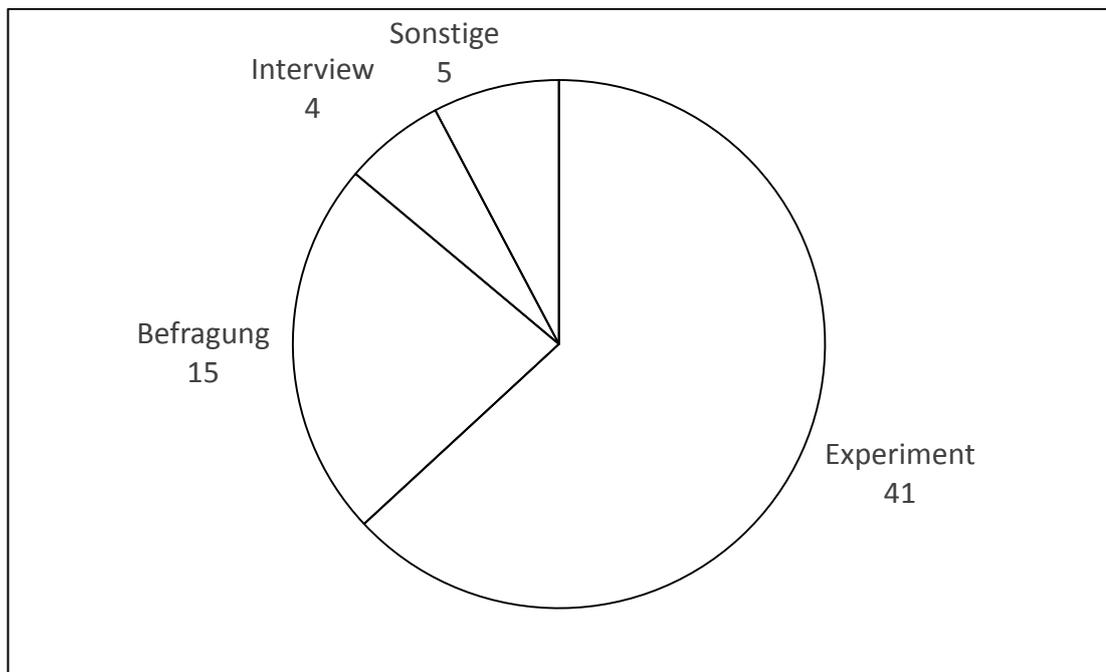
## **2.7 METHODISCHES VORGEHEN**

Um die Ergebnisse der Studien zusammenzufassen und miteinander vergleichen zu können, ist es wichtig, zu wissen, wie die Daten erhoben wurden. Die

Forschungsanlage wird bestimmt durch die Wahl der Forschungsmethode, die Wahl des konkreten Forschungsgegenstands und die Wahl der Studienteilnehmer.

Zunächst widme ich mich den von den Autoren benutzten Methoden (siehe Abbildung 5). Viele Autoren orientieren sich nicht nur theoretisch sondern auch methodisch an der Psychologie, die meisten untersuchen den Einfluss einer unabhängigen Variable auf Vertrauen mit einem klassischen psychologischen Experiment. Diese Methode wird in über der Hälfte der Studien gewählt. Das Experiment ist auch in der sonstigen eWOM-Forschung die üblichste Erhebungsmethode (Cheng & Ho, 2015). Darauf folgen 15 Studien mit einer standardisierten Befragung. Fast alle dieser Studien finden online-vermittelt statt. Neben diesem Großteil der Studien, die einen quantitativen Ansatz verfolgen, gibt es nur neun Studien, bei denen die Autoren qualitativ vorgehen. Vier davon nutzen offene Interviews zur Datenerhebung.

Abbildung 4: Studien pro Methode<sup>7</sup>



Die übliche Vorgehensweise bei den drei genannten Methoden stelle ich nun genauer vor, orientiert an drei Dimensionen. Zum einen unterscheide ich, ob die der

---

<sup>7</sup> In manchen Artikeln kommen zwei Studien mit verschiedenen Methoden vor, daher ist n = 65.

Studie zugrunde liegende Bewertung natürlich ist oder manipuliert wurde. Zum anderen teile ich die Studien danach ein, ob die Wahrnehmung der Bewertung unbekannt oder bekannt ist. Unbekannt hieße, dass die Wahrnehmung derselben Bewertung durch verschiedene Teilnehmer nicht untersucht wird. Bekannt bedeutet, dass die Wahrnehmung derselben Bewertung durch verschiedene Teilnehmer untersucht und verglichen wird. Die dritte Dimension ist die des Faktors, der entweder vorher festgelegt oder nicht vorher festgelegt wurde.

Beim Experiment lesen die Teilnehmer Online-Bewertungen, die der Forscher zuvor hinsichtlich eines bestimmten Faktors manipuliert hat. Die Bewertung ist häufig angelehnt an originale Bewertungen. Hier wird sichergestellt, dass die Manipulation des Faktors erfolgreich ist. Das bedeutet, dass der Faktor vom Leser so wahrgenommen wird, wie vom Forscher intendiert. Die Leser nehmen dieselbe Bewertung hinsichtlich der Ausprägung dieses Faktors demnach gleich wahr. So sind Aussagen darüber möglich, wie eine bestimmte, vorher vom Forscher festgelegte Eigenschaft einer manipulierten Bewertung, die als solche wahrgenommen wird, auf Vertrauen wirkt. Allerdings sind keine Aussagen darüber möglich, ob die Eigenschaft auch in einer natürlichen Bewertung wahrgenommen wird und ob es weitere Eigenschaften einer Bewertung gibt, die mit Vertrauen zusammenhängen.

Die zweithäufigste Erhebungsmethode ist die standardisierte Befragung ohne vorigen Stimulus. Hierbei werden die Teilnehmer z. B. zur letzten Bewertung, die sie gelesen haben, oder allgemein zu Bewertungen befragt. Die Teilnehmer werden zum Zusammenhang eines vorher festgelegten Faktors mit Vertrauen befragt. Die Antworten der Teilnehmer beziehen sich auf natürliche, aber dem Forscher unbekannt Bewertungen. Es sind Aussagen darüber möglich, wie eine bestimmte, vorher vom Forscher festgelegte Eigenschaft einer Bewertung, wenn sie als solche wahrgenommen wird, mit Vertrauen zusammenhängt. Es sind keine Aussagen darüber möglich, wie sich die Wahrnehmung von Bewertungen unterscheidet und ob es weitere Eigenschaften einer Bewertung gibt, die mit Vertrauen zusammenhängen.

In den Befragungen wird z. B. die vom Leser wahrgenommene Argumentation des Bewerter gemessen mit drei (Cheung et al., 2009, 2012; Fang et al., 2014) bis sechs (Luo et al., 2014) Items der Skala von Zhang (1996). Die Autoren fragen die Teilnehmer, ob die Argumente in der Bewertung „convincing“, „good“, „strong“,

„persuasive“, „reasonable“ und „effectively supported“ sind. Diese Befragung birgt das Problem, dass die Autoren die Bewertungen nicht kennen und daher nicht sagen können, was für den Leser denn nun eine „starke“ Argumentation ausmacht. Es kann sein, dass jeder Leser darunter etwas anderes versteht. Genau dasselbe gilt z. B. für die „Vollständigkeit“ einer Bewertung. Bei den Befragungen bleibt unklar, was für den Leser denn genügend oder die richtigen bzw. wichtigen Informationen sind.

Neben dem Experiment und den standardisierten Befragungen gibt es nur selten offene Methoden. In vier Studien werden teilstrukturierte Interviews genutzt. In einer Studie wird noch vor dem Interview eine Think-Aloud-Aufgabe durchgeführt (Llamero, 2014). Bei einer anderen Studie werden Artefakte der Teilnehmer und Forscher wie z. B. gespeicherte Suchen und Bewertungen zum Interview hinzugezogen (O’Reilly et al., 2016). In den Interviews werden die Teilnehmer gefragt, auf welche Faktoren sie achten, wenn sie die Vertrauenswürdigkeit einer Bewertung beurteilen. Die Antworten der Teilnehmer beziehen sich auf natürliche, aber dem Forscher unbekannt Bewertungen. Die Faktoren werden vom Forscher nicht vorher festgelegt. Dadurch sind Aussagen darüber möglich, wie nicht vorher festgelegte Eigenschaften einer Bewertung, wenn sie als solche wahrgenommen wird, mit Vertrauen zusammenhängen. Es sind keine Aussagen darüber möglich, wie sich die Wahrnehmung von Bewertungen unterscheidet.

Nun zum Gegenstand der Online-Bewertung. Die vorgestellten Studien handeln meist von Online-Bewertungen auf Plattformen, die von Händlern bereitgestellt werden (z. B. *amazon* oder *ebay*), oder um Online-Bewertungen auf reinen Bewertungsplattformen (wie *Tripadvisor*, *Epinions* oder *Yelp*). Die Ergebnisse beziehen sich daher manchmal auf Communities mit angemeldeten Nutzern und pseudonymen Profilen oder Bewertern mit Klarnamen, manchmal auf öffentlich einsehbare anonyme Bewertungen. Die meisten Studien untersuchen Vertrauen in Online-Bewertungen zu einem bestimmten Produkt (hauptsächlich Elektronik oder aber Filme, Games und Nahrungsmittel) oder einer bestimmten Dienstleistung (hauptsächlich Hotels und Restaurants).

Die untersuchten Gegenstände sind also nicht besonders vielfältig. Der Fokus der Marketing-Forschung auf diese Gegenstände lässt sich eventuell mit der Profitorientierung angewandter Forschung erklären. Vertrauen in Bewertungen von

Dienstleistungen aus anderen gesellschaftlichen Teilbereichen sind bisher kaum erforscht. Es wäre daher interessant, die bisherige Forschung um Erkenntnisse über das Vertrauen in Online-Bewertungen zu erweitern, in denen besonders relevante Produkte und Dienstleistung bewertet werden. Zu Dienstleistungen aus dem Gesundheitssystem gibt es bisher z. B. nur drei Studien. Cox et al. (2017) untersuchen Vertrauen in Bewertungen von einem Medikament, Reimer und Benkenstein (2016) forschen über Bewertungen einer zahnärztliche Behandlung und Grabner-Kräuter und Waiguny (2015) behandeln Arztbewertungen.

Im letzten Abschnitt zum methodischen Vorgehen wird die Frage beantwortet, welche Menschen an den bisherigen Studien teilgenommen haben. In 29 Studien und damit im Großteil des Korpus wird auf Studenten als Teilnehmer zurückgegriffen. Ansonsten werden in 19 Studien Nutzer ausgewählter Bewertungsplattformen angeschrieben. Noch bilden pro Gig bezahlte Teilnehmer, sogenannte „Amazon Mechanical Turks“ die Ausnahme (nur bei Munzel, 2015). Die Größe der Teilnehmer-Samples ist bei Experimenten und standardisierten Befragungen etwa gleich. Die Anzahl der Teilnehmer reicht bei Experimenten von 96 (Willemsen et al., 2012) bis 1.207 (Flanagin & Metzger, 2013), im Mittel 311. Für standardisierte Befragungen werden 95 (Cheung et al., 2012) bis 1.219 (Hsiao et al., 2010) Menschen befragt, der Mittelwert ist 408. Die offenen Interviews untersuchen 9 (O'Reilly & Marx, 2011) bis 38 (Fileri, 2016) Teilnehmer, 20 im Mittel. 26 der Studien wurde in den USA durchgeführt, 17 in Asien (v.a. Taiwan, Singapur und China), zwölf in Europa (z. B. Niederlande, UK, Spanien, Deutschland) und eine in Südafrika.

So uneinig sich die Studien in der Wahl ihrer Messinstrumente zu sein scheinen, so einig sind sie sich in der Wahl ihrer Methode. Eine typische Studie des vorliegenden Korpus besteht aus einem Experiment. Für dieses Experiment lesen 300 US-amerikanische Studenten als Stimulus online eine Bewertung zu einer Digitalkamera. Diese Bewertung wurde vorher hinsichtlich eines Faktors in verschiedenen Ausprägungen manipuliert. Im Anschluss beantworten die Teilnehmer weiterhin online einige standardisierte Fragen zu ihrem Vertrauen in diese Bewertung. Die Beschreibung der bisher verwendeten Methoden anhand der drei Dimensionen zeigt, dass es bisher keine Studie gibt, in der allen Teilnehmern

dieselben natürlichen Bewertungen vorgelegt werden. Außerdem zeigt sich ein Fokus auf Marketing-relevante Bewertungsgegenstände und den US-amerikanischen Raum.

Die hier erfolgte Darstellung und Kritik des methodischen Vorgehens bildet den Abschluss der Unterkapitel über die Hintergründe der Studien. Nun sind alle relevanten Voraussetzungen bekannt, unter denen die Ergebnisse der Studien zustande gekommen sind. Daher kann jetzt der Hauptteil dieses Kapitels folgen, die Darstellung der Ergebnisse zu den Faktoren des Vertrauens in Online-Bewertungen.

## **2.8 ERGEBNISSE ZU DEN FAKTOREN**

In diesem Unterkapitel geht es um alle bisher untersuchten Faktoren, die mit dem Vertrauen in Online-Bewertungen zusammenhängen. Da die meisten Studien dieses Korpus einer quantitativen Forschungslogik folgen, konzentriere ich mich zunächst auf das Vorgehen bei der Deduktion der Faktoren. In den wenigen qualitativen Studien werden die induktiv ermittelten Faktoren der Forschungslogik entsprechend erst im Nachhinein erklärt. Der Übersicht halber führe ich hier aber sowohl die im Vorhinein geleisteten theoretischen Erklärungen deduktiv abgeleiteter Faktoren auf, als auch die im Nachhinein gebrauchten Erklärungen induktiv vorgefundener Faktoren.

Der Deduktion der Faktoren liegen verschiedene theoretische Ansätze zugrunde (siehe Tabelle 6). Einige Studien verzichten gänzlich auf eine dezidierte theoretische Fundierung der von ihnen ausgewählten erklärenden Variable. Die meisten Autoren, die sich um eine theoretische Basis bemühen, um Hypothesen über den Zusammenhang ihrer Variablen zu bilden, greifen auf einen der zwei „Platzhirsche [...] im Wald der Persuasionsforschung“ (Klimmt, 2011, S. 98) zurück. Das „Elaboration Likelihood Model“ (ELM) von Petty und Cacioppo (1986) wird in 13 Studien verwendet. Auch in der eWOM-Forschung wird hauptsächlich das ELM genutzt (Cheng & Ho, 2015). Dieses Zwei-Prozess-Modell erklärt ursprünglich, den Einfluss einer Mitteilung auf eine Einstellungsänderung des Empfängers dieser Mitteilung. In der vorgestellten Forschung wird es dafür verwendet, um zu erklären, wie der Leser zu einem Urteil über die Glaubwürdigkeit einer Online-Bewertung gelangt. Dafür wird die grundlegende Idee des Modells übernommen, dass ein Leser auf zwei verschiedene Weisen mit einer Mitteilung umgehen kann: Wenn der Leser

sehr involviert ist, nutzt er den „zentralen“ Weg, d. h. er setzt sich mit den in einer Mitteilungen enthaltenen Argumenten auseinander. Wenn der Leser nicht sehr involviert ist, dann verzichtet er auf diese „Elaboration“ und nimmt den „peripheren“ Weg. Er orientiert sich dafür an Merkmalen, die über den Inhalt der Mitteilung hinausgehen, z. B. an senderbezogenen Hinweisreizen (wie z. B. dem Expertenstatus eines Senders) oder an formalen Eigenschaften der Mitteilung (wie z. B. der Anzahl der Argumente). Schon bei der bereits erwähnten Literaturübersicht von Cheung und Thadani (2012) wird deutlich, dass eine der beiden Dual-Process-Theorien den meisten Studien als Grundlage dienen. Wie in Tabelle 5 klar wird, greifen die Autoren der Studien aber auch noch auf eine Vielzahl weiterer Theorien zurück. Bei den meisten theoretischen Erklärungen handelt es sich übrigens nur dem Namen nach um „Theorien“. Es sind Theorien sehr geringer Reichweite, die sich vielmehr als „Ansätze“ oder „Modelle“ bezeichnen lassen, da sie nur dazu dienen, einen sehr kleinen Ausschnitt sozialer Realität zu erklären.

*Tabelle 5: Theoretische Grundlagen der Studien (sortiert nach Häufigkeit)*

Ansatz/ Modell/ Theorie	Anzahl der Studien
Elaboration Likelihood Model (Petty & Cacioppo (1986)	13
Source Credibility (Louie & Obermiller, 2002; Ohanian, 1990; Tripp et al., 1994)	9
Attribution Theory	5
Perceptual Homophily (Gilly et al. 1998)	5
Heuristics (Metzger & Flanagin, 2000), Heuristic-Systematic Processing Model	6
Uncertainty Reduction (Berger & Calabrese, 1975)	4
Dual-Process Theory (Deutsch & Gerrard, 1955)	4
Warranting Principle (Walther & Parks, 2002)	4
Prominence - Interpretation theory	3
Signaling Theory (Donath, 2007)	3
Expectancy Violations Theory (Afifi & Burgoon, 2000, Burgoon & Miller, 1985)	2
Social Identity Theory	2
Social Information Processing	2
Informational Social Influence Theories (z. B. Deutsch & Gerard, 1955)	2
Social Influence Theory (Kelman, 1961)	2
Theory of Planned Behavior	2

Covariation Attribution Model (Kelley, 1967)	2
Persuasion Knowledge Model	2
ABC Model of Attitude	1
Accessibility-Diagnosticity Model	1
Assimilation Contrast Theory	1
Cognitive Fit Theory	1
Correspondence Theory	1
Cue Combination Literature	1
Culture Theory (Hofstede, 1980)	1
Ceception Theory	1
Inoculation Theory	1
Regulatory Focus Theory (Higgins, 1997, 2002)	1
Self-Perception Theory	1
Sense of Membership	1
Sense of Virtual Community (Blanchard, 2007; Tonteri et al., 2011)	1
Social Distance Theory	1
Source-Attractiveness Model (McGuire, 1985)	1
Theory of Cognitive Authority (Wilson, 1983)	1
Theory of Reasoned Action	1
Theory of Social Power	1
Theory of Temporal Contiguity Cue	1
Unified Framework of Credibility Assessment (Hilligoss & Rieh, 2008)	1

Die Fülle an Ansätzen zusammen mit der Beobachtung, dass viele Ansätze nur einmal verwendet wurden, zeigt, dass die Autoren in allen Richtungen nach Faktoren suchen. Das kann daran liegen, dass es keine theoretische Vorstellung darüber gibt, woher die Faktoren deduziert werden können. Darauf weisen Cheung et al. schon 2012 hin: „While the importance of review credibility for consumer behaviour has been well established, knowledge about the factors that affect online review credibility *from a theoretical perspective* remains scant.“ (S. 620, Hervorhebung von mir). Die Herleitung des untersuchten Faktors erscheint daher oft recht zufällig. Die Autoren der behandelten Studien weisen auch bei ihrer Begründung der Wahl des Faktors häufig auf „previous research“ hin, ohne das weiter zu explizieren.

Unabhängig von den theoretischen Erklärungen, die die Autoren der verschiedenen Studien für die Faktoren heranziehen, sollen diese Faktoren mit ihren Ergebnissen nachfolgend genauer ausgeführt werden. Wie bei der Vorstellung der

---

Methode der integrativen Literaturübersicht bereits erklärt, werde ich für die Darstellung der Ergebnisse die über verschiedene Studien hinweg erhobenen Faktoren zunächst systematisch zusammenführen. Um alle erhobenen Faktoren sortieren sowie die Ergebnisse miteinander vergleichen und synthetisieren zu können, wird hier also die Arbeit geleistet, die Faktoren mit Hilfe eines theoretischen Rahmens zu ordnen. Für die Synthese gruppiere ich ähnliche Faktoren in Gruppen. Diese Gruppen werden zur Systematisierung anschließend theoretisch geordnet.

Die Ordnung der bisher untersuchten Faktoren erfolgt nach deren funktionaler Äquivalenz. Ich orientiere mich also z. B. nicht an äquivalenten Namen der untersuchten Faktoren, sondern an der äquivalenten Bedeutung. So ordne ich einen Faktor, selbst wenn er von den Autoren mit demselben Namen bezeichnet wird, unterschiedlichen Faktorengruppen zu, wenn die Autoren unterschiedliche Konzepte damit messen. Ein Beispiel dafür ist der Faktor „information quality“, der bei Filieri (2015) und Filieri et al. (2015) etwas anderes misst als bei Tsao und Hsieh (2015), aber von allen Autoren gleich genannt wird. Mit funktionaler Äquivalenz meine ich auch, dass die Faktoren, denen dasselbe abstrakte Konzept zugrunde liegt, in einer gemeinsamen Gruppe eingeordnet werden. Die zugrundeliegenden abstrakten Konzepte ermittle ich anhand der Definitionen und Operationalisierungen der Faktoren.

In einzelnen Fällen gestaltet sich die Zuordnung einfach. Das ist so, wenn der Faktor nicht nur gleich heißt, sondern auch gleich gemessen wird. Für den Faktor „Bewertung der Bewertung“ benutzen z. B. alle Studien dieselben Items wie Cheung et al. (2009), wahrscheinlich weil teilweise dieselben Autoren beteiligt sind. D. h. auch, dass ich mich, anders als die bisherige Forschung, nicht am konkreten Gegenstand der Erscheinung des Faktors orientiere. Die abstrakten Konzepte liegen manchmal quer zum konkreten Gegenstand, in dem sich das Konzept zeigt. So kann sich z. B. der Faktor „Ähnlichkeit des Bewerter“ auf verschiedenen Ebenen manifestieren: in der Sprache der Bewertungen oder in Informationen im Bewerterprofil. In der bisherigen Forschung würde dieser Faktor in zwei Faktoren getrennt auftauchen. Als ein Faktor, der vom Bewertungstext herrührt und als ein zweiter Faktor, der vom Bewerterprofil herrührt. Ich hingegen halte das übergeordnete abstrakte Konzept der Ähnlichkeit des Bewerter für ein sinnvolleres

Unterscheidungskriterium als den konkreten Gegenstand, in dem es sich in einem bestimmten Fall zeigt. Daher ordne ich die Faktoren dementsprechend. Ein weiteres Beispiel dafür ist das Profilfoto, das als konkreter Gegenstand in mehreren Studien eine Rolle spielt, aber jeweils als Träger verschiedener abstrakter Konzepte fungiert. In einer Studie steht das Profilfoto als Faktor für das „Vertrauen in den Bewerter“ (McGloin et al., 2014), in einer anderen für die „Selbstauskunft des Bewerter“ (Xu, 2014), in einer dritten Studie wiederum für die „Kompetenz des Bewerter“ (Chiou et al., 2014). Auch hier ist es also sinnvoll, nicht alle Faktoren, bei denen das Profilfoto eine Rolle spielt, in einer Gruppe zusammenzufassen, sondern die Faktoren nach der darunter liegenden Erklärung für die Bedeutsamkeit des Profilfotos zu sortieren.

Eine Systematisierung muss abstrakt genug sein, um die verschiedenen Faktoren zusammenzufassen, aber konkret genug, um jede Gruppe aussagekräftig beschreiben zu können. Selbstverständlich geht die Zuteilung verschiedener Faktoren in eine Gruppe immer mit einer Vereinfachung einher. Was die Genauigkeit der Systematisierung aller Faktoren angeht, bleibt zu beachten, dass die Autoren auch innerhalb solcher von mir sortierten Faktorengruppen verschiedene Konstrukte nutzen und diese unterschiedlich operationalisieren. Überschneidungen zwischen diesen Faktorengruppen sind daher nicht auszuschließen (z. B. wenn bei der Messung des einen Faktors mit einer Skala ein gleiches Item benutzt wird, das der nächste Autor zur Abfrage eines anderen Faktors mit einer anderen Skala auch benutzt). Allgemein gilt, dass meine Abgrenzungen der Gruppen untereinander nur beschränkt passen. Wie jede Zusammenfassung sind auch die Grenzen dieser Gruppen nicht immer trennscharf. Es gibt Ungenauigkeiten zugunsten des Überblicks, z. B. an einigen Stellen inter-faktorielle oder intra-faktorielle Inkonsistenzen. Diese Einschränkungen bilden auch den Hintergrund vor dem die Ergebnisse zu interpretieren sind: Wenn die Ergebnisse innerhalb einer Faktorengruppe eindeutig sind oder sich widersprechen, kann das immer auch mit Unterschieden in der Konstruktion und Operationalisierung der einzelnen Faktoren derselben Gruppe zusammenhängen.

Die Benennung der Gruppen von gleichen Faktoren erfolgt, wenn möglich, mit einem passenden (auf Deutsch übersetzten) Begriff aus der bestehenden Literatur. So nutze ich z. B. den Begriff der „Übereinstimmung“, um den Faktor zu

beschreiben, der die tatsächliche oder wahrgenommene Übereinstimmung der einzelnen Bewertungen untereinander misst. „Übereinstimmung“ kann auch als allgemeine Heuristik verstanden werden, die auf verschiedene Bereiche angewandt wird. Eine Heuristik, d. h. Faustregel, nach der der Leser verfährt, kann als abstrakter Mechanismus quer zur hiesigen Sortierung der Gründe liegen. So kann sich z. B. die Bandwagon-Heuristik, die Orientierung an „Übereinstimmung“, zeigen, wenn für den Leser die *Übereinstimmung der Bewertungen* untereinander eine Rolle für sein Vertrauen spielt, oder auch dann, wenn der Leser wegen der *Übereinstimmung der Bewertung mit anderen Quellen* vertraut. Um diesen Zusammenhang zwischen den Faktoren deutlich zu machen, enthalten die Namen der Faktoren mit gleicher Heuristik auch den gleichen Begriff.

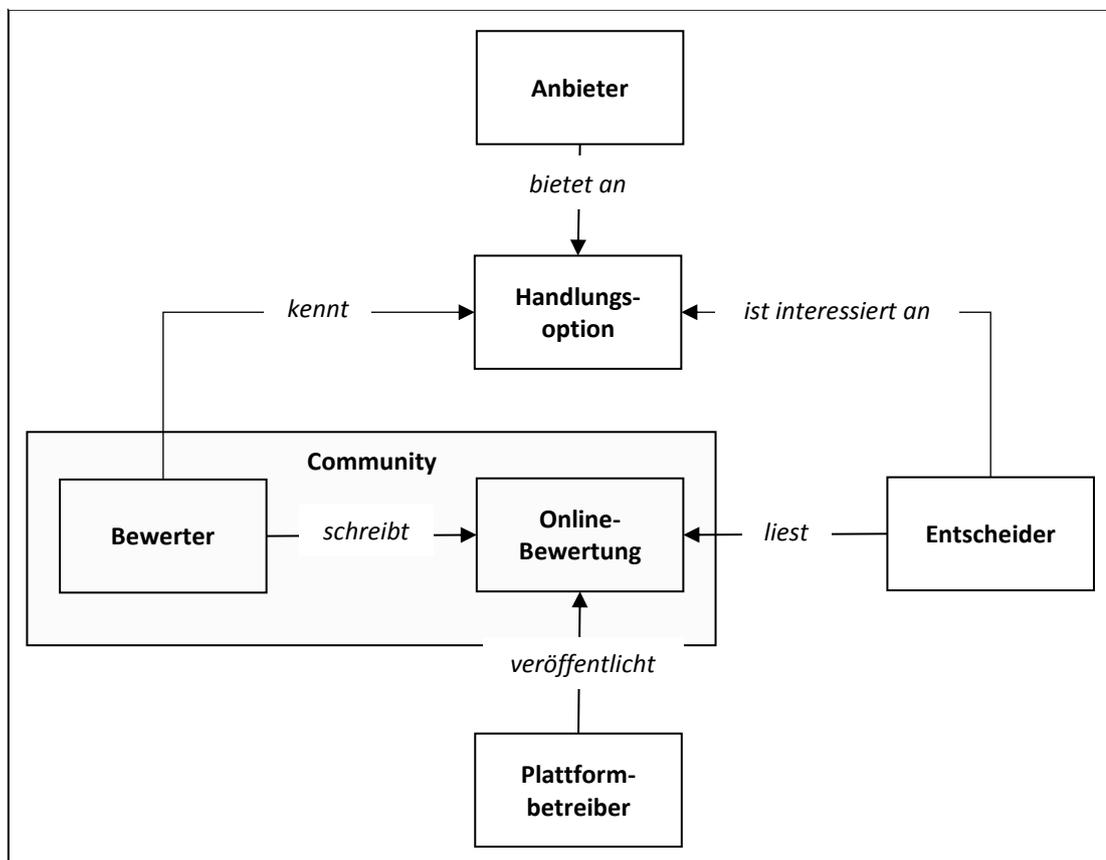
Nach der Synthetisierung und Benennung erfolgt die Systematisierung der Faktoren. So werden die Gruppen an Faktoren, die sich durch die funktionale Äquivalenz der Faktoren ergeben, geordnet. Die von Moran und Muzellec (2017) vorgeschlagene Systematisierung erweist sich aufgrund des atheoretischen Vorgehens der Autoren leider als unbrauchbar für mein Anliegen. Die Autoren machen „Community, Competence, Content, and Consensus“ als die Faktoren aus, die für die Glaubwürdigkeit von eWOM wichtig sind. Damit stellen sie einen Akteur (Community) gleich mit einer Eigenschaft des Bewerter (Competence), dem Inhalt der Bewertung (Content) und der Übereinstimmung verschiedener Bewertungen untereinander (Consensus). Die eingängige Alliteration im Titel des Modells tröstet leider nicht über die willkürlich erscheinende Mischung dieser auf verschiedenen Ebenen gelagerten Kategorien hinweg. Cheung und Thadani (2012) dagegen ordnen in ihrer Literaturübersicht die untersuchten Elemente nach Bewerter, Bewertung, Leser und Effekt. Auch sie mischen hier verschiedene Ebenen. Damit nähern sie sich aber der gängigen Vorgehensweise vieler Autoren an. Denn in den meisten Studien werden die Faktoren dort verortet, wo sie sich manifestieren, z. B. in den Eigenschaften des Bewerter oder der Bewertung.

Borchers (2016) bietet mit seinem theoretisch fundierten Modell zum Ablauf von Online-Bewertungs-Prozessen noch mehr an. Er identifiziert sowohl alle relevanten Akteure als auch alle relevanten Objekte, die bei Online-Bewertungen eine Rolle spielen (siehe Abbildung 6). Der Prozess beginnt mit dem (1) „Anbieter“,

der eine (2) Handlungsoption anbietet. Der (3) Bewerter hat bereits Erfahrung mit dieser Handlungsoption gemacht. Er verfasst darüber eine (4) Online-Bewertung. Der Bewerter und die Online-Bewertung sind beide Teil einer (5) Community. Die vom Bewerter verfasste Online-Bewertung über die Handlungsoption des Anbieters wird dann von einem (6) Plattformbetreiber veröffentlicht. So kann der (7) Entscheider, der an der Handlungsoption interessiert ist, aber noch keine Erfahrung damit gemacht hat, die Online-Bewertung auf der Plattform lesen. Im Modell ist bisher nicht vorgesehen, dass der Entscheider ebenfalls Teil der Community ist – doch denkbar ist das.

Im Rahmen dieser Arbeit bezeichne ich die „Handlungsoption“ von nun als „Bewertetes“ und den „Entscheider“ als „Leser“. Neben den in diesem Modell aufgeführten Elementen gibt es auch noch (8) externe Akteure, die nicht direkt am Online-Bewertungsprozess beteiligt sind. Das Modell von Borchers (2016) erlaubt es mit dieser kleinen Ergänzung, alle bisher untersuchten Faktorengruppen demjenigen Akteur bzw. Element zuzuordnen, auf den sich der Faktor bezieht.

Abbildung 5: Online-Bewertungs-Prozess



Die erhobenen Faktoren lassen sich zu 54 Faktoren(gruppen) synthetisieren. Diese 54 Faktorengruppen ordne ich den oben genannten acht Akteuren bzw. Elementen des Online-Bewertungsprozesses zu. Zur verbesserten Anschaulichkeit und Übersichtlichkeit stelle ich die Ergebnisse zu den Faktoren aus den verschiedenen Studien in Tabellenform dar. In der linken Spalte stehen der Name des Faktors sowie eine Erklärung des Faktors, ggf. mit Beispielen der Messung oder Operationalisierung. In der mittleren Spalte findet sich das Ergebnis der Untersuchung. Ein Plus („+“) steht für einen positiven Zusammenhang des Faktors mit Vertrauen, ein Minus („-“) steht für einen negativen Zusammenhang, eine Null („0“) dafür, dass kein Zusammenhang gefunden wurde. Je nach Forschungsanlage kann dieser Zusammenhang mal als direkter Einfluss des Faktors auf Vertrauen verstanden werden, mal als Korrelation. In der mittleren Spalte ergänze ich ggf. noch die Interaktionen mit anderen Faktoren. In der rechten Spalte werden dann die Studien aufgeführt, aus denen die Ergebnisse stammen, gekennzeichnet durch Autor und Jahr. Nach der Tabelle folgt jeweils eine Zusammenfassung der Ergebnisse zum direkten Zusammenhang. Hier wird kurz gezeigt, welche der Faktoren bisher nur einmal untersucht wurden. Daran schließt sich eine Darstellung der eindeutigen und der widersprüchlichen Ergebnisse an von den Faktoren, die mehrmals untersucht wurden.

In der ersten Tabelle werden die Faktoren aufgeführt, die sich auf den Anbieter der Handlungsoption beziehen (siehe Tabelle 6).

*Tabelle 6: Anbieter-bezogene Faktoren*

<b>Faktor</b>	<b>Zusammenhang</b>	<b>Autor (Jahr)</b>
<b>Markeneinstellung des Lesers</b>	0	Jensen & Yetgin (2017)
Einstellung des Lesers gegenüber der Marke des Bewerteten		Jensen et al. (2013)

Bisher wurde nur der Faktor der *Markeneinstellung des Lesers* untersucht. Beide Studien kommen hier zu dem Schluss, dass es keinen Zusammenhang zwischen der Einstellung des Lesers gegenüber der Marke des Bewerteten und dem Vertrauen des

Lesers in die Bewertung gibt. Es stellt sich die Frage, ob es noch weitere Vertrauensfaktoren gibt, die sich auf den Anbieter beziehen.

Auch in Bezug auf das Bewertete, d. h. die Handlungsoption beim Modell von Borchers (2016), wurde bisher nur ein Faktor untersucht (siehe Tabelle 7).

*Tabelle 7: Bewertetes-bezogene Faktoren*

<b>Faktor</b>	<b>Zusammenhang</b>	<b>Autor (Jahr)</b>
<b>Typ des Bewerteten</b>	bei Erfahrungsgut größer	Bae & Lee (2011)
	Bewertetes ist ein	Tsao & Hsieh (2015)
Erfahrungs- oder Suchgut	bei Erfahrungsgut höherer Zusammenhang wenn Plattformbetreiber Konsument statt Verkäufer ist, keine Interaktion bei Suchgut	Bae & Lee (2011)
	keine Interaktion mit Qualität der Bewertung und Betreiber der Plattform	Tsao & Hsieh (2015)

Der *Typ des Bewerteten* wird in beiden Studien nicht als Faktor mit einem direkten Zusammenhang modelliert, sondern als moderierende Variable. In zwei Studien kann gezeigt werden, dass das Vertrauen in Online-Bewertungen von Erfahrungsgütern generell höher ist als das Vertrauen in Online-Bewertungen von Suchgütern. Außerdem wurden Interaktionen mit einigen anderen Faktoren überprüft.

Faktoren, die sich auf den Bewerter beziehen, wurden bisher sehr häufig untersucht (siehe Tabelle 8). Bei den neun Faktoren geht es um den Zusammenhang von Vertrauen mit den vom Leser wahrgenommenen Eigenschaften des Bewerter. Die Wahrnehmung des Lesers kann aus verschiedenen Quellen gespeist werden. Der Leser enthält z. B. Informationen vom Profil des Bewerter, das auf der Plattform zur Verfügung gestellt wird. Außerdem kann der Leser von der Bewertung selbst auf Eigenschaften des Bewerter schließen. In der Bewertung kann der Bewerter persönliche Informationen direkt anbringen, z. B. indem er seine Hobbys erwähnt. Im Bewertungstext können aber auch indirekt Alter oder Geschlecht des Bewerter deutlich werden.

Tabelle 8: Bewerter-bezogene Faktoren

Faktor	Zusammenhang	Autor (Jahr)
<b>Ähnlichkeit des Bewerter</b> Bewerter hat hohe Ähnlichkeit mit Leser, z. B. bezüglich Geschlecht, Alter, Beruf, Geschmack, oder „very similar“	+	Ayeh et al. (2013) Llamero (2014) Shamhuyenzva et al. (2016)
	+ stärker bei Involviertheit des Lesers	Racherla et al. (2012) Smith et al. (2005)
	0	Lin & Xu (2017)
	0 aber stärker bei Involviertheit des Lesers	Lis (2013)
<b>Attribuierung</b> Leser schreibt die Bewertung dem Bewerteten zu (external) statt dem Bewerter (internal)	+	Qiu et al. (2012)
<b>Erfahrung des Bewerter</b> Bewerter hat viel Erfahrung mit Bewerten, z. B. viele eigene Bewertungen	+	Filieri (2016)
	+ wenn Übereinstimmung der Bewertung mit anderen Bewertungen und Vertrautheit des Lesers mit der Plattform hoch sind	van der Heide & Lim (2016)
	keine Interaktion mit Eingebundenheit des Bewerter	Lim & van der Heide (2015)
	keine Interaktion mit Vertrautheit des Lesers mit der Plattform	Lim & van der Heide (2015)
	kein größerer Zusammenhang als Eingebundenheit des Bewerter	van der Heide & Lim (2016)
<b>Identifikation des Bewerter</b> Bewerter gibt Informationen über sich selbst, z. B. Klarname, Wohnort, Vorlieben	+	Kusumasondjaja et al. (2012) Xie et al. (2011)
	<b>Kompetenz des Bewerter</b> Bewerter ist kompetent in Bezug auf das Bewertete, z. B. hohe formale Bildung, viele eigene Bewertungen im Bereich	+

des Bewerteten oder „competent“, „experienced“, „well-qualified“	+ stärker bei Involviertheit des Lesers	Lis (2013)
	+ stärker bei gutem Bewerterstatus	Wang, X. et al. (2015)
	-	Willemsen et al. (2012)
<b>Profilfoto des Bewerter</b> Bewerter hat ein Profilfoto (im Gegensatz zu keinem Profilfoto) oder einen menschlichen Avatar (im Gegensatz zu einem Zeichentrick-Dinosaurier)	+	Filieri (2016) McGloin et al. (2014) Xu (2014)
	+ bei physischer Attraktivität des Profildfotos	Lin & Xu (2017)
<b>Soziale Nähe des Bewerter</b> Leser spürt soziale Nähe mit dem Bewerter, z. B. hätte er Bewerter gerne als Freund oder würde gerne mit ihm arbeiten	0	Lin & Xu (2017)
<b>Unabhängigkeit des Bewerter</b> Bewerter wird nicht von Anbieter beeinflusst, z. B. durch Sponsoring, oder „will keep promises“	+	Dickinger (2011) Filieri (2016) Hsiao et al. (2010) Nekmat & Gower (2012) Jensen & Yetgin (2017) O'Reilly et al. (2016)
	bezogen auf Offenlegung der Einflussnahme des Anbieters durch den Bewerter: - bei prominenter Offenlegung, + bei nichtprominenter Offenlegung, umgekehrt bei negativer Valenz der Bewertung	Jensen & Yetgin (2017)
<b>Vertrauenswürdigkeit des Bewerter</b> Bewerter ist glaub- bzw. vertrauenswürdig, z. B. „dependable“, „honest“, „reliable“	+	Cheung et al. (2009) Cheung et al. (2012) Lis (2013) Luo et al. (2014) Shamhuyenhazva et al. (2016)
	keine Interaktion mit Kompetenz des Lesers	Cheung et al. (2012)
	keine Interaktion mit Involviertheit des Lesers	Cheung et al. (2012)
	geschwächt durch Involviertheit des Lesers	Lis (2013)

keine Interaktion mit Kultur des Lesers	Luo et al. (2014)
keine Interaktion mit Zugehörigkeit des Lesers	Luo et al. (2014)
niedrige Vertrauenswürdigkeit stärkt den Zusammenhang von Argumentation des Bewerter	Luo et al. (2013)
hohe Vertrauenswürdigkeit stärkt den Zusammenhang von Ausführlichkeit	Luo et al. (2013)
moderiert den Einfluss von Ähnlichkeit des Bewerter	Shamhuyenzva et al. (2016)
-	Filieri et al. (2015)

Drei Bewerter-bezogene Faktoren wurden bisher nur in einer Studie untersucht. Für die *Attribuierung des Bewerter* und die *Erfahrung des Bewerter* konnte dabei ein positiver Zusammenhang mit Vertrauen festgestellt werden. Die *Soziale Nähe des Bewerter* hängt hingegen nicht mit Vertrauen zusammen. Die anderen Faktoren wurden mehrmals untersucht. Bei drei dieser Faktoren herrscht über verschiedene Studien hinweg Uneinigkeit über den Zusammenhang. Während in mehreren Studien ein positiver Zusammenhang von Vertrauen mit der *Kompetenz des Bewerter* und der *Vertrauenswürdigkeit des Bewerter* festgestellt wurde, wird in je einer Studie ein negativer Zusammenhang entdeckt. Die *Ähnlichkeit des Bewerter* hängt laut drei Studien positiv mit Vertrauen zusammen, in einer Studie konnte wiederum gar kein Zusammenhang gefunden werden. Wiederum bei drei Faktoren sind sich die Studien allerdings einig. Sowohl die *Identifikation des Bewerter*, als auch das *Profilfoto des Bewerter*, sowie die *Unabhängigkeit des Bewerter* hängen positiv mit Vertrauen zusammen. Während für ein Drittel der Faktoren der Zusammenhang eindeutig ist, sind die Ergebnisse bei einem Drittel widersprüchlich. Zu diesem und zum letzten Drittel der nur einmal erforschten Bewerter-bezogenen Faktoren ist weitere Forschung nötig, um die Widersprüche zu erklären bzw. die ersten Ergebnisse zu validieren.

In der bisherigen Forschung wurden Faktoren, die sich auf die Bewertung beziehen, am häufigsten untersucht (Tabelle 9). Hier geht es um den Zusammenhang von Eigenschaften der Bewertung mit Vertrauen.

*Tabelle 9: Bewertungs-bezogene Faktoren*

<b>Faktor</b>	<b>Zusammenhang</b>	<b>Autor (Jahr)</b>
<b>Argumentation der Bewertung</b> Bewertung ist gut begründet	+	Cheung et al. (2009) Cheung et al. (2012) Luo et al. (2014) Racherla et al. (2012)
	+ stärker bei niedriger Vertrauenswürdigkeit des Bewerter	Luo et al. (2013)
	+ stärker bei hoher Zugehörigkeit des Lesers	Luo et al. (2015)
	+ schwächer bei skeptischer Einstellung des Lesers gegenüber Online-Bewertungen	Reimer & Benkenstein (2016)
	+ keine Interaktion mit Kompetenz des Lesers	Cheung et al. (2012)
	+ keine Interaktion mit Involviertheit des Lesers	Cheung et al. (2012) Racherla et al. (2012)
	keine Interaktion mit Kultur des Lesers	Luo et al. (2014)
	<b>Ausführlichkeit der Bewertung</b> Bewertung ist ausführlich und enthält viele Informationen	+
+ bei moderatem Umfang (- bei zu hohem oder zu niedrigem Umfang)		Furner et al. (2016)
0		Tsang & Prendergast (2009)
hängt von der Art der Informationen ab		Filieri (2016)
<b>Ausgewogenheit der Bewertung</b> Bewertung enthält sowohl positive als auch negative Erfahrungen mit dem Bewerteten	+	Cheung et al. (2012) Filieri (2016) Lee & Koo (2012) Luo et al. (2014)
	+ mediiert durch positive Enttäuschung der Spracherwartung	Jensen et al. (2013)
	+ stärker bei hoher Kompetenz des Lesers	Cheung et al. (2012) Lee & Koo (2012)
	+ stärker bei niedriger Involviertheit des Lesers	Cheung et al. (2012)

	+ stärker bei individualistischer Kultur des Lesers	Luo et al. (2014)
	+ stärker bei hoher Zugehörigkeit des Lesers	Luo et al. (2015)
	0	Cheung et al. (2009)
<b>Emotionalität der Bewertung</b>	+	Wang, S. et al. (2015)
Bewertung enthält z. B. viele emotionsgeladene Wörter wie „lieben“, viele Großbuchstaben, Emoticons und Ausrufezeichen, keine quantifiziert beobachtbare Kriterien	-	Hong & Park (2012) Jensen et al. (2013) Llamero (2014)
	- schwächer bei hoher Zugehörigkeit des Lesers	Luo et al. (2015)
	+ bei hoher Anzahl der Bewertungen	Grabner-Kräuter & Waiguny (2015)
	- bei niedriger Anzahl der Bewertungen	
<b>Fotos des Bewerteten</b>	+	Filieri (2016) Llamero (2014)
Bewertung enthält Fotos des Bewerteten		
<b>Qualität der Bewertung</b>	+	Filieri (2015) Filieri (2016) Filieri et al. (2015)
Bewertung hat eine gute Qualität		
	+ nur für Bewertungen mit positiver Valenz	Tsao & Hsieh (2015)
	+ stärker bei Unabhängigkeit der Plattform	Tsao & Hsieh (2015)
	+ keine Interaktion mit Typ des Bewerteten	Tsao & Hsieh (2015)
	0	Wang, S. et al. (2015) Willemsen et al. (2012)
<b>Rechtschreibung und Grammatik der Bewertung</b>	0 bezogen auf hohe lexikalische Komplexität	Jensen et al. (2013)
Rechtschreibung, Grammatik und Ausdruck der Bewertung, Bewertung enthält z. B. orthografische Fehler, Fachbegriffe und Fremdwörter, oder ist „easy to read“, „well written“	+ bezogen auf korrekte Rechtschreibung	McGloin et al. (2014) O'Reilly & Marx (2011)
	0 bezogen auf korrekte Rechtschreibung	Cox et al. (2017)
	0 bezogen auf Rechtschreibfehler bei niedrigem allgemeinen Vertrauen des Lesers	Cox et al. (2017)
	- bei typografischen Fehlern (vs. orthografischen) bei hohem allgemeinen Vertrauen des Lesers	Cox et al. (2017)

<b>Übereinstimmung der Zahlen- und Textbewertung</b> Bewertung im Kommentartext stimmt überein mit Zahlenbewertung	+	Tsang & Prendergast (2009)
<b>Unmittelbarkeit der Bewertung</b> Bewerter hat die Bewertung während (statt nach) der Nutzung des Bewerteten geschrieben	+ bei niedrigem Personal Sense of Power des Lesers, - bei hohem Personal Sense of Power des Lesers	Wu et al. (2017)
<b>Valenz der Bewertung (negativ)</b> Bewertung ist negativ	+	Lee & Koo (2012)
	+ stärker bei statistischer (statt narrativer) Form der Bewertung	Hong & Park (2012)
	+ bei Bewertetem des Typs Erfahrungsgut, 0 bei Bewertetem des Typs Vertrauensgut	Pan & Chiou (2011)
	+ bei Identifikation des Bewerter, 0 bei keiner Identifikation des Bewerter	Kusumasondjaja et al. (2012)
	+ bei niedriger Eingebundenheit des Bewerter, 0 bei hoher Eingebundenheit des Bewerter	Pan & Chiou (2011)
	keine Interaktion mit Motivation des Lesers	Lee & Koo (2012)
	0	Cheung et al. (2009) Willemsen et al. (2012)
<b>Valenz der Bewertung (positiv)</b> Bewertung ist positiv	+	Lim & van der Heide (2015) Lin & Xu (2017)
	+ bei hoher Qualität der Bewertung und Unabhängigkeit der Plattform	Tsao & Hsieh (2015)
	+ stärker für Bewertetes des Typs Vertrauensgut als für Bewertetes des Typs Suchgut	Tsao & Hsieh (2015)
	Interaktion mit Unabhängigkeit des Bewerter	Nekmat & Gower (2012)

Interaktion mit Anzahl der Bewertungen	O'Reilly & Marx (2011)
keine Interaktion mit Übereinstimmung der Bewertung mit anderen Bewertungen	Munzel (2015)
keine Interaktion mit statistischer oder narrativer Form der Bewertung	Hong & Park (2012)
-	Filieri (2016)

Von den elf Bewertungs-bezogenen Faktoren wurden zwei Faktoren erst einmal untersucht. Sowohl für die *Übereinstimmung der Zahlen- und Textbewertung* als auch für die *Unmittelbarkeit der Bewertung* findet jeweils eine Studie einen positiven Zusammenhang mit Vertrauen. Über mehrere Studien hinweg wird auch der *Argumentation der Bewertung* und den *Fotos des Bewerteten* ein positiver Zusammenhang mit Vertrauen attestiert. Für den Großteil der untersuchten Faktoren gibt es allerdings widersprüchliche Ergebnisse. Für die *Ausführlichkeit der Bewertung* sowie die *Rechtschreibung und Grammatik der Bewertung* wird mal ein positiver, mal ein negativer und mal gar kein Zusammenhang mit Vertrauen festgestellt. Bei den zwei Faktoren *Ausgewogenheit der Bewertung* und *Qualität der Bewertung* finden mehr Studien einen positiven als keinen Zusammenhang. Für die *Emotionalität der Bewertung* wird über einige Studien hinweg ein negativer Zusammenhang entdeckt, aber einmal auch ein positiver. Die *Valenz der Bewertung* wird als Faktor sehr häufig erforscht. In dieser Darstellung unterscheide ich zwischen den Studien, bei welchen die Autoren denken, dass eine negative Valenz der Bewertung stärker mit Vertrauen zusammenhängt als eine positive Valenz, und den Studien, die den Zusammenhang genau andersherum modellieren. Selbst wenn die Ergebnisse zu den beiden Faktoren jeweils eindeutig wären, würden sich die Ergebnisse also hinsichtlich des Zusammenhangs von Valenz und Vertrauen widersprechen. Doch die Ergebnisse sind nicht einmal innerhalb der Faktorengruppen eindeutig. Während die meisten Studien einen positiven Zusammenhang mit Vertrauen jeweils für die negative bzw. für die positive Valenz einer Bewertung feststellen, sieht eine Studie keine bzw. einen negativen Zusammenhang. Hinsichtlich der meisten Bewertungs-bezogenen Faktoren herrscht

also bisher Unklarheit, die nur in weiteren empirischen Studien aufgelöst werden kann.

Bewerter und Bewertung sind laut Borchers (2016) Teil der Community. Bisher wurden fünf Faktoren untersucht, die sich auf die Community beziehen (siehe Tabelle 10).

*Tabelle 10: Community-bezogene Faktoren*

<b>Faktor</b>	<b>Zusammenhang</b>	<b>Autor (Jahr)</b>
<b>Anzahl der Bewertungen</b> hohe Gesamtanzahl an Bewertungen oder Bewertern, z. B. 357 statt 1 Bewertungen, 30 statt 3 Bewerter oder „many members“	+	Flanagin & Metzger (2013) Hsiao et al. (2010) O'Reilly & Marx (2011)
	+ stärker bei Identifikation des Lesers schwächt den Zusammenhang von Nicht-Emotionalität	Flanagin & Metzger (2013) Grabner-Kräuter & Waiguny (2015)
<b>Bewertung der Bewertung</b> Community bewertet die Bewertung positiv, z. B. durch die Markierung als „hilfreich“, „highly rated by other members“	+	Cheung et al. (2009) Lis (2013) Llamero (2014) Luo et al. (2014) Shamhuyenzhanva et al. (2016)
	+ schwächer bei Involviertheit des Lesers	Lis (2013)
	+ stärker bei kollektivistischer Kultur des Lesers	Luo et al. (2014)
	+ stärker bei Zugehörigkeit des Lesers	Luo et al. (2014)
<b>Eingebundenheit des Bewerter in Community</b> Bewerter ist stark in die Community eingebunden, hat z. B. viele Freunde oder Follower	+	Xu (2014)
	bei hoher Eingebundenheit keine Interaktion mit Valenz, bei niedriger Eingebundenheit sind negative Bewertungen vertrauenswürdiger als positive	Pan & Chiou (2011)
	keine Interaktion mit der Erfahrung des Bewerter	Lim & van der Heide (2015)
	keine Interaktion mit der Vertrautheit des Lesers mit der Plattform	Lim & van der Heide (2015)
	-	Pan & Chiou (2011)

<b>Übereinstimmung mit anderen Bewertungen</b> Bewertung stimmt mit anderen Bewertungen und Kommentaren überein, ist z.B. „consistent“, „similar“ oder „seem to say the same thing“	+	Cheung et al. (2009) Cheung et al. (2012) Doh & Hwang (2009) Filieri (2016) Hong & Park (2012) Llamero (2014) Luo et al. (2014) van der Heide & Lim (2016)	
	+ stärker bei kollektivistischer Kultur des Lesers	Luo et al. (2014)	
	+ bei hoher Vertrautheit des Lesers mit der Plattform ist das Vertrauen nur dann hoch, wenn die Erfahrung des Bewerter mit dem Bewerteten und die Übereinstimmung der Bewertung mit anderen Bewertungen hoch sind	van der Heide & Lim (2016)	
	+ keine Interaktion mit Kompetenz des Lesers	Cheung et al. (2012) Doh & Hwang (2009)	
	+ keine Interaktion mit Involviertheit des Lesers	Cheung et al. (2012) Doh & Hwang (2009)	
	keine Interaktion mit Unabhängigkeit des Bewerter	Jensen & Yetgin (2017)	
	+ keine Interaktion mit Zugehörigkeit des Lesers	Luo et al. (2014)	
	-	Munzel (2015)	
	<b>Zugehörigkeit des Lesers</b> Leser fühlt sich der Community zugehörig, z. B. „I feel membership to this community“ oder „I feel that I belong“	+	Filieri (2016) Hsu et al. (2016)
		keine Interaktion mit Vertrauenswürdigkeit des Bewerter	Luo et al. (2015)
keine Interaktion mit Übereinstimmung mit anderen Bewertungen		Luo et al. (2015)	
stärkt den Zusammenhang von Ausgewogenheit		Luo et al. (2015)	
stärkt den Zusammenhang von Bewertung der Bewertung		Luo et al. (2015)	
schwächt den Zusammenhang von Nicht-Emotionalität		Luo et al. (2015) Luo et al. (2015)	

Alle Community-bezogenen Faktoren wurden in mehr als einer Studie untersucht. Einig sind sich die Studien bezüglich des positiven Zusammenhangs von Vertrauen mit der *Anzahl der Bewertungen*, der *Bewertung der Bewertung* und der *Zugehörigkeit des Lesers zur Community*. Hinsichtlich des Zusammenhangs von Vertrauen und der *Eingebundenheit des Bewerter in Community* gibt es widersprüchliche Ergebnisse. Die *Übereinstimmung der Bewertung* mit anderen Bewertungen hängt in sehr vielen Studien positiv mit Vertrauen zusammen, in einer Studie allerdings negativ.

In einigen Studien werden auch Faktoren untersucht, die sich auf die Plattform beziehen, auf der die Online-Bewertung veröffentlicht wird (siehe Tabelle 11). Mit Plattform-bezogen ist gemeint, dass der untersuchte Faktor die Plattform als Ursprung hat, es ihn also ohne die Plattform nicht gäbe. Die von der Plattform bereitgestellte Information kann sich selbst wiederum z. B. auf den Bewerter oder die Bewertung beziehen.

Tabelle 11: Plattform-bezogene Faktoren

Faktor	Zusammenhang	Autor (Jahr)
<b>Bewerterstatus</b> Plattform weist den Status des Bewerter als gut aus, z. B. „Top Reviewer“	+	Wang, X. et al. (2015)
<b>Qualität der Plattform</b> Plattform hat eine gute Qualität, z. B., „well-organized“, „guarantees user privacy“	+	Filieri et al. (2015) Hsiao et al. (2010)
<b>Vertrauenswürdigkeit der Plattform</b> Plattform ist glaub- bzw. vertrauenswürdig oder z. B. „trustworthy“, „believable“, „credible“	+	Hsiao et al. (2010) Lee et al. (2011)
<b>Vertrautheit mit Plattform</b> Leser kennt Plattform oder ist mit ihr vertraut	+ + stärkt den Zusammenhang, bei hoher Erfahrung des Bewerter mit Bewerten und hoher Übereinstimmung der Bewertung mit anderen Bewertungen	Casaló et al. (2015) van der Heide & Lim (2016)

	keine Interaktion mit Eingebundenheit des Bewerter in Community und Erfahrung des Bewerter mit Bewertungen	Lim & van der Heide (2015)
<b>Warnhinweis der Plattform</b>	-	Munzel (2015)
Plattform weist darauf hin, dass die Bewertung gefälscht sein könnte		

Von den Plattform-bezogenen Faktoren wurden die meisten nur einmal untersucht, wenige zweimal. Je eine Studie findet für den *Bewerterstatus* und die *Vertrautheit mit der Plattform* einen positiven und für den Warnhinweis der Plattform einen negativen Zusammenhang. Über zwei Studien hinweg sind sich die Autoren einig über einen positiven Zusammenhang von Vertrauen mit der *Qualität der Plattform* und der *Vertrauenswürdigkeit der Plattform*. Hier scheint es sinnvoll, die Ergebnisse mit einigen weiteren Studien zu überprüfen.

Faktoren, die sich auf den Leser beziehen wurden sehr häufig untersucht (siehe Tabelle 12).

Tabelle 12: Leser-bezogene Faktoren

Faktor	Zusammenhang	Autor (Jahr)
<b>Allgemeines Vertrauen des Lesers</b> Leser hat allgemein ein hohes Vertrauen	+ bei hohem allgemeinem Vertrauen des Lesers Zusammenhang schwächer bei typografischen als bei orthografischen Fehlern, keine Interaktion bei niedrigem allgemeinem Vertrauen	Cox et al. (2017)
<b>Alter des Lesers</b> Alter des Lesers	0	Jensen & Yetgin (2017) Reimer & Benkenstein (2016) Wang, X. et al. (2015)
<b>Bildung des Lesers</b> formale Bildung des Lesers, z. B. Schul- oder Studienabschluss	0	Cox et al. (2017) Reimer & Benkenstein (2016)

<b>Einstellung des Lesers gegenüber Online-Bewertungen</b>	0	Grabner-Kräuter & Waiguny (2015) Jensen et al. (2013) Qiu et al. (2012)
Leser ist skeptisch gegenüber Online-Bewertungen	-	Reimer & Benkenstein (2016)
	schwächt den Zusammenhang mit Argumentation des Bewerter	Reimer & Benkenstein (2016)
<b>Erfahrung des Lesers mit Online-Bewertungen</b>	0	Filieri et al. (2015)
Leser hat viel Erfahrung mit dem Lesen oder Schreiben von Online-Bewertungen	+	López & Sicilia (2014)
	verstärkt den Zusammenhang von Warnhinweis	Munzel (2015)
<b>Erfahrung des Lesers mit Online-Shopping</b>	0	Bae & Lee (2011)
Leser hat viel oder wenig Erfahrung mit Online-Shopping		
<b>Ethnizität<sup>8</sup> des Lesers</b>	0	Xu (2014)
Ethnizität des Lesers		
<b>Fokus des Lesers</b>	keine Interaktion mit Valenz der Bewertung	Lee & Koo (2012)
Fokus des Lesers beim Lesen der Bewertung, d. h. „prevention“ vs. „promotion“		
<b>Geschlecht des Lesers</b>	0	Reimer & Benkenstein (2016) Wang, X. et al. (2015) Xu (2014)
männliches oder weibliches Geschlecht des Lesers		
<b>Internetnutzung des Lesers</b>	0	Wang, X. et al. (2015)
Häufigkeit der Internetnutzung des Lesers		

8 Zur Übersetzung von „race“ erklärt Dietrich (2018): „Die migrationshistorischen Unterschiede zwischen Deutschland und den USA haben verschiedene Sprachkonventionen in den Wissenschaften hervorgebracht: *Race* als analytischer Begriff bezieht sich auf eine andere historische Konstellation und ist anders besetzt als ‚Rasse‘ in Deutschland, weswegen die deutschsprachige Intersektionalitätsforschung den Begriff durch „Ethnizität“ ersetzt.“ (S. 8). Diesem Vorgehen folge ich in meiner Arbeit.

<b>Involviertheit des Lesers</b> hohe Involviertheit des Lesers bei der Suche nach dem Bewerteten oder beim Umgang mit Bewertungen, z. B. „greatly involved“, „interested“, „very important“	0	Jensen & Yetgin (2017) Jensen et al. (2013) Nekmat & Gower (2012) Reimer & Benkenstein (2016) Xu (2014)
	stärkt den Zusammenhang von Kompetenz des Bewerter	Lis (2013)
	stärkt den Zusammenhang von Ähnlichkeit des Bewerter	Lis (2013) Racherla et al. (2012)
	niedrige Involviertheit stärkt den Zusammenhang von Ausgewogenheit	Cheung et al. (2012)
	schwächt den Zusammenhang von Vertrauenswürdigkeit des Bewerter	Lis (2013)
	schwächt den Zusammenhang von Bewertung der Bewertung	Lis (2013)
	keine Interaktion mit Vertrauenswürdigkeit des Bewerter	Cheung et al. (2012)
	keine Interaktion mit Übereinstimmung der Bewertung mit anderen Bewertungen	Cheung et al. (2012) Doh & Hwang (2009)
	keine Interaktion mit Argumentation	Racherla et al. (2012) Cheung et al. (2012)
	<b>Kultur des Lesers</b> individualistische vs. kollektivistische Kultur des Lesers	individualistische Kultur stärkt den Zusammenhang mit Ausgewogenheit der Bewertung
kollektivistische Kultur stärkt Zusammenhang von Übereinstimmung mit anderen Bewertungen		Luo et al. (2014)
kollektivistische Kultur stärkt Zusammenhang von Bewertung der Bewertung		Luo et al. (2014)

	keine Interaktion mit Argumentation des Bewerterers	Luo et al. (2014)
	keine Interaktion mit Vertrauenswürdigkeit des Bewerterers	Luo et al. (2014)
<b>Kompetenz des Lesers</b> hohe Kompetenz des Lesers bezüglich des Bewerteten	0	Bae & Lee (2011) Dickinger (2011) Flanagin & Metzger (2013) Wang, X. et al. (2015) Willemsen et al. (2012)
	+ wenn die Bewertung mit dem Wissen des Lesers übereinstimmt	Cheung et al. (2009)
<b>Kontrollillusion des Lesers</b> Leser schätzt sich selbst so ein, dass er die Vertrauenswürdigkeit des Bewerterers richtig einzuschätzen kann oder „I have the ability to assess the truthfulness“	+	Lee (2014)
<b>Motivation des Lesers</b> Motivation des Lesers beim Lesen der Bewertung	+	Chih et al. (2013)
	zielorientierte höher als erfahrungsorientierte Motivation	Dickinger (2011)
<b>Orientierung des Lesers an anderen</b> Leser orientiert sich stark an Freunden oder Autoritäten	+	Hsu et al. (2016) Shamhuyenhazva et al. (2016)
<b>Personal Sense of Power des Lesers</b> Leser nimmt seine eigene Macht, etwas zu gestalten, als gering oder groß wahr	Zusammenhang mit Unmittelbarkeit der Bewertung stärker bei niedrigem, schwächer bei hohem personal sense of power	Wu et al. (2017)
<b>Selbstoffenbarung des Lesers</b> Leser gibt viele Informationen über sich im Internet preis	stärkt den Zusammenhang von Anzahl der Bewertungen	Flanagin & Metzger (2013)
<b>Trägheit des Lesers</b> hoher Unwille des Lesers, sein Verhalten zu ändern	+	Lee (2014)

Für drei Leser-bezogene Faktoren, die nur einmal untersucht wurden, konnte kein Zusammenhang mit Vertrauen festgestellt werden: *Erfahrung des Lesers mit Online-Shopping*, *Ethnizität des Lesers* und *Internetnutzung des Lesers*. Für weitere drei einmal erforschte Faktoren dagegen sieht der Zusammenhang positiv aus: *Kontrollillusion des Lesers*, *Motivation des Lesers* und *Trägheit des Lesers*. Keinen Zusammenhang mit Vertrauen finden mehrere Studien für das *Alter des Lesers*, die *Bildung des Lesers*, das *Geschlecht des Lesers*, die *Involviertheit des Lesers* und die *Kompetenz des Lesers*. Die *Orientierung des Lesers an anderen* scheint dagegen einen positiven Zusammenhang mit Vertrauen zu haben. Widersprüchlich sind die Ergebnisse über mehrere Studien hinweg bei dem Zusammenhang des Vertrauens mit dem *Allgemeinen Vertrauen des Lesers*, der *Einstellung des Lesers gegenüber Online-Bewertungen* und der *Erfahrung des Lesers mit Online-Bewertungen*.

Nur in sehr wenigen Fällen wurden auch Faktoren erforscht, die sich auf externe Akteure beziehen (siehe Tabelle 13).

Tabelle 13: Externer Akteur-bezogene Faktoren

Faktor	Zusammenhang	Autor (Jahr)
<b>Gütesiegel für Plattform</b> Plattform wird durch externen Akteur ausgezeichnet, z. B. als Testsieger	+	Munzel (2015)
<b>Reputation der Plattform</b> Plattform hat einen guten Ruf, z. B. „popular“, „widely known for having a good reputation“	+	Hsiao et al. (2010)
<b>Übereinstimmung der Bewertung mit externen Quellen</b> Bewertung stimmt mit anderen Quellen überein, z. B. Blogs, Diskussionsforen	+	O'Reilly & Marx (2011)

Für jeden der drei nur einmal untersuchten Faktoren *Gütesiegel für Plattform*, *Reputation der Plattform* und *Übereinstimmung der Bewertung mit externen Quellen* finden die Autoren einen positiven Zusammenhang mit Vertrauen.

Insgesamt lässt sich festhalten, dass bisher schon eine große Anzahl sehr verschiedener Faktoren des Vertrauens in Online-Bewertungen untersucht wurde. Vor dem Hintergrund der bereits formulierten Einschränkungen durch das synthetisierende Vorgehen kann ich nun einige zusammenfassende Aussagen über den bisherigen Stand der Forschung treffen.

Über mindestens zwei Studien hinweg zeigt sich ein positiver Zusammenhang von Vertrauen mit den zwölf Faktoren *Anzahl der Bewertungen*, *Argumentation der Bewertung*, *Bewertung der Bewertung*, *Fotos des Bewerteten*, *Identifikation des Bewerbers*, *Orientierung des Lesers an anderen*, *Profilfoto des Bewerbers*, *Qualität der Plattform*, *Typ des Bewerteten (Erfahrungsgut)*, *Unabhängigkeit des Bewerbers*, *Vertrauenswürdigkeit der Plattform* und *Zugehörigkeit des Lesers zur Community*. Keinen Zusammenhang scheint es dagegen mit den sechs Faktoren *Alter des Lesers*, *Bildung des Lesers*, *Geschlecht des Lesers*, *Involviertheit des Lesers*, *Kompetenz des Lesers* und *Markeneinstellung des Lesers* zu geben. Bisher wurde kein Faktor untersucht, bei dem sich über mehrere Studien hinweg ein negativer Zusammenhang mit Vertrauen zeigt.

18 Faktoren wurden bisher nur einmal untersucht. Davon hängen positiv mit Vertrauen zusammen: *Attribuierung des Bewerbers*, *Bewerterstatus*, *Erfahrung des Bewerbers*, *Gütesiegel für Plattform*, *Kontrollillusion des Lesers*, *Motivation des Lesers*, *Reputation der Plattform*, *Trägheit des Lesers*, *Übereinstimmung der Bewertung mit externen Quellen*, *Übereinstimmung der Zahlen- und Textbewertung*, *Unmittelbarkeit der Bewertung* und *Vertrautheit mit der Plattform*. Ein negativer Zusammenhang wird gezeigt für *Warnhinweise der Plattform*. Keinen Zusammenhang gibt es hier mit *Erfahrung des Lesers mit Online-Shopping*, *Ethnizität des Lesers*, *Internetnutzung des Lesers* und *Soziale Nähe des Bewerbers*.

Bei 14 Faktoren zeigen sich über mehrere Studien hinweg widersprüchliche Ergebnisse: *Ähnlichkeit des Bewerbers*, *Allgemeines Vertrauen des Lesers*, *Ausführlichkeit der Bewertung*, *Ausgewogenheit der Bewertung*, *Eingebundenheit des Bewerbers in Community*, *Einstellung des Lesers gegenüber Online-Bewertungen*, *Emotionalität der Bewertung*, *Erfahrung des Lesers mit Online-Bewertungen*, *Kompetenz des Bewerbers*, *Qualität der Bewertung*, *Rechtschreibung und Grammatik*

*der Bewertung, Übereinstimmung der Bewertung mit anderen Bewertungen, Valenz der Bewertung (negativ oder positiv) und Vertrauenswürdigkeit des Bewerbers.*

Diese Beobachtungen decken sich mit der Feststellung, die Cheung und Thadani schon 2012 in ihrer Literaturschau zu eWoM machten: „we have identified a wide range of variables related to the key elements of social communication [...] However, we notice that a majority of these variables have only been studied once. Of those which have been studied more often, their impact has often been inconclusive.“ (S. 468).

Während die einstimmigen Ergebnisse mehrerer Studien ein Hinweis darauf sind, dass die Faktoren als gesichert gelten können, lassen sich aus den Ergebnissen der anderen Faktoren noch keine klaren Schlüsse ziehen. Die Ergebnisse zu den Faktoren, die bisher nur einmal untersucht wurden, sollten repliziert werden. Die widersprüchlichen Ergebnisse der restlichen Faktoren lassen sich als Aufforderung verstehen, diese Faktoren weiter zu untersuchen. Daneben bleibt stets die Frage offen, ob es noch weitere Faktoren geben könnte, die mit dem Vertrauen in Online-Bewertungen zusammenhängen, und wie diese herausgefunden werden könnten.

## 2.9 FAZIT

In diesem Kapitel legte ich mit der Definition von Online-Bewertung zunächst den konkreten Gegenstand der vorliegenden Arbeit fest. Die Definition erwies sich im Verlauf des Kapitels als sehr brauchbar, um hinsichtlich der Faktoren für Vertrauen vergleichbare Studien zu Online-Bewertungen zu identifizieren. Aus der Argumentation für die Notwendigkeit einer ausführlichen Literaturübersicht folgerte ich Fragen zum bisherigen Forschungsstand. Diese Fragen halfen dabei, die anschließenden Unterkapitel zu strukturieren. Anschließend erläuterte und begründete ich mein Vorgehen eingehend, um die Literaturrecherche nachvollziehbar und reproduzierbar zu gestalten. Die Fülle an gefundenen Studien scheint die Zweckmäßigkeit des Suchterms und der Datenbanken zu bestätigen. Nach der Korpusentwicklung stellte ich die von mir genutzte Methode der integrativen Literaturübersicht vor. Diese Methode zeigt sich im Folgenden als sehr nützlich, um die Vielfalt an Studienergebnissen handhabbar zu machen. Ich näherte mich der aktuellen Forschungslandschaft, indem ich zunächst einen Blick auf die Entwicklung des Themas als eigenes Forschungsfeld warf. Diese historische

Perspektive ermöglichte es mir, die aktuelle Forschung zu verorten und später einige Charakteristika des untersuchten Korpus zu erklären. Der Überblick über die publizierenden Autoren und Fachzeitschriften machte deutlich, dass die Ergebnisse aus vielen verschiedenen Disziplinen stammen. Das mag eine Erklärung für die mosaikartige Anmutung der Forschungslandschaft sein. Außerdem zeigte sich, dass das Thema auch langsam als Phänomen für die Kommunikationswissenschaft erkannt wird. Die ausführliche Beschäftigung mit dem Fundament und Rahmen der Forschung bildete die Voraussetzung für die kritische Auseinandersetzung mit ihren Inhalten.

Diese begann mit einem genauen Blick auf die in den Studien verwendeten Konzepte von Vertrauen. Dabei kam heraus, dass meist verschiedene Aspekte des Vertrauens gemessen werden. Schließlich erörterte ich das sich häufig sehr ähnelnde empirische Vorgehen der Autoren. Den größten Teil des Kapitels nahm die ausführliche Darstellung der bisherigen Ergebnisse der Studien ein. Ich gruppierte die Faktoren und konnte allen mithilfe des Modells von Borchers (2016) einen theoretischen Platz zuordnen. So synthetisiert und systematisiert konnten die vielen Ergebnisse miteinander verglichen werden. Die intensive Arbeit an einem tiefgehenden Vergleich ergab eindeutige Ergebnisse zu einigen Faktoren, offenbarte aber auch Lücken und widersprüchliche Ergebnisse bei anderen Faktoren. Mit dieser integrativen Literaturübersicht konnte nun endlich der aktuelle Forschungsstand zum Vertrauen in Online-Bewertung erschöpfend erfasst werden. Dieses Kapitel eignet sich als ausführliche Grundlage für ein „Systematic Review“, in dem mit statistischen Mitteln die Ergebnisse der Forschung zusammengeführt werden.

Aus der Zusammenfassung und Diskussion der bisherigen Forschung lassen sich einige offene Fragen folgern. Zunächst möchte ich auf die theoretischen Lücken eingehen. Der Gegenstand des Vertrauens wird theoretisch nicht reflektiert. Dadurch bleibt z. B. unklar, was die Autoren genau messen. Außerdem herrscht daher Uneinigkeit darüber, ob sich das Vertrauen des Lesers auf die Bewertung als Objekt oder den Bewerter als Akteur bezieht. Die mangelnde theoretische Reflexion zeigt sich auch bei der Interpretation der Faktoren. Mithilfe von Borchers' (2016) Modell ließen sich alle Faktoren übersichtlich den Elementen bzw. Akteuren zuordnen, auf die sie sich beziehen. Das Modell schließt an die Logik der bisherigen Forschung an

---

und erweitert sie. Das Modell gibt aber wie auch die bisherige Forschung keinen Aufschluss darüber, wie die Faktoren im Rahmen des Vertrauens theoretisch verstanden werden können. Es bleibt unklar, welche theoretische Bedeutung den Faktoren zukommt, wie der Zusammenhang zwischen Vertrauen und Faktoren also interpretiert werden kann. Hilfreich wäre hier demnach ein theoretisches Modell, das umfassend genug ist, um allen bisher erhobenen Faktoren eine theoretisch fundierte Position im Vertrauensprozess zuzuweisen. Zusätzlich müsste das Modell erklären können, was Vertrauen genau ist und auf wen sich das Vertrauen bezieht.

Empirisch sind ebenfalls einige Fragen offen. Insgesamt lässt sich zwar sagen, dass das Thema Vertrauen in Online-Bewertungen mittlerweile in der Forschung angekommen ist. Eine erste Annäherung an die Faktoren, die mit Vertrauen in Online-Bewertungen zusammenhängen, fand auch statt, das Forschungsfeld wurde umgrenzt. Zu einigen Faktoren gibt es bisher nur vereinzelte Studien, vor allem aber sind die Ergebnisse zu einem großen Teil der Faktoren widersprüchlich. Nur weitere empirische Forschung vermag diese Ergebnisse zu erklären.

Angeleitet von diesen Fragen besteht das Ziel meiner Arbeit daher darin, auf theoretischer und empirischer Ebene für mehr Klarheit im Forschungsfeld zu Vertrauen in Online-Bewertungen zu sorgen. Im nächsten Kapitel stelle ich dafür ein theoretisches Modell vor und zeige die Verknüpfungen mit dem hier ermittelten Forschungsstand auf.



### **3 THEORETISCHES MODELL: VERTRAUEN IN ONLINE-BEWERTUNGEN**

#### **3.1 EINFÜHRUNG**

In diesem Kapitel schlage ich ein Modell zum theoretischen Verständnis von Vertrauen in Online-Bewertungen vor. Damit sollen die theoretischen Fragen, die sich aus der bisherigen Forschung ergeben, geklärt werden. Das hier angebotene theoretische Modell beruht auf der Theorie des Vertrauens in Journalismus von Kohring (2004). Ich erkläre seine Definition von „Vertrauen“ angewandt auf den Gegenstand der Online-Bewertungen. Anschließend stelle ich die verschiedenen Bestandteile des Modells vor. Auf diese Ausführungen folgt die Verknüpfung des Modells mit dem bisherigen Forschungsstand. Dafür führe ich die Ergebnisse früherer Forschung mit den theoretischen Erkenntnissen aus dem Modell zusammen. Dabei gehe ich u. a. darauf ein, auf wen oder was sich Vertrauen beziehen kann. Außerdem zeige ich auf, wie die bisher erhobenen Faktoren aus der Perspektive dieses Modells verstanden werden können. Im Fazit fasse ich das Kapitel zusammen und diskutiere die Nützlichkeit dieses Modells.

#### **3.2 DEFINITION VON VERTRAUEN**

Vertrauen bedeutet laut Kohring (2004, aufbauend auf Luhmann, 1989), dass ein Mensch handelt, als ob er wisse, wie ein anderer Mensch handelt (S. 86, 88). Für den hiesigen Kontext bedeutet das, dass der Leser einer Online-Bewertung handelt, als ob er wisse, wie der Bewerter handelt. Vertrauen ist die „freiwillige Übernahme einer Fremdselektion als Eigenselektion“ (Kohring, 2004, S. 129f). Der Leser übernimmt also die Auswahl des Bewerter als seine eigene Auswahl. Weiterhin beschreibt Kohring (2004) Vertrauen als „eine beobachtete Delegation von Handlungsverantwortung“ (S. 138), sie

simuliert im Moment ihres Vollzugs, dass eine bestimmte Zukunft schon eingetreten sei. Sie reagiert auf die Kontingenz des ja immer auf die Zukunft gerichteten selektiven Handelns, d. h. auf die als Risiko bewertete Möglichkeit, dass eine andere Folge als die erwartete eintritt, indem sie dieses Handeln trotzdem ermöglicht oder akzeptiert und so tut, als könne sich nur eine bestimmte Folge daraus ergeben (S. 94).

Am Beispiel von Online-Bewertungen heißt das, dass der Leser seine Auswahl eines bestimmten Produkts oder einer bestimmten Dienstleistung von der Auswahl des Bewerbers abhängig macht. Wenn der Bewerter das Bewertete empfiehlt, dann wird der Leser das Bewertete kaufen wollen; wenn der Bewerter davon abrät, dann wird der Leser das Bewertete meiden. Vertrauen in Online-Bewertungen heißt, dass der Leser auf der Grundlage der Online-Bewertung des Bewerbers handelt.

Kohring (2004, S. 165) beschreibt diesen Vorgang in Anlehnung an Luhmann als Vergegenwärtigung, d. h. Vorwegnahme zukünftiger Gegenwarten. Mit dem Begriff der „zukünftigen Gegenwarten“ meint er die verschiedenen Möglichkeiten, die der Bewerter hat, zu handeln. Der Bewerter hat z. B. die Möglichkeit, eine geschönte Bewertung, eine Schmähbewertung oder eine ehrliche Bewertung zu verfassen. Bei Vertrauen aber erwartet der Leser – er geht sogar davon aus – dass der Bewerter sich unter all diesen Handlungsmöglichkeiten für die Möglichkeit entschieden hat, eine ehrliche Bewertung zu verfassen. Der Leser weiß nicht, ob der Bewerter wirklich so gehandelt hat, aber tut so, als wisse er es. Der Leser handelt, als ob der Bewerter seine Vertrauenserwartung bereits erfüllt hätte. Der Leser vergegenwärtigt damit diese eine bestimmte zukünftige Gegenwart. Kohring (2004) schreibt dazu: „Vertrauensrelationen katalysieren sozusagen Zukunft.“ (S. 138). Bei Luhmann (1989) heißt das: „Der vertrauensvoll Handelnde engagiert sich so, als ob es in der Zukunft nur bestimmte Möglichkeiten gäbe“ (S. 20). Zusammengefasst bedeutet das, dass der Leser handelt, weil der Bewerter eine richtige Bewertung gegeben haben wird (vgl. Kohring, 2009, S. 79).

Aufbauend auf diesem Verständnis von Vertrauen hat Kohring (2004) ein Modell zu Vertrauen in Journalismus entwickelt, das ich an den Gegenstand der Online-Bewertungen anpasse. In den weiteren Unterkapiteln beschreibe ich die einzelnen Bestandteile des hier grafisch dargestellten Vertrauensmodells (nach Kohring, 2004, eigene Darstellung).

Abbildung 6: Vertrauensmodell



### 3.3 VORAUSSETZUNGEN FÜR VERTRAUEN

Den Bedarf nach Vertrauen gibt es laut Kohring aufgrund der *Komplexität* der modernen Gesellschaft (siehe Abbildung 6). Komplexität bildet den Hintergrund, vor dem Vertrauen als soziales Phänomen auftritt. Die Funktion von Vertrauen liegt darin, diese Komplexität zu reduzieren. Kohring geht davon aus (2004), dass Menschen in der heutigen Gesellschaft mehr Möglichkeiten haben zu handeln als früher. Mit dem größeren Handlungsspielraum steigt aber auch die Komplexität (Luhmann, 1989, S.5-6). Die heutige Gesellschaft ist zu unübersichtlich, um als Mensch alles zu wissen, was man wissen müsste, um sicher handeln zu können. Mit

---

der generellen Komplexität der gesamten Gesellschaft ist das *ungenügende Wissen* des einzelnen Menschen über das Bewertete (siehe Abbildung 6) verknüpft.

So gibt es z. B. eine riesige Auswahl an Produkten und Dienstleistungen über deren Qualität ein Mensch aber nur wenig weiß. Um sich für eine der Möglichkeiten zu entscheiden, kann er Online-Bewertungen heranziehen. Doch auch hier zeigt sich die Komplexität der modernen Gesellschaft. Es gibt eine Vielzahl an Bewertungen, manche Bewertungen zum selben Bewerteten widersprechen sich, die Bewerter sind unbekannt und anonym. Der Leser weiß nicht, welche der Bewertungen stimmen. Das Problem des Lesers besteht darin, dass er handeln muss, ohne wissen zu können, wie die Bewerter handeln.

Vertrauen wird also dann relevant, wenn es dem Leser an problemspezifischem Wissen mangelt. Im Falle von Online-Bewertungen heißt mangelndes problemspezifisches Wissen, dass ein Mensch über das Bewertete nicht genug weiß, um handeln zu können. Daher ist er auf das Wissen anderer Menschen über das Bewertete angewiesen. Hier wiederum besteht das Problem darin, dass der Leser vor der Übernahme der Online-Bewertung nicht weiß, ob diese richtig oder falsch ist. Er kann die Bewertung nicht überprüfen, bevor er diese nutzt. Denn um die Bewertung zu überprüfen, bräuchte er genügend Wissen darüber. Hätte er genügend Wissen, dann wäre er von vornherein gar nicht auf die Bewertungen anderer angewiesen. Würde er z. B. einen guten Arzt in seinem neuen Wohnort kennen, so müsste er nicht online nach Bewertungen suchen. Der Leser kann vor seiner Handlung nicht entscheiden, ob die Bewertung, die er seiner Handlung zugrunde legt, richtig oder falsch ist. Damit ist das ungenügende Wissen des Lesers eine Voraussetzung, damit es überhaupt zu Vertrauen in den Bewerter kommen kann.

Neben ungenügendem Wissen gibt es laut Kohring (2004) noch eine weitere Bedingung, damit Vertrauen relevant wird. Dabei handelt es sich um *Risikobewusstsein* (siehe Abbildung 6). Der Leser kann nur vertrauen, wenn er das Risiko wahrnimmt, welches durch „doppelte Kontingenz“ (Kohring, 2004) entsteht. Mit Kontingenz ist gemeint, dass ein Mensch unterschiedliche Möglichkeiten hat, zu handeln. Sein zukünftiges Handeln ist ungewiss. Dasselbe gilt aber auch für den anderen Menschen. Auch er hat unterschiedliche Möglichkeiten zu handeln und

auch sein Handeln ist ungewiss. Diese Kontingenz auf beiden Seiten wird als doppelte Kontingenz bezeichnet.

Doppelte Kontingenz bedeutet im Fall von Online-Bewertungen, dass das zukünftige Handeln des Lesers nicht vorgegeben ist (= eigene Kontingenz) und der Leser selbst über sein Handeln entscheiden kann (= Eigenselektion). Aber auch der Bewerter hat verschiedene Möglichkeiten zu handeln (= fremde Kontingenz) und zu entscheiden (= Fremdselektion). Der Leser kann sich dafür entscheiden, dem Bewerter zu vertrauen. Das hieße, dass er die Bewertung als Grundlage für sein Handeln nutzt und z. B. das positiv Bewertete kauft. Er kann sich aber auch dagegen entscheiden. Genauso kann der Bewerter sich dafür entscheiden, eine angemessene Bewertung abzugeben oder auch einem Konkurrenten mit einer gefälschten schlechten Bewertung zu schaden.

Wenn sich der Leser dieser doppelten Kontingenz bewusst ist, verfügt er über Risikobewusstsein. Denn wie bereits dargestellt kann der Leser nicht wissen, ob die Bewertungen stimmen. Ein Bewerter mag andere Kriterien an ein Smartphone anlegen als der Leser selbst; die Online-Bewertung eines Hotels kann eingekauft sein. Das Risiko besteht darin, dass der Leser eine Bewertung nutzt, die sich später als falsch herausstellt. Es besteht die Gefahr, dass er eine Fehlinformation zur Grundlage seiner Handlung macht. Das mag mehr oder weniger schwere Folgen haben. So kann der Leser z. B. Geld verlieren durch ein fades Essen in einem gut bewerteten Restaurant oder Zeit verlieren durch die Wahl eines unorganisierten Handwerksbetriebs, der im Internet empfohlen wurde. Dieses Risikos kann sich der Leser bewusst sein. Er erkennt das Risiko, das darin liegt, die Information des Bewerters zu übernehmen. Der Leser weiß, dass es riskant ist, auf Grundlage der Bewertung zu handeln. Das Risikobewusstsein beinhaltet auch das Bewusstsein des Lesers darüber, dass eine Chance darin besteht, die Bewertung zu übernehmen. Denn es besteht immer die Möglichkeit, dass die Bewertung stimmt.

Ist sich der Mensch nicht über das Risiko im Klaren, so handelt es sich laut Kohring (2004, S. 96ff) um „Vertrautheit“. Vertrauen dagegen ist „eine beobachtete Delegation von Handlungsverantwortung bei gegebenem Risikobewusstsein“ (Kohring, 2004, S. 138) und eine „selektive Verknüpfung von Fremdhandlungen mit

Eigenhandlungen unter der Bedingung einer rational nicht legitimierbaren Tolerierung des wahrgenommenen Risikos“ (Kohring, 2004, S. 178).

Warum aber sollte ein Mensch bei solch einem Risiko vertrauen? Vertrauen ist ein Weg, um mit der ungewissen Zukunft umzugehen. Da, wie bereits erläutert, soziales Handeln in der modernen Gesellschaft immer komplexer wird, muss diese Komplexität erst verringert werden, damit der Mensch handeln kann. Vertrauen reduziert Komplexität, indem es einem Menschen dabei hilft, zu entscheiden, wie er handelt (Kohring, 2004, S. 129). Der Mensch kann durch Vertrauen mit dem Problem der doppelten Kontingenz umgehen; es ermöglicht ihm, zu handeln. Vertrauen bietet dem Menschen die Möglichkeit zu handeln, obwohl er nicht weiß, wie der andere Mensch handelt.

So hat der Leser die Möglichkeit durch die Übernahme einer Bewertung, die sich als richtig herausstellt, Zeit oder Geld zu sparen. Mit einer zutreffenden Bewertung erspart er sich den Besuch verschiedener Frisöre oder findet gleich den geeigneten Laptop. Indem der Leser dem Bewerter vertraut, spart er zeitliche und finanzielle Ressourcen. Die Zeit bräuchte er sonst, um z. B. den Bewerter und dessen Aussagen zu überprüfen oder sich selbst problemspezifisches Wissen anzueignen, indem er das Bewertete kauft oder nutzt. Dem Bewerter zu vertrauen ist eine ressourcenschonende Möglichkeit für den Leser, die Selektion des Bewerbers zu übernehmen.

Neben ungenügendem Wissen und Risikobewusstsein formuliert Kohring (2004, S. 125f) fünf *strukturelle Voraussetzungen für Vertrauen* (siehe Abbildung 6): Dazu gehört Freiwilligkeit, d. h. die als eigenständig wahrgenommene Entscheidung des Lesers für oder gegen Vertrauen in den Bewerter. Der Leser muss die Möglichkeit haben, selbst zu entscheiden, ob er dem Bewerter vertraut oder nicht. Er darf sich nicht dazu unter Druck gesetzt oder gezwungen fühlen. Eine weitere strukturelle Voraussetzung ist Erlernbarkeit. Damit ist gemeint, dass der Leser positive Erfahrungen mit dem Vertrauen in den Bewerter verallgemeinern können muss. Bei Online-Bewertungen ist das z. B. möglich, wenn sich über die Zeit eine Community herausbildet, in welcher Leser und Bewerter einander kennen und einschätzen können. Eine dritte Voraussetzung ist Adressierbarkeit. Das bedeutet, dass „auf Vertrauensprobleme genügend spezifisch eingegangen werden muss“

(Kohring, 2004, S. 126). Der Bewerter muss erkennen können, dass der Leser die doppelte Kontingenz wahrnimmt und der Leser bestimmte Erwartungen an ihn als Bewerter richtet. Kontrollierbarkeit ist eine weitere Voraussetzung. Damit ist gemeint, dass der Leser das Vertrauen in den Bewerter überprüfen kann. Der Leser kann das Vertrauen nicht überprüfen, indem er die Bewertung überprüft (denn genau dafür fehlt ihm ja das Wissen). Aber der Leser kann sich z. B. auf die Nutzungsregeln der Bewertungsplattform und deren Durchsetzung verlassen. Die letzte Voraussetzung ist Sanktionierbarkeit. Das heißt, dass der Leser enttäushtes Vertrauen in den Bewerter bestrafen können muss. Das kann entweder durch ihn selbst geschehen, z. B. indem der Leser die Bewertung oder den Bewerter bewertet bzw. kommentiert. Oder aber der Plattformbetreiber übernimmt die Sanktionierung, indem er z. B. auf gefälschte Bewertungen mit der Sperrung des Bewerter reagiert.

### 3.4 OBJEKT UND DIMENSIONEN DES VERTRAUENS

Nachdem der Hintergrund für die Relevanz von Vertrauen sowie die strukturellen Voraussetzungen für Vertrauen nun geklärt sind, widme ich mich dem *Objekt des Vertrauens* (siehe Abbildung 6). Laut Kohring (2004) erfordert eine Vertrauenssituation die Selektivität eines Akteurs, also seine Entscheidungsmöglichkeit, entweder so oder anders zu handeln. Eine Bewertung aber kann sich nicht zwischen verschiedenen Handlungen entscheiden. Außerdem muss sich das Vertrauen des Lesers auf ein Objekt beziehen, von dem der Leser enttäuscht werden kann und das um die Vertrauenserwartung des Lesers wissen muss (Kohring, 2004). Daher kann sich Vertrauen letztlich nur auf einen sozialen Akteur beziehen. Darum kann es kein Vertrauen in eine Bewertung bzw. Vertrauenswürdigkeit von einer Bewertung geben. Die Online-Bewertung als Vertrauensobjekt scheidet mit Kohrings (2004) Vertrauensmodell aus.

Nach dem für die Literaturübersicht verwendeten Modell von Borchers (2016) lassen sich verschiedene soziale Akteure als potenzielle Vertrauensobjekte im Kontext von Online-Bewertungen ausmachen. Im Modell tauchen der Anbieter, der Bewerter, die Community, der Leser und der Plattformbetreiber auf; auch externe Akteure sind als Vertrauensobjekt denkbar. Im Kommunikationsprozess ist allerdings das Verhältnis von Leser und Bewerter zentral. Daher liegt der Fokus der vorliegenden Arbeit auf dem Bewerter als

Vertrauensobjekt. Die anderen Akteure kommen im Folgenden nur insofern vor, als dass der Leser von ihnen bereitgestellte Informationen als Grund für oder gegen Vertrauen in den Bewerter heranzieht. Genaugenommen müsste der Titel der vorliegenden Arbeit und des Modells daher „Vertrauen in Online-Bewerter“ lauten. Der Anschlussfähigkeit an bisherige Forschung halber bleibt der Titel jedoch beim ungenaueren aber eingängigeren „Vertrauen in Online-Bewertungen“.

Doch worin genau besteht das Vertrauen in den Bewerter? Kohring (2004) schlägt vor, die Dimensionen des Vertrauens theoretisch von der Funktion des Vertrauensobjekts abzuleiten (S. 41). Es braucht daher eine Funktionsbestimmung von Bewertern. So lässt sich klären, was die Vertrauenswürdigkeit speziell eines Bewerters ausmacht im Gegensatz zu „generellen Faktoren von Kommunikation“ (Kohring, 2004, S. 41). Laut Kohring „spezifizieren [Dimensionen des Vertrauens] die Referenz des Vertrauens und verweisen damit auf die Funktion des Vertrauensobjekts“ (2004, S. 120). Die Dimensionen machen deutlich, was das Spezifische an dieser Form der Kommunikation ist, das nicht auf jede Kommunikation zutrifft (Kohring, 2004). Sie erklären demnach, worin sich das Vertrauen in einen Schulfreund von dem Vertrauen in einen Taxifahrer oder dem Vertrauen in einen Bewerter unterscheidet. In der bisherigen Forschung zu Vertrauen in Online-Bewertungen findet sich keine Antwort darauf, worin die spezifische Vertrauenswürdigkeit eines Bewerters besteht. Weder werden die in der eWOM-Forschung bestehenden Analysen zur Funktion von Online-Bewertungen für den Leser (siehe z. B. „usage functions for eWOM“ bei Reichelt et al., 2014) aufgegriffen, noch die bereits erwähnte Studie von Borchers (2016), in der er sich mit dem theoretischen Verständnis und der gesellschaftlichen Funktion von eWOM auseinandersetzt. In den meisten Studien wird nur der Gegenstand der Online-Bewertungen kurz definiert, häufig mit Rückgriff auf die populärsten Definitionen von eWOM von Hennig-Thurau et al. (2004, siehe S. 7 in vorliegender Arbeit) oder auf die von Dellarocas (2003) oder Mudambi & Schuff (2010). Der Gegenstand wird also anhand formaler, nicht funktionaler Charakteristika abgegrenzt. Fast immer wird anschließend auf die spezielle Gefahr bzw. das Risiko für den Nutzer dieser Kommunikationsform hingewiesen, die die Eigenheiten des World Wide Web mit sich bringen: die Masse an Information und die Anonymität der Verfasser. Es wird

nicht darauf eingegangen, welche Funktion die Bewertung für den Nutzer hat und welche Erwartungen des Nutzers sich daraus an die Bewertungen ergeben.

Ich schlage vor, Online-Bewertungen als Teil des Systems Öffentlichkeit zu verstehen. Bewertungen sind damit funktional äquivalent zum Journalismus. Dies ermöglicht es mir, Kohrings Modell von Vertrauen in Journalismus auf den Gegenstand der Online-Bewertung zu übertragen. Kohring beschränkt sich „[f]ür die spätere Entwicklung der Theorie und vor allem wegen der späteren Operationalisierung [...] auf die bei weitem wichtigste Form öffentlicher Kommunikation, den organisierten und professionalisierten Journalismus“ (Kohring, 2004, S. 12). Andere Formen öffentlicher Kommunikation wie Online-Bewertungen sind damit aber nicht ausgeschlossen.

Die Funktion von Öffentlichkeit besteht darin, Beobachtungen von Interdependenzen mitzuteilen. Die Funktion des Bewerter liegt also darin, die Beobachtungen über das Bewertete mitzuteilen, die für den Leser relevant sind. Das ist die Vertrauenserwartung, die der Leser an den Bewerter hat. Die spezifische Selektivität des Bewerter, also seine Entscheidung darüber, was er in seiner Bewertung berichtet, bildet damit die Referenz des Vertrauens des Lesers.

Von der Funktion des Bewerter, die die Referenz des Vertrauens des Lesers darstellt, lassen sich nun vier *Dimensionen des Vertrauens* ableiten (siehe Abbildung 6). Diese spezifischen Erwartungen des Lesers an den Bewerter grenzen sich ab von seinen allgemeinen Vertrauenserwartungen an jeden Interaktionspartner. Diese vier öffentlichkeits-spezifischen Vertrauensdimensionen sind Themenselektivität, Faktenselektivität, Richtigkeit von Beschreibungen und explizite Bewertungen (Kohring, 2004, S. 214f und 229).

Themenselektivität bedeutet, dass der Leser die Erwartung hat, dass der Bewerter für seine Bewertung relevante Themen auswählt. Bei Online-Bewertungen steht allerdings anders als beim Journalismus schon das Bewertete als Thema fest. Daher ist davon auszugehen, dass Themenselektivität als Vertrauensdimension eher eine geringe Rolle spielt.

Die zweite Dimension der Faktenselektivität steht für die Erwartung, dass der Bewerter alle relevanten Informationen liefert, damit der Leser die angesprochenen Themen einordnen kann (Kohring, 2014, S. 214). Das betrifft die Vollständigkeit der

Fakten, d. h., dass alle wesentlichen Informationen berücksichtigt werden. Außerdem ist damit die Gewichtung der Fakten gemeint, d. h. dass die wichtigen Fakten im Vordergrund stehen. Des Weiteren steht Faktenselektivität noch für die Vielfalt der Fakten, d. h. die Berichterstattung aus mehreren Perspektiven.

Bei der Richtigkeit von Beschreibungen handelt es sich um die dritte Dimension des Vertrauens. Damit ist gemeint, dass die Beschreibungen des Bewerter nachweislich stimmen. Die Informationen sind also empirisch überprüfbar, korrekt und wahr.

Die letzte Vertrauensdimension bezieht sich auf den Kern einer Online-Bewertung, das Vertrauen in die expliziten Bewertungen. Der Leser erwartet, dass der Bewerter das Bewertete angemessen beurteilt. Dies bedeutet, dass Kritik adäquat geäußert und mit einer nachvollziehbaren Begründung verknüpft wird.

Diese Dimension zeigen die Erwartungen des Lesers an. Sie beschreiben, worin der Leser vertraut. Der Leser vertraut dem Bewerter darin, dass er die relevanten Themen und Fakten auswählt, diese korrekt darstellt und angemessen bewertet. Dabei ist es wichtig, im Hinterkopf zu behalten, dass sich die Leser dahingehend unterscheiden können, was sie als relevant, korrekt und angemessen empfinden.

### 3.5 GRÜNDE FÜR VERTRAUEN

Angesichts des bereits erwähnten Risikobewusstseins scheint es zunächst nicht naheliegend, zu vertrauen. Um sein Vertrauen zu legitimieren, zieht der Leser laut Kohring (2004) daher *Gründe für Vertrauen* (siehe Abbildung 6) heran. Diese Gründe für Vertrauen sind Indikatoren für die Vertrauenswürdigkeit des Bewerter, sie gleich „unzureichendes Wissen aus“ (Kohring, 2004, S. 113). Gründe für Vertrauen sind „Merkmale des Vertrauensobjekts, die für seine Vertrauenswürdigkeit stehen“ (S. 113), sie dienen dem Leser daher als Anhaltspunkte. Da „Vertrauen [...] letztendlich immer unbegründbar [ist]“ (Luhmann, 1989, S. 26), sind Gründe für Vertrauen nicht als kausale Legitimation zu verstehen, sondern nur als die Darstellung einer kausalen Konditionierung (Kohring, 2004, S. 178).

Kohring (2004) unterscheidet zwischen unspezifischen und spezifischen Gründen für Vertrauen (siehe Tabelle 14). Spezifische Gründe für Vertrauen kann der Leser heranziehen, wenn er auf Erfahrungen mit dem Bewerter zurückgreifen kann.

Das kann eine selbst erlebte gute oder enttäuschende Erfahrung sein, eine bei anderen beobachtete oder eine durch andere mitgeteilte Erfahrung. Wenn ein Leser z. B. aufgrund einer Bewertung ein gutes Hotel gefunden hat, kann er dies als Grund heranziehen, um sich bei der nächsten Hotelbuchung auf denselben Bewerter zu verlassen. Eine durch andere mitgeteilte Erfahrung wäre z. B. der dankende Kommentar eines anderen Lesers unter einer Bewertung.

Unspezifische Gründe beruhen hingegen nicht auf bereits gemachten Erfahrungen mit dem Bewerter. Die unspezifischen Gründe werden von Kohring (2004, S. 181) eingeteilt in situative und situationsunabhängige Gründe. Die situativen Gründe betreffen alle vom Bewerter gegebenen Informationen, hier zieht der Leser einen Rückschluss von der Mitteilung auf die Selektion. Ein Beispiel dafür ist ihr sprachlicher Stil, aber auch, ob der Leser sich durch die Bewertung „ein direktes Bild von dem Geschehen machen“ kann und „Fakten und Meinungen“ auseinandergehalten werden (Kohring, 2004, S. 246f).

Bei den situationsunabhängigen Gründen schließt der Leser vom Kontext der Mitteilung auf die Selektion. Das kann z. B. die Unabhängigkeit des Bewerter von anderen Akteuren betreffen oder, dass der Plattformbetreiber auf die Einhaltung der Regeln achtet.

Tabelle 14: Gründe für Vertrauen

	Spezifische Gründe	Unspezifische Gründe
<b>Beschreibung</b>	aus bestätigten Vertrauenswartungen abgeleitete Gründe	nicht aus bestätigten Vertrauenswartungen abgeleitete Gründe
<b>Typen</b>	Bestätigung durch <i>Eigenerfahrung</i>	Rückschluss von <i>Mitteilung</i> auf Selektion
	Bestätigung durch <i>selbst beobachtete Erfahrungen anderer</i>	Rückschluss von <i>Kontext der Mitteilung</i> auf Selektion
	Bestätigung durch <i>mitgeteilte Erfahrungen anderer</i>	

(Kohring, 2004, S. 181, Tab. 14)

### 3.6 RESULTAT UND KONSEQUENZ DES VERTRAUENS

Nach der eigentlichen Vertrauensbeziehung lassen sich das *Resultat des Vertrauens* und die *Konsequenz des Vertrauens* beobachten (siehe Abbildung 6). Beim Resultat des Vertrauens handelt es sich um die tatsächliche Leistung des Bewerter für den Leser. Ein mögliches Resultat wäre, dass der Bewerter alle Erwartungen des Lesers erfüllt hat. Der Leser würde sein Vertrauen dann bestätigt sehen. Wenn der Bewerter den Erwartungen des Lesers allerdings nicht oder nicht vollständig nachgekommen ist, ist das Vertrauen des Lesers enttäuscht. Beide Erfahrungen können wiederum als spezifische Gründe für oder gegen Vertrauen in die nächste Entscheidung des Lesers miteinbezogen werden.

Mit der Konsequenz des Vertrauens sind Anschlusshandlungen des Lesers in der Umwelt öffentlicher Kommunikation gemeint. Es geht darum, ob der Leser die vom Bewerter erhaltenen Informationen nutzt und darauf basierend handelt, z. B. indem er das empfohlene Produkt kauft, eine negative bewertete Dienstleistung nicht in Anspruch nimmt oder den Bewerter weiterempfiehlt. Hier geht es also um die Wirkung einer Online-Bewertung. Kohring (2004) fasst dies wie folgt zusammen:

„Eine Vertrauensrelation [...] liegt [...] dann vor, wenn ein Rezipient eine Selektion vornimmt, die nur durch die [Bewerter-] Selektion (der Mehrsystemzugehörigkeit eines Ereignisses) kausal ermöglicht wird, deren Kontingenz der Rezipient wahrnimmt und als Risiko bewertet. Diese vom Rezipienten vorgenommene Selektion ist nicht mit der *Konsequenz* von Vertrauen gleichzusetzen. Sie *ist* das Vertrauen. Erst wenn der Rezipient auf der Basis dieser selektierten Information in der Umwelt [der Bewertung] handelt (Vertrauen als Handlung) bzw. auf eine diese Handlung bezogene Einstellung ausgebildet hat (Vertrauen als Einstellung [...]), kann von einer Konsequenz der Vertrauensrelation gesprochen werden.“ (S. 182)

Kohring weist an dieser Stelle aber auch darauf hin, dass neben Vertrauen in den Bewerter auch noch andere Einflussfaktoren auf das Handeln des Lesers bestehen (2004, S. 182f). So kann es z. B. sein, dass ein Leser einem Bewerter vertraut, aber aus Bequemlichkeit oder finanziellen Gründen doch ein anderes als das positiv bewertete Produkt wählt. Mit dem Resultat und den Konsequenzen des Vertrauens sind die letzten Elemente des Vertrauensmodells erläutert.

### 3.7 LITERATURÜBERSICHT IM LICHT DES THEORETISCHEN MODELLS

In diesem Unterkapitel betrachte ich nun die Ergebnisse und offenen Fragen aus der integrativen Literaturübersicht im Licht des vorgestellten theoretischen Modells zu Vertrauen in Online-Bewertungen.

Die oben dargebotene Definition von Vertrauen kann mit der gesamten bisherigen Forschung in Einklang gebracht werden. Mit dem Vertrauensmodell lässt sich auch die Frage beantworten, auf wen oder was sich das Vertrauen des Lesers bezieht. Mit Kohring (2004) wird klar, dass sich das Vertrauen genau genommen nur auf den Bewerter, nicht auf die Bewertung beziehen kann. Es ist theoretisch unsinnig, zwischen dem Bewerter und der Bewertung als Vertrauensobjekt zu unterscheiden. Deutlich wird das z. B. wenn der Leser die Kompetenz des Bewerter einschätzt. Dabei kann er sich entweder auf Informationen aus dem Bewertungstext, dem Bewerterprofil beziehen oder aus beidem – gemeint bleibt immer die Kompetenz des Bewerter. Laut Kohring (2004) ist es daher ein Zirkelschluss, dass „information credibility“ zu „source credibility“ führt. Es mag zwar sein, dass der Leser das Gefühl hat, dass er der Bewertung vertrauen kann. Doch analytisch gesehen ist die Bewertung kein selektiver Akteur. Der Wissenschaftler muss die Bezugseinheit des Vertrauens, d. h. das Vertrauensobjekt, theoretisch eindeutig bestimmen können. Laut Kohring (2004) ist aber nicht zu erwarten, dass ein Studienteilnehmer begrifflich zwischen der Mitteilung, d. h. der Bewertung, und der Quelle, d. h. dem Bewerter, als Objekt seines Vertrauens unterscheiden kann. Daher kann die Abfrage des theoretisch korrekten Vertrauensobjekts während der Messung vernachlässigt und somit können die Ergebnisse von Studien zum Vertrauen in Bewertung und in Bewerter weiterhin zusammengefasst werden. In einigen Fällen ist es sogar sinnvoll, dass sich einzelne Items auf verschiedene Referenzen beziehen. Aber es sollte klar sein, dass sich das damit gemessene Vertrauen nur auf den Bewerter selbst beziehen kann und daher dementsprechend zu interpretieren ist. Da fünf von sechs Studien einen eindeutigen positiven Zusammenhang zwischen der *Vertrauenswürdigkeit des Bewerter* und dem Vertrauen in die Bewertung sehen, spricht der Großteil der Empirie nicht gegen diese theoretische Argumentation.

Wie bereits erwähnt kann sich das Vertrauen des Lesers auch auf andere Akteure innerhalb des Online-Bewertungsprozesses richten. So beziehen sich einige

Faktoren der bisherigen Forschung auf die Community, den Plattformbetreiber oder externe Akteure. Der Leser zieht das Vertrauen in diese jeweiligen Akteure aber letztlich nur als Grund für sein Vertrauen in den Bewerter heran. Der Leser verlagert also sein Vertrauen (Kohring, 2004, S. 179). Bei diesen Gründen durch Verlagerung vertraut der Leser genau genommen zunächst anderen Vertrauensobjekten, nimmt ihre Handlung aber dann als Grund, um dem Bewerter zu vertrauen. Einige Beispiele dafür nenne ich weiter unten bei den unspezifischen Gründen für Vertrauen.

Auch die bisherigen Messungen von Vertrauen werden durch das theoretische Modell erhellt. Die für eine Messung verwendete Definition des Vertrauens (siehe Kapitel 2.6 Verständnis von Vertrauen) deckt sich sogar mit der Definition von Vertrauen im hiesigen Modell, nämlich der Übernahme einer Fremdselektion als Eigenselektion. Allerdings wird das Vertrauen in den Bewerter mithilfe der genannten vier Dimensionen hier noch spezifischer gefasst als in der bisherigen Forschung.

Zusätzlich kann die große Gruppe der restlichen Faktoren zur Messung von Vertrauen nun differenziert werden. Bei den Items handelt es sich entweder um Dimension oder um Gründe für Vertrauen. In vielen Fällen erfolgt die Messung anhand der Dimensionen von Vertrauen. Auf die Themenselektivität bezieht sich keines der Items. Das liegt wohl daran, dass die Gewichtung, Frequenz und Kontinuität des Berichtens über ein Thema nicht von dem einzelnen Bewerter abhängig sind. Der Dimension der Faktenselektivität lassen sich Items wie „complete“ und „depth“, die sich auf die Vollständigkeit der Bewertung beziehen, zuordnen. Andere Teile der Faktenselektivität (Gewichtung und Vielfalt) wurden bisher nicht gemessen. Die meisten der Items beziehen sich auf die Dimension der Richtigkeit von Beschreibung, d. h. auf die Überprüfbarkeit. Darunter fallen Abfragen wie „accurate“, „authentic“, „factual“, „honest“, „no false information included“, „realistic“ und „sincere“. Die Dimension der expliziten Bewertung umfasst Items, die zur Adäquatheit gehören, wie „fair“ und „unbiased“ und solche, die sich auf die Nachvollziehbarkeit richten, wie „convincing“ und „reasonable“. Die Messung von Vertrauen anhand der Dimensionen von Vertrauen ist prinzipiell möglich und gültig, wird aber meist nur unvollständig (d. h. nur bezogen auf einzelnen Dimensionen) gemacht.

Die am dritthäufigsten vorkommenden Items zur Messung von Vertrauen lassen sich als Gründe für Vertrauen verstehen. Die auf Hovland et al. (1951) zurückgehende häufig verwendete Dimension von Vertrauen „expertise“ verstehe ich, Kohrings (2004) Argumentation folgend, als Grund für Vertrauen. Auf das Problem zur Unterscheidung von „predictors“ und „dimensions“ weist schon Llamero (2014) hin. Sie schreibt dazu:

„One of the difficulties lies in the extensive range of indicators used to examine credibility issues (fairness, accuracy, factuality, truthfulness, completeness, objectivity, lack of bias, trustworthiness, etc.). The quantitative studies performed tend to rank the importance of these measures, but the interpretation of them remains unclear, not distinguishing whether they are predictors of perceived credibility or underlying dimensions of the credibility concept itself (Rieh and Danielson, 2007). This is due to a lack of consensus about the attributes that define credibility.“  
(Llamero, 2004, S. 957)

Die Items der auf Hovland et al. (1951) aufbauenden Skala von Ohanian (1990), die in den hier untersuchten Studien am häufigsten verwendet wird, gelten daher nach Kohring (2004) als Gründe für Vertrauen. Die Items der Studien lassen sich außerdem danach unterscheiden, ob sie sich auf Eigenschaften des Bewerbers oder auf Eigenschaften der Bewertung beziehen. Unter den gemessenen Eigenschaften des Bewerbers spielt die Kompetenz des Bewerbers die größte Rolle, wie in „bright“, „competent“, „informed“, „trained“, „intelligent“, „experienced“, „knowledgable“, „expert“, „qualified“, „skilled“. Aber auch andere Eigenschaften des Bewerbers werden abgefragt, wie z. B. „good“, „attractive“, „open-minded“, „ethical“, „moral“, „caring“, „concerned with me“, „warm“, „sensitive“, „not self-centered“, „understanding“, „constructive“, „very open“, „honorable“, „genuine“, „has my interests at heart“ und „I can talk freely with him“. Items, die Eigenschaften der Bewertung beschreiben, handeln meist vom Ausdruck, z. B. „well written“, „easy to read“, „communicated clearly“, „logic of this review is easy to follow“, aber auch von weiteren Eigenschaften wie „good“, „high quality“, „interesting“, „real“ und „enjoy reading“. Llamero (2014) findet heraus, dass ihre Teilnehmer unter Vertrauen verstehen, dass der Bewerter kompetent und die Bewertung plausibel ist. In Tabelle 15 zeige ich die aktualisierte Version der Items für Vertrauen. Hier wird die Gruppe der restlichen Items aufgeteilt in solche Items, die als Dimensionen von Vertrauen und jene Items, die als Gründe für Vertrauen zu verstehen sind.

Tabelle 15: Items für Vertrauen

Studie	Abstrak- tum	Dimen- sion	Grund	Konse- quenz	Defini- tion
Ayeh et al. (2013)	3	2	5		
Bae & Lee (2011)	1	3			
Casaló et al. (2015)	1	2			
Chang (2014)	1	3			
Cheung et al. (2009)	1	2			
Cheung et al. (2012)	2	2			
Chih et al. (2013)	2		2		
Cox et al. (2017)	2	2	1		
Dickinger (2011)	1				
Doh & Hwang (2009) <sup>9</sup>					
Dou et al. (2012)	3	2	5		
Filieri (2015)	3		1		
Filieri (2016)	1		1		
Filieri et al. (2015)	3		1		
Flanagin & Metzger (2013)	2				
Furner et al. (2016)	2			1	
Grabner-Kräuter & Waiguny (2015)	3	2	5		
Grabner-Kräuter & Waiguny (2015)	1	1			
Hong & Park (2012)	2	3			
Hsiao et al. (2010)	3				
Hsu et al. (2016)	1	2			
Jensen & Averbeck (2013)	3	1			
Jensen & Yetgin (2017)	3	1			
Kusumasondjaja et al. (2012)	2	3			
Lee & Koo (2012)	1	1			
Lee (2014)	4			2	
Lee et al. (2011) <sup>10</sup>					
Lim & van der Heide (2015)	1	1	16		
Lin & Xu (2017)	2	2			
Lis (2013)	1	2			
Lis (2013)	3	2	5		
Llamero (2014)	1	1	1		
Lopéz & Sicilia (2014)	1	1			
Luo et al. (2013)	3	1			
Luo et al. (2014)	3	1			
Luo et al. (2015)	3	1			
McGloin et al. (2014) <sup>11</sup>		1	1	1	

9 Anzahl und Items nicht genannt

10 Items nicht genannt

Studie	Abstrak- tum	Dimen- sion	Grund	Konse- quenz	Defini- tion
Munzel (2015)	3				
Nekmat & Gower (2012)	3	2		4	
O'Reilley & Marx (2011)	1	2			
O'Reilley et al. (2016)	1			1	
Pan & Chiou (2011)	3				
Qui et al. (2013)	3				
Racherla et al. (2012) <sup>12</sup>					
Reimer & Benkenstein (2016)	2	2			
Shamhuyenhazva et al. (2016)	2			2	
Shamhuyenhazva et al. (2016)	5				
Smith et al. (2005)	1				1
Tsang & Prendergast (2009)	1			1	1
Tsao & Hsieh (2015)	3	2		5	
van der Heide & Lim (2016)	1	1		10	
Wang, S. et al. (2015)	3	1		3	
Wang, X. et al. (2015)	2	1		3	
Willemsen et al. (2012) <sup>13</sup>		2		2	
Wu et al. (2017)	1	1			
Xie et al. (2011)	1	1			
Xu (2014)	3				

Das Problem der Messung von Vertrauen mithilfe von Gründen für Vertrauen liegt in der Verwechslung von Vertrauen mit Indikatoren für Vertrauen. Mit Gründen für Vertrauen als Items entsteht ein Zirkelschluss, da es sich sowohl bei der zu erklärenden (abhängigen) Variable als auch bei den erklärenden (unabhängigen) Variablen (häufig um Gründe für Vertrauen handelt. Später passiert diese Verwechslung noch einmal, wenn einzelne Items als Gründe für Vertrauen dargestellt werden, es sich dabei aber um Dimensionen des Vertrauens handelt. Der Zusammenhang ist dann statistisch erkennbar, aber theoretisch „falschherum“ interpretiert. Es lässt sich außerdem vorstellen, dass Leser dem Bewerter vertrauen, aber aus verschiedenen Gründen. Wenn die für ihn gültigen Gründe nicht als Item

---

11 6 Items nicht genannt

12 Anzahl und Items nicht genannt

13 2 Items nicht genannt

angezeigt würden, würde seine Ablehnung der anderen Gründe hier missverstanden werden als kein Vertrauen.

Nun soll es darum gehen, wie die bisher erhobenen Faktoren theoretisch zu verstehen sind. Ich schaue danach, wie der Zusammenhang der Faktoren mit Vertrauen interpretiert werden kann. Dafür gehe ich die einzelnen Elemente des Vertrauensmodells durch und suche in der integrativen Literaturübersicht nach den entsprechenden Faktoren.

Zum ersten Element, der Komplexität der modernen Welt, wurde im Rahmen des Forschungsfelds bisher nichts untersucht. Dies verwundert nicht, da es sich hier um eine sehr weitgehende, allgemeine Voraussetzung handelt, die nicht nur die Nutzung von Online-Bewertungen betrifft. Genauso verhält es sich mit den strukturellen Voraussetzungen für Vertrauen, sowie dem Resultat und den Konsequenzen des Vertrauens. Diese Elemente betreffen keine Faktoren, die das Vertrauen in den Bewerter beeinflussen und sind wohl daher nicht Teil der Studien aus der Literaturübersicht.

Das ungenügende Wissen des Lesers spielt in der bisherigen Forschung bei drei Faktoren eine Rolle. Diese Faktoren beziehen die sich auf verschiedene Objekte: Die *Markeneinstellung des Lesers* bezieht sich auf den Anbieter, die *Kompetenz des Lesers* auf den Leser und die *Übereinstimmung mit der Bewertung mit externen Quellen* auf externe Akteure. Alle drei Faktoren stehen in einem theoretischen Zusammenhang mit dem Wissen des Lesers. Dieses muss laut dem theoretischen Modell ungenügend sein, damit Vertrauen für den Leser überhaupt relevant wird. Bei der *Markeneinstellung des Lesers* stellt sich daher die theoretische Frage, ob z. B. bei einer positiven Einstellung des Lesers gegenüber der Marke des Bewerteten der Leser überhaupt Bewertungen lesen muss. Es könnte sein, dass der Leser seine positive oder negative Markeneinstellung direkt auf das Bewertete überträgt und daher über genügend Informationen verfügt, um zu handeln. Dagegen braucht es eine niedrige *Kompetenz des Lesers* in Bezug auf das Bewertete, damit Vertrauen theoretisch notwendig wird. Denn eine hohe Kompetenz des Lesers hieße, dass er genug über das Bewertete weiß und daher keinen Bewertern vertrauen „muss“. Dasselbe gilt für den Faktor *Übereinstimmung der Bewertung mit externen Quellen*. Wenn der Leser Informationen zum Bewerteten von externen Akteuren

---

erhält, hat er genügend Wissen. Bei allen drei Faktoren zeigen die empirischen Ergebnisse keinen Zusammenhang mit Vertrauen bzw. einen positiven Zusammenhang, wenn die Bewertung mit dem Wissen übereinstimmt, das der Leser hat. Hier wird genau genommen nur gemessen, ob der Leser seine Informationen bestätigt sieht – vertrauen muss er dafür nicht. Nach meinem oben dargelegten Verständnis handelt es sich bei den hier untersuchten Phänomenen, je nach Ausprägung, daher nicht um Vertrauen. Nur wenn eine negative *Markeneinstellung des Lesers*, eine geringe *Kompetenz des Lesers* und keine *Übereinstimmung der Bewertung mit externen Quellen* vorliegen, kann Vertrauen überhaupt auftreten.

Zur Risikowahrnehmung des Lesers wurden zwei Faktoren untersucht. Bei einem hohen *Allgemeinen Vertrauen des Lesers* liegt die Überlegung nahe, dass der Leser nicht über Risikobewusstsein verfügt. Es kann sein, dass der Leser in Vertrautheit lebt, d. h. kein Risiko wahrnimmt und sich für ihn daher das Problem des Vertrauens gar nicht stellt. Der empirisch uneindeutige Zusammenhang kann daher nicht als Vertrauen verstanden werden. Dagegen ist eine skeptische *Einstellung des Lesers gegenüber Online-Bewertungen* theoretisch als Risikobewusstsein interpretierbar. Die generell skeptische Einstellung hat nach bisherigen Studien meist keinen Zusammenhang mit Vertrauen. Das ist theoretisch einleuchtend, da diese Einstellung nur das Risikobewusstsein anzeigt. Das wiederum sagt aber noch nichts darüber aus, ob der Leser dem konkreten Bewerter vertraut oder nicht.

Einige der erhobenen Faktoren lassen sich als Dimensionen des Vertrauens verstehen. Die *Ausführlichkeit der Bewertung* und die *Ausgewogenheit der Bewertung* gehören zur Faktenselektivität, die *Argumentation der Bewertung*, die *Attribuierung* und die *Übereinstimmung der Zahlen- und Textbewertung* gehören zur Dimension der expliziten Bewertungen. Die meisten Studien stellen für diese Faktoren jeweils einen positiven Zusammenhang mit Vertrauen fest. Das spricht dafür – oder zumindest nicht dagegen –, dass es sich bei diesen Faktoren um Dimensionen des Vertrauens handelt. Die Autoren messen hier als unabhängige Variable Teile ihrer abhängigen Variable, d. h. auf beiden Seiten dasselbe. Das erklärt den positiven Zusammenhang.

Bei den meisten bisher erhobenen Faktoren handelt es sich nach dem theoretischen Modell um Gründe für Vertrauen. Einige der Faktoren lassen sich als

spezifische Gründe verstehen, d. h. sie werden aus bestätigten Vertrauenserwartungen abgeleitet. Bisher wurden keine Faktoren erhoben, die zur Gruppe der Gründe der Bestätigung durch Eigenerfahrung gehören. Das hängt wohl damit zusammen, dass Leser auf einer Bewertungsplattform meist auf Online-Bewertungen von Bewertern zurückgreifen, die sie selbst nicht kennen. Eine Bestätigung durch selbst beobachtete Erfahrungen anderer wurde als Grund jedoch bereits untersucht, nämlich mit dem Faktor *Eingebundenheit des Bewerbers in Community*. Hier zieht der Leser Schlüsse auf die Vertrauenswürdigkeit des Bewerbers aus der Beziehung des Bewerbers mit der Community. Auch die Bestätigung durch mitgeteilte Erfahrungen anderer wurde untersucht. Bei den Faktoren *Warnhinweis der Plattform* und *Bewertung der Bewertung* kann der Leser die Erfahrungen der Plattform bzw. der Community mit dem Bewerber als Grund heranziehen, um über dessen Vertrauenswürdigkeit zu entscheiden.

Bei den meisten untersuchten Faktoren handelt es sich aber um unspezifische Gründe. Das ist dadurch erklärbar, dass es sich beim Vertrauen eines Lesers in einen Bewerber meist um den Beginn einer kurzen Vertrauensbeziehung handelt. Insofern kann der Leser nicht auf Vertrauenserwartungen zurückgreifen, die sich bereits bestätigt haben. Die Vertrauensbeziehung steht noch am Anfang, Gründe für oder gegen das Vertrauen können nicht aus bestätigten Vertrauenserwartungen abgeleitet werden. Einige Faktoren behandeln den Rückschluss des Lesers von der Mitteilung, d. h. der Bewertung, auf die Selektion des Bewerbers: *Emotionalität der Bewertung*, *Fotos des Bewerteten*, *Qualität der Bewertung*, *Rechtschreibung und Grammatik der Bewertung*, *Unmittelbarkeit der Bewertung*, *Valenz der Bewertung (negativ)* und *Valenz der Bewertung (positiv)*. Auch *Kompetenz des Bewerbers* gehört in diese Gruppe, wenn die Mitteilung der Kompetenz über die Bewertung erfolgt. Wenn die Mitteilung der Kompetenz dagegen über den Kontext erfolgt, dann zählt *Kompetenz des Bewerbers* zu den unspezifischen Gründen der Gruppe „Rückschluss von Kontext der Mitteilung auf Selektion“. Dazu gehören auch *Ähnlichkeit des Bewerbers*, *Erfahrung des Bewerbers*, *Identifikation des Bewerbers*, *Profildfoto des Bewerbers*, *Soziale Nähe des Bewerbers* und *Unabhängigkeit des Bewerbers*. In diese Gruppe gehört außerdem eine Reihe von Faktoren, in denen der Bewerber zunächst anderen Akteuren vertrauen muss.

Das Vertrauen in diese Akteure ermöglicht ihm wiederum, deren Selektion zu übernehmen, um dem Bewerter zu vertrauen. Diese Faktoren, die als Gründe durch Verlagerung zu verstehen sind, sind: *Bewerterstatus, Qualität der Plattform, Anzahl der Bewertungen, Zugehörigkeit des Lesers, und Vertrauenswürdigkeit der Plattform, Übereinstimmung mit anderen Bewertungen, Vertrautheit mit Plattform, Erfahrung des Lesers mit Online-Bewertungen, Gütesiegel für Plattform und Reputation der Plattform*. Die Ergebnisse zum Zusammenhang von Vertrauen mit all diesen Gründen sind bisher häufig nicht eindeutig.

Das kann an weiteren Faktoren liegen, die mit diesen Gründen interagieren. Unter den bisher untersuchten Faktoren finden sich einige, bei welchen theoretisch denkbar ist, dass sie mit verschiedenen Elementen des Vertrauensmodells interagieren, z. B. mit den Faktoren, die das Risikobewusstsein, die Dimensionen oder eben die Gründe des Vertrauens betreffen. Diese Faktoren beziehen sich meist auf den Leser selbst: *Alter des Lesers, Bildung des Lesers, Erfahrung des Lesers mit Online-Shopping, Ethnizität des Lesers, Fokus des Lesers, Geschlecht des Lesers, Internetnutzung des Lesers, Involviertheit des Lesers, Kontrollillusion des Lesers, Kultur des Lesers, Motivation des Lesers, Orientierung des Lesers an anderen, Personal Sense of Power des Lesers, Selbstoffenbarung des Lesers und Trägheit des Lesers*. Aber auch der Faktor *Typ des Bewerteten* kann mit anderen Faktoren interagieren. So kann er z. B. einen Einfluss auf das Risikobewusstsein haben. Der Leser kann aufgrund des hohen Konkurrenzdrucks davon ausgehen, dass Restaurantbewertungen gefälscht sein können, für Filmbewertungen kommt ihm das aber nicht in den Sinn. Oder aber er hält die Übernahme von Fremdselektionen gerade bei Filmen für sehr riskant, weil es wenig objektive Kriterien gibt und sich die Geschmäcker der Kinogänger stark unterscheiden können. Diese unterschiedlichen Einflüsse auf das Risikobewusstsein könnten auch das Ergebnis erklären, warum das Vertrauen in Bewertungen des einen Typs größer ist als in Bewertungen eines anderen Typs. Zu möglichen Interaktionen lässt sich außerdem sagen, dass auch alle anderen Faktoren sich untereinander beeinflussen können, z. B. indem sie einander verstärken oder sich gegenseitig aufheben.

### 3.8 FAZIT

In diesem Kapitel schlug ich ein Modell zum theoretischen Verständnis von Vertrauen in Online-Bewertungen vor. Das Modell von Kohring (2004) von Vertrauen in Journalismus ließ sich gut auf den Gegenstand der Online-Bewertungen übertragen. Mithilfe des Modells konnten die theoretischen Fragen, die aus der Literaturübersicht folgten, beantwortet werden. Die allgemeine Definition von Vertrauen und die spezifischere Bestimmung von Vertrauen im Kontext von Online-Bewertungen schließen eine theoretische Lücke der bisherigen Forschung. Außerdem konnte die Bewertung als Objekt des Vertrauens ausgeschlossen und dagegen der Bewerter als das zentrale Objekt des Vertrauens des Lesers ausgemacht werden. Durch das Modell konnten außerdem die verschiedenen Items zur Messung von Vertrauen theoretisch genauer differenziert werden in Dimensionen von und Gründe für Vertrauen. Darüber hinaus konnte jedem einzelnen Faktor eine theoretisch fundierte Position im Vertrauensprozess zugewiesen werden. Das umfassende Vertrauensmodell zeigte sich abstrakt und gleichzeitig anwendbar genug, um alle bisherigen Ergebnisse zu integrieren und theoretisch zu deuten.

Wie aus dem bisherigen Forschungsstand hervorgeht, gab es vorher keine umfassende Vertrauens Theorie, die überzeugend angewandt wurde. Verschiedene Argumente sprechen dafür, ein umfassendes Modell zu Vertrauen in Online-Bewertungen auf der Basis von Kohring (2004) aufzubauen – wie hier geschehen. Zum einen entwickelt Kohring (2004) sein Vertrauensmodell in Auseinandersetzung mit anderen Vertrauenskonzepten, auf die sich auch einige Literatur aus dem hiesigen Korpus bezieht. Kohring (2004) führt diese Forschungsansätze in ein schlüssiges Konzept zusammen. Er betrachtet das Phänomen des Vertrauens soziologisch und geht damit über die psychologisch geprägte Glaubwürdigkeitsforschung hinaus. McLeod et al. (2017) meinen zu einer solchen Verknüpfung von Vertrauens- und Glaubwürdigkeitsforschung:

„The connection to a broader theory of trust is certainly an advancement that could help situate findings regarding perceptions of media within sociological theory [...]. At the present, media credibility work appears dominant in the U.S, while scholarship examining media trust is more evident in work by scholars outside the U.S. (Hellmueller, 2012). The ability for scholars to confidently draw on both traditions to make observations and inferences related to perceptions of media performance would be greatly enhanced by work to both distinguish and unify these approaches.“ (S. 45)

---

Im Gegensatz zu den behandelten Studien setzt sich Kohring ausführlich damit auseinander, was Vertrauen überhaupt ist. Diese Auseinandersetzung erfolgt auf einem so hohen abstrakten Niveau, dass das Modell auch allgemeiner anwendbar ist. So kann das Modell für Vertrauen in Journalismus sich auch auf andere Forschungsgegenstände aus dem System Öffentlichkeit übertragen lassen, wie z. B. Bewertungen. Obwohl Kohring (2004) nicht speziell auf den Online-Kontext eingeht, ist das Modell auch hierfür brauchbar. Möglich ist beides dadurch, dass sich Kohrings (2004) Modell auf Luhmanns Systemtheorie als makrotheoretische Basis stützt. Diese fokussiert auf eine funktionale Analyse (Borchers, 2015, S. 207-209). Damit tritt die konkrete Erscheinung eines Phänomens in den Hintergrund. Solange Journalismus und Bewertungen als funktional äquivalent verstanden werden, passt das Modell zu beiden Phänomenen. Ob Bewertungen schließlich online oder offline stattfinden, ändert auch nichts auf der theoretischen Ebene der Relation des Vertrauens. Der Unterschied zeigt sich nur auf der Ebene des konkreten Gegenstands, auf dem z. B. Gründe für Vertrauen sichtbar und greifbar werden. Folglich können auch Erkenntnisse aus der vorliegenden Studie zu Vertrauen in Online-Bewertungen wieder auf andere Gegenstände rückübertragen werden.

Mithilfe des Modells zu Vertrauen in Online-Bewertungen konnten die theoretischen Fragen, die aus dem Forschungsstand folgten, geklärt werden. Außerdem gab das Modell Hinweise darauf, warum die Ergebnisse zu den Faktoren, die als Dimensionen des Vertrauens zu verstehen sind, eindeutig sind. Nicht geklärt werden konnten bisher die empirischen Fragen danach, warum die Ergebnisse zu einem großen Teil der Faktoren widersprüchlich sind. Im Folgenden gilt es daher, das Vertrauen in Online-Bewertung auf der Basis der bisherigen Forschung und im Rahmen des theoretischen Modells weiter empirisch zu explorieren. Nach diesem Abschluss der theoretischen Zuweisung, kann außerdem geprüft werden, inwiefern sich die bisher gefunden Faktoren empirisch differenzieren lassen, um die widersprüchlichen Ergebnisse klären zu können.



## **4 EMPIRISCHE EXPLORATION: VERTRAUEN IN ARZTBEWERTUNGEN**

### **4.1 EINFÜHRUNG**

Nach der integrativen Literaturübersicht und einem Vorschlag für ein theoretisches Modell folgt nun die empirische Exploration zu Vertrauen in Online-Bewertungen. Damit werden die bisherigen Ergebnisse differenziert, um widersprüchliche Ergebnisse bisheriger Studien zu klären. Außerdem wird das theoretische Modell am Beispiel von Arztbewertungen empirisch untersucht.

Ich stelle erst den Gegenstand der Arztbewertung vor und erläutere die gesellschaftliche und wissenschaftliche Relevanz dieses Fallbeispiels. Dann präsentiere ich die Methoden zur Datenerhebung und -auswertung. Anschließend werden die Auswahl der Teilnehmer und der Bewertungen erläutert. Im Hauptteil des Kapitels stelle ich die Ergebnisse der empirischen Exploration vor. Dafür ordne ich diese in der vom theoretischen Modell vorgegebene Reihenfolge und stelle Bezüge zum bisherigen Forschungsstand her. Am Ende des Kapitels folgt das Fazit zur empirischen Studie.

Die Darstellung der Ergebnisse wird unterfüttert mit beispielhaften Zitaten der Teilnehmer. Für die Zitation gebe ich die pseudonymen Vornamen der Teilnehmer sowie die Absatznummer im jeweiligen Transkript an. Zur besseren Lesbarkeit und Verständlichkeit erfolgt die Wiedergabe der Zitate ohne Füllworte der gesprochenen Sprache, wie z. B. „eigentlich“, „also“, „halt“, „einfach“, „auch“, „wieder“, „irgendwie“, „dann“, „sage/ glaube/denke/meine ich“, „mal“, „manchmal“, „immer“, „so“, „jetzt“, „schon“, „ziemlich“ oder „vielleicht“.

### **4.2 RELEVANZ DES FALLBEISPIELS ARZTBEWERTUNG**

In diesem Unterkapitel stelle ich den Gegenstand der Arztbewertung vor. Dafür gehe ich kurz auf die Geschichte der Arztbewertungen ein, erkläre dann die Bedeutung von Arztbewertungen für die Gesellschaft und gebe einen Einblick in die bisherige wissenschaftliche Beschäftigung mit Arztbewertungen.

Arztbewertungen sind Online-Bewertungen von Ärzten (Emmert et al., 2011, S. 232; Reimann & Strech, 2010, S. 333). Ärzte können anhand verschiedener Kriterien beurteilt werden, meist mit Hilfe von Schulnoten oder Punkten und einem

Freitextfeld (Damman et al., 2010; Emmert et al., 2012; Emmert et al., 2011, S. 229; Kofahl & Horak, 2010, S. 133; Reimann & Strech, 2010; Schaefer et al., 2011; Strech & Reimann, 2012). Dazu werden vom Bewertungsportal häufig allgemeine Informationen zur Praxis veröffentlicht, z. B. die Adresse und die Öffnungszeiten (Emmert et al., 2011, S. 230–232). Der Leser kann in Arztbewertungsportalen nach einem bestimmten Namen, nach einem Arzt in einem bestimmten Ort(steil) oder nach einer bestimmten Fachrichtung suchen.

Online-Bewertungen von Ärzten gibt es in Deutschland seit 2001 (Friedl & Wietlisbach, 2008, S. 29), in anderen Ländern sind sie schon länger bekannt (Sander, 2007). Ab 2007 kam es „zu einem regelrechten Boom von Arztbewertungsseiten“ (Friedl & Wietlisbach, 2008, S. 29). Die Plattformen werden von verschiedenen Akteuren betrieben. In Großbritannien gibt es ein staatlich finanziertes Portal für Arztbewertungen (Greaves et al., 2012). In den USA betreibt eine der größten Versicherungen ein eigenes Arztbewertungsportal (Raniszewski Herrera, 2006, S. 1). In Deutschland existiert mit dem *Arztnavi* ein einziges nichtkommerzielles Arztbewertungsportal. Dabei handelt es sich um ein Angebot der zwei großen deutschen Krankenkassen *AOK* und *Barmer GEK* und der *Bertelsmann Stiftung* (DEGAM, 2011, S. 1; Stiftung Gesundheit, 2008, S. 2). Alle weiteren Arztbewertungsportale in Deutschland sind kommerziell, d. h. sie arbeiten profitorientiert und finanzieren sich über Werbung der Ärzte.

Arztbewertungen dienen in der folgenden empirischen Exploration als Fallbeispiel für die Untersuchung von Vertrauen in Online-Bewertungen. Meine Wahl fiel auf Arztbewertungen als Beispiel für Online-Bewertungen, weil ihnen eine größere gesellschaftliche und wissenschaftliche Relevanz zukommt als anderen Online-Bewertungen.

Gesellschaftlich relevant sind Arztbewertungen, da Ärzte als Gesundheitsdienstleister verstanden werden können und es sich bei der Gesundheit der Menschen aus normativer Perspektive um ein hohes Gut handelt. Mit der Auswahl von Produkten wie Smartphones oder Dienstleistungen wie beispielsweise Thai-Massagen sind vor allem zeitliche und finanzielle Risiken verbunden. Mit dem Arztbesuch können darüber hinausgehend gesundheitliche Risiken verbunden sein. Davon ist zwar erst einmal nicht auszugehen, da der Beruf des Arztes lizenziert ist

und starken gesetzlichen Voraussetzungen und Beschränkungen unterliegt – auszuschließen ist ein gesundheitliches Risiko aber nicht. Zudem lässt sich sagen, dass die Wahl des Arztes für viele Menschen von höherer persönlicher Bedeutsamkeit ist, v. a. weil „ärztliche Dienstleistungen durch einen hohen Anteil an Vertrauenseigenschaften geprägt“ sind (Hubert et al., 2011, S. 54f).

Die hohe gesellschaftliche Relevanz zeigt sich auch auf empirischer Ebene. Bei *Jameda*, dem größten Arztbewertungsportal Deutschlands (Kühl, 2018), stiegen die Nutzerzahlen innerhalb eines Jahres um zwölf Prozent auf 6,48 Millionen Nutzer (Mh, 2017). Im Jahr 2017 wurden 28 Prozent mehr Bewertungen abgegeben als 2016, insgesamt gibt es mittlerweile 1,9 Millionen Arztbewertungen auf *Jameda* (Mh, 2017). Neben der Anzahl der Bewerter weisen auch die Zahlen zu den Lesern einen Anstieg auf. Hubert et al. (2011, S. 54f) weisen darauf hin, dass jeder Mensch einmal Patient ist. Obwohl die Wahl eines Arztes daher für jeden Menschen irgendwann notwendig ist, bekommt er bei der Suche nach einem geeigneten Arzt wenig Unterstützung. Diese mangelnde Hilfestellung entspricht nicht dem Wunsch vieler Menschen nach „mehr Transparenz im Gesundheitswesen“ (Hubert et al., 2011, S. 54f). Online-Bewertungen von Ärzten füllen diese Lücke. Laut Blawat (2012) nutzen zwischen zehn und 20 Prozent der Bevölkerung in Deutschland Arztbewertungen. Einer neueren Umfrage von Emmert et al. (2013) zufolge haben 25 Prozent der Befragten schon einmal eine Arztbewertung gelesen. Akteure aus dem Gesundheitssektor sind sich darüber einig, dass die Nutzung von Arztbewertungen weiter steigen wird (Schaefer et al., 2011, DEGAM, 2011, S. 2, Sander & Schult, 2010, S. 46). Das wird u. a. darauf zurückgeführt, dass immer mehr ärztliche Leistungen vom Patienten selbst bezahlt werden müssen. Dies führt dazu, dass Patienten vor der Behandlung mehr über deren Erfolg wissen wollen (Emmert et al., 2011, S. 229; Kofahl & Horak, 2010, S. 110; Raniszewski Herrera, 2006, S. 1).

Ein weiterer Indikator für die gesellschaftliche Relevanz von Arztbewertungen sind die häufigen Diskurse darüber. Arztbewertungen werden immer wieder stark diskutiert, sowohl von Arzt- oder Patientenorganisationen, als auch in den Massenmedien (DEGAM, 2011, S. 1; Emmert et al., 2011, S. 229; Grün, 2011; Qaseem et al., 2010; Reimann & Strech, 2010, S. 333; Sander & Schult, 2010, S. 46; Thadeusz, 2009). Ärzte befürchten, dass Online-Bewertungen ihnen schaden, weil sie

gefälscht oder unsachlich sind (Coombes, 2009, S. 679; Friedl & Wietlisbach, 2008, S. 30; Kesmodel, 2005; Sander, 2007, Schwenkenbecher & Hütten, 2018). Andere hoffen, dass Online-Bewertungen ein Feedback-Mechanismus sind, der die Gesundheitsversorgung verbessert (DEGAM, 2011, S. 2; Bacon, 2009, S. 688; Baker, 1990, S. 490; Damman et al., 2010, S. 345; Emmert et al., 2011, S. 229; Friedl & Wietlisbach, 2008, S. 33; Jain, 2010, S. 6, 2007; Kofahl & Horak, 2010, S. 105-106, 113; Kolata, 2005; Lagu et al., 2010, S. 942, 945; Raniszewski Herrera, 2006, S. 2; Reimann & Strech, 2010, S. 344; Sander, 2007). Letzteres wurde bisher nicht empirisch bestätigt. Dagegen konnte schon gezeigt werden, dass Arztbewertungen die Leser bei der Wahl eines Arztes beeinflussen (Huber et al., 2011, S. 85; Kühl, 2018). Erst vor kurzem flammte das Interesse der Medien an Arztbewertungen wieder auf, angefacht durch das aktuelle Gerichtsurteil nach einem längeren Prozess gegen *Jameda*. Bei diesem Prozess „standen vor allem die Zweifel an der Neutralität des Portals im Vordergrund. Denn bei dem Portal gibt es zwei Klassen von Ärzten: Jene, die ein Premium-Paket abgeschlossen haben - und die Nichtzahler.“ (Jameda reagiert auf BGH-Urteil, 2018). Laut dem Urteil hat *Jameda* „die für Bewertungsportale gebotene Neutralität verlassen, weil es mit seinem Geschäftsmodell die für Werbung bezahlenden Ärzte begünstigt“ (Jameda reagiert auf BGH-Urteil, 2018).

Die hohe gesellschaftliche Relevanz von Arztbewertungen spiegelt sich bisher allerdings kaum wider in der wissenschaftlichen Beschäftigung damit. Bisher gibt es nur wenige Studien zu Arztbewertungen. In den meisten Studien geht es um den Inhalt von Arztbewertungen aus verschiedenen Perspektiven (z. B. Emmert et al., 2011, 2013, 2014; Geierhos & Bäumer, 2015; Geierhos et al., 2015b, c; Lagu et al., 2010; Strech & Reimann, 2012). Daneben gibt es einen Forschungsstrang zur Nutzung von Arztbewertungen (z. B. Beck, 2014; Bidmon et al., 2014; Borchers & Krömer, 2015; Geierhos et al., 2015a; Emmert et al., 2013; Reinhardt et al., 2017; Terlutter et al., 2014). Außerdem beschäftigt sich ein Teil der Wissenschaft mit Risiken und Nutzen von Arztbewertungen für Ärzte (z. B. Däumler & Noah, 2015; Halling et al., 2012; Hesse, 2015; Schwarz & Schäfer, 2015). Die meisten dieser Studien stammen aus dem Management (z. B. Emmert et al., 2011; Friedl & Wietlisbach, 2008; Galizzi et al., 2012; Gao et al., 2012), einige auch aus den

---

Gesundheitswissenschaften (z. B. Bleckmann et al., 2008; Greaves et al., 2012a; Greaves et al., 2012b; Kadry et al., 2011; Schaefer et al., 2011; Schwarz et al., 2012). Bisher sind mir nur zwei Studien zu Arztbewertungsportalen aus der Medien- und Kommunikationswissenschaft bekannt (Borchers & Krömer, 2015; Reinhardt et al., 2017).

Mit Grabner-Kräuter und Waiguny (2015) findet sich in der bisherigen Forschung nur eine Studie, die sich mit Vertrauen in Arztbewertungen beschäftigt. Denn in der Forschung zu Vertrauen in Online-Bewertungen standen bisher Bewertungen zu anderen Produkten und Dienstleistungen im Fokus. Wie in der integrativen Literaturübersicht deutlich geworden ist, wird die Forschung dominiert von Online-Bewertungen zu Hotels und Digitalkameras. Dienstleistungen aus anderen gesellschaftlichen Teilbereichen, wie dem Gesundheitssystem, sind bisher kaum erforscht. Auch Beldad et al. (2010) stellen fest, dass es eine „*apparent imbalance between studies on trust in e-commerce and trust in non-commercial transactions*“ (S. 877) gibt. Als Beispiel für nicht-kommerzielle Transaktionen nennen sie die Bereiche des e-government und der e-health (Beldad et al., 2010, S. 877). Obwohl Vertrauen gerade in diesen Bereichen sehr wichtig ist, gibt es, verglichen mit den Studien in kommerziellen Bereichen, wenig Forschung dazu (Beldad et al., 2010, S. 877). Sie schließen daraus: „*This indicates that trust in the aforementioned area is still on its infancy phase.*“ (Beldad et al., 2010, S. 877). In den wenigen Studien aus der Online Health Forschung geht es bisher vor allem um allgemeine Gesundheitsinformationen, nicht um Bewertungen von Ärzten.

Bei Arztbewertungen handelt es sich also um einen Gegenstand, der gesellschaftlich relevant ist, aber bisher noch nicht ausreichend wissenschaftlich erforscht wurde. Daher eignen sich Arztbewertungen besonders als Fallbeispiel, um das Vertrauen in Online-Bewertungen zu untersuchen. Damit wird das Spektrum der bisherigen Forschung um einen relevanten Gegenstand erweitert.

Allerdings muss bei Arztbewertungen als Fallbeispiel stets bedacht werden, dass sie sich in zwei Punkten von vielen anderen Online-Bewertungen unterscheiden. Zum einen handelt es sich, wie bereits dargestellt, bei der Übernahme der Bewertung eines Arztes um eine riskantere Entscheidung als z. B. bei der Übernahme der Bewertung eines Smartphones. Zum anderen werden Arztbewertungen, wegen

des sehr sensiblen Themas der Bewertung und der brisanten Daten zur Gesundheit, meist von anonymen Bewertern verfasst. Die Funktionsweise von Vertrauen wird hier also in einer sehr zugespitzten, hochriskanten Situation untersucht. Daraus lassen sich Rückschlüsse ziehen auf Situationen mit weniger hohen Risiken. Daher sind die Ergebnisse zum Vertrauen in Arztbewertungen auch auf Vertrauen in Online-Bewertungen anderer Art übertragbar.

### **4.3 METHODISCHES VORGEHEN: THINK-ALLOUD, PZI UND QUAL. IA**

In diesem Unterkapitel stelle ich die Methoden der Datenerhebung und -auswertung vor. Zunächst erkläre ich den Forschungsaufbau mit Think-Aloud-Aufgabe und darauffolgendem Problemzentrierten Interview. Im Anschluss erläutere ich mein Vorgehen bei der Qualitativen Inhaltsanalyse der erhobenen Daten. Zum Schluss wird auf die Gründe für das präsentierte Vorgehen eingegangen.

Nach einer Begrüßung und einigen einführenden Worten (siehe 7.1 Einführung, S. 159) startet die Studie. Die Datenerhebung beginnt mit einer Think-Aloud-Aufgabe (Ericsson & Simon, 1984). Die Think-Aloud-Aufgabe besteht darin, dass der Teilnehmer am Computer zehn Arztbewertungen liest und entscheidet, ob er die vorliegende Arztbewertung einem Freund weitergibt, der einen Arzt sucht (siehe Anhang, 7.2 Aufgabenstellung für Think-Aloud, S. 159). Dabei soll der Teilnehmer laut sagen, was ihm durch den Kopf geht. Während der Bearbeitung der Think-Aloud Aufgabe notiere ich Auffälligkeiten und erinnere den Teilnehmer ggf. daran, weiter laut zu sprechen. Die Bearbeitung dieser Aufgabe dauerte in dieser Studie durchschnittlich 15 Minuten.

Der Aufgabenstellung liegt die von Kohring überprüfte Annahme zu Grunde, dass der Teilnehmer nur Informationen weitergibt, deren Urheber er vertraut (Kohring, 2004, S. 248). Es ist aber wichtig, dass der Begriff des „Vertrauens“ nicht vor der Bearbeitung der Aufgabe auftaucht. Damit wird verhindert, dass die Forschungsanlage Risikobewusstsein beim Teilnehmer induziert. Denn bei bisherigen Studien lässt sich methodisch kritisieren, dass durch die Frage nach Vertrauen bzw. Glaubwürdigkeit überhaupt erst Risiko induziert wird. Mit dieser methodisch bedingten Risikoinduktion ist gemeint, dass sich für einen Leser vielleicht gar nicht die Frage nach der Glaubwürdigkeit einer Mitteilung stellt, wenn er nicht danach gefragt wird. Erst durch diese Frage wird der Leser auf die Möglichkeit gestoßen,

einer Mitteilung nicht zu vertrauen. Dem Leser wird also erst eine Situation als ein Problem vergegenwärtigt, für das er dann eine Lösung finden muss. Die Reaktion des Lesers auf diese Frage wäre dann ein unbrauchbares Artefakt.

Die Aufgabenstellung mit der Weitergabe der Mitteilung an einen Freund führte dazu, dass alle Teilnehmer ihr Vertrauen in die Bewertung und die Gründe dafür thematisierten. Bei der Weitergabe der Bewertungen verwendeten die Teilnehmer verschiedene Formulierungen wie „vertrauen“ bzw. „Vertrauen“ oder „Vertrauenswürdigkeit“ (z. B. Dorothee: 55, Ina: 131, Quentin: 21) und „glauben“ bzw. „glaubwürdig“ (Bernd: 136, Charlotte: 75, Quentin: 85, Renata: 114). Teilweise wurden die Begriffe auch gleichzeitig verwendet, z. B.: „schenke ich dieser Bewertung nicht überwiegend Glauben und vertraue dann auf die anderen.“ (Renata: 114). Weitere von den Teilnehmern verwendete Begriffe waren sich auf die Bewertung zu „verlassen“ (Bernd: 7, Nadja: 224, Paula: 67), oder aber sie „ernst“ (Charlotte: 73) bzw. „für bare Münze“ (Ina: 123) zu nehmen. All diese Formulierungen können als Hinweis darauf aufgefasst werden, dass die Aufgabenstellung richtig gewählt wurde und gut funktioniert hat. Mit der Weitergabe der Bewertung an einen Freund konnte das Wort „Vertrauen“ und damit die Induktion des Risikobewusstseins bei den Teilnehmern vermieden und trotzdem eine Einsicht in die Vertrauenszuschreibung der Teilnehmer gewonnen werden.

In Ausnahmefällen waren Anlaufschwierigkeiten zu bemerken. Selten musste ich auf Nachfrage zu Beginn der Think-Aloud-Aufgabe erläutern, dass es nicht um die Weiterempfehlung des bewerteten Arztes, sondern um die Weiterempfehlung der Bewertung geht. Ab und zu musste ich die Teilnehmer darauf hinweisen, weiter laut zu sprechen, weil sie mit der Zeit verstummten. Häufiger zeigte sich hingegen das Phänomen, dass die Teilnehmer zunächst prinzipiell alle Bewertungen weitergeben wollten. Die Teilnehmer gaben an, ihrem arztsuchenden Freund ein „Gesamtbild“ liefern zu wollen (Anna: 66, 144, Charlotte: 49, 65, 83, 85, Julia: 56, Otto: 42, Quentin: 175, Sabine: 22, 88). Hier wird eine Grenze der Aufgabenstellung sichtbar. Wie im theoretischen Modell bereits erwähnt, ist Vertrauen allein nicht immer handlungsleitend. Es kann – trotz Vertrauen in die Mitteilung – Gründe geben, diese nicht weiterzugeben. Genau so kann es – trotz fehlenden Vertrauens in die Mitteilung – Gründe geben, diese weiterzugeben. Deutlich wird dies auch in einer

Aussage von Renata. Sie gibt eine Bewertung nicht weiter, obwohl sie ihr vertraut: „hier würde ich es nicht weiterleiten, um diesen Freund nicht zu bespammen [...] muss ja nicht das Gleiche x-mal schicken.“ (Renata: 62). Das ist ein weiteres Beispiel dafür, dass sich die Weitergabe der Bewertung nicht nur auf Vertrauen bzw. die Nicht-Weitergabe auf Nicht-Vertrauen zurückführen lässt. Die Bewertung mit den meisten Weitergaben als vertrauenswürdigste Bewertung zu interpretieren, wäre daher verkürzt und falsch. Die zahlreichen Kommentare der Teilnehmer zum Vertrauen machen aber deutlich, dass sich die Aufgabenstellung nichtsdestotrotz sehr gut eignet, um das Vertrauen in Online-Bewertungen zu explorieren. Tatsächlich wurden alle zehn Bewertungen schließlich nur von vier Teilnehmern weitergegeben (Charlotte, Franziska, Otto, Sabine). Damit handelten einige Teilnehmer entgegen ihrer eigenen Ankündigung, ein „Gesamtbild“ geben zu wollen. Sie sortierten die Bewertungen im Laufe der Aufgabenbearbeitung doch selbst vor.

Unklar bleiben drei Fälle, in denen die Teilnehmerinnen die Bewertung weiterleiteten, obwohl sie nur Gründe gegen Vertrauen und nicht dafür verbalisieren (Dorothee: 47, Gesa: 53, Ina: 23). Das kann ein Hinweis darauf sein, dass die Teilnehmerinnen sich verklickt haben. Oder aber diese Fälle zeigen einen weiteren Nachteil der Methode. Denn bei Think-Aloud kommt nur auf, was der Teilnehmer von sich aus verbalisiert. Alle Gedanken, die der Teilnehmer hat, aber nicht ausspricht, bleiben dem Forscher während des Think-Aloud unbekannt.

Neben den aufgezeigten Grenzen hat die Methode des Think-Aloud einige Vorteile, die ihre Anwendung für diese Studie legitimieren. Die Methode eignet sich für die Untersuchung, da Faktoren des Vertrauens wirken, ohne dass der Teilnehmer sie auf direkte Nachfrage nachträglich konkret verbalisieren könnte. Beim Think-Aloud sollen die Teilnehmer die Faktoren laufend mitteilen. Für die Wahl dieser Methode spricht zudem, dass hier nicht nur festgestellt werden kann, ob ein Teilnehmer einer Arztbewertung (nicht) vertraut, sondern auch offengelegt werden kann, warum er (nicht) vertraut. Die Think-Aloud-Methode ist außerdem offen genug, um nicht vorhersehbare Aspekte zu erfassen. Dabei ist es von Vorteil, dass der Faktor direkt in dem Moment aufgenommen wird, in dem er relevant wird. So kommt es nicht zu einem Report möglicherweise falscher Erinnerungen oder zu

anderen Verzerrungseffekten bei einer nachgelagerten Erhebung. Mit der Wahl dieser Methode folge ich auch einer Empfehlung von Vasquez (2014, S. 184) für künftige Forschung über Vertrauen in Online-Bewertungen. Sie schreibt:

„it would be useful to approach more systematically the collection of introspective data with both review users and review writers. For example, while there are some studies from the field of communication that examine the features of reviews that readers attend to, most of these are controlled experiments, rather than investigations of actual consumer behavior. A complementary perspective could be to develop a protocol for observing individuals as they read reviews, perhaps combined with introspective reports about which features of reviews they are attending to and which they find most useful in their decision-making.” (S. 184)

Für Think-Aloud spricht außerdem, dass die Methode im Bereich der Online Health Forschung, auch bei älteren Teilnehmern (Bolle et al., 2016) sowie in Verbindung mit Glaubwürdigkeit (Rieh, 2000; Liao & Fu, 2014) bereits erfolgreich angewendet wurde. Bolle et al. (2016) weisen auf die Kombination von Think-Aloud und Interview hin. Zur Untersuchung von Vertrauen in Online-Bewertungen hat auch Llamero (2014) Think-Aloud mit semi-strukturierten Interviews erfolgreich verbunden. Diesem Beispiel folge ich in meiner Forschungsanlage.

Nach der Think-Aloud-Aufgabe führte ich daher ein Problemzentriertes Interview (Witzel, 2012) mit den Teilnehmern. Die Themen des Interviews gab ich durch meinen Leitfaden grob vor (siehe Anhang, 7.4 Leitfaden für Problemzentriertes Interview, S. 171). Der Leitfaden beruht auf meinem Vorwissen aus der integrativen Literaturübersicht und dem theoretischen Modell. Beim Problemzentrierten Interview handelt es sich um eine Variante des Qualitativen Interviews, die Fragen sind also vergleichsweise offen. Das führt dazu, dass sowohl die Forscherin als auch der Teilnehmer zu gleichen Teilen das Gespräch strukturieren (Froschauer & Lueger, 2003). Ich steige in das Problemzentrierte Interview mit Fragen zur Nutzung von Arztbewertungen ein. Dann folgt der offene Teil. Hier spreche ich die Themen an, die mir während der Think-Aloud-Aufgabe aufgefallen sind. Ich bitte den Teilnehmer gezielt um Kommentare dazu. Nach den bisher untersuchten Faktoren des Vertrauens werden die Teilnehmer nicht explizit gefragt. Diese werden nur dann relevant, wenn der Teilnehmer sie selbst als Erklärung für sein Verhalten heranzieht. Das Problemzentrierte Interviews schließt mit einem kurzen standardisierten

Fragebogen zu soziodemografischen Daten des Teilnehmers ab. Insgesamt dauerte das Problemzentrierte Interview durchschnittlich etwa 15 Minuten pro Teilnehmer.

Die Think Aloud Aufgabe wird mit einem Problemzentrierten Interview ergänzt, da die bei Think Aloud verbalisierten Informationen unvollständig sein können (Duncker, 1945, S. 4). Diese Annahme wird durch meine oben bereits geschilderten Beobachtungen bestätigt. Someren et al. (1994, S. 33) meinen, dass kognitive Prozesse manchmal schneller seien, als man sie laut verbalisieren könne und es daher zu ‚Löchern‘ in den Think-Aloud-Protokollen komme. Die Teilnehmer werden im Problemzentrierten Interview daher zu den Kommentaren befragt, die mir während der Bearbeitung der Think-Aloud-Aufgabe aufgefallen sind und mir klärungsbedürftig schienen.

Beim Problemzentrierten Interview steht im Vordergrund, was die Teilnehmer der Untersuchung selbst für wichtig halten (Froschauer & Lueger, 2003, S. 16). Es geht darum zu verstehen „was Menschen in einem sozialen Kontext dazu bringt, in einer bestimmten Weise zu handeln“ (Froschauer & Lueger, 2003, S. 217). Somit passt die Methode zum Forschungsgegenstand, da die subjektive Wahrnehmung des Lesers die Basis für sein Vertrauen in den Bewerter bildet. Mithilfe des Problemzentrierten Interviews konnte ich gezielt Erkenntnisse aus der Think-Aloud-Aufgabe vertiefen. Die Teilnehmer hatten im Problemzentrierten Interview die Möglichkeit, auf die Faktoren des Vertrauens in die einzelnen Bewertungen näher einzugehen und ich konnte systematisch nachfragen.

Die Bearbeitung der Think-Aloud-Aufgabe und das Problemzentrierte Interview wurden aufgenommen und transkribiert<sup>14</sup>. Die Transkripte wertete ich mithilfe der Qualitativen Inhaltsanalyse nach Schreier (2012) aus. Die Qualitative Inhaltsanalyse erlaubt es, theoretische Vorarbeiten mit einer offenen Vorgehensweise zu verknüpfen. Ich folge dabei Schreiers Vorschlag einer Verbindung der „concept-driven“- mit einer „data-driven“-Strategie (2012, S. 125). Ein Teil der Coding-Kategorien wird hierbei schon vor der Analyse festgelegt, während zusätzliche Kategorien weiterhin aus den Daten gewonnen werden können. Für das

---

14 An dieser Stelle bedanke ich mich ganz herzlich bei der studentischen Mitarbeiterin Juliane Queck und dem studentischen Mitarbeiter Marius Neuberger, die die Transkription der Aufnahmen sehr sorgfältig, sehr gewissenhaft und sehr zuverlässig erledigten.

---

Codebuch entnehme ich die Kategorien aus der integrativen Literaturübersicht und dem theoretischen Modell (siehe Anhang, 7.5 Codebuch für Qualitative Inhaltsanalyse, S. 172). Das Codebuch folgt im Aufbau der Struktur des theoretischen Modells von Vertrauen in Arztbewertungen. Schreier (2012) formuliert für die Qualitative Inhaltsanalyse den Anspruch, transparent, systematisch und nachvollziehbar zu verfahren. Daraus folgt ihre Forderung, dass die Qualitative Inhaltsanalyse generell erst nach abgeschlossener Datenerhebung erfolgt. Während der gesamten Analyse bleibt das Codebuch offen für weitere Kategorien, die sich induktiv aus den Daten ergeben. Werden solche Kategorien gefunden, müssen alle Transkripte noch einmal mit dem neuen Codenbuch codiert werden. Für die Strukturierung der verbalen Daten benutze ich die speziell für die Qualitative Inhaltsanalyse entwickelte Software MaxQDA.

Die Qualitative Inhaltsanalyse eignet sich sowohl für die Auswertung der Transkripte der Think-Aloud-Aufgabe als auch der Transkripte der Problemzentrierten Interviews. Think-Aloud-Daten werden zwar von den Vätern der Methode Ericsson und Simon stets mit der von ihnen entwickelten Protokollanalyse ausgewertet. Bei ihnen steht allerdings die Analyse von Prozessen im Vordergrund, nicht das Ergebnis eines Prozesses (Someren et al., 1994, S. 14). In der Forschung mit Think-Aloud ist es daher nicht unüblich, die Auswertung der erhobenen Daten an den eigenen Gegenstand und die eigene Fragestellung anzupassen (Someren et al., 1994; MacNeela et al., 2010; Pottier et al., 2010; McCabe et al., 2011; Ferguson et al. 2012). Daher ist eine Auswertung dieser Daten mit der Qualitativen Inhaltsanalyse prinzipiell möglich. Für Problemzentrierte Interviews sind prinzipiell vielfältige Auswertungsmethoden denkbar. Für diese Studie erwies sich die Qualitative Inhaltsanalyse als sehr brauchbar. In das Codebuch konnten alle Ergebnisse aus der bisherigen empirischen Forschung sowie die Ergebnisse aus dem vorgestellten theoretischen Modell einfließen. Gleichzeitig war das Codebuch prinzipiell offen für alle möglichen weiteren Faktoren, die mit Vertrauen in Online-Bewertungen zusammenhängen, sowie für die Ausdifferenzierung der bereits bekannten Faktoren.

#### **4.4 AUSWAHL DER BEWERTUNGEN UND DER TEILNEHMER**

Nach der Vorstellung der verwendeten Methoden geht es nun um die bei der Untersuchung benutzten Bewertungen und die Teilnehmer der Studie.

In meiner Studie legte ich allen Teilnehmern dieselben zehn natürlichen Arztbewertungen vor. Bei diesen Bewertungen handelte es sich um zehn zufällig ausgewählte Online-Bewertungen desselben Arztes von dem bekanntesten deutschen Arztbewertungsportal *Jameda*. Die vollständigen Arztbewertungen finden sich im Anhang (siehe 7.3 Arztbewertungen, S. 161). Unten findet sich ein tabellenförmiger Überblick über die verwendeten Arztbewertungen (siehe Tabelle 16: Überblick Arztbewertungen, S. 98). In der Tabelle sind das Datum der Erstellung der Bewertung aufgeführt sowie ggf. freiwillige Angaben der Bewerter. Bewerter können z. B. Angaben zur Art ihrer Krankenversicherung machen. Bei den hier verwendeten Bewertungen machen die Bewerter entweder keine Angabe zum Versicherungstyp oder sie sind gesetzlich krankenversichert (GKV); eine private Krankenversicherung gibt kein Bewerter an. Ein Bewerter kann außerdem sein grobes Alter angeben. In der Tabelle finden sich daneben die Gesamtnote und Angaben zum Umfang des Bewertungstexts. Zum Schluss wird die Anzahl der Weitergaben durch die Teilnehmer notiert. Für zwei der benutzten Bewertungen wurden außerdem auf dem Bewertungsportal Stimmen dazu abgegeben, zu wieviel Prozent die Bewertung hilfreich ist.

*Tabelle 16: Überblick Arztbewertungen*

	Datum	KV	Alter	Note	Überschrift Zeichen/ Wörter	Bewertung Zeichen/ Wörter	Weiter- gaben
<b>1</b>	08.12.2016	-	-	1,2	25/5	484/87	11 (58%)
<b>2</b>	10.08.2016	GKV	-	1,0	23/4	444/75	16 (84%)
<b>3</b>	18.11.2015	GKV	30-50	4,2	27/3	163/27	11 (58%)
<b>4</b>	12.10.2015	GKV	ü50	1,0	32/5	935/151	17 (90%)
<b>5*</b>	28.08.2015	GKV	-	1,0	36/4	312/51	15 (79%)
<b>6**</b>	27.05.2015	-	-	5,4	20/4	111/19	6 (32%)
<b>7</b>	16.07.2015	GKV	u30	1,2	13/3	68/14	7 (37%)
<b>8</b>	07.01.2015	GKV		1,0	42/6	285/43	18 (95%)
	Datum	KV	Alter	Note	Überschrift Zeichen/ Wörter	Bewertung Zeichen/ Wörter	Weiter- gaben
<b>9</b>	11.10.2014	GKV	u30	1,0	17/3	418/70	16 (84%)
<b>10</b>	05.03.2014	GKV	u30	1,0	32/5	153/21	15 (79%)

\* hilfreich mit 1 Stimme zu 100%  
\*\* hilfreich mit 2 Stimmen zu 20%

---

Alle genannten messbaren Eigenschaften der Bewertungen stehen in dieser Untersuchung im Hintergrund. Es ist auch nicht möglich, repräsentative – im Sinne von typischen – Arztbewertungen als Grundlage für diese Studie zu nehmen. Denn Baechler (2017) stellt mithilfe einer linguistischen Untersuchung deutscher Arztbewertungen fest, dass sich bisher keine verschiedenen Typen von Arztbewertungen durchgesetzt haben. Die Textsorte der Arztbewertung ist vielmehr von großen inhaltlichen und sprachlichen Unterschieden geprägt (Baechler, 2017, S. 209). Sie erklärt die geringe Standardisierung damit, dass es sich bei Arztbewertungen um eine noch sehr junge Textsorte handelt, die sich noch nicht etabliert hat (Baechler, 2017, S. 209). Da ohnehin nicht die objektiven Eigenschaften einer Bewertung im Vordergrund stehen, kann dieser Punkt jedoch vernachlässigt werden. Im Fokus der Studie steht die Wahrnehmung der Arztbewertungen durch die Rezipienten.

Als Stimulus dienen hier also reale Bewertungen anstelle eines künstlich kreierte Stimulus. Damit will die Untersuchung der Komplexität des alltäglichen Umgangs mit Online-Bewertungen gerecht werden, d. h. externe Validität gewährleisten. So soll Artefakten vorgebeugt und eine bisherige Forschungslücke geschlossen werden. Bisher besteht eine Schwachstelle der überwiegend experimentell arbeitenden Studien in der extremtypischen Manipulation der Quelle bzw. Mitteilung (zur Kritik daran siehe auch Kohring, 2004, S. 26). Bei einer starken Manipulation, d. h. extremen Zuspitzung des Stimulusmaterials auf offensichtliche Gegensätze stellt sich die Frage der Übertragbarkeit der Ergebnisse auf reales Material, also tatsächlich empirisch vorfindbare Bewertungen aus dem Alltag der Leser. Die Untersuchung der Rezeption extrem manipulierter Bewertungen macht sicher grobe Tendenzen sichtbar, doch die Übertragbarkeit dieser Ergebnisse auf weniger extreme und z. B. auch hinsichtlich eines Faktors uneindeutige natürliche Bewertungen ist damit nicht gesichert. Bisher ist nicht klar, wie natürlich vorkommende, alltägliche Online-Bewertungen wahrgenommen werden, in denen verschiedene, eher gering ausgeprägte Eigenschaften um Aufmerksamkeit kämpfen. Die bisherige Forschung kann vor allem Aussagen darüber treffen, welchen Einfluss eine Eigenschaft einer Bewertung hat, wenn sie so extrem manipuliert wurde, dass sie sehr wahrscheinlich wahrgenommen wird.

In der bisherigen Forschung kommen selten natürliche Bewertungen vor. Wenn sich Leser auf natürliche Bewertungen beziehen, dann sind diese dem Autor der jeweiligen Studie unbekannt. Die Studienteilnehmer werden dafür nach ihrem Umgang mit z. B. der letzten Online-Bewertung, die sie gelesen haben, gefragt. Diese liegt dem Autor der jeweiligen Studie aber nicht selbst vor. In meiner Studie werden erstmals natürliche Bewertungen verwendet, die dem Forscher bekannt sind. Außerdem werden allen Teilnehmer dieselben Bewertungen vorgelegt. So möchte ich die Wahrnehmungen derselben (natürlichen und bekannten) Bewertung durch unterschiedliche Teilnehmer beobachten.

Die Auswahl der Bewertungen beschränkt aber auch den möglichen Erkenntnisgewinn in einigen Bereichen. Einige Faktoren, die in der integrativen Literaturübersicht ermittelt wurden, können in dieser Studie nicht vorkommen. Dazu gehört der Faktor *Übereinstimmung der Bewertung mit externen Quellen* als Indikator für *Ungenügendes Vorwissen*. In dieser Studie werden den Teilnehmer lediglich die Bewertungen vorgelegt. Sie können sich daher nicht auf externe Quellen verlassen. Bei allen weiteren Faktoren, die in dieser Studie nicht vorkommen können, handelt es sich um Gründe für Vertrauen. Bei der Bewertungsplattform *Jameda* handelt es sich, wie bei den meisten Arztbewertungsplattformen, nicht um eine Community. Damit sind alle Faktoren, die den Akteur *Community* betreffen, ausgeschlossen: *Eingebundenheit des Bewerter in Community* und *Zugehörigkeit des Lesers*. Wie bereits erläutert werden Arztbewertungen meist anonym veröffentlicht. Daher sind alle Faktoren, die sich auf die Identität eines bestimmten Bewerter beziehen, ebenfalls nicht Teil dieser Studie: *Identifikation des Bewerter*, *Profilfoto des Bewerter*, *Bewerterstatus* und *Erfahrung des Bewerter*. Da *Jameda* nicht vor möglichen Fälschungen warnt, kann auch der Faktor *Warnhinweis der Plattform* nicht vorkommen.

Die Auswahl der Teilnehmer erfolgte wie die Auswahl der Bewertungen ebenfalls zufällig. Über einen breit gestreuten Aufruf im erweiterten Bekanntenkreis bat ich um Kontakt zu Menschen, die ich nicht kenne. So wurde sichergestellt, dass die Teilnehmer der Studien vorher nichts über diese Studie wissen und somit kein Risikobewusstsein induziert wird. Die Datenerhebung endete aus forschungspragmatischen Gründen nach 19 Teilnehmern. Auf das Transkript eines

Teilnehmers muss leider verzichtet werden, da das Aufnahmegerät nicht funktionierte. Die Ergebnisse legen nahe, dass schon mit der Analyse von 18 Transkripten eine theoretische Sättigung erreicht wurde.

Bei den Teilnehmern handelt es sich um vier Männer und 15 Frauen. Bei der Erstsprache der meisten Teilnehmer (14) handelt es sich um Deutsch. Vier Teilnehmer wuchsen mit noch einer weiteren Sprache als Erstsprache auf. Nur eine Teilnehmerin sprach nicht Deutsch als Erstsprache. Bis auf einen Schwarzen Teilnehmer, waren alle Teilnehmer Weiß. Das Alter der Teilnehmer lag zwischen 24 und 78 Jahren. Der Mittelwert lag bei 38 Jahren, davon waren acht Teilnehmer unter 30, fünf über 30, drei über 40, eine über 50 und zwei über 70 Jahre alt. Zwei Teilnehmer absolvierten nach der Mittleren Reife eine Ausbildung, die anderen Teilnehmer haben mindestens Abitur, 13 davon einen Hochschulabschluss. Unter den Teilnehmern befanden sich sechs Studenten. Desweiteren tauchten drei technische Berufe auf, es gab vier Mitarbeiter aus dem (Hoch-)Schulbereich sowie je einen Teilnehmer aus Gastronomie, Recht und Sozialberatung. Vier Teilnehmer waren privat, alle anderen gesetzlich krankenversichert.

Mit dieser anfallenden Stichprobe wird kein repräsentativer Anspruch erhoben. Dementsprechend kann keine Aussage über Häufigkeiten getroffen werden. Als Alternative zu einer willkürlichen Auswahl der Teilnehmer gilt z. B. das gezielte, theoretische Sampling (Patton, 2002). Dabei wird a priori festgelegt, welche Eigenschaften die Teilnehmer der Studie haben müssen. Damit soll eine möglichst hohe Varianz an Teilnehmern sichergestellt werden. So möchte man zu Ergebnissen gelangen, die möglichst breit gültig sind. Um von vorneherein ein hohe Varianz auf der Seite der Teilnehmer festlegen zu können, muss man aber die entscheidenden Merkmale kennen, die sich bei den Teilnehmern unterscheiden müssten. Das ist hier nicht der Fall. In der Forschung konnten bisher keine eindeutigen Erkenntnisse darüber gewonnen werden, welche Eigenschaften des Lesers mit seinem Vertrauen zusammenhängen. Typische demografische Eigenschaften, nach denen das Sample hätte ausgewählt werden können – wie *Geschlecht*, *Alter* und *Bildung des Lesers* – scheinen nicht direkt mit Vertrauen in Online-Bewertungen zusammenzuhängen. Bisher wird überhaupt nur in wenigen Studien von einer Varianz auf der Seite der Leser ausgegangen. Falls doch, zeigen manche Ergebnisse, dass bestimmte

Eigenschaften eines Lesers moderierend wirken. Die *Involviertheit des Lesers* führt zu einem geringeren Zusammenhang von Vertrauen mit *Übereinstimmung* (Cheung et al., 2009) und *Vertrauen in den Bewerter* (Lis, 2013), aber verstärkt den Zusammenhang mit *Ähnlichkeit* (Racherla et al., 2012) und *Kompetenz des Bewerter* (Lis, 2013). Sie erhöht den Zusammenhang zwischen *Argumentation* und Vertrauen nicht (Cheung et al., 2012; Racherla et al., 2012). Niedrige Involviertheit verstärkt nicht den Zusammenhang von Vertrauen mit *Vertrauen in den Bewertung* oder *Übereinstimmung*, aber sie verstärkt den Zusammenhang von *Ausgewogenheit* und Vertrauen (Cheung et al., 2012). Diese Ergebnisse sind ein Hinweis darauf, dass es durchaus eine Varianz zwischen den Teilnehmern gibt. Das trifft selbst dann zu, wenn die Stichprobe demografisch recht homogen ist, wie bei den Experimenten mit Studenten als Teilnehmern häufig der Fall. Alles in allem ist davon auszugehen, dass es Unterschiede zwischen Lesern gibt, die aber nicht im Vorhinein festgelegt werden können. Bei dem hier gewählten Vorgehen handelt es sich dementsprechend um eine erste Exploration. Damit soll die Komplexität des Phänomens Vertrauen in Online-Bewertung nach der bisher recht breiten Forschung nun auch in seiner Tiefe erfasst werden.

#### 4.5 VORAUSSETZUNGEN FÜR VERTRAUEN IN ARZTBEWERTUNGEN

Zu den Voraussetzungen für Vertrauen gehören *Ungenügendes Wissen* und *Risikobewusstsein*. Die Ergebnisse der damit verbundenen Faktoren werden hier, gemeinsam mit den entsprechend zugeordneten Äußerungen der Teilnehmer, aufgeführt. Zunächst geht es um die Faktoren, die mit dem *ungenügenden Wissen* zusammenhängen: *Markeneinstellung des Lesers* und *Kompetenz des Lesers*.

Die *Kompetenz des Lesers* im Bereich des Gesundheitswesens ließe sich z. B. über ihren Beruf feststellen. Sieben Teilnehmer waren entweder selbst im medizinischen Bereich tätig oder hatten Freunde oder enge Verwandte, die dort tätig sind. Allerdings zog keiner der Teilnehmer Wissen aus diesem Beruf heran, um die Vertrauenswürdigkeit einer Arztbewertung zu ermitteln. Dagegen gab es andere Aussagen, die als Hinweis auf das im theoretischen Modell formulierte *Ungenügende Wissen* der Leser zu verstehen sind. Sie zeigen sich in den Gründen für die Nutzung von Arztbewertungen. Vier Teilnehmer nutzten Arztbewertungen, als sie in eine neue Stadt gezogen sind (Anna: 34-40, Gesa: 59, Karin: 37, Sabine: 48). Zwei

Teilnehmer besuchten ein Arztbewertungsportal, als die Wahl eines neuen Facharztes anstand (Bernd: 234, Julia: 89-97). Die Teilnehmer befanden sich also in einer Situation, in der ihnen geeignete Ärzte unbekannt waren. Daher waren sie auf das Wissen anderer Menschen angewiesen, um zu handeln. Die Teilnehmer wandten sich dafür an Menschen, die schon Erfahrungen mit einem Arzt gesammelt hatten und diese als Arztbewertungen weitergaben.

Einen weiteren Grund für die Nutzung von Arztbewertungen nennt Karin. Sie liest Bewertungen von Ärzten, die sie bereits besucht hat. Dabei möchte sie abgleichen, ob ihre Empfindung bei der Behandlung von Anderen geteilt wird oder ob sie die einzige ist, der z. B. das Verhalten eines Arztes nicht gefiel (Karin: 43). Die Bewertung wird in diesem Fall nicht dafür benutzt, sich für oder gegen den Besuch bei einem Arzt zu entscheiden. Vielmehr dient die Meinung des Bewerter hier dazu, die eigene Bewertung der Teilnehmerin zu überprüfen. Auch hier fehlt dem Teilnehmer Wissen, daher ist auch in diesem Fall Vertrauen seitens der Leserin nötig, um die Bewertung des Bewerter zum Vergleich heranzuziehen.

Die *Markeneinstellung des Lesers* zeigt sich im Kontext von Arztbewertungen in der Einstellung der Teilnehmer gegenüber Ärzten an sich. Damit ist nicht gemeint, dass ein Teilnehmer direkt den bewerteten Arzt kennt, so wie es bei der Markeneinstellung nicht darum geht, dass der Leser genau das bewertete Produkt kennt. Gemeint ist vielmehr die allgemeine Einstellung gegenüber der Ärzteschaft. In meiner Studie haben zwei Teilnehmerinnen ihre generelle Einstellung gegenüber Ärzten geäußert. Franziska meint zu einer schlechten Bewertung, dass sie sich nicht vorstellen könne, „dass ein praktizierender Arzt, dass alles so schlecht ist“ (Franziska: 51). Nadja stellt einer negativen Bewertung ebenfalls ihre Annahme gegenüber, dass es „keinen Hausarzt [gibt], der sich gar nicht kümmert“ (Nadja: 140). Diese zwei Teilnehmerinnen übertragen ihre grundsätzlichen Annahmen über alle Ärzte auf den hier bewerteten Arzt. Für sie ist klar, dass kein Arzt sich nur schlecht verhalten kann. Mit dieser festen Überzeugung ist für beide Teilnehmerinnen das Wissen der Bewerter nicht notwendig. Dem theoretischen Modell zufolge kommen die Teilnehmerinnen dadurch nicht in die Position, den Bewertern vertrauen zu müssen. Ihre Ablehnung der Bewertungen ist damit nicht als mangelndes Vertrauen

zu verstehen, da eine Voraussetzung für Vertrauen überhaupt nicht erfüllt ist. Die Teilnehmerinnen haben kein *Ungenügendes Wissen*.

Zu den Faktoren, die das Risikobewusstsein beeinflussen, gehören das *Allgemeine Vertrauen des Lesers* und die *Einstellung des Lesers gegenüber Online-Bewertungen*. Zum *Allgemeinen Vertrauen des Lesers* wurden keine Äußerungen der Teilnehmer gefunden.

Die *Einstellung der Leser gegenüber Online-Bewertungen* unterscheidet sich bei den Teilnehmern. Zwei Teilnehmerinnen sind Online-Bewertungen gegenüber grundsätzlich offen (Franziska: 45, Dorothee: 82). Dorothee zieht Bewertungen der Werbung vor. Das führt sie auf ihre Erfahrung als Praktikantin in einer Werbefilmproduktion zurück. Dort hat sie erlebt, dass in der Werbung „alles manipuliert“ (Dorothee: 82) wird. Dorothee weist allerdings darauf hin, dass sie unter den Bewertungen wiederum solche „aus erster Hand“ (Dorothee: 82) den Online-Bewertungen vorzieht. Drei Teilnehmer teilen eine abwägende Haltung gegenüber Online-Bewertungen. Sie empfinden Online-Bewertungen zwar als „prinzipiell [...] gut“ (Renata: 216) bzw. als „Hilfestellung“ (Quentin: 83), legen aber Wert auf einen kritischen, hinterfragenden Umgang damit (Quentin: 85) – oder mahnen zur Vorsicht (Erika: 159), insbesondere bei negativen Bewertungen (Renata: 216).

Die meisten Teilnehmer sind gegenüber Online-Bewertungen grundsätzlich eher skeptisch eingestellt (Bernd: 74, Franziska: 129, Gesa: 108, Nadja: 84, Renata: 70). Sie nennen dabei verschiedene Gründe für ihre generell skeptische Haltung. Manche Teilnehmer stört, dass sie den Bewerter nicht kennen (Franziska: 131). Das erklärt Bernd ausführlich:

„Wenn ich jemanden persönlich gut kenne und der gibt mir eine Bewertung über einen Arzt, dann kann ich das ja einstufen. Dann kann ich aufgrund der Kenntnis der Person dann sagen, das ist für mich auch glaubhaft. Ich hab Bekannte, da weiß ich, die sind so negativ eingestellt, die erzählen immer nur das Schlechte. Und wieder andere, sind ganz wenige, die sind da eher etwas übertrieben freundlich und übertreiben das wieder. Und das kann man ja dann bewerten und kann das dann entsprechend einstufen und sagen, ‚Ja, wenn der mir das sagt, dann ist das für mich gut und richtig und wichtig.‘“ (Bernd: 220 - 226)

Ähnlich geht es auch Gesa. Dadurch, dass sie bei Personen, die sie kenne, wisse, wie diese interagieren und reagieren, könne sie deren Bewertungen besser einschätzen

(Gesa: 150). Für diese drei Teilnehmer ist es demnach wichtig, den Bewerter und damit seine Maßstäbe persönlich zu kennen, um über das Vertrauen in dessen Bewertung entscheiden zu können. Da dies bei Online-Bewertungen grundsätzlich nicht gegeben ist, sind sie diesen gegenüber misstrauisch.

Zwei Teilnehmerinnen nennen weitere Gründe für ihr Misstrauen. Franziska stört sich daran, dass sie bei den Bewertern nicht nachfragen kann (Franziska: 131, 295). Nadja hält Bewertungen nicht für repräsentativ, da sie davon ausgeht, dass nur Social Media-affine und unzufriedene Patienten Bewertungen schreiben (Nadja: 100).

Viele Teilnehmer vermuten gefälschte Bewertungen (Dorothee, Gesa, Ina, Julia, Karin, Nadja, Otto, Quentin, Renata, Sabine). Julia meint dazu lachend „heutzutage muss man [sich eher] fragen: ‚Welche ist nicht gefälscht?‘ als ‚Welche ist gefälscht?‘“ (Julia: 113). Als mögliche Fälscher werden z. B. Mitarbeiter des Arztes (Gesa: 100), Konkurrenten (Quentin: 45), Bots (Otto: 241) oder andere Freunde bzw. Feindes des Arztes (Bernd: 316) vermutet.

Für zwei andere Teilnehmerinnen sind gerade wegen des geringen Fälschungsrisikos Arztbewertungen vertrauenswürdiger als andere Online-Bewertungen. Bei diesen Teilnehmerinnen sinkt das Risikobewusstsein im Falle von Arztbewertungen, weil sie denken, dass diese nicht gefälscht werden. Gesa vermutete gefälschte Bewertungen eher bei Hotels (Gesa: 106). Renata glaubt nicht, dass Ärzte Schmähbewertungen über konkurrierende Ärzte verfassen (Renata: 226). In anderen Bereichen hält sie solche gefälschten Bewertungen für möglich, doch im Gesundheitswesen „sind wir schon fast in dem Thema Moralsachen. [...] Das wäre bisschen heftig, wenn es so ist.“ (Renata: 228). Der Arzt als *Typ des Bewerteten* ist hier ausschlaggebend dafür, das Risikobewusstsein zu senken. Bei drei Teilnehmern verhält es sich genau gegenteilig. Hier sorgt der Arzt als *Typ des Bewerteten* für mehr Risikobewusstsein.

Drei weitere Teilnehmerinnen sind misstrauisch gegenüber Online-Bewertungen von Ärzten. Sie meinen, dass die Bewertungen anderer Menschen über einen Arzt nicht übertragbar sind. Das führen sie darauf zurück, dass das Verhältnis zu einem Arzt „persönlich“ (Erika: 161, Nadja: 274) ist. Damit meinen die Teilnehmerinnen, dass es bei Arztbewertungen um „subjektive Empfindungen von

unterschiedlichsten Menschen“ (Renata: 218) geht. Dieses subjektive Gefühl im Umgang mit dem Arzt (Nadja: 274) ist nicht verallgemeinerbar.

Was die Voraussetzungen des Vertrauens in Arztbewertungen angeht, können mehrere Ergebnisse festgehalten werden. Als *Kompetenz des Lesers* scheint hier nicht die medizinische Kompetenz relevant zu sein, sondern sein Wissen über einen geeigneten Arzt. Das mangelnde Wissen darüber führen einige Teilnehmer als Grund an, um Arztbewertungen zu nutzen. Bei der *Markeneinstellung des Lesers* handelt es sich in diesem Rahmen um positive Grundannahmen zweier Teilnehmerinnen über Ärzte. Die ungenügende Kompetenz der Teilnehmer kann als Hinweis auf *Ungenügendes Wissen* verstanden werden, die positive Markeneinstellung der Teilnehmer hingegen zeugt von ausreichendem Wissen. Im ersten Fall ist Vertrauen also möglich, im zweiten nicht. Als Faktor, der sich auf das Risikobewusstsein bezieht, tauchte in dieser Studie nur die *Einstellung des Lesers gegenüber Online-Bewertungen* auf. Hier zeigten sich Unterschiede darin, ob und warum Online-Bewertungen generell vertraut wird. Dabei wurde deutlich, dass der *Typ des Bewerteten* für einige Teilnehmer mit ihrem Risikobewusstsein zusammenhängt. Allerdings war dieser Zusammenhang nicht eindeutig: Einige sahen besonders bei Arztbewertungen kein Risiko der Fälschung, andere sahen gerade bei Arztbewertungen verschiedene Risiken.

#### 4.6 DIMENSIONEN VON VERTRAUEN IN ARZTBEWERTUNGEN

In diesem Kapitel gehe ich auf die Faktoren ein, die mit den Dimensionen des Vertrauens in Arztbewertungen zusammenhängen. Im Zusammenspiel von integrativer Literaturübersicht und theoretischem Modell konnten hier die Faktoren *Ausführlichkeit der Bewertung* und *Ausgewogenheit der Bewertung* der Dimension Faktenselektivität sowie die Faktoren *Argumentation der Bewertung*, *Attribuierung* und *Übereinstimmung der Zahlen- und Textbewertung* der Dimension der expliziten Bewertung herausgearbeitet werden.

Die *Ausführlichkeit der Bewertung* bringen die Teilnehmer dieser Studie sehr oft zur Sprache. Viele finden es wichtig, dass die Bewertung ausführlich ist, d. h. alle relevanten Details konkret aufgeführt werden (Franziska: 97, 171; Hanna: 107, Julia: 113, Nadja: 236). Allerdings unterschieden sich die Vorstellungen der Teilnehmer darüber, welche Details relevant sind. Das wird deutlich beim Vergleich

der unterschiedlichen Reaktionen auf dieselben Bewertungen. Die erste Bewertung finden einige Teilnehmer nicht ausführlich genug (Dorothee: 108, Hanna: 49, Karin: 18, Otto: 117, Paula: 25), eine Teilnehmerin entdeckt hingegen „Aspekte, die bei einem Hausarzt wichtig sind“ (Franziska: 46). Genau den richtigen Grad der Ausführlichkeit traf auch der vierte Bewerter laut vieler Teilnehmer (Dorothee: 137, Gesa: 98, 117, Ina: 27, Nadja: 198, Paula: 35, 77, 107). Zwei Teilnehmerinnen fanden diese Bewertung wiederum zu detailliert (Charlotte: 77, Julia: 129). Otto fand die Bewertung zunächst zu ausführlich (Otto: 111), korrigiert sein eigenes Urteil aber später zu: „im Prinzip liebe ich konkrete Beschreibungen“ (Otto: 163). Dieselbe Unentschiedenheit legt Gesa bei Bewertung 8 offen. Sie fand diese zunächst „nicht so wahnsinnig aussagekräftig“ (Gesa: 49), lenkte aber später um zu: „eigentlich ist die schon recht aussagekräftig, wenn man darüber nachdenkt“ (Gesa: 122). Auch die anderen Teilnehmer verhalten sich hier nicht gleich. Nadja findet die Bewertung nicht ausführlich genug (Nadja: 58), Paula schon (Paula: 39). Bei der letzten Bewertung sind sich die Teilnehmer ebenfalls nicht einig. Während zwei Teilnehmer diese nicht ausführlich genug finden (Anna: 102, Quentin: 69), finden drei andere Teilnehmerinnen sie ausführlich (Dorothee: 73, Hanna: 176, Paula: 41). Bei zwei Bewertungen zeigt sich hohe Einigkeit mit jeweils einem Ausreißer. So finden viele Teilnehmer Bewertung 6 nicht ausführlich genug (Gesa: 43, Hanna: 64, Karin: 24, Otto: 188, 190, Renata: 51, 154). Franziska dagegen zeigt Verständnis für die wenigen Notenkategorien, denn „wenn man so enttäuscht ist von einem Arzt, dann will man natürlich auch nicht die irrelevanten Dinge bewerten.“ (Franziska: 67).

Einigkeit darüber, dass die Bewertung nicht ausführlich genug ist, scheint lediglich bei Bewertung 3 (Dorothee: 133, Franziska: 27, Gesa: 28, Ina: 23, Nadja: 32, 168, Otto: 159, Paula: 32, 67) und Bewertung 7 (Anna: 88, Franziska: 70, 183, Hanna: 69, 158, 160, Karin: 25, Nadja: 51, Paula: 38, Quentin: 62, Renata: 53, 172) zu herrschen. Nicht vergleichen lässt sich hingegen die Wahrnehmung der Ausführlichkeit bei der zweiten, fünften und neunten Bewertung, da sich hier jeweils nur ein Teilnehmer dazu geäußert hat. Bewertung 2 war für Karin ausführlich genug (Karin: 20), Bewertung 5 ist laut Dorothee nicht ausführlich (Dorothee: 52) genauso wie Bewertung 9 laut Otto (Otto: 205, 207).

Die *Ausführlichkeit der Bewertung* umfasst neben der Fülle an Details auch die Auswahl der genannten Details. Einige Teilnehmer nennen die Themen, die für sie in einer Bewertung relevant sind. Fast alle Teilnehmer wollen wissen, um welche Krankheit es sich handelt, bzw. welches konkrete Anliegen der Bewerter beim Arztbesuch hatte (Anna: 78, Charlotte: 75, Dorothee: 47, 121, 123, Erika: 51, Franziska: 177, Gesa: 96, 154, 156, Hanna: 152, Ina: 85, 87, Julia: 173, Karin: 69, 81, Nadja: 26, 37, 47, 188, 190, Otto: 147, Paula: 63, 91, Quentin: 143, 145, 147, Renata: 156).

Sprechzeiten sind ebenfalls vielen Teilnehmern wichtig (Dorothee: 47, 61, 207, Gesa: 124, 132, Ina: 87, Karin: 20, 26, 101, Nadja: 24, 154, 244, Otto: 147, Quentin: 23, 36, 69, Sabine: 20). Für Nadja handelt es sich bei der Angabe der Sprechzeiten um „konkret[e]“ bzw. „sachlich[e]“ (Nadja: 244) Informationen. Diese Wortwahl kann als Hinweis darauf verstanden werden, dass sie überprüfbare Informationen erwartet. Paula schränkt die Relevanz der Nennung von Sprechzeiten ein, da ihrer Meinung nach Praxen ohnehin meist so geöffnet haben, dass Berufstätige diese besuchen können (Paula: 31).

Auch die Wartezeiten spielen für einige Teilnehmer eine wichtige Rolle (Gesa: 132, Julia: 93, Karin: 101, Nadja: 71, Paula: 65, Quentin: 147). Außerdem möchten die Teilnehmer wissen, wieviel Zeit sich der Arzt genommen hat (Dorothee: 61, 207, Gesa: 117, Hanna: 186, Karin: 26, Nadja: 71, Quentin: 69, 147), ob er fachlich kompetent (Dorothee: 47, 207, Hanna: 12, Nadja: 242, Paula: 65, Quentin: 64, 69, 149), freundlich (Bernd: 139, Dorothee: 47, Hanna: 125, 186) und vertrauensvoll (Dorothee: 47, 207, Quentin: 69) ist. Die Betreuung bzw. Behandlung durch den Arzt ist zwei Teilnehmerinnen ebenfalls wichtig (Ina: 117, Nadja: 71, 152). Eine Teilnehmerin möchte Informationen haben über die Ausstattung (Nadja: 242), die Barrierefreiheit (Nadja: 246) und über beide Ärzte aus der Gemeinschaftspraxis (Nadja: 24). Drei Teilnehmer sind sich darüber einig, dass Hinweise zum Entertainment in der Praxis unwichtig sind (Hanna: 135, Nadja: 25, Quentin: 52).

Bei zwei Aspekten herrscht Uneinigkeit. Die Internationalität des Personals zu erwähnen findet Dorothee wichtig (Dorothee: 61). Quentin hätte auf diese Information verzichten können, da er davon ausgeht, dass der Freund, dem er die Bewertungen weiterleiten soll, deutsch ist (Quentin: 64). Auch das Erwähnen der

Parkplatzsituation ist für zwei Teilnehmerinnen wichtig (Nadja: 242, 244, Quentin:23), aber für zwei andere nicht wichtig (Erika: 42, Hanna: 186). Quentin ergänzt, dass die Relevanz dieser Information für ihn vom besagten Freund abhängt. Er findet solche Details dann wichtig, wenn er weiß, dass sie seinem Freund wichtig sind (Quentin: 36).

Eine Teilnehmerin erläutert, dass sie in einer Bewertung nur Informationen zu den beruflichen Leistungen des Arztes lesen möchte. Als Beispiel für ein Detail, das sie nicht interessiert, nennt sie private Schulden des Arztes (Charlotte: 83). In der Studie ist deutlich geworden, dass die Teilnehmer vor allem die Relevanz der Themen beurteilen, welche von dem Arztbewertungsportal selbst als Kategorien zur Notenbewertung vorgegeben werden. Über die vorgegebenen Kategorien hinaus nennen die Teilnehmer nur die Krankheit des Bewerter als eine Information, die sie bei einer Arztbewertung erwarten.

Die Länge der Bewertung ist ein weiterer Aspekt der Ausführlichkeit, der öfter thematisiert wurde. Auch hier zeigt ein Blick auf die Wahrnehmung einzelner Bewertungen die Unterschiede zwischen den Teilnehmern. So sind sich die Teilnehmer z. B. bei Bewertung 10 uneinig. Für zwei Teilnehmerinnen ist sie zu kurz (Anna: 104, Erika: 153). Zwei andere Teilnehmerinnen dagegen finden diese Bewertung „Kurz und knapp, aber gut.“ (Dorothee: 73) bzw. „auf den Punkt gebracht“ (Karin: 28). Dasselbe gilt für die siebte Bewertung. Zwei Teilnehmerinnen finden diese zu kurz (Erika: 56, Julia: 135), Otto findet sie „kurz und bündig“ (Otto: 198). Für je eine Teilnehmerin wirkt Bewertung 3 kurz und daher „impulsiver“ (Ina: 103), Bewertung 4 dagegen „fast ein bisschen zu lang“ (Erika: 155). Einig sind sich vier Teilnehmerinnen darüber, dass die sechste Bewertung zu kurz ist (Anna: 78, Erika: 54, 131, 133, Nadja: 47, Paula: 37). Allgemein scheint es aber so, dass die Teilnehmer unterschiedliche Präferenzen haben, was die Länge der Bewertung angeht. Einige Teilnehmerinnen sagen, dass sie generell dazu bereit sind, längere Bewertungen zu lesen (Erika: 78, Franziska: 175, Julia: 131, Karin: 101) und bei kürzeren Bewertungen eher kritisch sind (Sabine: 168, Nadja: 198). Franziska meint aber auch, dass es „wahrscheinlich auch eine kritische Menge“ gibt, „wo man’s dann gar nicht mehr liest“ (Franziska: 173). Paula hingegen kann nicht festlegen, wie lang

eine Bewertung generell sein muss, da es ihr dabei auf den genauen Inhalt ankommt (Paula: 71).

Die Ergebnisse zum Faktor der *Ausführlichkeit der Bewertung* zeigen zweierlei. Zum einen wird deutlich, dass sich dieser Faktor aufteilen lässt in drei Unterkategorien. Die Teilnehmer äußern sich darüber, ob eine Bewertung ausführlich genug ist oder nicht. Außerdem machen sie Angaben dazu, welche Informationen sie erwarten und welche Länge für eine Bewertung angemessen ist. Damit kann das theoretische Modell bezüglich der Dimensionen des Vertrauens weiter ausdifferenziert werden. Zum anderen wird klar, dass sich die Wahrnehmung und die Beurteilung der Teilnehmer stark unterscheiden können.

Zur Dimension der Faktenselektivität gehört der Faktor *Ausgewogenheit der Bewertung*. Drei Teilnehmer erklären, dass sie einer Arztbewertung nicht vertrauen, wenn darin besonders gut oder besonders schlecht bewertet wurde. Für sie sind extreme Bewertungen ein Zeichen dafür, dass der Bewerter nicht differenziert ist, was von den Teilnehmern wiederum negativ bewertet wird (Dorothee: 55, 70, 73, 88, 179, 181, 195, Nadja: 134-144, Otto: 117, 123, 175). Einige Teilnehmer halten extreme oder starke Ausdrücke (Erika: 141) und Bewertungen für unrealistisch, da nicht alles schlecht (Franziska: 51, 84, 95, 149, 163, 293, Julia: 125, Quentin: 133) bzw. nicht alles perfekt sein kann (Franziska: 58, 82, 95, Julia: 117, Quentin: 133). Franziska mutmaßt, dass bei solchen Bewertungen die „Wahrheit wahrscheinlich irgendwo dazwischen“ (Franziska: 97) liegt. Zwei Teilnehmer können extreme und unausgewogene Bewertungen nicht nachvollziehen. Bernd würde bei Bewertungen prinzipiell niemals die beste oder schlechteste Note geben (Bernd: 244). Dorothee würde auch extreme Noten verwenden, aber nicht durchgehend (Dorothee: 88). Der Faktor der *Ausgewogenheit der Bewertung* umfasst demnach zwei Aspekte: die extreme Ausprägung der Bewertung (z. B. eine Note 1 oder eine Note 6) als auch die durchgehende Bewertung mit ein und derselben Note (z. B. nur die Note 3 in allen Kategorien). Bei diesem Faktor scheint es Einigkeit darüber zu geben, dass eine Bewertung ausgewogen sein muss. Allerdings stellt sich hier die Frage, ob die Teilnehmer Ausgewogenheit alle gleich wahrnehmen.

Im Folgenden geht es um die Ergebnisse zur Dimension der expliziten Bewertung. Der Faktor der *Argumentation der Bewertung* lässt sich ebenfalls in

verschiedene Aspekte aufteilen. Vier Teilnehmerinnen haben den Anspruch, dass die Bewertung überhaupt begründet wird (Hanna, 109; Anna: 110, 112, Franziska: 97, Karin: 51). Eine gute Begründung gleicht für Charlotte sogar Fehler in *Rechtschreibung und Grammatik* aus (Charlotte: 73). Sechs Teilnehmern fehlt z. B. eine Begründung bei Bewertung 6 (Dorothee: 55, 169, Erika: 131, Franziska: 84, Hanna: 152, Nadja: 146, Otto: 194). Otto erwartet hier „ein bisschen mehr Argumente, ein bisschen mehr warum, wieso, weshalb“ (Otto: 194). Dorothee kann dem Bewerber deswegen „überhaupt nicht vertrauen“ (Dorothee: 169). Zwei Teilnehmerinnen finden die Begründung bei Bewertung 3 (Gesa: 111, 115, Karin: 75), bei Bewertung 7 (Ina: 143, Gesa: 152) und bei Bewertung 10 (Erika: 153, Dorothee: 209) unzureichend. Renata schließt bei der siebten Bewertung von der nicht vorhandenen Begründung darauf, dass es sich um einen jungen Bewerber handelt (Renata: 168-172). Dorothee ergänzt bei Bewertung 10, dass sie diese nicht einzeln weitergeben würde, sondern nur mit mehreren anderen Bewertungen, da ihr die Begründung nicht ausreicht (Dorothee: 209). Für je einen Teilnehmer fehlen Begründungen bei Bewertung 1 (Nadja: 21), bei Bewertung 2 (Dorothee: 37) und bei Bewertung 9 (Julia: 197). Interessanterweise findet sich in dieser Studie also keine Bewertung, die einige Teilnehmer begründet finden und andere nicht. Die Wahrnehmung, ob eine Bewertung begründet wurde oder nicht scheint also über alle Teilnehmer hinweg homogen.

Ein weiterer Aspekt der Argumentation der Bewertung liegt in der Nachvollziehbarkeit der Begründung. Dafür wollen die Teilnehmer die von den Bewertern verwendeten Kriterien und Maßstäbe kennen (Sabine: 96) und teilen (Erika: 175). Wenn Franziska feststellt, dass die genannten Themen für sie nicht relevant sind, „dann kann [sie] die Bewertungen auch einfach ausblenden“ (Franziska: 99). Das Kriterium „Entertainment“ findet Hanna z. B. nicht relevant (Hanna: 30). Sie und Dorothee meinen außerdem, dass die Freundlichkeit eines Arztes kein angemessenes Kriterium für dessen Bewertung darstellt, weil es nichts über die Kompetenz des Arztes aussagt (Dorothee: 35, Hanna: 142). Zwei Teilnehmerinnen können die dritte Bewertung nicht nachvollziehen. Der Bewerber hege falsche Erwartungen an den Arzt, denn „manche Schmerzen gehen auch nicht weg, wenn der Arzt die Hand aufgelegt hat“ (Nadja: 166). Schließlich sei der Arzt

„kein Zauberer, kein Gott“ (Renata: 128). Bewertung 4 ist dagegen für einige Teilnehmer gut begründet, für andere Teilnehmer nicht. Zwei Teilnehmerinnen halten es für eine angemessene Bewertungsgrundlage, wie der Arzt mit der Peinlichkeit der Situation umgegangen ist (Erika: 119, 121, Franziska: 58) und, dass er die Akutheit der Situation erfasst hat (Erika: 114). Die Meinung darüber, ob die Entfernung einer Zecke eine relevante Kategorie zur Bewertung eines Arztes ausmacht, gehen auseinander. Nadja meint, dass es sich dabei um ein gängiges und damit bewertungswürdiges Problem handelt (Nadja: 37). Zwei Teilnehmer meinen dagegen, dass dieser Fall nicht geeignet wäre, um die medizinische Kompetenz eines Arztes zu bewerten (Anna: 18, Quentin: 52).

Auch bei der Wartezeit als Kriterium gehen die Meinungen der Teilnehmer auseinander. Karin stimmt Bewerter 5 zu, dass eine kurze Wartezeit ein guter Indikator für die Qualität der Praxis ist (Karin: 23). Dorothee dagegen meint, dass eine kurze Wartezeit dafür spricht, dass ein Arzt wenige Patienten habe. Das wiederum sei ein Zeichen dafür, dass der Arzt nicht gut sei (Dorothee: 145). Drei Teilnehmer meinen zwar, dass die Wartezeit wichtig ist, finden aber, dass Bewerter 5 einen falschen Maßstab anlegt. Der Bewerter schrieb im Kommentartext, dass er innerhalb von einem Tag einen Termin bekam, vergab aber als Note lediglich eine 2,0. Bernd und Franziska verglichen die Erfahrung des Bewerter mit ihren eigenen Erfahrungen und fanden den Bewerter zu anspruchsvoll (Bernd: 141). Franziska formuliert das so: „Wenn ich da anrufe und am selben Tag einen Termin bekomme und dann nur fünf Minuten warten muss [...] idealer geht's doch eigentlich gar nicht.“ (Franziska: 187). Auch für Quentin ist das nicht nachvollziehbar, er „würde deswegen der Bewertung eher nicht so viel Gewicht beimessen“ (Quentin: 213).

Entgegen der Meinung der genannten Teilnehmer ist die Wartezeit für Sabine an sich kein geeignetes Kriterium zur Bewertung eines Arztes. Sie weist darauf hin, dass es auch unvorgesehene Notfälle geben kann, die die Wartezeit verlängern (Sabine: 98). Die *Argumentation der Bewertung* umfasst also zum einen, dass eine Bewertung überhaupt begründet wird, zum anderen, anhand welcher Kriterien und mit welchem Maßstab die Bewertung geschieht. In dieser Studie zeigen sich in allen drei Aspekten große Unterschiede zwischen den Teilnehmern.

Ein weiterer relevanter Faktor ist die *Attribuierung der Bewertung*. Zwei Teilnehmerinnen weisen darauf hin, dass die Ausprägung einer Arztbewertung manchmal nicht auf den Arzt zurückzuführen ist. Charlotte meint z. B.: „Es gibt Menschen, die haben immer etwas zu meckern“ (Charlotte: 85). Hanna weist darauf hin, dass manchmal die Umstände und nicht der Arzt selbst für eine Bewertung verantwortlich sein können (Hanna: 235). Zu Bewertung 3 äußerten sich besonders viele Teilnehmer. Zwei Teilnehmerinnen waren sich unsicher, ob die Bewertungen „am Patienten liegt oder am Arzt“ (Dorothee: 133, ähnlich auch Gesa: 74). Viele Teilnehmerinnen waren sich sicher, dass diese Bewertung dem Bewerter zuzuschreiben ist. Sie halten den Bewerter für „überempfindlich“ (Erika: 88) bzw. „sensibler“ (Erika: 92), generell unzufrieden (Ina: 23, Franziska: 51), gestresst bzw. „genervt“ (Nadja: 168-172) oder „mies drauf“ (Karin: 75). Zwei Teilnehmerinnen mutmaßten, dass der Bewerter große Schmerzen hatte und sehr litt (Renata: 136, Nadja: 168). Renata beschrieb den Bewerter außerdem als „Hitzkopf“ (Renata: 112), Paula meinte, dieser würde „nur gern Dampf ablassen“ (Paula: 67). Nadja führt ihre Attribuierung auf die fehlerhafte Sprache des Bewerter zurück. Für sie sind Fehler in Rechtschreibung und Grammatik „ein Zeichen [dafür], dass die Person nie alle Details beachtet oder nicht das Gesamtbild im Kopf hat“ bzw. „verärgert ist und [...] dann sehr emotional so eine Bewertung schreibt.“ (Nadja: 130). Neun Teilnehmerinnen sind sich also darüber einig, dass die Bewertung mit dem Bewerter zu tun hat und nicht mit dem Arzt. Ähnlich nahmen zwei Teilnehmerinnen auch die sechste Bewertung war (Hanna: 154, Julia: 65). Interessanterweise gab es bei dieser Bewertung aber auch eine Teilnehmerin, die die Bewertung dem Arzt zuschrieb: „Kann passieren. Ist natürlich nicht toll. Wenn das der Fall ist, dann liegt die Verantwortung wohl beim Arzt.“ (Sabine: 28). Festzuhalten bleibt, dass es sich in beiden Fällen, in denen die Attribuierung der Bewertung von den Teilnehmern angesprochen wurde, um sehr negative Bewertungen handelt. Bei positiven Bewertungen wurde nie thematisiert, ob die Bewertung dem Bewerter statt dem Arzt zuzuschreiben ist.

Als letzter Faktor dieser Dimension spielt auch die *Übereinstimmung der Zahlen- und Textbewertung* eine Rolle. Um die Übereinstimmung zu überprüfen, muss dem Teilnehmer die Zahlenbewertung überhaupt bekannt sein. Ein Indikator

dafür, dass die Teilnehmer die Zahlenbewertungen wahrgenommen haben, ist die Verbalisierung dieser. Manche Teilnehmer haben die Zahlenbewertungen bei jeder Bewertung (Bernd, Dorothee, Erika, Franziska, Ina, Julia), eine Teilnehmerin nur bei manchen Bewertungen (Karin) vorgelesen. Eine Teilnehmerin äußerte, dass sie immer erst auf die Noten blicke (Ina: 121, 139). Eine andere Teilnehmerin hielt die Zahlenbewertung für relevanter als die Textbewertung, v.a. bei aussagelosen Texten (Franziska: 86). Für zwei Teilnehmerinnen dagegen spielt die Textbewertung eine größere Rolle als die Zahlenbewertung (Paula, Sabine: 104). Ein Teilnehmer äußerte explizit, dass er darauf achte, dass die Zahlen- und Textbewertung stimmig sind (Quentin: 209). Bei drei Bewertungen fallen einigen Teilnehmern Widersprüche zwischen Zahlen- und Textbewertung auf. Einige Teilnehmer wundern sich, dass die Noten zu dem sehr gut bewertenden Kommentar bei Bewertung 1 (Bernd: 49-51, Franziska: 11, 17, 121, Nadja: 20, 43, 118) und bei Bewertung 2 (Hanna: 42, 137-140) schlechter ausfallen. Franziska erklärt sich dies dann damit, dass es wohl bei der Notenbewertung eher um Rahmenbedingungen ginge, als um den Arzt selbst (Franziska: 125-127). Wie bereits bei der Nachvollziehbarkeit der Begründung erwähnt, sind bei der fünften Bewertung außerdem Teilnehmer überrascht, dass die Terminvergabe und die Wartezeit mit der Note 2 statt mit einer 1 bewertet wurden. Insgesamt fällt der Faktor der *Übereinstimmung der Zahlen- und Textbewertung* aber nicht sehr häufig ins Gewicht.

Die Dimensionen des Vertrauens in Arztbewertungen wurden in dieser Studie häufig thematisiert. Die bisher erforschten Faktoren fanden sich in der Studie alle wieder. Auf der Ebene dieser Faktoren ergaben sich keine Ergänzungen. Wie bereits im Kapitel zum theoretischen Modell erwartet, gibt es keine Faktoren zur Dimension der *Themenselektivität*. Unklar bleibt aber, warum auch die Dimension der *Richtigkeit von Beschreibungen* nicht vorkommt. Das mag darauf zurückzuführen sein, dass es bei Online-Arztbewertungen nicht um Beschreibungen geht, die überprüfbar sind, sondern um subjektive Meinungen und persönliche Erfahrungen der Bewerter. Das theoretische Modell müsste dann dementsprechend modifiziert werden. Für das Fallbeispiel der Arztbewertungen scheinen nur die Dimension der *Faktenselektivität* und der *Expliziten Bewertungen* relevant zu sein. Das kann auch als Hinweis darauf verstanden werden, dass die Erwartungen an Arztbewertungen als

öffentliche Kommunikation spezifischer sind als die Erwartungen an andere Formen der öffentlichen Kommunikation. Die zwei übrigen Dimensionen wurden dafür sehr häufig thematisiert. Fast alle Faktoren, die sich auf die Dimensionen der *Faktenselektivität* bzw. der *Expliziten Bewertungen* bezogen konnte in dieser Studie noch genauer ausdifferenziert werden. Zunächst wurde bestätigt, dass viele Teilnehmer z. B. Wert legen auf die *Ausführlichkeit*, *Ausgewogenheit* oder *Argumentation der Bewertung*. Anschließend konnte gezeigt werden, dass sich die Wahrnehmung der Teilnehmer davon, ob eine Bewertung ausführlich, ausgewogen oder gut begründet ist, sich unterscheiden. Das ist z. B. darauf zurückzuführen, dass die Teilnehmer verschiedene Informationen für relevant halten und unterschiedliche Maßstäbe zur Bewertung anlegen. Es wurde deutlich, welche verschiedenen Aspekte die Faktoren umfassen und wie heterogen und teilweise entgegengesetzt die Teilnehmer auf dieselbe Bewertung reagieren.

#### 4.7 GRÜNDE FÜR VERTRAUEN IN ARZTBEWERTUNGEN

Bei den meisten bisher untersuchten Faktoren handelt es sich um Gründe für Vertrauen. Auch in dieser Studie spielen die Gründe für Vertrauen eine große Rolle. Meistens handelt es sich dabei um Faktoren aus der Gruppe der unspezifischen Gründe. Dem theoretischen Modell folgend, stelle ich nun aber zunächst die spezifischen Gründe vor.

Wie bereits gezeigt, wurden bisher keine Faktoren der Gruppe *Bestätigung durch Eigenerfahrung* erhoben. Die Teilnehmer der vorliegenden Studien teilen auf Nachfrage im Interview ihre eigenen Erfahrungen mit Arztbewertungen mit. In den meisten Fällen bestätigten sich positive Bewertungen (Bernd: 240, Hanna: 121, Karin: 35, Otto: 79-81, Renata: 78, Sabine: 62). Zwei Teilnehmer haben einen Arzt trotz negativer Bewertungen besucht, da bei diesem schnell ein Termin frei war. Doch die negativen Bewertungen bestätigten sich (Franziska: 267, Nadja: 94). Nur eine Teilnehmerin hat die Erfahrung gemacht, dasss Arztbewertungen nicht zutrafen. Sie war enttäuscht von einem gut bewerteten Arzt (Renata: 68), aber fand eine schlecht bewertete Ärztin gut (Renata: 100). Keiner der Teilnehmer zieht allerdings seine bisherige Erfahrung mit Arztbewertungen als Grund für Vertrauen in die nun gelesenen Arztbewertungen heran.

Wie bereits erläutert, ist aufgrund des Fallbeispiels Arztbewertungen nur noch einer der bisher erhobenen Faktoren als *spezifischer Grund* durch *Bestätigung durch mitgeteilte Erfahrungen anderer* zu verstehen: die *Bewertung der Bewertung*. Bei *Jameda* können Leser abstimmen, wie „hilfreich“ sie eine Bewertung fanden. Die von *Jameda* unter jeder Bewertung gestellte Frage „Wie hilfreich fanden Sie diese Bewertung?“ wird nur von drei Teilnehmern verbalisiert (Bernd: 114, Erika: 22, Quentin: 24). Die anderen Teilnehmer nahmen diese Frage entweder nicht wahr oder lasen sie nicht vor. Nur ein Teilnehmer äußert sich zu der *Bewertung der Bewertung*. Er bezieht diese aber nicht als Grund für Vertrauen ein, weil sie „sehr zufallsabhängig ist“ (Quentin: 58). Außerdem hält er das Kriterium „hilfreich“ für zu „unbestimmt“ (Quentin: 225-227) und eher für eine „Standardfrage im Netz“ (Quentin: 231).

Die Gruppe der *spezifischen Gründe* spielt für die Teilnehmer dieser Studie demnach keine Rolle. Das kann verschiedene Ursachen haben. Zum einen kann es daran liegen, dass die meisten Teilnehmer bisher recht wenig Erfahrung mit Arztbewertungen gemacht haben. Es kann aber auch sein, dass sie ihre bisherigen Erfahrungen nicht auf die aktuelle Situation übertragen wollen, da es sich um andere Bewerter handelt. Dass die *Bewertung der Bewertung* nur für einen Teilnehmer relevant ist, lässt sich eventuell dadurch erklären, dass diese Funktion sehr unauffällig platziert ist. Dass diese Funktion auch außerhalb der Studie kaum genutzt wird, zeigt sich darin, dass für die meisten Bewertungen auf dieser Plattform keine Bewertung abgegeben wurde.

Wesentlich häufiger werden dagegen *unspezifische Gründe* für Vertrauen genannt. Am häufigsten ziehen die Teilnehmer einen *Rückschluss von der Mitteilung auf die Selektion*. Dabei fällt vielen Teilnehmern die *Emotionalität der Bewertung* auf. Einige Teilnehmer erkennen an der fehlerhaften *Rechtschreibung und Grammatik* (Julia: 123, Nadja: 130, Sabine: 110) oder an der extremen *Valenz der Bewertung* die „emotionale Erregtheit“ (Julia:123) des Bewerter. Die *Emotionalität der Bewertung* wird von Teilnehmern bei den beiden negativen Bewertungen, 3 und 6, thematisiert. Einige Teilnehmerinnen sind sich einig darüber, dass der Bewerter „aggro“ (Dorothee: 55), „sehr, sehr emotional“ (Franziska: 51), „impulsiver“ (Ina: 105) bzw.

„bisschen sauer“ (Julia: 65) ist. Für einen Teilnehmer wirkt die Bewertung wie „ein kleiner Wutausbruch“ (Otto: 153).

Diese Emotionalität fassen zwei Teilnehmerinnen als Hinweis darauf auf, dass die Bewertung „sehr subjektiv“ (Karin: 21) bzw. „nicht objektiv“ ist (Franziska: 51). Eine Teilnehmerin vergleicht Bewertung 3 mit anderen Bewertungen. Sie stellt fest, dass die Parkmöglichkeiten bei anderen mit der Note 3 bewertet wurden, der Verfasser dieser Bewertung aber eine 5 oder 6 dafür gegeben hat. Daraus schließt sie, dass die Bewertung „ein bisschen gefärbt durch die Emotionen“ (Franziska: 149) ist. Dasselbe stellt sie auch bei Bewertung 6 fest. Sie meint, dass sich die Wut des Bewerterers auf andere Bewertungsaspekte übertragen hat. Dazu sagt sie: „Barrierefreiheit, daran ändert sich auch nichts, wenn der Arzt gut war oder nicht“. Die Beobachtung, dass ein Bewerter alle Aspekte negativ bewertet, selbst solche, die andere positiv bewerten, zeigt für sie, „dass der [Bewerter] noch nicht mit der Situation ein Schritt zurück gehen konnte und eine realistische Bewertung abgegeben hat“ (Franziska: 167). Für sie ist die *Emotionalität der Bewertung* daher ein Grund, dem Bewerter nicht zu vertrauen. Ein anderer Teilnehmer sieht das genau anders. Er meint. „solche [impulsiven] Personen kennt man eigentlich auch dadurch, dass sie sagen, was Sache ist.“ (Otto: 192).

Die *Emotionalität der Bewertung* wird nur bei einer einzigen positiven Bewertung thematisiert. Die neunte Bewertung wird von vier Teilnehmerinnen als „sehr persönliche Bewertung“ bzw. als „subjektiv“ wahrgenommen (Dorothee: 69, Gesa: 53, Paula: 40, Renata: 56). Allerdings ziehen diese Teilnehmerinnen aus derselben Wahrnehmung sehr unterschiedliche Schlüsse bezüglich der Vertrauenswürdigkeit des Bewerterers. Eine Teilnehmerin glaubt dem Bewerter, da hier eine „persönliche Ebene angesprochen wird [...] Ich weiß nicht, ob sich das jemand aus den Fingern saugen würde.“ (Dorothee: 69). Für sie ist die Emotionalität also ein Zeichen der Authentizität der Bewertung. Eine andere Teilnehmerin findet die Bewertung zunächst „bisschen pathetisch“ (Gesa: 138), beschließt später aber dennoch: „sehr persönlich, aber definitiv weitergebbar.“ (Gesa: 53). Dagegen hält eine Teilnehmerin nichts von der Bewertung. Sie glaubt, dass der Bewerter einfach sehr glücklich ist, dass sein Vater noch lebt und diese Freude einfach auf den Arzt übertrage (Paula: 99). Auch eine andere Teilnehmerin findet, dass die Bewertung

nichts aussagt (Renata: 56). Die *Emotionalität der Bewertung* wird bei beiden negativen Bewertungen und einer positiven Bewertung thematisiert. Für einige Teilnehmer wirkt eine hohe Emotionalität nicht vertrauenswürdig, für wenige schon.

Da die Bewerber bei *Jameda* keine Fotografien hochladen können, kommt der Faktor *Fotos des Bewerteten* in dieser Studie nicht auf Arztbewertungen bezogen vor. Nur eine Teilnehmerin erwähnt, dass sie das Risiko einer Fälschung ausschließen könne, wenn der Online-Bewertung ein Video hinzugefügt ist (Julia: 113-117).

Dagegen wird die *Rechtschreibung und Grammatik* einer Bewertung häufig thematisiert. Sprachliche Eigenheiten können an sich Gründe für Vertrauen sein. Die Sprache einer Bewertung dient den Teilnehmern aber auch häufig als Indikator für einen Grund. Sie überträgt also Informationen über den Bewerber, die ihn mehr oder weniger vertrauenswürdig erscheinen lassen. Wenn die Sprache nur herangezogen wird, um auf einen dahinterliegenden Grund zu schließen, dann ordne ich die Äußerung der Teilnehmer direkt diesem Grund zu. Alle anderen Fälle finden sich hier.

Einige Teilnehmer halten Bewertungen mit vielen Fehlern bei *Rechtschreibung und Grammatik* nicht für vertrauenswürdig (Charlotte: 75, Franziska: 45, Ina: 93, Julia: 125, 155; Quentin: 24, 35, 121). Gleichzeitig stellt eine Teilnehmerin aber fest, dass sie nur selten Online-Bewertungen sieht, die grammatisch korrekt formuliert sind (Franziska: 45). In einigen Fällen halten Teilnehmer eine Bewertung aber auch für besonders vertrauenswürdig, gerade weil sie Fehler enthält (Franziska: 251, Otto: 38, Quentin: 21). Alle drei Teilnehmer ziehen aber auch die Möglichkeit in Betracht, dass diese Fehler mit Absicht eingebaut wurden, um eine gefälschte Bewertung authentischer erscheinen zu lassen (Franziska: 251, Otto: 229, Quentin: 21).

Anderen Teilnehmer fallen sprachliche Fehler zwar auch auf, diese spielen aber für ihre Beurteilung der Vertrauenswürdigkeit des Bewerbers keine Rolle (Gesa: 22, Nadja: 28, 72). Drei Teilnehmer sind dann mit Fehlern nachsichtig, wenn sie denken, dass es sich bei den Bewertern um Nicht-Erstsprachler handelt (Anna: 62, Gesa: 80, Otto: 149). Die verschiedenen Erwartungen der Teilnehmerinnen möchte ich anhand zweier Zitate verdeutlichen. Eine Teilnehmerin meint, man könne „von jedem, der so Bewertungen abgeben kann, darf, will, erwarten, dass [er] Hochdeutsch spricht.“ (Ina: 93). Eine andere Teilnehmerin sagt

dagegen, sie „erwarte kein 1A Deutsch“ (Hanna: 164) bei Online-Bewertungen. Dementsprechend fallen dann ihre Vertrauensurteile aus.

Auffällig ist die Formulierung eines Teilnehmers, er verwendet den Begriff der „warme[n] Glaubwürdigkeit“ (Otto: 213). Otto erkennt diese in der Formulierung einer Bewertung: „Dieses doppelte Verstärken: „Sehr, sehr lange. Sehr, sehr zufrieden.“ verleiht dem eine gewisse warme Glaubwürdigkeit [...] Diese persönliche Verstärkung, wenn man nach Worten sucht und einfach dass „sehr“ mit einem „sehr“ potenziert“ (Otto: 213-217). Der sprachliche Stil der Bewertung ist für ihn also vertrauensrelevant.

Vier Teilnehmer begründen ihren Fokus auf die Sprache des Bewerbers mit ihrem Beruf. Sie setzen sich z. B. wissenschaftlich mit Sprache auseinander (Gesa: 78), korrigieren als Nebenjob Texte (Nadja: 128) oder haben viel mit schriftlicher Kommunikation zu tun (Quentin: 125). Eine Teilnehmerin sagt, ihr fallen Fehler auf, weil sie „in einem Feld tätig [ist], wo es wichtig ist, auf Details zu achten“ (Franziska: 109). Sowohl Franziska weist darauf hin, dass sie es damit „ein bisschen zum Extrem treibe“, als auch Quentin meint, dass er seine Kollegen damit „genervt [habe], dass [er] das immer ziemlich genau nehme.“ (Quentin: 119). Dieser Teilnehmer reflektiert aber auch kritisch, ob die *Rechtschreibung und Grammatik* einer Bewertung tatsächlich ein geeignetes Kriterium ist, um über die Vertrauenswürdigkeit eines Bewerbers zu entscheiden. Er meint: „Das ist vielleicht ein bisschen idiotisch, weil auch jemand, der nicht so gutes Deutsch kann und schreibt, natürlich als Patient einen Arzt auch sehr gut beurteilen kann.“ (Quentin: 125). Solch eine Infragestellung dieses Kriteriums findet bei den anderen Teilnehmern nicht statt.

Zusammengefasst lässt sich sagen, dass die *Rechtschreibung und Grammatik* einer Bewertung für viele Teilnehmer relevant ist. Sprachliche Fehler werden aber sowohl als Grund für wie auch als Grund gegen Vertrauen in den Bewerter herangezogen. Interessant ist außerdem, dass die Teilnehmer sich die Relevanz dieses Faktors für sich selbst zu erklären versuchen durch ihre berufliche Beschäftigung mit Sprache. Im nächsten Abschnitt taucht der Faktor der *Rechtschreibung und Grammatik* häufig als Indikator für die *Kompetenz des Bewerbers* auf.

Die *Kompetenz des Bewerter*s tritt in dieser Studie als Faktor auf mithilfe dessen die Teilnehmer einen Rückschluss von der Mitteilung auf die Selektion ziehen. Ein Aspekt der *Kompetenz des Bewerter*s ist das Bemühen des Bewerter. Sechs Teilnehmerinnen fällt negativ auf, wenn der Bewerter sich keine „Mühe gegeben“ bzw. „Gedanken gemacht“ hat. Einige Teilnehmerinnen machen dieses mindere Bemühen an den Fehlern in *Rechtschreibung und Grammatik* fest (Charlotte: 69, 73, Erika: 108, Franziska: 113-115, Hanna: 162, Sabine: 108, Sabine: 147). Einer Teilnehmerin kommt es dabei auf die Art und Anzahl der Fehler an. Solange es nur einige Fehler bei „Rechtschreibung, Kommasetzung und Grammatik“ gebe, schiebe sie das auf das schnelle Schreiben des Bewerter. Sobald aber „jemand auch bei der Rechtschreibung so gar nicht aufpasst, das Lautschrift ist, dann würde [sie] das schon skeptischer angucken“ (Ina: 91). Für durchgehendes Kleinschreiben hat sie dagegen Verständnis, weil das am Handy einfacher sei (Ina: 91). Drei Teilnehmer schließen von Fehlern in *Rechtschreibung und Grammatik* auf die geringe Reflektiertheit des Bewerter (Dorothee: 55, Quentin: 141, Sabine: 106-110, 149). Ähnlich wie Quentin es schon in seiner Selbstreflexion zu Fehlern in *Rechtschreibung und Grammatik* bemerkte, findet eine Teilnehmerin, dass Fehler in der Sprache nicht ausschlaggebend sind. Denn sie „würde jetzt nicht sagen, nur weil die Person kein vernünftiges Deutsch beherrscht, kann sie schlechter entscheiden.“ (Hanna: 170-172).

Neben der *Rechtschreibung und Grammatik* der Bewertung stellt der Inhalt der Bewertung für einige Teilnehmer einen Indikator für die Kompetenz des Bewerter dar. Eine Teilnehmerin macht die Bemühen des Bewerter z. B. an der *Ausführlichkeit der Bewertung* fest. Wenn ihr bestimmte Informationen fehlen, dann denkt sie, dass der Bewerter sich keine „Gedanken dazu gemacht hat“ (Paula: 32).

Einem Bewerter wird außerdem häufig Kompetenz beim Bewerten zugesprochen, wenn er sagt, dass er schon lange bzw. regelmäßig den Arzt besucht. Eine Teilnehmerin sieht das beim ersten Bewerter gegeben (Charlotte: 69), drei Teilnehmer bei dem achten und neunten Bewerter (Julia: 70, 125, 167, Nadja: 55, Paula: 39). Ein Teilnehmer erklärt diesen Grund für Vertrauen genauer. Er denkt, dass jemand, der nur sehr selten zu dem Arzt geht, dessen Kompetenz nicht beurteilen kann. Dagegen habe jemand, der schon seit fünfzehn Jahre Patient

desselben Arztes ist, schon mehrere Behandlungen erlebt und könne den Arzt daher besser beurteilen (Quentin: 97). Dementsprechend finden einige Teilnehmer die vierte Bewertung nicht fundiert, da sie nur auf einem Arztbesuch beruht (Dorothee: 47, Quentin: 52, 113-117, Julia: 77). Dasselbe gilt für einen Teilnehmer für Bewertung 5 (Quentin: 56). Für eine andere Teilnehmerin trübt die langjährige Erfahrung des neunten Bewerter allerdings dessen Urteilsfähigkeit. Sie glaubt, dass der Bewerter dann eher den Arzt als Person bewertet, da „da auch eine persönliche Bindung“ (Paula: 101) ist.

Die *Valenz der Bewertung* wirkt ebenfalls unterschiedlich auf die Teilnehmer. Zwei Teilnehmer finden negative Bewertungen verdächtig (Renata: 216, Quentin: 45, 133). Für zwei Teilnehmerinnen sind hingegen positive Bewertungen verdächtig, weil sie glauben, dass die bewerteten Ärzte diese selbst verfasst haben (Dorothee: 133, 195; Ina: 125-127). Auch andere Teilnehmer glauben positiven Bewertung weniger als negativen Bewertungen (Franziska: 82, Quentin: 70, 133). Die Teilnehmer finden Einser-Bewertungen z. B. „übertrieben“ (Bernd: 244, Franziska: 58), „dubios“ (Bernd: 292), „suspekt“ (Otto: 204) oder nicht differenziert (Dorothee: 55, Otto: 123). Eine Teilnehmerin meint, sie würde sich wundern, wenn ein Arzt nur positiv bewertet würde, da das unrealistisch sei (Sabine: 176). Ein anderer Teilnehmer sieht das genau so und führt ein konkretes Beispiel an: „‘Telefonische Erreichbarkeit‘, 1,0, das gibt es ja eigentlich nicht, [...] da ist doch irgendwann mal belegt.“ (Bernd: 279-280).

Zwei Teilnehmer führen diese vielen Einser-Benotungen darauf zurück, dass der Bewerter einfach schnell durchgeklickt hat (Nadja: 122, Quentin: 67). Einer der Teilnehmer erläutert das genauer:

„Man muss bei Bewertungen daran denken, wie man selber Bewertungen macht. Da passiert es mal, dass man genervt ist von den verschiedenen Kategorien und die auch gar nicht richtig einschätzen kann aufgrund der eigenen Erfahrung. Und dann haut man überall, wenn einem die zu bewertende Person oder Institution sympathisch ist, eine Bestnote rein, auch wenn man's wirklich gar nicht bewerten kann.“ (Quentin: 269-271)

Eine Teilnehmerin denkt dagegen, dass eine durchgehend positive Bewertung daran liegt, dass es „viele Leute [gibt], die [ihre Sichtweise] optimieren“ (Dorothee: 62) wollen. Eine andere Teilnehmerin mutmaßt dagegen, dass der Bewerter sich scheut, Kritik zu üben (Franziska: 239-245).

Unabhängig von der konkreten Bewertung halten es zwei Teilnehmerinnen für wichtig, auch negative Bewertungen an ihren Freund weiterzugeben. (Anna: 66, Charlotte: 65). Eine Teilnehmerin denkt, dass man „generell [...] eher dazu [neigt], die negativen Sachen weiterzuleiten“ (Dorothee: 45). Das vermutet eine andere Teilnehmerin auch bei den Bewertern. Sie führt es auf die „menschliche Natur“ (Renata: 116) zurück, dass man Bewertungen nicht verfasst, wenn man zufrieden ist, sondern dann, wenn man „sauer [...] und enttäuscht“ (Renata: 118) ist. Auch bei der *Valenz der Bewertung* zeigen sich also große Unterschiede zwischen den Teilnehmern. Deutlich wird hier auch, dass die *Valenz der Bewertung* häufig im Zusammenhang steht mit der *Ausgewogenheit der Bewertung*. So thematisieren die Teilnehmer meist nicht nur, dass eine Bewertung positiv bzw. negativ ist, sondern dass sie durchgehend positiv oder durchgehend negativ ist.

Es wurden also vier Faktoren gefunden, die zu den *unspezifischen Gründen* für Vertrauen gehören, in denen Teilnehmer einen *Rückschluss von der Mitteilung auf die Selektion* ziehen. Die ebenfalls in der bisherigen Forschung untersuchten Faktoren *Qualität der Bewertung* und *Unmittelbarkeit der Bewertung* zeigten sich in dieser Studie nicht. Den vorkommenden Faktoren ist gemein, dass die Teilnehmer sich meist nicht hinsichtlich der Wahrnehmung der Faktoren unterscheiden aber hinsichtlich ihrer Beurteilung des Beobachteten im Kontext des Vertrauens in den Bewerter.

Die letzte Gruppe der *Gründe für Vertrauen* gehört zu den *spezifischen Gründen für Vertrauen*, bei denen die Teilnehmer einen *Rückschluss vom Kontext der Mitteilung auf die Selektion* ziehen. Diese Faktoren werden selten erwähnt.

So kommt es z. B. nur einmal vor, dass ein Teilnehmer die *Ähnlichkeit des Bewerbers* explizit als möglichen Grund für Vertrauen anführt. Eine Teilnehmerin meint, dass es für ihr Vertrauen wichtig sein könne, dass sie sich „in die Person hineinversetzen kann“ (Ina: 97). Sie könne sich z. B. besser mit einem Bewerter identifizieren, wenn er einen ähnlichen Schreibstil benutzt wie sie selbst (Ina: 97). Auch eine andere Teilnehmerin thematisiert den Schreibstil einer Bewertung. So sagt sie über Bewertung 7:

„ich würde nie jemandem sagen ‚ey, ich hab so nen netten Zahnarzt, der ist voll cool und korrekt‘. [...] Ich sage: ‚Der war total nett, ich musste nicht so

viel warten und hat das kompetent rübergebracht.' Aber ich würde nie sagen ‚ist korrekt‘“(Gesa: 92-94)

Danach gibt die Teilnehmerin allerdings zu bedenken, dass die Bewertung für jemand anderen, der dem Bewerter ähnlicher ist, durchaus vertrauenswürdig sein kann, z. B. für ihren jüngeren Bruder (Gesa: 94). In dieser Studie werden aber keine konkreten Beispiele aus den Bewertungen genannt, in denen die Teilnehmer die *Ähnlichkeit des Bewerter*s als Grund für ihr Vertrauen in den Bewerter heranziehen.

Zum Kontext der Bewertung gehört außerdem die *Anzahl der Bewertungen*. Dabei formulieren mehrere Teilnehmerinnen das Problem, dass es bei Ärzten häufig keine (Franziska: 137, Nadja: 218) oder nur sehr wenige Bewertungen gibt (Dorothee: 92, Ina: 71, Julia: 93). Einige Teilnehmer betonen, dass sie sich nicht auf eine einzige Bewertung verlassen, sondern viele Bewertungen durchlesen (Franziska: 52, 129, 131). Dafür führen sie verschiedene Gründe an. Eine Teilnehmerin möchte „gucken, wo die Allgemeinheit hinstrebt“ (Charlotte: 85), eine andere tut das, „um einen differenzierten Blick zu entwickeln“ (Dorothee: 86) und eine Teilnehmerin möchte sich „annähernd ein objektives Bild machen“ (Hanna: 109). Auf die Frage danach, wie hoch die *Anzahl der Bewertungen* sein sollte, geben die Teilnehmer unterschiedliche Antworten. Zwei Teilnehmerinnen reichen fünf Bewertungen (Hanna: 111, Renata: 204), vier Teilnehmer wünschen sich mindestens zehn Bewertungen (Dorothee: 94, Nadja: 102, 264, Otto: 261, Paula: 83), zwei lesen gerne mehr als 50 (Julia: 95) und eine Teilnehmerin sogar bis zu 100 (Karin: 51-53) Online-Bewertungen.

Eine Teilnehmerin kann die benötigte Anzahl nicht festlegen. Sie sagt, es handle sich dabei um ein „Gefühl“ (Franziska: 133). Dieses könne unterschiedlich ausfallen (Franziska: 134). Die zehn in dieser Studie vorgelegten Arztbewertungen hält sie aber z. B. für ausreichend (Franziska: 134-135). Für eine andere Teilnehmerin gibt es einen Zusammenhang zwischen der *Anzahl an Bewertungen*, die sie liest, und der *Übereinstimmung mit anderen Bewertungen*. Sie lese sich so viele Bewertungen durch, bis sie „beruhigt“ (Charlotte: 87) sei. Das könnten „manchmal zehn sein, wenn alle gut sind. Wenn dann aber drei schlecht und sieben gut sind, dann lese ich mir noch mal zehn oder zwanzig weitere durch.“ (Charlotte: 87). Beide Teilnehmerinnen scheinen sich hier also auf eine unbestimmbare Intuition zu verlassen was die *Anzahl der Bewertungen* angeht. Allgemein zeigt sich, dass es für einige Teilnehmer relevant

ist, mehrere Bewertungen zu lesen, wenn sich auch die gewünschte Anzahl zwischen fünf und 100 Bewertungen bewegt oder gar nicht genau festlegbar ist.

Der eben bereits erwähnte Faktor der *Übereinstimmung mit anderen Bewertungen* ist für viele Teilnehmer wichtig. Die Voraussetzungen dafür, um eine Übereinstimmung oder einen Widerspruch festzustellen, sind zum einen, dass es mehrere Bewertungen gibt (siehe *Anzahl an Bewertungen*) und zum anderen, dass die Leser die Bewertungen untereinander vergleichen und zueinander in Beziehung setzen. In dieser Studie ist das bei vielen Teilnehmern der Fall (Dorothee: 121, Franziska: 48, Gesa: 34, Ina: 141, Julia: 205, Nadja: 44, Otto: 255, Quentin: 159, Sabine: 24). Zwei Teilnehmer achten dabei auf die Verteilung der Bewertungen (Julia: 107, Karin: 57), ein Teilnehmer sortiert die Ausreißer aus (Quentin: 161). Für einige Teilnehmerinnen wirken Wiederholungen vertrauenswürdig (Dorothee: 155, Gesa: 124, Karin: 51, Nadja: 43, Paula: 81). Eine Teilnehmerin erklärt das damit, dass mehrmals vorkommende Informationen ein Hinweis darauf sind, dass etwas „kein Einzelfall [sondern] wahrscheinlich wirklich so“ (Anna: 50-52) ist. Für andere Teilnehmerinnen werden die Aussagen einer einzelnen Bewertung bestätigt, wenn die Mehrheit der Bewerter auch dieser Meinung ist (Franziska: 137, Nadja: 202, Renata: 45).

Bei Widersprüchen zwischen den Bewertungen zieht eine Teilnehmerin, ähnlich wie in dem Beispiel oben, weitere Bewertungen zu Rate (Dorothee: 94). Ein einziger Teilnehmer weist darauf hin, dass Widersprüche zwischen den Bewertungen auch berechtigt sein können. Als Beispiel dafür nennt er die Benotung der Barrierefreiheit. Wer einen Fahrstuhl brauche, weil er schlecht zu Fuß sei, bewerte die Barrierefreiheit eben anders als jemand, der gut zu Fuß sei und daher nicht auf einen Fahrstuhl angewiesen ist (Bernd: 290).

Neben diesen Faktoren, die als *unspezifische Gründe für Vertrauen* herangezogen werden, erwähnen die Teilnehmer noch zwei andere Faktoren. Diese scheinen für sie aber keine Rolle in der Zuschreibung von Vertrauen zu spielen. Dazu gehört zum einen die *Erfahrung des Lesers mit Online-Bewertungen*, zum anderen die *Vertrautheit mit der Plattform*. Dazu wurden die Teilnehmer dieser Studie im Problemzentrierten Interview befragt. Bis auf zwei Teilnehmer (Otto: 53, Quentin: 153) haben alle Teilnehmer schon einmal Arztbewertungen genutzt. Acht

Teilnehmer kennen das Arztbewertungsportal *Jameda*, keiner der Teilnehmer kennt weitere Arztbewertungsportale wie *Weisse Liste* oder *Docinsider*. Eine Teilnehmerin liest „manchmal bis selten“ (Erika: 72) Arztbewertungen und wenn, dann eher zufällig. Dieses zufällige Lesen geschieht dadurch, dass sie nach der Telefonnummer einer Praxis sucht und daraufhin auch Bewertungen für den Arzt angezeigt werden (Erika: 80). Ähnlich geht es zwei weiteren Teilnehmerinnen, die daraufhin dann die angezeigten Arztbewertungen überfliegen (Ina: 69-75, Gesa: 63, 71).

Neben der Nutzung von Arztbewertungen als Leser wurde außerdem erhoben, ob die Teilnehmer schon selbst als Bewerter tätig waren. Nur zwei Teilnehmer haben schon Erfahrungen mit dem Verfassen von Arztbewertungen gesammelt (Bernd, Ina). Bernd schrieb zwei Arztbewertungen, weil er seinen Ärzten „einen Gefallen tun“ (Bernd: 202) wollte. Diese Ärzte waren von anderen schlecht bewertet worden. Daraufhin hat er positive Arztbewertungen über sie verfasst (Bernd: 212, 214, 318). Der Teilnehmer fügt aber scherzhaft hinzu, dass ihm seine positiven Bewertungen letztendlich selbst schaden, da die Ärzte dann mehr Patienten und damit weniger Termine für ihn frei hätten (Bernd: 326). Ina berichtete, dass sie in einer Arztpraxis eine Arztbewertung auf einem Tablet abgegeben habe (Ina: 59). Ihr Arzt hatte sie dazu aufgefordert; die Abgabe der Bewertung wurde in ihrer Akte vermerkt (Ina: 67). In keinem der Fälle thematisierten die Teilnehmer ihre Erfahrung mit dem Verfassen der Arztbewertung allerdings im Kontext des Vertrauens in den Bewerter, dessen Bewertung sie gerade gelesen hatten.

Einige Teilnehmer äußern sich dazu, warum sie keine Arztbewertung verfasst haben. Eine Teilnehmerin sagte, sie sei dafür „zu faul“ (Dorothee: 112) und versteht daher auch nicht, dass andere Menschen Arztbewertungen abgeben. Sie selbst würde nur eine Bewertung abgeben, „wenn [sie] sehr begeistert wäre oder sehr enttäuscht“ (Dorothee: 114). Letzterem stimmt eine andere Teilnehmerin zu (Renata: 118). Ein anderer Teilnehmer würde dagegen genau dann, wenn er unzufrieden mit einem Arzt ist, keine Bewertung abgeben (Bernd: 320-322).

Zwei Teilnehmerinnen haben zwar keinen Arzt, aber dafür schon andere Produkte online bewertet (Karin: 91, Sabine: 74). Ein Teilnehmer hat noch keine Online-Bewertungen verfasst, aber verfügt über langjährige Erfahrung mit Offline-Bewertungen. Diese Erfahrung beschreibt er so:

„in meinem Berufsleben wurde ich Opfer diverser Umfragen zu allem Möglichen [...] Das heißt, ich war häufig in der Situation, so einen Mist zu beantworten und hab' dann halt irgendwie dededededet [...] Je nachdem, ob ich fand, das ist eine gute Sache oder nicht, habe ich dann halt schlecht oder gut benotet. Und dann auch wirklich sehr willkürlich mal eine andere Note gegeben.“ (Quentin: 261 - 265)

Er ging also vor allem so vor, dass er seinen generellen Eindruck – positiv oder negativ – in Einzelnoten umsetzte, diese aber mit zufällig gewählten entgegengesetzten Noten ergänzte. Dieses eigene Erfahrungswissen darüber, wie Bewertungen zustande kommen, wendeten Teilnehmer jedoch nicht explizit an, um die Vertrauenswürdigkeit der Arztbewertungen einzuschätzen.

Die Perspektive des Bewerteten kennen nur zwei Teilnehmer. Eine Teilnehmerin wurde in ihrem Job in der Gastronomie online bewertet (Hanna: 251). Eine andere Teilnehmerin las ab und an die Bewertungen im Offline-Kummerkasten des Krankenhauses, in dem sie tätig war (Paula: 127). Auch diese Erfahrung mit dem bewertet Werden schien für die Teilnehmerinnen bei der Vertrauenszuschreibung keine Rolle zu spielen.

Mehrere Faktoren aus früheren Studien könnten theoretisch vorkommen, finden sich in meiner Studie aber nicht. Dazu gehören *Soziale Nähe des Bewerter*, *Unabhängigkeit des Bewerter*, *Qualität der Plattform*, *Vertrauenswürdigkeit der Plattform*, *Gütesiegel für Plattform* und *Reputation der Plattform*. Das kann zum einen daran liegen, dass die Leser sehr wenige Informationen zum Bewerter haben. Zum anderen kannten viele Teilnehmer zwar die Plattform, hatten aber keine dezidierte Meinung zu ihr. Die Teilnehmer konzentrierten sich außerdem auf die Online-Bewertungen an sich, keiner von ihnen erkundete die dazugehörige Plattform. Das wäre innerhalb der Think-Aloud-Aufgabe möglich gewesen. Es ist aber nicht auszuschließen, dass die Aufgabenstellung mit dem starken Fokus auf die einzelnen Bewertungen dazu geführt hat, dass die Teilnehmer das Umfeld dieser Bewertungen nicht genauer betrachtet haben.

Es konnten aber einige Faktoren gefunden werden, die in der Forschung bisher noch nicht vorkamen. Dazu gehört *Authentifizierung des Bewerter*, *Ethnizität des Bewerter*, *Geschlecht des Bewerter*, *Alter des Bewerter*, *Bildung des Bewerter*, *Krankenversicherung des Bewerter* und *Alter der Bewertung*.

Die *Authentifizierung des Bewerter*s durch das Portal wurde von einer Teilnehmerin als Grund für Vertrauen in Online-Bewertungen formuliert (Nadja: 266-268). Das bedeutet, dass die Plattform überprüft, dass der Bewerter tatsächlich das Bewertete gekauft hat. Hier handelt es sich also um verlagertes Vertrauen von der Plattform zum Bewerter. So eine Funktion gibt es bei dem Arztbewertungsportal Jameda allerdings nicht.

Von einer Teilnehmerin wird außerdem die *Ethnizität des Bewerter*s thematisiert. Sie vermutet bei der ersten Bewertung eine für diese Studie manipulierte Bewertung: „Das war wahrscheinlich extra gemacht, ob man dann den Leuten, die Migranten-Background haben, vertraut oder nicht.“ (Dorothee: 98). Diese Wahrnehmung führte sie allerdings selbst nicht als Grund für oder gegen Vertrauen an. Wie bereits erwähnt, sind drei Teilnehmer bei Nicht-Erstsprachlern nachsichtiger mit Fehlern in *Rechtschreibung und Grammatik* (Anna: 62, Gesa: 80, Otto: 149). Eine mögliche Erklärung dafür kann sein, dass zwei dieser Teilnehmer selbst einen Migrationshintergrund haben. Ansonsten scheint die Ethnie des Bewerter)s keine Rolle für die Teilnehmer zu spielen.

Über das *Geschlecht des Bewerter*s stellen drei Teilnehmer Mutmaßungen an. Eine Teilnehmerin hält die vierte Bewerterin für eine Frau (Ina: 89). Den siebten Bewerter halten zwei Teilnehmerinnen für einen Mann, wegen seiner Verwendung des Wortes „cool“ (Paula: 115, Renata: 174). Keiner der Teilnehmer zieht das *Geschlecht des Bewerter*s jedoch als Grund heran, um sich für oder gegen Vertrauen zu entscheiden.

Häufiger als das Geschlecht wird das *Alter des Bewerter*s thematisiert. Ein Teilnehmer kombiniert das Alter aus der Angabe über Bewertung 2 mit der Aussage im Text der Bewertung (Quentin: 38). Eine Teilnehmerin schätzt aufgrund der Sprache den dritten Bewerter auf „zwischen 25 und 30“ (Renata: 140) Jahre alt. Fast alle Aussagen der Teilnehmer die mit dem *Alter des Bewerter*s zusammenhängen, beziehen sich auf Bewertung 7. Viele Teilnehmer erkennen hier einen jugendlichen Sprecher (Anna: 82, Dorothee: 57, Franziska: 105, 183; Gesa: 86, Nadja: 51, 76, Otto: 198). Die Teilnehmer leiten das Alter vom Inhalt und Ausdruck des Bewertungstexts ab (Paula: 115, Renata: 172). Sie halten den Bewerter für jung z. B.

wegen der Ausrufezeichen (Dorothee: 185) oder wegen des Wortes „cool“ (Renata: 174).

Das *Alter des Bewerter*s führt bei den Teilnehmern zu unterschiedlichen Urteilen. Für zwei Teilnehmer ist das Alter irrelevant (Anna: 86, Quentin: 40). Eine Teilnehmerin ist „bei den Jungen [...] etwas nachsichtiger mit der Ausdrucksweise“ (Nadja: 176). Hier gibt es also einen Zusammenhang mit der Sprache der Bewertung. Für zwei Teilnehmerinnen ist das jugendliche Alter wiederum ein Grund, um der Bewertung nicht zu vertrauen. Eine Teilnehmerin erklärt das damit, dass der Bewerber aufgrund seines Alters „andere Hintergründe, andere Maßstäbe“ (Gesa: 82-84) habe. Damit bringt sie das *Alter des Bewerter*s in Zusammenhang mit der *Argumentation der Bewertung*. Eine andere Teilnehmerin sieht dagegen einen Zusammenhang mit der *Kompetenz des Bewerter*s. Sie sagt:

„Wenn er wirklich einfach so ein Jugendlicher, der bei dem Arzt war, diese Bewertung geschrieben hat, dann ist da ja nichts dagegen zu sagen, das ist einfach sein Stil. Aber man glaubt ja doch irgendwie so ein bisschen eher den älteren Leuten, die da regelmäßig hingehen.“ (Julia: 125)

An diesen Zitaten zeigt sich deutlich, dass das *Alter des Bewerter*s mit anderen Faktoren in Verbindung stehen kann. Zum einen wird von der *Rechtschreibung und Grammatik* auf das jugendliche *Alter des Bewerter*s geschlossen, zum anderen wird davon dann wieder auf die niedrige *Kompetenz des Bewerter*s deduziert. Für manchen Teilnehmer dagegen ist das *Alter des Bewerter*s nicht wichtig.

Ein weiterer Faktor liegt in der *Bildung des Bewerter*s. Einige Teilnehmer schließen von einer fehlerhaften *Rechtschreibung und Grammatik* auf die geringe Bildung des *Bewerter*s (Hanna: 168-172, Julia: 147, Nadja: 130). Zwei Teilnehmerinnen sind sich darüber einig, dass der Bewerber aufgrund seiner geringen Bildung nicht angemessen bewerten kann. Eine Teilnehmerin denkt, dass solche Bewerber „nicht so differenziert auseinander halten können“ (Nadja: 130), bei was ihnen der Hausarzt helfen kann und bei was nicht. Eine andere Teilnehmerin sagt, sie „verfalle [...] in Klischees“ und denke, „jemand mit einem höheren Bildungsabschluss kann das differenzierter beurteilen“ (Julia: 125). Einige Teilnehmerinnen schließen also von der *Rechtschreibung und Grammatik der Bewertung* auf die niedrige *Bildung des Bewerter*s und von dort auf die geringe *Kompetenz des Bewerter*s.

Nur von zwei Teilnehmern wird die *Krankenversicherung des Bewerbers* angesprochen. Eine Teilnehmerin fragte sich kurz, ob der Arzt nur Privatpatienten behandelt, „weil er so viel Zeit hat“ (Sabine: 38). Ein Teilnehmer erwähnte die angezeigt Versicherungsform eines Bewerbers. Diese hielt er jedoch nicht für relevant, um sein Vertrauen in den Bewerter einzuschätzen. Seiner Meinung nach gilt: „Patient ist Patient“ (Quentin: 293).

Ein einziger Teilnehmer hat sowohl bei der ersten als auch bei der dritten Bewertung auf das *Alter der Bewertung* geachtet. Für ihn war das Alter wichtig, um zu beurteilen, ob die Bewertung noch gültig ist, „unabhängig von der Bewertung“ (Quentin: 22) an sich. Er erklärt dazu, dass er einer z. B. zehn Jahre alten Bewertung nicht vertrauen würde, sondern eher den aktuelleren Bewertungen (Quentin: 103, 109). Bei der dritten Bewertung schränkt er diese Aussage allerdings ein. Hier meint er, dass die Bewertung zwar etwas älter ist, „aber wenn jemand tatsächlich so negativ zu bewerten ist, dann wird sich das auch in zwei Jahren nicht geändert haben.“ (Quentin: 47). Außer diesem Teilnehmer thematisiert aber niemand das *Alter der Bewertung*.

In dieser Studie zeigen sich also einige vormals unbekannte unspezifische Gründe für Vertrauen. Außerdem wird deutlich, dass es sich bei den meisten genannten Faktoren um *unspezifische Gründe für Vertrauen* handelt, bei denen die Teilnehmer einen *Rückschluss von der Mitteilung auf die Selektion* des Bewerbers ziehen. Bei fast allen Faktoren werden große Unterschiede zwischen den Teilnehmern deutlich.

#### 4.8 WEITERE FAKTOREN UND ZUSAMMENHÄNGE ZWISCHEN DEN FAKTOREN

In diesem Kapitel geht es um weitere Faktoren und um Zusammenhänge zwischen den einzelnen Faktoren. Im theoretischen Modell wurde bereits erwähnt, dass die bisher zum Leser untersuchten Faktoren (*Alter des Lesers, Bildung des Lesers, Erfahrung des Lesers mit Online-Shopping, Ethnizität des Lesers, Fokus des Lesers, Geschlecht des Lesers, Internetnutzung des Lesers, Involviertheit des Lesers, Kontrollillusion des Lesers, Kultur des Lesers, Motivation des Lesers, Orientierung des Lesers an anderen, Personal Sense of Power des Lesers, Selbstoffenbarung des Lesers und Trägheit des Lesers*) mit allen anderen Faktoren interagieren können. Von allen diesen Faktoren thematisierten die Teilnehmer nur zwei selbst.

Bei dem einen Faktor handelt es sich um die *Kontrollillusion des Lesers*. Das Vermögen bzw. Unvermögen, die Vertrauenswürdigkeit des Bewerter einzuschätzen wird von den meisten Teilnehmern nicht thematisiert. Nur ein Teilnehmer denkt, dass er sehr gut einschätzen kann, ob ein Bewerter vertrauenswürdig ist. Er meint: „Habe nicht in Erinnerung, dass ich irgendwo mal mit meiner Entscheidung richtig daneben gelegen hätte.“ (Quentin: 79). Diese Selbsteinschätzung des Teilnehmers scheint aber keinen weiteren Faktor zu beeinflussen.

Im Gegensatz dazu gibt es Hinweise darauf, dass die *Bildung des Lesers* und die *Ethnizität des Lesers* mit anderen Faktoren zusammenhängen. Auch an anderen Stellen zeigt sich, dass Faktoren einander verstärken, sich gegenseitig abschwächen oder einander ausgleichen. Das mag auch der Grund dafür sein, dass einige Teilnehmer bei ihren Entscheidungen schwanken (Franziska, Julia, Otto). Eine Teilnehmerin sagt dazu: „es gab Gründe zum Weitergeben und nicht Weitergeben.“ (Erika: 137). In diesem Unterkapitel werden die einzelnen Fälle, in denen Faktoren zusammenhängen, noch einmal gesammelt aufgeführt.

Neben den Faktoren, die sich auf den Leser beziehen wird im theoretischen Kapitel auch noch erklärt, dass ein Faktor, der sich auf das Bewertete bezieht, mit anderen Faktoren interagieren kann: der *Typ des Bewerteten*. Wie bereits im Unterkapitel zu den Voraussetzungen für Vertrauen beschrieben, hat der *Typ des Bewerteten* einen Einfluss auf das *Risikobewusstsein* mancher Teilnehmer. Für einige Teilnehmer ist das *Risikobewusstsein* bei Arztbewertungen höher als bei anderen Online-Bewertungen, weil sie denken, dass so subjektive Erfahrungen mit einem Arzt schwierig übertragbar sind. Für andere Teilnehmer ist das *Risikobewusstsein* bei Arztbewertungen dagegen niedriger als bei anderen Online-Bewertungen, da sie davon ausgehen, dass bei solch einem *Typ des Bewerteten* niemand Bewertungen fälscht.

Der *Typ des Bewerteten* hat außerdem einen Einfluss auf die *Nutzung von Online-Bewertungen*. Zwei Teilnehmerinnen lesen bei anderen *Typen des Bewerteten* Online-Bewertungen, verlassen sich aber bei Ärzten lieber auf die Urteile von Fachpersonen oder Freunden (Dorothee: 16, Franziska: 89).

Die meisten Zusammenhänge mit anderen Faktoren werden aber bei dem Faktor *Rechtschreibung und Grammatik* deutlich. Von einigen Teilnehmern wird dieser Faktor in Zusammenhang gebracht mit drei Dimensionen von Vertrauen. So wird die schlechte *Rechtschreibung und Grammatik* einer Bewertung ausgeglichen mit einer hohen *Ausführlichkeit der Bewertung* (Hanna: 176) oder einer guten *Argumentation der Bewertung* (Charlotte: 73). Bei einer anderen Teilnehmerin wirkt sich die fehlerhafte *Rechtschreibung und Grammatik* auf die *Attribuierung der Bewertung* aus. Sie schreibt die Ursachen für die negative Bewertung dann dem Bewerter selbst zu (Nadja: 130). Eine weitere Teilnehmerin zieht aus der Kombination von einer schlechten *Rechtschreibung und Grammatik* sowie einer schlechten *Argumentation der Bewertung* auf die geringe *Kompetenz des Bewerter*s (Dorothee: 55). Anderen Teilnehmern reicht schon die schlechte *Rechtschreibung und Grammatik* alleine, um die *Kompetenz des Bewerter*s als niedrig zu interpretieren (Charlotte: 69, 73, Erika: 108, Franziska: 113-115 Hanna: 162, Quentin: 141, Sabine: 108, 110, 147, 149). Andere schließen daraus auf die niedrige *Bildung des Bewerter*s (Hanna: 168-172, Julia: 147, Nadja: 130).

Einen Zusammenhang zwischen der schlechter *Rechtschreibung und Grammatik* und der *Ethnizität des Bewerter*s wird von drei Teilnehmern hergestellt. Sie sind nachsichtiger bei Bewertern, die sie nicht für einen Erstsprachler halten (Anna: 62, Gesa: 80, Otto: 149-151). Hier könnte auch ein Zusammenhang mit der *Ethnizität des Lesers* hergestellt werden, da es sich bei zwei dieser Teilnehmer um solche handelt, die nicht nur Deutsch als Erstsprache haben. Dieser Einfluss des eigenen Migrationshintergrunds auf die Wahrnehmung einer Bewertung wird von den Teilnehmern allerdings selbst nicht thematisiert. Neben der *Ethnizität des Bewerter*s scheint auch das *Alter des Bewerter*s die Relevanz der *Rechtschreibung und Grammatik* zu verringern. So ist eine Teilnehmerin bei jungen Bewertern nachsichtiger mit der Sprache (Nadja: 176). Einige Teilnehmer setzen Fehler in *Rechtschreibung und Grammatik* in Verbindung mit einer hohen *Emotionalität der Bewertung* (Julia: 123, Nadja: 130, Sabine: 110). Die durchgehende Kleinschreibung eines Bewerter's sorgt für unterschiedliche Interpretationen seitens der Teilnehmer. Eine Teilnehmerin schließt davon auf eine niedrige *Bildung des Bewerter*s (Julia: 147). Ein anderer Teilnehmer findet darin einen Hinweis, dass der Bewerter

„unkonventionell“ (Quentin: 54) ist, „jemand, der sich dazu entschlossen hat, alternativ zu sein, zumindest in seinen schriftlichen Ausdrücken“ (Quentin: 197-207) und daher weniger vertrauenswürdig. Eine andere Teilnehmerin erkennt darin lediglich jemanden, der am Handy einfacher tippen möchte, wie sie selbst auch (Ina: 91).

Der Faktor *Rechtschreibung und Grammatik* steht außerdem im Zusammenhang mit der *Bildung des Lesers*. Denn vier Teilnehmer erklären die Relevanz, die *Rechtschreibung und Grammatik* der Bewertung für sie haben, mit ihrem eigenen Beruf (Franziska: 103, 109; Gesa: 78, Nadja: 128, Quentin: 119, 125).

Nicht direkt die *Rechtschreibung und Grammatik*, aber die Sprache einer Bewertung spielen für die Wahrnehmung der *Ähnlichkeit des Bewerbers* eine Rolle. Eine Teilnehmerin könnte sich z. B. besser mit einem Bewerter identifizieren, wenn dieser denselben sprachlichen Stil hat wie sie (Ina: 97). Eine andere Teilnehmerin weist ebenfalls darauf hin, dass es vertrauenswürdig sein kann, wenn Bewerter und Leser dieselben Formulierungen benutzen (Gesa: 94). Ebenfalls von der Sprache leiten viele Teilnehmer das *Alter des Bewerbers* ab (Anna: 82, Dorothee: 57, 185; Franziska: 105, 183; Gesa: 86, Nadja: 51, 76; Paula: 115, Renata: 140, 172, 174; Otto: 198).

Mit der Dimension *Ausführlichkeit der Bewertung* hängen zwei Gründe zusammen. Für zwei Teilnehmerinnen sinkt mit der steigenden *Anzahl an Bewertungen* der Anspruch an die *Ausführlichkeit der Bewertung* (Dorothee: 209, Julia: 135). Eine andere Teilnehmerin schließt von der *Ausführlichkeit der Bewertung* auf die *Kompetenz des Bewerbers*.

Die Dimension *Argumentation der Bewertung* hängt wiederum mit dem Grund *Alter des Bewerbers* zusammen. Eine Teilnehmerin schließt von einer nicht vorhandenen Begründung darauf, dass es sich um einen jungen Bewerter handelt (Renata: 168-172).

Die *Übereinstimmung mit anderen Bewertungen* spielt ebenfalls mit anderen Faktoren zusammen. Eine Teilnehmerin beschreibt eine Interaktion mit der *Anzahl der Bewertungen*. Sie erhöht die Anzahl der Bewertungen, die sie liest, wenn die Bewertungen nicht eindeutig übereinstimmen (Charlotte: 87). Eine andere

Teilnehmer stellt eine hohe *Emotionalität der Bewertung* dann fest, wenn die *Übereinstimmung mit anderen Bewertungen* gering ist (Franziska: 149).

Oben wurde schon besprochen, wie Teilnehmer von *Rechtschreibung und Grammatik* der Bewertung auf das *Alter des Bewerbers* schließen. Das *Alter des Bewerbers* wird dann wiederum mit anderen Gründen für Vertrauen in Verbindung gebracht. Eine Teilnehmerin hält junge Bewerber nicht für vertrauenswürdig, weil sie dessen Maßstäbe für die *Argumentation der Bewertung* nicht teilt (Gesa: 82-84). Eine andere Teilnehmerin hält seine *Kompetenz des Bewerbers* für geringer wegen seines Alters (Julia: 125).

Die *Valenz der Bewertung* wird sowohl mit der *Emotionalität der Bewertung* als auch mit der *Ausgewogenheit der Bewertung* in Verbindung gebracht. Eine Teilnehmerin erkennt in einer extremen Valenz die „emotionale Erregtheit“ (Julia: 123) des Bewerbers. Der Zusammenhang von *Valenz der Bewertung* und *Ausgewogenheit der Bewertung* wird von den Teilnehmern nicht expliziert. Es fällt jedoch auf, dass die Teilnehmer meist nicht nur darauf hinweisen, dass eine Bewertung positiv bzw. negativ ist (und sich damit auf die Valenz beziehen), sondern dazu, dass sie durchgehend positiv oder durchgehend negativ ist (und sich damit auf die Ausgewogenheit beziehen).

In diesem Unterkapitel wird deutlich, dass es bisher nicht untersuchte Zusammenhänge zwischen verschiedenen Faktoren gibt. Diese treten bisher meist nur vereinzelt auf.

#### 4.9 KONSEQUENZEN DES VERTRAUEN IN ARZTBEWERTUNGEN

In dieser Studie werden verschiedene *Konsequenzen des Vertrauens* in eine Arztbewertung sichtbar. Die in der Think-Aloud-Aufgabe angelegte Konsequenz von Vertrauen war das Weitergeben der Arztbewertung. Jeder Teilnehmer gab mindestens eine Bewertung weiter. Im Durchschnitt gaben die Teilnehmer sechs Bewertungen weiter. Ein Teilnehmer gab nur eine einzige Bewertung weiter, vier Teilnehmer gaben alle zehn Bewertungen weiter. Die Grenzen des Vorgehens, die Weitergabe der Arztbewertungen als Konsequenz des Vertrauens zu verstehen, wurden schon in anderen Unterkapiteln verhandelt (siehe 4.3 Methodisches Vorgehen: Think-Aloud, PZI und Qual. IA, S. 92 und 4.4 Auswahl der Bewertungen und der Teilnehmer, S. 97). Neben der Weitergabe an sich verbalisierten die

Teilnehmer verschiedene weitere *Konsequenzen des Vertrauens* in eine Arztbewertung.

Für viele Teilnehmer liegt eine *Konsequenz des Vertrauens* darin, dass der Teilnehmer der Bewertung entsprechend den Arzt besucht oder nicht besucht (Anna, Bernd, Charlotte, Franziska, Gesa, Hanna, Nadja). Dabei handelt es sich um die Übernahme der Fremdselektion als Eigenselektion wie ihm theoretischen Modell beschrieben. Einige Teilnehmer sagen, dass der Besuch eines Arztes für sie nicht von der Arztbewertung abhängt (Erika: 86, Franziska: 52, Gesa:43, Ina, 82, Nadja:88). So sagt Anna: „Ich bin auch schon zu Ärzten gegangen, die gar nicht mal so gute Bewertungen hatten.“ (Anna: 46). Vor dem Hintergrund des theoretischen Modells ist das so zu verstehen, dass diese Teilnehmer kein Vertrauen in bestimmte oder alle Arztbewertungen haben. In beiden Fällen wird jedenfalls die Entscheidung für oder gegen einen Besuch dieses Arztes als mögliche Folge des Vertrauens oder Nicht-Vertrauens thematisiert.

Wie in der Aufgabenstellung bereits angelegt, wird als weitere *Konsequenz des Vertrauens* in eine Online-Bewertung gesehen, dass die Bewertung weitererzählt oder weitergegeben wird (Bernd: 248, Franziska: 52, Nadja: 27, Nadja: 37, 58, 198, Otto: 38). Ein Teilnehmer sagt z. B. „Dem glaub ich das, diese Story würde ich sogar erzählen.“ (Bernd: 136). Auch die direkte Weiterempfehlung eines Arztes ohne den Umweg über die Bewertung anderer, wird als mögliche *Konsequenz des Vertrauens* beschrieben (Franziska: 50-52, Gesa: 55, 94, Karin: 18, Nadja: 27, 77).

Damit finden sich die vorher im theoretischen Modell formulierten *Konsequenzen von Vertrauen* auch in den Aussagen der Teilnehmer wieder.

#### 4.10 FAZIT

Mit der an die integrative Literaturübersicht und das theoretische Modell angeschlossenen empirischen Exploration können die bisherigen empirischen Ergebnisse ergänzt sowie das theoretische Modell ausdifferenziert werden. In der Studie wurde deutlich, dass die Teilnehmer, wie erwartet, zwischen dem Bewerter und der Bewertung als Vertrauensobjekt wechseln. Außerdem konnte jede Aussage der Teilnehmer einem Bestandteil des theoretischen Modells zugeordnet werden. Umgekehrt konnten auch zu fast jedem Bestandteil des theoretischen Modells Aussagen der Teilnehmer gefunden werden. Die Teilnehmer kommentierten die

Faktoren, die mit den *Voraussetzungen für Vertrauen* zusammenhängen, d. h. das *ungenügende Wissen* und das *Risikobewusstsein des Lesers* betrafen. Sie äußerten sich außerdem zahlreich zu *Faktenselektivität* und *Expliziten Bewertungen* als zwei *Dimensionen des Vertrauens*. Dabei griffen sie alle Faktoren, die bisher dazu erforscht wurden, auf. Die meisten Kommentare der Teilnehmer drehten sich jedoch um *unspezifische Gründe für Vertrauen*, bei denen die Teilnehmer einen *Rückschluss von der Mitteilung auf die Selektion* des Bewerbers ziehen. Daneben wurden einige neue Faktoren gefunden, die mit dem Vertrauen in Arztbewertungen zusammenhängen. Außerdem wurde gezeigt, dass einzelne Faktoren auf vorher unbekannte Weise miteinander interagieren. Nichtzuletzt fanden sich auch verschiedene *Konsequenzen des Vertrauens* wieder, die zuvor im theoretischen Modell ausgemacht worden waren.

Einige vorher untersuchte Faktoren kommen bei der empirischen Exploration von Vertrauen in Arztbewertungen nicht vor. Das liegt teilweise daran, dass bestimmte Funktionen auf dem untersuchten Arzbewertungsportal nicht gegeben sind, wie *Profilfoto des Bewerbers*, *Identifikation des Bewerbers*, *Erfahrung des Bewerbers*, *Warnhinweis der Plattform*, *Bewerterstatus* und *Eingebundenheit des Bewerbers in Community*. Mehrere Faktoren aus früheren Studien könnten theoretisch vorkommen, finden sich in meiner Studie aber nicht. Dazu gehören *Soziale Nähe des Bewerbers*, *Unabhängigkeit des Bewerbers*, *Qualität der Plattform*, *Vertrauenswürdigkeit der Plattform*, *Gütesiegel für Plattform* und *Reputation der Plattform*. Das kann daran liegen, dass die Leser sehr wenige Informationen zum Bewerter haben. Zum anderen kannten viele Teilnehmer zwar die Plattform, hatten aber keine dezidierte Meinung zu ihr. Die Teilnehmer konzentrierten sich außerdem auf die Online-Bewertungen an sich, keiner von ihnen erkundete die dazugehörige Plattform. Das wäre innerhalb der Think-Aloud-Aufgabe möglich gewesen. Es ist aber nicht auszuschließen, dass die Aufgabenstellung mit dem starken Fokus auf die einzelnen Bewertungen dazu geführt hat, dass die Teilnehmer das Umfeld dieser Bewertungen nicht genauer betrachtet haben. Absichtlich nicht codiert wurde, wie im theoretischen Teil erläutert, die *Vertrauenswürdigkeit des Bewerbers*. Die *Qualität der Bewertung* schien außerdem in anderen Faktoren aufzugehen, wie z. B.

*Argumentation* bzw. *Ausführlichkeit der Bewertung* als Dimensionen des Vertrauens sowie z. B. in der *Rechtschreibung und Grammatik der Bewertung*.

Auffällig ist, dass die meisten Aussagen der Teilnehmer den vorher bereits theoretisch deduzierten Faktoren zugeordnet werden konnten. Eine mögliche Erklärung ist, dass ich aufgrund meines Vorwissens aus der bisherigen Arbeit meine Fragen im Problemzentrierten Interview nur auf mir bereits bekannte Faktoren bezog. Durch eine suggestive Fragestellung hätten die Teilnehmer kaum eine Möglichkeit, andere Faktoren zu nennen. Dem Problemzentrierten Interview war allerdings die Think-Aloud-Aufgabe vorangestellt. Vor und während der Think-Aloud-Aufgabe hatte ich keinerlei Einfluss auf die Aussagen der Teilnehmer. Sie konnten frei sagen, was ihnen durch den Kopf geht und hätten hier alle möglichen Faktoren nennen können, die bisher nicht in der wissenschaftlichen Arbeit auftauchten. Für meine anschließenden Fragen an die Teilnehmer orientierte ich mich jeweils an dem dort Gesagten. Eine Überprüfung der Transkripte im Nachhinein ergab, dass ich alle dort verbalisierten Faktoren im folgenden Problemzentrierten Interview aufgriff. Es gab vor allem in den anfänglichen Interviews einige Stellen, an denen ich tiefer hätte nachfragen können, aber mit der Anzahl der Interviews wuchsen die Erfahrung und damit die Genauigkeit meiner Nachfragen. Eine weitere Erklärung dafür, dass in dieser Studie keine weiteren Faktoren gefunden werden konnten, kann sein, dass die Auswahl der Bewertung oder der Teilnehmer zu wenig vielfältig war. Die wahrscheinlichste Erklärung ist aber, dass in der bisherigen Forschung und im theoretischen Modell tatsächlich ein Großteil der relevanten Faktoren einbezogen wurde.

Allerdings konnten mithilfe der induktiven Datenerhebung auch einige Faktoren gefunden werden, die in der bisherigen Forschung nicht vorkamen. Die bisher zumeist verwendete Methode, Teilnehmern Antworten für die Faktoren des Vertrauens vorzugeben, schließt alle diejenigen Faktoren aus, welche die Autoren nicht vorher manipulieren und nicht abfragen. Es kann also nur gefunden werden, was erfragt wird. Vor dieser Studie gab es nur drei qualitativ-induktive vorgehende Studien. Bei der fast durchgängigen methodischen Homogenität blieb die Frage offen, ob mit einer induktiven Methode geklärt werden kann, warum die Ergebnisse der bisherigen Studien sich teilweise widersprechen. Der vom Forschungsstand

abgeleitete Forschungsbedarf sprach daher für eine offene, explorative Erhebungsmethode mit tatsächlich vorgefundenen Bewertungen als Stimulusmaterial.

Für gewöhnlich werden explorative Verfahren angewendet, wenn ein Phänomen erstmals untersucht wird. Im Fall von Vertrauen in Online-Bewertungen zeigte sich eine offene Erhebungsmethode auch nach über 50 publizierten Studien noch sinnvoll. Zum einen hätten die relevanten Faktoren für Vertrauen nicht vorher umfassend abgeleitet werden. Beldad et al. (2010) meinen „even if online trust research in e-commerce has somehow reached its level of maturity, as evidenced by the profusion of empirical studies, research in this domain is not yet saturated.“ (S. 868). Sie stellen fest:

„Only a handful of studies cited in this paper used a comprehensive model in predicting the factors that contribute to the development of Internet users' trust. Using a more complete framework that incorporates all possible personality-based (e.g., experience with the Internet and with online transactions), socio-psychological (e.g., propensity to trust, perception of risks involved in the exchange), sociological (e.g., reputation, third-party certification), socio-cultural (e.g., culture, education), and technical (e.g., website quality) factors would be very beneficial in empirical attempts at understanding online trust development.“ (S. 867)

Doch in den bisherigen Studien gibt es kein Modell, das alle möglicherweise relevanten Faktoren vorhersagen kann. Die Auswahl der untersuchten Einflussfaktoren scheint ziemlich beliebig. Innerhalb des recht allgemeinen Rahmens, den z. B. ELM bzw. HSM als Modelle vorgeben, ziehen die Autoren beliebige Ansätze heran, um neue Einflussfaktoren zu finden. Es scheint keine grundlegende Theorie zu geben, um zu erklären, woher die Einflussfaktoren stammen. Statt das Problem zu lösen, indem man sich dem Forschungsfeld mit offenen Methoden nähert, halten die Autoren aber weiter an der standardisierten Erhebung vorab deduktiv festgelegter isolierter Faktoren fest. Die Autoren der Studien blicken innerhalb dieser Rahmung von peripheren versus zentralen Verarbeitungsmechanismen scheinbar recht unsystematisch in verschiedene Richtungen, um zu verstehen, was die Wahrnehmung von Glaubwürdigkeit denn nun beeinflussen könnte. Jeder Autor zieht nach Bedarf einen anderen Ansatz heran und testet dann die daraus abgeleiteten Hypothesen (z. B. ob der „Negativity Bias“ vielleicht auch bei der Rezeption von Online-Bewertungen eine Rolle spielt). So wandert die Forschung

recht beliebig von einem Ansatz zum nächsten. Bisher gibt es noch kein überzeugendes theoretisches Modell mit zufriedenstellender Vorhersagekraft. Daher bot sich aus theoretischer Perspektive ein exploratives Vorgehen an.

Bei der Untersuchung war es auch wichtig, dass der Leser im Mittelpunkt der Untersuchung steht. In den bisherigen Studien geht es immer um die Wahrnehmung der Vertrauenswürdigkeit des Bewerbers durch den Leser. Keiner der Autoren meint, dass Vertrauenswürdigkeit eine objektive Eigenschaft des Bewerbers ist, sondern alle gehen davon aus, dass der Rezipient selbst die Vertrauenswürdigkeit zuschreibt. Diese Perspektive ist zu würdigen, da sie nicht an den theoretischen unhaltbaren „beobachterunabhängigen Wahrheits[...]begriff“ (Kohring, 2004, S. 25) gebunden ist. Vertrauen wird immer aus der Sicht des Rezipienten, des Lesers, verstanden. Diese Vorannahme und theoretische Perspektive wird aber bisher nicht in das empirische Vorgehen übersetzt. Sobald es an die Erhebung geht, verliert der Rezipient in den Studien seine tragende Rolle. Der Fokus liegt auf dem Einfluss eines Faktors auf eine Masse an Rezipienten. Konsequenz wäre dagegen, den Fokus auf dem Rezipienten und seiner Wahrnehmung zu belassen. Das würde bedeuten, dass Unterschiede beim Rezipienten miteinbezogen werden. Flanagin und Metzger (2013) meinen dazu:

„people’s knowledge or experience may influence the extent to which they understand a particular signal or warrant and may therefore impact their credibility judgments. More specifically, the ability to understand the warranting value of a signal (e.g., the extent to which a piece of information might be successfully manipulated) may be influenced by the knowledge a person has about how that information is derived, suggesting that warrants or signals may not be interpreted uniformly by all information receivers. [...] Accordingly, signaling theory argues that signals can be ambiguous, and their interpretation may vary between individuals depending on their understanding of the factors guaranteeing a signal’s honesty (Donath, 2011). Thus, different people may comprehend signals differently, based on their experiences, background, or group memberships. Indeed, in some cases, only those with particular experiences or backgrounds (“the initiated”) can understand some signals accurately, or may understand them in ways that differ from people with other backgrounds or experiences (Donath, 2011).“  
(S. 1629)

Auch die vereinzelt Erhebungen, in denen verschiedene Eigenschaften des Lesers als Moderatoren für den Zusammenhang von Vertrauen mit einem anderen Faktor modelliert wurden, zeigen, dass die Unterschiede zwischen den Lesern sich auf den Zusammenhang des anderen untersuchten Faktors mit Vertrauen auswirken. Diese Unterschiede auf der Seite des Lesers können allerdings nur sichtbar werden, wenn

---

die Perspektive auf dem Leser bleibt. Daher war es sinnvoll, die Teilnehmer der Studie eingehend zu ihrer Wahrnehmung zu interviewen.

In der Studie konnte sich so die heterogene Rezeption der Teilnehmer zeigen. Die Teilnehmer nahmen die Arztbewertungen oft unterschiedlich, manchmal sogar gegensätzlich wahr. Selbst dann, wenn die Wahrnehmung der Teilnehmer sehr ähnlich war, konnte die Beurteilung des Wahrgenommenen durch die Teilnehmer wiederum sehr unterschiedlich ausfallen. Diese großen Unterschiede zwischen den Teilnehmern wurden bei fast allen Faktoren deutlich. Außerdem konnten mithilfe der empirischen Exploration die bisher untersuchten Faktoren weiter ausdifferenziert werden. Es wurde deutlich, welche verschiedenen Aspekte einzelne Faktoren umfassen. Überdies konnten noch unbekannte Faktoren ermittelt werden, die mit dem Vertrauen in Arztbewertungen zusammenhängen. Dazu wurden einige bisher nicht erforschte Interaktionen aufgedeckt. Diese vielfältigen empirischen Ergebnisse vertiefen die Erkenntnisse der bisherigen Forschung.



## 5 SCHLUSS

In dieser Arbeit wurde das Vertrauen in Online-Bewertungen erforscht. Zunächst wurde die gesamte bisherige Forschung zu diesem Thema mithilfe einer integrativen Literaturübersicht systematisch erschlossen. Nach einem Überblick über die Geschichte der Forschung wurde die gegenwärtige Forschungslandschaft genauer beschrieben. Darauf folgte die Darstellung der bisher verwendeten Theorie und Methoden. Vor diesem Hintergrund wurden alle bisher untersuchten Faktoren, die mit Vertrauen in Online-Bewertungen zusammenhängen sortiert und systematisiert. Mithilfe dieser Integration wurden theoretische Leerstellen sowie empirische Widersprüche der breit gefächerten Forschung sichtbar.

Um die theoretischen Lücken zu schließen wurde anschließend ein theoretisches Modell zu Vertrauen in Online-Bewertungen entwickelt. Hier wurde der Begriff des Vertrauens genau definiert sowie verschiedene Elemente des Vertrauens in Online-Bewertungen identifiziert. Anschließend wurden die bisher untersuchten Faktoren des Vertrauens mit dem theoretischen Modell verbunden. So wurde deutlich, welchen Aspekt des Vertrauens die Faktoren jeweils gemessen haben und wie die bisherigen Ergebnisse zu Vertrauen in Online-Bewertungen demnach zu interpretieren sind. Das legte den Grund dafür offen, dass bei manchen Faktoren über bisherige Studien hinweg Einigkeit herrscht, andere Ergebnisse dagegen miteinander unvereinbar waren. Die theoretische Bestimmung der Stellung der Faktoren im Vertrauensprozess klärte diese Beobachtung auf. Bei den Faktoren, die Dimensionen für Vertrauen darstellen, herrscht über alle Studien hinweg ein großer Zusammenhang mit Vertrauen. Das überrascht nun nicht mehr, da mit diesen Faktoren das Vertrauen selbst gemessen wird. Im Hinblick auf andere Faktoren gab es allerdings weiterhin Widersprüche.

In der empirischen Exploration sollten diese weiter ergründet werden. Am Beispiel von Arztbewertungen wurde das Wissen über die Wahrnehmung von Online-Bewertungen durch die Leser vertieft. Mithilfe einer offenen Methode konnten die Ergebnisse der bisherigen Forschung weiter differenziert und ergänzt werden. So stellte sich heraus, dass die Teilnehmer bei derselben Bewertung verschiedene Gründe heranziehen, um ihr Vertrauen zu bewerten. Das erklärt die

widersprüchlichen Ergebnisse. Hier wurde allerdings auch deutlich, dass sich die Teilnehmer zwar einig sind im Hinblick auf die abstrakten Dimensionen des Vertrauens, aber die Erfüllung dieser Vertrauenserwartung ebenfalls bei unterschiedlichen Bewertungen und aufgrund unterschiedlicher Kriterien bestätigt oder enttäuscht sehen. Deutlich wurde auch, dass viele Faktoren aus der empirischen Exploration in den bisherigen Studien schon identifiziert wurden. Einige weitere Faktoren konnten dennoch ergänzt werden. Außerdem wurden bisher nicht erforschte Interaktionen zwischen den Faktoren deutlich.

Diese Ergebnisse können in der zukünftigen Forschung weiter untersucht werden. Die neu entdeckten Faktoren sowie die hier beobachteten Interaktionen sollten in folgende Forschung integriert werden. Zukünftige Forschung könnte dann auch untersuchen, wie sich die deskriptiven Ergebnisse dieser Studie zum Vertrauen in Online-Bewertungen erklären lassen. Erklärende Faktoren könnten Eigenschaften des Lesers sein, wie z. B. Alter (Bol et al., 2016), Geschlecht, Erfahrung mit dem WWW und Online-Bewertungen, Bildung oder seine generelle Vertrauenseinstellung. Beldad et al. (2010) formulieren mögliche Unterschiede der Rezipienten, die sich auf ihr Vertrauen auswirken:

„A number of factors still merit attention. Questions such as such ‘how do cultural differences influence online trust’ and ‘does person’s economic standing impact his disposition to trust within the context of online commercial exchanges’ can be logical catalysts for further studies on trust in different online transactions.“ (S. 868)

Eine weitere Erklärung könnten Eigenschaften der Bewertung sein: Man könnte untersuchen, wie diese subjektive Zuschreibung durch die Leser mit objektiv bestimmbaren Eigenschaften der Bewertung zusammenhängt (z. B. Länge, Vokabular, Grammatik, Rechtschreibung). Nur dann lassen sich nämlich Aussagen folgender Art treffen: Die Bewertung hat bestimmte objektive Eigenschaften, die der Leser auf bestimmte Weise wahrnimmt bzw. bewertet und damit sein Vertrauen begründet (Bsp.: „Wenn eine Bewertung unter 50 Wörtern hat, wird sie meist als zu kurz und damit als nicht vertrauenswürdig wahrgenommen.“). Würde man die Bewertungen selbst nicht untersuchen, könnte man nur sagen, dass die Interpretation der Bewertung durch den Lesers auf eine bestimmte Weise zu Vertrauen in diese Bewertung führt (z. B. „Wenn ein Leser eine Bewertung als zu

---

positiv wahrnimmt, hält er sie für nicht vertrauenswürdig.“). Diese empirischen Erkenntnisse könnten wiederum verwendet werden, um das theoretische Modell zu Vertrauen in Online-Bewertung weiter zu entwickeln.



## 6 LITERATURVERZEICHNIS

- Afifi, W. A. & Burgoon, J. K. (2000). The impact of violations on uncertainty and the consequences for attractiveness. *Human Communication Research*, 26 (2), 203-233.
- Andrews, J. C., Netemeyer, R. G. & Burton, S. (1998). Consumer generalization of nutrient content claims in advertising. *Journal of Marketing*, 62 (4), 62.
- Ayeh, J. K., Au, N. & Law, R. (2013). "Do We Believe in TripAdvisor?". Examining credibility perceptions and online travelers' attitude toward using user-generated content. *Journal of Travel Research*, 52 (4), 437-452.
- Bacon, N. (2009). Will doctor rating sites improve standards of care? Yes. *British Medical Journal*, 338, 688-689.
- Bae, S. & Lee, T. (2011). Product type and consumers' perception of online consumer reviews. *Electronic Markets*, 21 (4), 255-266.
- Baechler, C. (2017). *trés pro et sympa... un médecin que je recommande. Textlinguistische und handlungstheoretische Untersuchung deutscher und französischer Online-Arztbewertungen*. Dissertationsschrift, Universität Mannheim.
- Baker, R. (1990). Development of a questionnaire to assess patients' satisfaction with consultations in general practice. *British Journal of General Practice*, 40 (341), 487-490.
- Bates, M. J. & Maack, M. N. (Hrsg.). (2010). *Encyclopedia of library and information sciences* (3. Aufl.). Boca Raton, Florida: Taylor & Francis. Verfügbar unter <http://www.tandfonline.com/doi/book/10.1081/E-ELIS3>
- Beck, A. J. Nutzung und Bewertung deutscher Arztbewertungsportale durch Patienten in deutschen Hausarztpraxe.
- Beldad, A., Jong, M. de & Steehouder, M. (2010). How shall I trust the faceless and the intangible? A literature review on the antecedents of online trust. *Computers in Human Behavior*, 26 (5), 857-869.
- Berger, C. R. & Calabrese, R. J. (1975). Some explorations in initial interaction and beyond. Toward a developmental theory of interpersonal communication. *Human Communication Research*, 1 (2), 99-112.
- Berlo, D. K., Lemert, J. B. & and Mertz, R. J. (1969). Dimensions for evaluating the acceptability of message sources. *Public Opinion Quarterly*, 33 (4), 563-576.
- Bettman, J. R. & Park, C. W. (1980). Effect of prior knowledge and experience and phase of the choice process on consumer decision process. *Journal of Consumer Research*, 7, 234-248.
- Bidmon, S., Terlutter, R. & Röttl, J. (2014). What explains usage of mobile physician-rating apps? Results from a web-based questionnaire. *Journal of medical Internet research*, 16 (6), e148.
- Blanchard, A. L. (2007). Developing a sense of virtual community measure. *CyberPsychology & Behavior*, 10 (6), 827-830.
- Blawat, K. (2012, 15. Mai). "Skeptisch bin ich bei allen". *Mediziner im Internet. Süddeutsche.de*. Zugriff am 27.03.2017. Verfügbar unter <http://www.sueddeutsche.de/gesundheit/mediziner-im-internet-skeptisch-bin-ich-bei-allen-1.1348848>

- Bleckmann, A., Kolada, C., Loskill, H., Nachtigäller, C., Englert, G., Heß, K. et al. (2008, 12. März). *Woran erkennt man eine gute Arztpraxis? Checkliste für Patientinnen und Patienten* (3. Aufl.) (Ärztliches Zentrum für Qualität in der Medizin, Hrsg.) (äzq Schriftenreihe Nr. 34), Neukirchen. Zugriff am 14.10.2011. Verfügbar unter [http://www.bundesaerztekammer.de/downloads/Checkliste\\_Arztpraxis\\_3\\_Auflage.pdf](http://www.bundesaerztekammer.de/downloads/Checkliste_Arztpraxis_3_Auflage.pdf)
- Block, L. & Keller, P. (1995). When to accentuate the negative: The effects of perceived efficacy and message framing on intentions to perform a healthrelated behavior. *Journal of Marketing Research*, 32 (2), 192-203.
- Bol, N., van Weert, J. C. M., Loos, E. F., Romano Bergstrom, J. C., Bolle, S. & Smets, E. M. A. (2016). How are online health messages processed? Using eye tracking to predict recall of information in younger and older adults. *Journal of Health Communication*, 21 (4), 387-396.
- Bolle, S., Romijn, G., Smets, E. M. A., Loos, E. F., Kunneman, M. & van Weert, J. C. M. (2016). Older cancer patients' user experiences with web-based health information tools. A think-aloud study. *Journal of Medical Internet Research*, 18 (7), e208.
- Borchers, N. S. (2016). *Grundlagen einer Makrotheorie der Online-Bewertungen*. Conference of the Digital Communication Division in the DGpuK, Braunschweig.
- Borchers, N. S. & Krömer, N. (2015). Die Rolle von Arztbewertungsportalen im Prozess der Arztwahl – Patienten-Empowerment durch Laienbewertungen? In M. Schäfer, O. Quiring, C. Rossmann, M. R. Hastall & E. Baumann (Hrsg.), *Gesundheitskommunikation im gesellschaftlichen Wandel* (Medien + Gesundheit, 1. Aufl., S. 153-164). Baden-Baden: Nomos.
- Bundeszentrale für politische Bildung (Hrsg.). (2018). *Rap* (Aus Politik und Zeitgeschichte, Bd. 9). Bonn.
- Burgoon, M. , & Miller, G. R. (1985). An expectancy interpretation of language and persuasion. In H. Giles & R. St. Clair (Hrsg.), *The social and psychological contexts of language* (199-229). London: Lawrence Erlbaum.
- Casaló, L. V., Flavián, C., Guinalíu, M. & Ekinci, Y. (2015). Do online hotel rating schemes influence booking behaviors? *International Journal of Hospitality Management*, 49, 28-36.
- Chang, C. (2009). Effectiveness of promotional premiums. The moderating role of affective state in different contexts. *Psychology & Marketing*, 26 (2), 175-194. Verfügbar unter <https://onlinelibrary.wiley.com/doi/full/10.1002/mar.20266>
- Cheng, Y.-H. & Ho, H.-Y. (2015). Social influence's impact on reader perceptions of online reviews. *Journal of Business Research*, 68 (4), 883-887.
- Cheung, C. M. K. & Thadani, D. R. (2012). The impact of electronic word-of-mouth communication. A literature analysis and integrative model. *Decision Support Systems*, 54 (1), 461-470.
- Cheung, C. M. K., Lee, M. K.O. & Rabjohn, N. (2008). The impact of electronic word-of-mouth. The adoption of online opinions in online customer communities. *Internet Research*, 18 (3), 229-247.
- Cheung, C. M.-Y., Sia, C. L. & Kuan, K. K. Y. (2012). Is this review believable? A study of factors affecting the credibility of online consumer reviews from an ELM

- perspective. *Journal of the Association for Information Systems*, 13 (8), 618-635.
- Cheung, M. Y., Luo, C., Sia, C. L. & Chen, H. (2009). Credibility of electronic word-of-mouth. Informational and normative determinants of on-line consumer recommendations. *International Journal of Electronic Commerce*, 13 (4), 9-38.
- Chih, W.-H., Wang, K.-Y., Hsu, L.-C. & Huang, S.-C. (2013). Investigating electronic word-of-mouth effects on online discussion forums. The role of perceived positive electronic word-of-mouth review credibility. *Cyberpsychology, Behaviour, and Social Networking*, 16 (9), 658-668.
- Choi, W. & Stvilia, B. (2015). Web credibility assessment. Conceptualization, operationalization, variability, and models. *Journal of the Association for Information Science and Technology*, 66 (12), 2399-2414.
- Coombes, R. (2009). Plans for patient feedback criticised. *British Medical Journal*, 338 (17), 679.
- Cox, D., Cox, J. G. & Cox, A. D. (2017). To Err is human? How typographical and orthographical errors affect perceptions of online reviewers. *Computers in Human Behavior*, 75, 245-253.
- Damman, O. C., Hengel, Y. K. A. van den, Loon, A. J. M. van & Rademakers, J. (2010). An international comparison of web-based reporting about health care quality. Content analysis. *Journal of Medical Internet Research*, 12 (2).
- Däumler, M. & Noah, E. M. (2015). Public Relations für Ärzte – Tops und Flops. *Journal für Ästhetische Chirurgie*, 8 (1), 7-11.
- Deutsch, M. & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, 51 (3), 629-636.
- Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin (Hrsg.). (2011). *Positionspapier der Deutschen Gesellschaft für Allgemeinmedizin und Familienmedizin (DEGAM) zu Arztbewertungsportalen*. Zugriff am 15.10.2011. Verfügbar unter [http://www.degam.de/fileadmin/user\\_upload/degam/Positionen/110921\\_Stellungnahme\\_Arztbewertungsportale\\_Sektion\\_QF20110913\\_redigiert.pdf](http://www.degam.de/fileadmin/user_upload/degam/Positionen/110921_Stellungnahme_Arztbewertungsportale_Sektion_QF20110913_redigiert.pdf)
- Dickinger, A. (2011). The trustworthiness of online channels for experience- and goal-directed search tasks. *Journal of Travel Research*, 50 (4), 378-391.
- Dietrich, M. (2018). Rap als Forschungsgegenstand. In Bundeszentrale für politische Bildung (Hrsg.), *Rap* (Aus Politik und Zeitgeschichte, Bd. 9, S. 4-10). Bonn.
- Doh, S.-J. & Hwang, J.-S. (2009). How consumers evaluate eWOM (electronic word-of-mouth) messages. *Cyberpsychology & Behavior*, 12 (2), 193-197.
- Donath, J. (2007). Signals in social supernets. *Journal of Computer-Mediated Communication*, 13 (1), 231-251.
- Dou, X., Walden, J. A., Lee, S. & Lee, J. Y. (2012). Does source matter? Examining source effects in online product reviews. *Computers in Human Behavior*, 28 (5), 1555-1563.
- Duncker, K. (1945). On problem-solving. *Psychological Monographs*, 58 (5), i-113.
- Emmert, M. & Meier, F. (2013). An analysis of online evaluations on a physician rating website. Evidence from a German public reporting instrument. *Journal of Medical Internet Research*, 15 (8), e157.
- Emmert, M., Meier, F., Heider, A.-K., Dürr, C. & Sander, U. (2014). What do patients say about their physicians? an analysis of 3000 narrative comments posted on

- a German physician rating website. *Health policy (Amsterdam, Netherlands)*, 118 (1), 66-73.
- Emmert, M., Meier, F., Pisch, F. & Sander, U. (2013). Physician choice making and characteristics associated with using physician-rating websites. Cross-sectional study. *Journal of Medical Internet Research*, 15 (8), e187.
- Emmert, M., Sander, U. & Pisch, F. (2013). Eight questions about physician-rating websites: a systematic review. *Journal of Medical Internet Research*, 15 (2), e24.
- Emmert, M., Sander, U., Esslinger, A. S., Maryschok, M. & Schöffski, O. (2012). Public reporting in Germany: the content of physician rating websites. *Methods of information in medicine*, 51 (2), 112-120.
- Emmert, M., Sander, U., Esslinger, S., Maryschok, M. & Schöffski, O. (2011). Informationen über die Qualität von Ärzten auf deutschen Arztbewertungsportalen. In G. Schreier, D. Hayn & E. Ammenwerth (Hrsg.), *Tagungsband der eHealth2011. Health Informatics meets eHealth - Grenzen überwinden - Continuity of Care* (S. 229-234). OCG Books. Zugriff am 15.10.2011.
- Ericsson, K. A. & Simon, H. A. (1984). *Protocol analysis. Verbal reports as data*. Cambridge, Massachusetts: MIT Press.
- Fan, Y. & Miao, Y. (2012). Effect of electronic word-of-mouth on consumer purchase intention. The perspective of gender differences. *International Journal of Electronic Business Management*, 10 (3), 175-181.
- Ferguson, L. E., Bråten, I. & Strømsø, H. I. (2012). Epistemic cognition when students read multiple documents containing conflicting scientific evidence. A think-aloud study. *Learning and Instruction*, 12, 103-120.
- Filieri, R. (2015). What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in e-WOM. *Journal of Business Research*, 68 (6), 1261-1270.
- Filieri, R. (2016). What makes an online consumer review trustworthy? *Annals of Tourism Research*, 58, 46-64.
- Filieri, R., Alguezaui, S. & McLeay, F. (2015). Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. *Tourism Management*, 51, 174-185.
- Fisher, C. D., Ilgen, D. R. & Hoyer, W. D. (1979). Source credibility, information favorability, and job offer acceptance. *Academy of Management Journal*, 22 (1), 94-103.
- Flanagin, A. J. & Metzger, M. J. (2000). Perceptions of internet information credibility. *Journalism & Mass Communication Quarterly*, 77 (3), 515-540.
- Flanagin, A. J. & Metzger, M. J. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media & Society*, 9 (2), 319-342.
- Flanagin, A. J. & Metzger, M. J. (2008). The credibility of volunteered geographic information. *GeoJournal*, 72, 137-148.
- Flanagin, A. J. & Metzger, M. J. (2013). Trusting expert- versus user-generated ratings online. The role of information volume, valence, and consumer characteristics. *Computers in Human Behavior*, 29 (4), 1626-1634.

- Flanagin, A. J. & Metzger, M. J. (2013). Trusting expert- versus user-generated ratings online. The role of information volume, valence, and consumer characteristics. *Computers in Human Behavior*, 29 (4), 1626-1634.
- Friedl, A. & Wietlisbach, I. (2008). Ärztebewertung im Internet. Megatrend oder Gigaflop? *Care Management*, 1 (5), 29-33.
- Friedlandt, J. (2013). *Glaubwürdigkeit und Nützlichkeit kundengenerierter Produktbewertungen im Internet* (Studien zum Konsumentenverhalten, Bd. 48, 1., Aufl). Hamburg: Dr. Kovac Verlag.
- Froschauer, U. & Lueger, M. (2003). *Das qualitative Interview. Zur Praxis interpretativer Analyse sozialer Systeme* (UTB Soziologie, Bd. 2418, 1. Aufl.). Wien: WUV. Verfügbar unter <http://www.utb-studi-e-book.de/9783838524184>
- Furner, C. P., Racherla, P. & Zhu, Z. (2012). Uncertainty, trust and purchase intention based on online product reviews. An introduction to a multinational study. *International Journal of Networking and Virtual Organisations*, 11 (3/4), 260.
- Furner, C. P., Zinko, R. & Zhu, Z. (2016). Electronic word-of-mouth and information overload in an experiential service industry. *Journal of Service Theory and Practice*, 26 (6), 788-810.
- Galizzi, M. M., Miraldo, M., Stavropoulou, C., Desai, M., Jayatunga, W., Joshi, M. et al. (2012). Who is more likely to use doctor-rating websites, and why? A cross-sectional study in London. *BMJ open*, 2 (6).
- Gao, G. G., McCullough, J. S., Agarwal, R. & Jha, A. K. (2012). A changing landscape of physician quality reporting. Analysis of patients' online ratings of their physicians over a 5-year period. *Journal of Medical Internet Research*, 14 (1), e38.
- Gefen, D. & Straub, D. W. (2004). Consumer trust in B2C e-commerce and the importance of social presence. Experiments in e-products and e-services. *Omega*, 32 (6), 407-424.
- Gefen, D. (2000). E-commerce. The role of familiarity and trust. *Omega*, 28 (6), 725-737.
- Geierhos, M. & Bäumer, F. S. (2015a). Erfahrungsberichte aus zweiter Hand. Erkenntnisse über die Autorschaft von Arztbewertungen in Online-Portalen. Graz: DhD Tagung.
- Geierhos, M., Bäumer, F., Schulze, S. & Stuß, V. (2015b). "I grade what I get but write what I think.". Inconsistency analysis in patients' reviews. ECIS 2015 Completed Research Papers (Paper 55).
- Geierhos, M., Schulze, S. & Bäumer, F. S. S. (2015c). Analyse anonymer Arztbewertungen im Web 2.0 (Verbraucherzentrale NRW/Kompetenzzentrum Verbraucherforschung NRW, Hrsg.) (Working Paper und Fact Sheet, Nr. 3).
- Gershoff, A. D., Mukherjee, A. & Mukhopadhyay, A. (2003). Consumer acceptance of online agent advice. Extremity and positivity effects. *Journal of Consumer Psychology*, 13, 161-170.
- Giles, H. & St. Clair, R. (Hrsg.). (1985). *The social and psychological contexts of language*. London: Lawrence Erlbaum.
- Gilly, M. C., Graham, J. L., Wolfinger, M. F. & Yale, L. J. (1998). A dyadic study of interpersonal information search. *Journal of the Academy of Marketing Science*, 26 (2), 83-100.

- Grabner-Kräuter, S. & Waiguny, M. K. J. (2015). Insights into the impact of online physician reviews on patients' decision making: randomized experiment. *Journal of Medical Internet research*, 17 (4).
- Greaves, F., Pape, U. J., King, D., Darzi, A., Majeed, A., Wachter, R. M. et al. (2012). Associations between Internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. *BMJ quality & safety*, 21 (7), 600-605.
- Greaves, F., Pape, U. J., Lee, H., Smith, D. M., Darzi, A., Majeed, A. et al. (2012). Patients' ratings of family physician practices on the internet: usage and associations with conventional measures of quality in the English National Health Service. *Journal of Medical Internet Research*, 14 (5), e146.
- Grün, G.-C. (2011, 26. Juli). Medizinportale taugen wenig. *Zeit Online*. Verfügbar unter <http://www.zeit.de/digital/internet/2011-07/arztbewertungsportale-test>
- Halling, F., Halling, C., Pohl-Dernick, K. & Emmert, M. (2012). Arztbewertungsportale aus zahnärztlicher Perspektive. Physician rating websites from the perspective of dentists. *Deutsche Zahnärztliche Zeitschrift*, 76 (11), 716-723.
- Hellmüller, L. & Trilling, D. (2012, Juni) The credibility of credibility measures. A meta-analysis in leading communication journals, 1951 to 2011. Vortrag auf der 65. WAPOR Jahreskonferenz in Hong Kong.
- Hennig-Thurau, T., Gwinner, K. P., Walsh, G. & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms. What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing*, 18 (1), 38-52.
- Hennig-Thurau, T., Hansen, U., Eifler, V. & Bornemann, D. (2002). Vertrauen in Kundenartikulationen auf virtuellen Meinungsplattformen. In M. Bruhn & B. Stauss (Hrsg.), *Electronic Services. Dienstleistungsmanagement Jahrbuch 2002* (S. 461-487). Wiesbaden: Gabler.
- Higgins, E. T. (1997). Beyond pleasure and pain. *The American psychologist*, 52 (12), 1280-1300.
- Higgins, E. T. (2000). Making a good decision. Value from fit. *American Psychologist*, 55 (11), 1217-1230.
- Hilligoss, B. & Rieh, S. Y. (2008). Developing a unifying framework of credibility assessment. Construct, heuristics, and interaction in context. *Information Processing & Management*, 44 (4), 1467-1484.
- Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage.
- Hong, S. & Park, H. S. (2012). Computer-mediated persuasion in online reviews. Statistical versus narrative evidence. *Computers in Human Behavior*, 28 (3), 906-919.
- Hovland, C. I. & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, 15 (4), 635.
- Hovland, C. I., Janis, I. L. & Kelley, H. H. (1953). *Communication and persuasion. Psychological studies of opinion change*. New Haven: Yale University Press.
- Hsiao, K.-L., Lin, J. C.-C., Wang, X.-Y., Lu, H.-P. & Yu, H. (2010). Antecedents and consequences of trust in online product recommendations. An empirical study in social shopping. *Online Information Review*, 34 (6), 935-953.

- Hsu, L.-C., Chih, W.-H. & Liou, D.-K. (2016). Investigating community members' eWOM effects in Facebook fan page. *Industrial Management & Data Systems*, 116 (5), 978-1004.
- Huang, J.-H. & Chen, Y.-F. (2006). Herding in online product choice. *Psychology & Marketing*, 23 (5), 413-428.
- Huber, F., Krönung, S., Meyer, F. & Vollmann, S. (2011). *Akzeptanz von Bewertungsportalen als Basis von Electronic Word-of-Mouth. Eine empirische Studie zur interpersonellen Kommunikation im Web 2.0* (Reihe, Bd. 54, 1. Aufl.). Lohmar, Rheinl: Eul.
- Jain, S. (2010). Googling ourselves. What physicians can learn from online rating sites. *The New England Journal of Medicine*, 362 (1), 6-7.
- Jameda reagiert auf BGH-Urteil. *Süddeutsche.de*. Zugriff am 21.02.2018. Verfügbar unter <http://www.sueddeutsche.de/wirtschaft/jameda-bundesgerichtshof-urteil-1.3874842>
- Jensen, M. L. & Yetgin, E. (2017). Prominence and interpretation of online conflict of interest disclosures. *Management Information Systems Quarterly*, 41 (2), 629-A13.
- Jensen, M. L., Averbek, J. M., Zhang, Z. & Wright, K. B. (2013). Credibility of anonymous online product reviews. A language expectancy perspective. *Journal of Management Information Systems*, 30 (1), 293-323.
- Johnson, D. & Grayson, K. (2005). Cognitive and affective trust in service relationships. *Journal of Business Research*, 58 (4), 500-507.
- Johnson, T. J. & Kaye, B. K. (2004). Wag the Blog. How reliance on traditional media and the internet influence credibility perceptions of weblogs among blog users. *Journalism & Mass Communication Quarterly*, 81 (3), 622-642.
- Johnson, T. J. & Kaye, B. K. (2009). In blog we trust? Deciphering credibility of components of the internet among politically interested internet users. *Computers in Human Behavior*, 25 (1), 175-182.
- Johnson, T. J., Kaye, B. K., Bichard, S. L. & Wong, W. J. (2007). Every blog has its day. Politically-interested internet users' perceptions of blog credibility. *Journal of Computer-Mediated Communication*, 13 (1), 100-122.
- Kadry, B., Chu, L. F., Kadry, B., Gammas, D. & Macario, A. (2011). Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. *Journal of Medical Internet Research*, 13 (4), e95.
- Kelley, H. H. (1967). Attribution theory in social psychology. *Nebraska Symposium on Motivation*, 15, 192-238.
- Kelman, H. C. (1961). Processes of opinion change. *Public Opinion Quarterly*, 25, 57-78.
- Kesmodel, D. (2005, 14. September). As angry patients vent online, doctors sue to silence them. *Wall Street Journal Online*. Verfügbar unter [http://online.wsj.com/public/article/0,,SB112568868274630508-LCeMjwsVCTdLbpUXciwtBVJsek\\_20060914,00.html?mod=blogs#printMode](http://online.wsj.com/public/article/0,,SB112568868274630508-LCeMjwsVCTdLbpUXciwtBVJsek_20060914,00.html?mod=blogs#printMode)
- Kiecker, P. & Cowles, D. (2002). Interpersonal Communication and Personal Influence on the Internet. A Framework for Examining Online Word-of-Mouth. *Journal of Euromarketing*, 11 (2), 71-88.
- King, R. A., Racherla, P. & Bush, V. D. (2014). What we know and don't know about online word-of-mouth. A review and synthesis of the literature. *Journal of Interactive Marketing*, 28 (3), 167-183.

- Klimmt, C. (2011). *Das Elaboration-Likelihood-Modell* (Konzepte, Bd. 5, 1. Aufl.). Baden-Baden: Nomos. Verfügbar unter <http://dx.doi.org/10.5771/9783845260327>
- Knapp, M. L. & Daly, J. A. (Hrsg.). (2002). *Handbook of interpersonal communication* (3. Aufl.): Sage.
- Kofahl, C. & Horak, I. (2010). Arztbewertungsportale. Neue Wege im Gesundheitswesen bei Information, Bewertung und Suche im Internet. In C. Koch (Hrsg.), *Achtung: Patient online! Wie Internet, soziale Netzwerke und kommunikativer Strukturwandel den Gesundheitssektor transformieren* (S. 105-126). Wiesbaden: Gabler.
- Kohring, M. (2004). *Vertrauen in Journalismus. Theorie und Empirie* (Kommunikationswissenschaft). Konstanz: UVK Verlagsgesellschaft.
- Kolata, G. (2005, 30. November). When the doctor is in, but you wish he weren't. *New York Times*. Verfügbar unter <http://www.nytimes.com/2005/11/30/health/30patient.html?pagewanted=print>
- Kühl, E. (2018). 10/10, hier gerne wieder krank. *Zeit Online*. Zugriff am 21.02.2018. Verfügbar unter <http://www.zeit.de/digital/internet/2018-02/jameda-bewertungsportal-aerzte-empfehlungen-urteil/komplettansicht>
- Kusumasondjaja, S., Shanka, T. & Marchegiani, C. (2012). Credibility of online reviews and initial trust. The roles of reviewer's identity and review valence. *Journal of Vacation Marketing*, 18 (3), 185-195.
- Lagu, T., Hannon, N. S. & Rothberg, M. B. (2010). Patients' evaluations of health care providers in the era of social networking. An analysis of physician-rating websites. *Journal of General Internal Medicine*, 25 (9), 942-946.
- Lee, J., Park, D.-H. & Han, I. (2008). The effect of negative online consumer reviews on product attitude. An information processing view. *Electronic Commerce Research and Applications*, 7 (3), 341-352.
- Lee, J., Park, D.-H. & Han, I. (2011). The different effects of online consumer reviews on consumers' purchase intentions depending on trust in online shopping malls. *Internet Research*, 21 (2), 187-206.
- Lee, K.-T. & Koo, D.-M. (2012). Effects of attribute and valence of e-WOM on message adoption. Moderating roles of subjective knowledge and regulatory focus. *Computers in Human Behavior*, 28 (5), 1974-1984.
- Lee, Y.-C. (2014). Impacts of decision-making biases on eWOM retrust and risk-reducing strategies. *Computers in Human Behavior*, 40, 101-110.
- Leserdiskussion: Vertrauen Sie Bewertungsportalen? (2018, 20. Februar). *Süddeutsche.de*. Zugriff am 27.03.2018. Verfügbar unter <http://www.sueddeutsche.de/gesundheit/leserdiskussion-vertrauen-sie-bewertungsportalen-1.3875024>
- Liao, Q. V. & Fu, W.-T. (2014). Age differences in credibility judgments of online health information. *ACM Transactions on Computer-Human Interaction*, 21 (1), 1-23.
- Lim, Y.-s. & van der Heide, B. (2015). Evaluating the wisdom of strangers. The perceived credibility of online consumer reviews on Yelp. *Journal of Computer-Mediated Communication*, 20 (1), 67-82.

- Lin, C. A. & Xu, X. (2017). Effectiveness of online consumer reviews. The influence of valence, reviewer ethnicity, social distance and source trustworthiness. *Internet Research*, 27 (2), 362-380.
- Lindzey, G. & Aronson, E. (Hrsg.). (1985). *The Handbook of Social Psychology* (2. Aufl.). Hillsdale, New Jersey: Erlbaum.
- Lis, B. (2013). In eWOM we trust a framework of factors that determine the eWOM credibility. *Business & Information Systems Engineering*, 5 (3), 129-140.
- Litman, J. (2005). Curiosity and the pleasures of learning. Wanting and liking new information. *Cognition & Emotion*, 19 (6), 793-814.
- Llamero, L. (2014). Conceptual mindsets and heuristics in credibility evaluation of e-Word of Mouth in tourism. *Online Information Review*, 38 (7), 954-968.
- López, M. & Sicilia, M. (2014). eWOM as source of influence. The impact of participation in eWOM and perceived source trustworthiness on decision making. *Journal of Interactive Advertising*, 14 (2), 86-97.
- Louie, T. A. & Obermiller, C. (2002). Consumer response to a firm's endorser (dis) association decisions. *Journal of Advertising*, 31 (4), 41-52.
- Luhmann, N. (2000 [1989]). *Vertrauen. Ein Mechanismus der Reduktion sozialer Komplexität* (4. Aufl.). Stuttgart: Lucius und Lucius.
- Luo, C., Luo, X., Schatzberg, L. & Sia, C. L. (2013). Impact of informational factors on online recommendation credibility. The moderating role of source credibility. *Decision Support Systems*, 56, 92-102.
- Luo, C., Luo, X., Xu, Y., Warkentin, M. & Sia, C. L. (2015). Examining the moderating role of sense of membership in online review evaluations. *Information & Management*, 52 (3), 305-316.
- Luo, C., Wu, J., Shi, Y. & Xu, Y. (2014). The effects of individualism–collectivism cultural orientation on eWOM information. *International Journal of Information Management*, 34 (4), 446-456.
- MacNeela, P., Clinton, G., Place, C., Scott, A., Treacy, P. & Hyde, A. (2010). Psychosocial care in mental health nursing. A Think-Aloud study. *Journal of Advanced Nursing*, 66 (6), 1297-1307.
- McAllister, D. J. (1995). Affect- and cognition based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, 38 (1), 24-59.
- McCabe, D. P., Geraci, L., Boman, J. K. & Sensenig, A. E. (2011). On the validity of remember-know judgments. Evidence from Think-Aloud protocols. *Consciousness and Cognition*, 20, 1625-1633.
- McCroskey, J. C. & Teven, J. J. (2009). Goodwill. A reexamination of the construct and its measurement. *Communication Monographs*, 66 (1), 90-103.
- McGloin, R., Nowak, K. & Watt, J. (2014). Avatars and expectations. Influencing perceptions of trustworthiness in an online consumer setting. *PsychoNology Journal*, 1, 7-28.
- McGuire, W. J. (1985). Attitudes and attitude change. In G. Lindzey & E. Aronson (Hrsg.), *The Handbook of Social Psychology* (2. Aufl., S. 262-276). Hillsdale, New Jersey: Erlbaum.
- McKnight, D. H. & Chervany, N. L. (2014). What trust means in e-commerce customer relationships. An interdisciplinary conceptual typology. *International Journal of Electronic Commerce*, 6 (2), 35-59.

- McLeod, D. M., Wise, D., & Perryman, M. (2017). Thinking about the media: A review of theory and research on media perceptions, media effects perceptions, and their consequences. *Review of Communication Research*, 5(2), 35–83.
- Mh. (2017, 11. Oktober). Jameda wächst bei Nutzerzahlen. *ÄrzteZeitung.de*. Verfügbar unter [https://www.aerztezeitung.de/praxis\\_wirtschaft/internet\\_co/article/945086/arztbewertung-jameda-waechst-nutzerzahlen.html](https://www.aerztezeitung.de/praxis_wirtschaft/internet_co/article/945086/arztbewertung-jameda-waechst-nutzerzahlen.html)
- Moran, G. & Muzellec, L. (2017). eWOM credibility on social networking sites. A framework. *Journal of Marketing Communications*, 23 (2), 149-161.
- Mudambi, S. M. & Schuff, D. (2010). What makes a helpful review? A study of customer reviews on amazon.com. *MIS Quarterly*, 34 (1), 185-200.
- Munzel, A. (2015). Malicious practice of fake reviews. Experimental insight into the potential of contextual indicators in assisting consumers to detect deceptive opinion spam. *Recherche et Applications en Marketing (English Edition)*, 30 (4), 24-50.
- Nekmat, E. & Gower, K. K. (2012). Effects of disclosure and message valence in online word-of-mouth (eWOM) communication. Implications for integrated marketing communication. *International Journal of Integrated Marketing Communications*, 4 (1), 85-98.
- Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising*, 19 (3), 39-52.
- Ohanian, R. (1991). The impact of celebrity spokespersons' perceived image on consumers' intention to purchase. *Journal of Advertising Research*, 31 (1), 46-54.
- Ong, B. S. (2011). Online shopper reviews. Ramifications for promotion and website utility. *Journal of Promotion Management*, 17 (3), 327-344.
- O'Reilly, K. & Marx, S. (2011). How young, technical consumers assess online WOM credibility. *Qualitative Market Research: An International Journal*, 14 (4), 330-359.
- O'Reilly, K., MacMillan, A., Mumuni, A. G. & Lancendorfer, K. M. (2016). Extending our understanding of eWOM impact. The role of source credibility and message relevance. *Journal of Internet Commerce*, 15 (2), 77-96.
- Pan, L.-Y. & Chiou, J.-S. (2011). How much can you trust online information? Cues for perceived trustworthiness of consumer-generated online information. *Journal of Interactive Marketing*, 25 (2), 67-74.
- Park, D.-H. & Kim, S. (2008). The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews. *Electronic Commerce Research and Applications*, 7 (4), 399-410.
- Petty, R. E. & Cacioppo, J. T. (1986). *Communication and persuasion. Central and peripheral routes to attitude change*. New York: Springer.
- Pottier, P., Hardouin, J.-B., Hodges, B. D., Pistorius, M.-A., Connault, J., Durant, C. et al. (2010). Exploring how students think. A new method combining think-aloud and concept mapping protocols. *Medical Education*, 44, 926-935.
- Qaseem, A., Snow, V., Gosfield, A., Gregg, D., Michl, K., Wennberg, D. et al. (2010). Pay for performance through the lens of medical professionalism. *Annals of Internal Medicine*, 152 (6), 366-369.

- Qiu, L., Pang, J. & Lim, K. H. (2012). Effects of conflicting aggregated rating on eWOM review credibility and diagnosticity. The moderating role of review valence. *Decision Support Systems*, 54 (1), 631-643.
- Racherla, P., Mandviwalla, M. & Connolly, D. J. (2012). Factors affecting consumers' trust in online product reviews. *Journal of Consumer Behavior*, 11 (2), 94-104.
- Raniszewski Herrera, C. (2006). Giving doctors a check-up. *American Legislative Exchange Council Issue Analysis*.
- Reichelt, J., Sievert, J. & Jacob, F. (2014). How credibility affects eWOM reading. The influences of expertise, trustworthiness, and similarity on utilitarian and social functions. *Journal of Marketing Communications*, 20 (1/2), 65-81-81. Verfügbar unter doi:10.1080/13527266.2013.797758
- Reimann, S. & Strech, D. (2010). The representation of patient experience and satisfaction in physician rating sites. A criteria-based analysis of English- and German-language sites. *BioMed Central Health Services Research*, 10, 332. Zugriff am 17.10.2011. Verfügbar unter <http://www.biomedcentral.com/1472-6963/10/332>
- Reimer, T. & Benkenstein, M. (2016). When good WOM hurts and bad WOM gains. The effect of untrustworthy online reviews. *Journal of Business Research*, 69 (12), 5993-6001.
- Reinhardt, A., Weber, W. & Rossmann, C. (2017). Effects of exemplars and base-rate information on online physician rating sites. *Studies in Communication and Media*, 6 (4), 420-443.
- Rieh, S. Y. (2010). Credibility and cognitive authority of information. In M. J. Bates & M. N. Maack (Hrsg.), *Encyclopedia of library and information sciences* (3. Aufl., S. 1337-1344). Boca Raton, Florida: Taylor & Francis.
- Sander, R. (2007, 8. November). Wo es Doktoren weh tun kann. Ärztebewertung im Internet. *Stern.de*. Zugriff am 17.10.2011. Verfügbar unter <http://www.stern.de/digital/online/aerztebewertung-im-internet-wo-es-doktoren-weh-tun-kann-602081.html>
- Sander, U. & Schult, T. J. (2010) Eyetracking-Analyse von Arztsuch- und Arztbewertungsportalen. In *eHealth Niedersachsen. Exzellente Leistungen und innovative Lösungen in Medizin und IT* (S. 46-47). Hannover.Projektbüro eHealth.Niedersachsen.
- Sander, U., Emmert, M. & Grobe, T. G. (2013). Effektivität und Effizienz der Arztsuche mit Arztsuch- und Bewertungsportalen und Google. *Gesundheitswesen (Bundesverband der Ärzte des Öffentlichen Gesundheitsdienstes (Germany))*, 75 (6), 397-399.
- Schaefer, C. & Schwarz, S. (2010). Wer findet die besten Ärzte Deutschlands? Arztbewertungsportale im Internet. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 104 (572-577).
- Schaefer, C., Sängler, S., Dehmlow, R., Balke, K., Berner, B., Butz, N. et al. (2011, 28. Januar). *Bewertungsportale. Qualitätsanforderungen Arztbewertungsportale* (2. Aufl.) (Bundesärztekammer & Kassenärztliche Bundesvereinigung, Hrsg.) (Gute Praxis), Berlin. Zugriff am 17.10.2011.
- Schäfer, M., Quiring, O., Rossmann, C., Hastall, M. R. & Baumann, E. (Hrsg.). (2015). *Gesundheitskommunikation im gesellschaftlichen Wandel* (Medien + Gesundheit, 1. Aufl.). Baden-Baden: Nomos.
- Schreier, M. (2012). *Qualitative content analysis in practice*. London: Sage.

- Schwarz, S. & Schaefer, C. (2015). Arztbewertungsportale. Was können sie leisten? *Public Health Forum*, 23 (3), 685.
- Schwarz, S., Schaefer, C., Sanger, S., Dehmlow, R., Balke, K., Berner, B. et al. (Februar 2012). *Bewertungsportale. Qualitatsanforderungen fur Zahnarzbewertungsportale* (2. Aufl.) (Bundeszahnartekammer, Kassenzahnartzliche Bundesvereinigung, Bundesartekammer, Kassenartzliche Bundesvereinigung & Arztliches Zentrum fur Qualitat in der Medizin, Hrsg.) (Gute Praxis).
- Schwenkenbecher, J. & Hutten, F. (2018, 20. Februar). "Irgendein Patient ist immer unzufrieden - Jameda lebt davon". Drei Arzte uber ihre Erfahrungen. *Suddeutsche.de*. Verfugbar unter <http://www.sueddeutsche.de/gesundheit/jameda-urteil-protokolle-1.3870914>
- Senecal, S. & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, 80 (2), 159-169.
- Seo, D. & Lee, J. (2014). Experts versus Friends: To Whom Do I Listen More? The Factors That Affect Credibility of Online Information. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, A. Kobsa, F. Mattern et al. (Hrsg.), *HCI in Business* (Lecture Notes in Computer Science, Bd. 8527, S. 245-256). Cham: Springer International Publishing.
- Shamhuyenzva, R. M., van Tonder, E., Roberts-Lombard, M. & Hemsworth, D. (2016). Factors influencing Generation Y consumers' perceptions of eWOM credibility. A study of the fast-food industry. *International Review of Retail, Distribution & Consumer Research*, 26 (4), 435-455.
- Smith, D., Menon, S. & Sivakumar, K. (2005). Online peer and editorial recommendations, trust, and choice in virtual markets. *Journal of Interactive Marketing*, 19 (3), 15-37.
- Smith, R. E. & Vogt, C. A. (1995). The effects of integrating advertising and negative word-of-mouth communications on message processing and response. *Journal of Consumer Psychology*, 4 (2), 133-151.
- Someren, M. W. v., Barnard, Y. F. & Sandberg, J. A. C. (1994). *The Think-Aloud method. A practical guide to modelling cognitive processes*. London: Academic Press.
- Stiftung Gesundheit (Hrsg.). (2010). *Clearingverfahren Arztbewertungsportale - "Arzt-Auskunft" Stellungnahme der Stiftung Gesundheit*. Zugriff am 24.10.2011. Verfugbar unter <http://www.aeqz.de/aezq/arzbewertungsportale/bewertete-portale/bewertung/pdf/stellungnahme-arzt-auskunft.pdf>
- Strech, D. & Reimann, S. (2012). Deutschsprachige Arztbewertungsportale. *Gesundheitswesen (Bundesverband der Arzte des Offentlichen Gesundheitsdienstes (Germany))*, 74 (8-9), e61-7.
- Strech, D. (2011). Ethical principles for physician rating sites. *Journal of Medical Internet Research*, 13 (4). Verfugbar unter <http://www.jmir.org/2011/4/e113/>
- Terlutter, R., Bidmon, S. & Rottl, J. (2014). Who uses physician-rating websites? Differences in sociodemographic variables, psychographic variables, and health status of users and nonusers of physician-rating websites. *Journal of*

- Medical Internet Research, 16 (3), e97. Zugriff am 05.09.2014. Verfügbar unter <http://www.jmir.org/2014/3/e97/>
- Thadeusz, F. (2009). Klagen der Zermürbten. *Der Spiegel* (3), 115. Zugriff am 19.10.2011. Verfügbar unter <http://www.spiegel.de/spiegel/print/d-63546838.html>
- Tonteri, L., Kosonen, M., Ellonen, H.-K. & Tarkiainen, A. (2011). Antecedents of an experienced sense of virtual community. *Computers in Human Behavior*, 27 (6), 2215-2223.
- Torraco, R. J. (2005). Writing integrative literature reviews. Guidelines and examples. *Human Resource Development Review*, 4 (2), 356-367.
- Trenz, M. & Berger, B. (2013). Analyzing online customer reviews. An interdisciplinary literature review and research agenda. In *Proceedings of the 21st European Conference on Information Systems. Completed Research*. Verfügbar unter [http://aisel.aisnet.org/ecis2013\\_cr/83](http://aisel.aisnet.org/ecis2013_cr/83)
- Tripp, C., Jensen, T. D. & Carlson, L. (1994). The effects of multiple product endorsements by celebrities on consumers' attitudes and intentions. *Journal of Consumer Research*, 20 (4), 535-547.
- Tsang, A. S.L. & Prendergast, G. (2009). Is a "star" worth a thousand words? *European Journal of Marketing*, 43 (11/12), 1269-1280.
- Tsao, W.-C. & Hsieh, M.-T. (2015). eWOM persuasiveness. Do eWOM platforms and product type matter? *Electronic Commerce Research*, 15 (4), 509-541.
- Van der Heide, B. & Lim, Y.-s. (2016). On the conditional cueing of credibility heuristics. The case of online influence. *Communication Research*, 43 (5), 672-693.
- Vásquez, C. (2014). *The discourse of online consumer reviews* (Bloomsbury discourse series). London: Bloomsbury.
- Walther, J. & R. Parks, M. (2002). Cues filtered out, cues filtered in computer-mediated communication and relationships. In M. L. Knapp & J. A. Daly (Hrsg.), *Handbook of interpersonal communication* (3. Aufl.). Sage.
- Wang, S., Cunningham, N. R. & Eastin, M. S. (2015). The impact of eWOM message characteristics on the perceived effectiveness of online consumer reviews. *Journal of Interactive Advertising*, 15 (2), 151-159.
- Wang, X., Teo, H.-H. & Wei, K. K. (2015). Simultaneity and interactivity of the effects of communication elements on consumers' decision making in ewom systems. *Journal of Electronic Commerce Research*, 16 (3), 153-174.
- Wathen, C. N. & Burkell, J. (2002). Believe it or not. Factors influencing credibility on the Web. *Journal of the American Society for Information Science and Technology*, 53 (2), 134-144.
- Whittemore, R. & Knafl, K. (2005). The integrative review. Updated methodology. *Journal of Advanced Nursing*, 52 (5), 546-553.
- Willemsen, L. M., Neijens, P. C. & Bronner, F. E. (2011). Perceived expertise vs. perceived trustworthiness. The suppressed effect of source type on review attitude. In S. Okazaki (Hrsg.), *Advances in Advertising Research* (S. 423-436). Wiesbaden: Gabler.
- Willemsen, L. M., Neijens, P. C. & Bronner, F. E. (2012). The ironic effect of source identification on the perceived credibility of online product reviewers. *Journal of Computer-Mediated Communication*, 18 (1), 16-31.

- 
- Wilson, P. (1983). *Second-hand knowledge. An inquiry into cognitive authority*. Westport, Connecticut: Greenwood.
- Witzel, A. & Reiter, H. (2012). *The problem-centred interview. Principles and practice*. London: Sage.
- Wu, L., Shen, H., Li, M. & Deng, Q. (2017). Sharing information now vs later. The effect of temporal contiguity cue and power on consumer response toward online reviews. *International Journal of Contemporary Hospitality Management*, 29 (2), 648-668.
- Xie, H., Miao, L., Kuo, P.-J. & Lee, B.-Y. (2011). Consumers' responses to ambivalent online hotel reviews. The role of perceived source credibility and pre-decisional disposition. *International Journal of Hospitality Management*, 30 (1), 178-183.
- Xu, Q. (2014). Should I trust him? The effects of reviewer profile characteristics on eWOM credibility. *Computers in Human Behavior*, 33, 136-144.
- Yoon, S. N. (2008). *The effects of electronic word-of-mouth systems (ewoms) on the acceptance of recommendation*. Dissertation, University of Nebraska. Nebraska.
- Zhang, J. Q., Craciun, G. & Shin, D. (2010). When does electronic word-of-mouth matter? A study of consumer product reviews. *Journal of Business Research*, 63 (12), 1336-1341.
- Zhang, W., and Watts, S. Knowledge adoption in online communities of practice. In *ICIS 2003 Proceedings* (S. 96-109).
- Zhang, Y. (1996). Responses to humorous advertising. The moderating effect of need for cognition. *Journal of Advertising*, 25 (1), 15-32.

## 7 ANHANG

### 7.1 EINFÜHRUNG

Ich würde gerne erst mal einige Worte über mich und über mein Projekt sagen. Mein Name ist Sevda Arslan und ich mache eine Studie zum Thema Arztbewertungen. Die Studie besteht aus zwei Teilen: erst arbeiten Sie alleine am Computer, danach möchte ich Sie interviewen. Die Studie ist streng vertraulich. Nur ich habe Zugang zu den Daten und alles wird sofort anonymisiert. Damit ich mich voll darauf konzentrieren kann, was Sie sagen, würde ich das Gespräch gerne auf Tonband aufnehmen. Später wird meine studentische Mitarbeiterin das transkribieren, also abtippen, was ich aufgenommen habe. In dem Text werden dann alle Namen durch ausgedachte Namen ersetzt. Die Aufnahmen werden danach gelöscht. Haben Sie noch irgendwelche Fragen bevor wir beginnen? Sie können sonst auch jederzeit später fragen.

### 7.2 AUFGABENSTELLUNG FÜR THINK-ALLOUD

#### Studie zu Arztbewertungen

Sehr geehrte Teilnehmerin, sehr geehrter Teilnehmer,

gleich werden Sie eine Aufgabe erhalten, die Sie an diesem Bildschirm bearbeiten sollen.

**Bitte sagen Sie während der gesamte Aufgabe laut, was Ihnen durch den Kopf geht.**

Sie lesen laut mit und sprechen laut aus, was Ihnen dabei durch den Kopf geht - so als würden Sie mit sich selbst sprechen.

Falls Sie zwischendurch vergessen, zu sprechen, werde ich Sie mit dem Satz „Sprechen Sie weiter.“ darauf hinweisen, mit dem Sprechen fortzufahren.

**Bitte lesen Sie ab jetzt laut mit und sagen Sie laut, was Ihnen durch den Kopf geht.**

Hier ist die Aufgabe:

Stellen Sie sich vor, ein guter Freund sucht einen Hausarzt.

Im Internet stoßen Sie auf mehrere Bewertungen zu einem Hausarzt.

Bitte klicken Sie den blauen Link an. Dann öffnet sich ein weiteres Tab mit einer Bewertung zu diesem Arzt. **Bitte lesen Sie sich diese Bewertung durch.** Wenn Sie dieses Tab mit einem Klick auf das kleine Kreuzchen schließen, gelangen Sie wieder zurück zur Umfrage.

**Entscheiden Sie, ob Sie diese Bewertung an Ihren guten Freund weitergeben würden oder nicht.** Wählen Sie dafür mithilfe des Pfeils Ihre Antwort beim Drop-Down-Menü aus. Sie können sich sofort oder später entscheiden.

Verfahren Sie genauso mit den weiteren neun Bewertungen zu diesem Arzt.

Wenn Sie die Aufgabe abgeschlossen haben, klicken Sie auf "Fertig".

\* 1. [Hat immer ein offenes Ohr](#)

Ich würde die Bewertung an meine Freundin weitergeben. ▾

\* 2. [Sehr nett und kompetent](#)

Ich würde die Bewertung nicht an meine Freundin weitergeben. ▾

\* 3. [Freundlich aber inkompetent](#)

▾

\* 4. [Erster Besuch in fremder Praxis](#)

▾

\* 5. [überaus freundlich und interessiert](#)

▾

\* 6. [Geht am Thema vorbei](#)

▾

\* 7. [Best Doc ever](#)

▾

\* 8. [Ein sehr menschlicher und kompetenter Arzt](#)

▾

\* 9. [sehr zu empfehlen](#)

▾

\* 10. [sehr kompetenter und netter Arzt](#)

▾

Fertig

## 7.3 ARZTBEWERTUNGEN

Im folgenden Kapitel sind die zehn verwendeten Arztbewertungen mit Quelle und Anzahl der Weitergaben aufgeführt.

### 7.3.1 BEWERTUNG 1: „HAT IMMER EIN OFFENES OHR“

11 Weitergaben

Note

1,2



Bewertung vom 08.12.2016

#### Hat immer ein offenes Ohr

Bin seit mein Übergang vom Kinderarzt zum HausArzt bei Dr. [REDACTED] und das ist jetzt fast 15 Jahre . Mach mir nur sorgen was ich machen soll wenn Dr. [REDACTED] in Rente geht. Er macht seine Arbeit aus Leidenschaft und " der Mensch " ist im Mittelpunkt bei seiner Untersuchungen. Kann ihn nur weiterempfehlen . Ich schätze ihn sehr , denn die Erfahrungen meiner Bekannten mit ihren Hausärzten wundern mich , dann bin Ich echt froh so einen guten Hausarzt zu haben.  
Vielen vielen Dank Dr. [REDACTED]"

#### Notenbewertung dieses Patienten

Behandlung	2,0	<b>Gesamtnote</b> <b>1,2</b>
Aufklärung	1,0	
Vertrauensverhältnis	1,0	
Genommene Zeit	1,0	
Freundlichkeit	1,0	

Wartezeit Termin	2,0
Wartezeit Praxis	2,0
Sprechstundenzeiten	1,0
Betreuung	3,0
Entertainment	3,0
alternative Heilmethoden	2,0
Barrierefreiheit	2,0
Praxisausstattung	3,0
Telefonische Erreichbarkeit	2,0
Parkmöglichkeiten	3,0
Öffentliche Erreichbarkeit	2,0

[Diesem Patienten anonym eine Frage stellen >](#)

[Bewertung kommentieren \(nur für Herr \[REDACTED\]\) >](#)

[Problem melden >](#)

Wie hilfreich fanden Sie diese Bewertung? ★★★★★

Diese Bewertung ist die subjektive Meinung eines Patienten und nicht die der jameda GmbH.

### 7.3.2 BEWERTUNG 2: „SEHR NETT UND KOMPETENT“

16 Weitergaben

Note

1,0



#### Sehr nett und kompetent

Bewertung vom 10.08.2016, Kassenpatient

Vor 7 Jahren bin ich von [REDACTED] nach [REDACTED] gezogen und bin dann dort auf die Suche nach einem neuen Hausarzt gegangen. Durch Zufall bin ich auf die Praxis von Dr. [REDACTED] und Dr. [REDACTED] gestoßen.

Egal mit welchen Symptomen ich zu den beiden gehe, bisher wurde ich immer gut betreut und geholfen.

Kann beide nur empfehlen.

Die Öffnungszeiten sind für berufstätige perfekt. Vielen Dank für diesen tollen Service.

Das bietet nicht jeder Arzt."

#### Notenbewertung dieses Patienten

Behandlung	1,0	Gesamtnote <b>1,0</b>
Aufklärung	1,0	
Vertrauensverhältnis	1,0	
Genommene Zeit	1,0	
Freundlichkeit	1,0	
Wartezeit Termin	2,0	
Wartezeit Praxis	2,0	
Sprechstundenzeiten	1,0	
Betreuung	3,0	
Entertainment	3,0	
Praxisausstattung	3,0	
Telefonische Erreichbarkeit	2,0	
Parkmöglichkeiten	2,0	
Öffentliche Erreichbarkeit	2,0	

[Diesem Patienten anonym eine Frage stellen](#) >

[Bewertung kommentieren \(nur für Herr \[REDACTED\]\)](#) >

[Problem melden](#) >

Wie hilfreich fanden Sie diese Bewertung? ★★★★★

Diese Bewertung ist die subjektive Meinung eines Patienten und nicht die der jameda GmbH.

### 7.3.3 BEWERTUNG 3: „FREUNDLICH ABER INKOMPETENT“

#### 11 Weitergaben

Note

4,2

Bewertung vom 18.11.2015, Kassenpatient, Alter: 30 bis 50

#### Freundlich aber inkompetent

Mit großen Schmerzen genauso wieder entlassen worden - und immer noch keine Ahnung, was los ist... weil er auch keine Ahnung hat.... muss mir neuen Hausarzt suchen"

#### Notenbewertung dieses Patienten

Behandlung	6,0	Gesamtnote <b>4,2</b>
Aufklärung	6,0	
Vertrauensverhältnis	3,0	
Genommene Zeit	4,0	
Freundlichkeit	2,0	
Wartezeit Termin	1,0	
Wartezeit Praxis	3,0	
Sprechstundenzeiten	1,0	
Betreuung	4,0	
Entertainment	5,0	
alternative Heilmethoden	5,0	
Barrierefreiheit	6,0	
Praxisausstattung	4,0	
Telefonische Erreichbarkeit	2,0	
Parkmöglichkeiten	5,0	
Öffentliche Erreichbarkeit	2,0	

[Diesem Patienten anonym eine Frage stellen](#) >

[Bewertung kommentieren \(nur für Herr \[REDACTED\]\)](#) >

[Problem melden](#) >

Wie hilfreich fanden Sie diese Bewertung? ★★★★★

Diese Bewertung ist die subjektive Meinung eines Patienten und nicht die der jameda GmbH.

### 7.3.4 BEWERTUNG 4: „ERSTER BESUCH IN FREMDER PRAXIS“

17 Weitergaben

Note

1,0



Bewertung vom 12.10.2015, Kassenpatient, Alter: über 50

#### Erster Besuch in fremder Praxis

Als ich im Juni eine Zecke in einem unangenehmen Bereich feststellte, die ich selbst nicht entfernen konnte dachte ich daran, dass die Praxis [REDACTED]/Dr. [REDACTED] auch zu ungewöhnlicheren Zeiten geöffnet ist und ging einfach hin, obwohl ich vorher noch nie da war. (Erst vor Kurzem in die Nachbarschaft gezogen.) Ich konnte bleiben und lernte so Herrn [REDACTED] kennen. Mir war der Sitz "meiner" Zecke etwas peinlich, aber mit einem Scherz entkrampfte er die Situation und entfernte sie souverän.

Später war ich dann noch mal dort, weil ich mir nicht sicher war, ob es nicht doch gut wäre, wenn er die Bissstelle noch einmal inspizieren würde, da ich selbst da nicht hinsehen konnte. Seine ruhige, freundliche Art schaffte sofort Vertrauen und ich hatte das Gefühl, dass es nichts gibt, über das ich nicht mit ihm reden könnte. Es ist ein gutes Gefühl zu wissen, dass in unmittelbarer Nachbarschaft ein so hervorragendes Ärzteteam erreichbar ist!"

#### Notenbewertung dieses Patienten

Behandlung	1,0	Gesamtnote <b>1,0</b>
Aufklärung	1,0	
Vertrauensverhältnis	1,0	
Genommene Zeit	1,0	
Freundlichkeit	1,0	
Wartezeit Termin	1,0	
Wartezeit Praxis	1,0	
Sprechstundenzeiten	1,0	
Betreuung	1,0	
Entertainment	2,0	
Barrierefreiheit	3,0	
Praxisausstattung	1,0	
Telefonische Erreichbarkeit	1,0	

[Diesem Patienten anonym eine Frage stellen >](#)

[Bewertung kommentieren \(nur für Herr \[REDACTED\]\) >](#)

[Problem melden >](#)

Wie hilfreich fanden Sie diese Bewertung? ★★★★★

Diese Bewertung ist die subjektive Meinung eines Patienten und nicht die der jameda GmbH.

### 7.3.5 BEWERTUNG 5: „ÜBERAUS FREUNDLICH UND INTERESSIERT“

15 Weitergaben

Note

1,0

Bewertung vom 28.08.2015, (zu 100 % hilfreich bei 1 Stimme)

„ überaus freundlich und interessiert

ich war heute das erste mal bei herr [REDACTED] und muss sagen, ich glaube meinen neuen hausarzt gefunden zu haben. ich habe heute morgen angerufen und einen termin um 12:30 bekommen. wartezeit ca 5 min.  
er konnte mir ein krankheitsbild sehr gut und umfassend erklären und macht einen überaus symphatischen eindruck."

#### Notenbewertung dieses Patienten

Behandlung	1,0	Gesamtnote <b>1,0</b>
Aufklärung	1,0	
Vertrauensverhältnis	1,0	
Genommene Zeit	1,0	
Freundlichkeit	1,0	
Wartezeit Termin	2,0	
Wartezeit Praxis	1,0	

[Diesem Patienten anonym eine Frage stellen >](#)

[Bewertung kommentieren \(nur für Herr \[REDACTED\]\) >](#)

[Problem melden >](#)

Wie hilfreich fanden Sie diese Bewertung? ★★★★★

Diese Bewertung ist die subjektive Meinung eines Patienten und nicht die der jameda GmbH.

### 7.3.6 BEWERTUNG 6: „GEHT AM THEMA VORBEI“

6 Weitergaben

Note

5,4



Geht am Thema vorbei

Bewertung vom 27.07.2015, (zu 20 % hilfreich bei 2 Stimmen)

Dieser Arzt hat mich einfach das Thema verfehlt und über etwas anderes geredet als wofür ich eigentlich da war."

#### Notenbewertung dieses Patienten

Behandlung	6,0	Gesamtnote <b>5,4</b>
Aufklärung	5,0	
Vertrauensverhältnis	5,0	
Genommene Zeit	6,0	
Freundlichkeit	5,0	

Dieser Patient möchte leider nicht anonym befragt werden.

[Bewertung kommentieren \(nur für Herr \[REDACTED\]\)](#) >

[Problem melden](#) >

Wie hilfreich fanden Sie diese Bewertung? ★★★★★

Diese Bewertung ist die subjektive Meinung eines Patienten und nicht die der jameda GmbH.

## 7.3.7 BEWERTUNG 7: „BEST DOC EVER“

## 7 Weitergaben

Note

1,2



Bewertung vom 16.07.2015, Kassenpatient, Alter: unter 30

Best Doc ever

Er ist cool, hilft mir bei allem.  
ist immer da !!  
Und ist korekt..."

## Notenbewertung dieses Patienten

Behandlung	1,0	Gesamtnote <b>1,2</b>
Aufklärung	2,0	
Vertrauensverhältnis	1,0	
Genommene Zeit	1,0	
Freundlichkeit	1,0	
Wartezeit Termin	2,0	
Wartezeit Praxis	2,0	
Betreuung	1,0	
Entertainment	1,0	
alternative Heilmethoden	1,0	
Barrierefreiheit	1,0	
Praxisausstattung	1,0	
Telefonische Erreichbarkeit	1,0	
Öffentliche Erreichbarkeit	1,0	

[Diesem Patienten anonym eine Frage stellen](#) >

[Bewertung kommentieren \(nur für Herr \[REDACTED\]\)](#) >

[Problem melden](#) >

Wie hilfreich fanden Sie diese Bewertung? ★★★★★

Diese Bewertung ist die subjektive Meinung eines Patienten und nicht die der jameda GmbH.

### 7.3.8 BEWERTUNG 8: „EIN SEHR MENSCHLICHER UND KOMPETENTER ARZT“

18 Weitergaben

Note

1,0



Bewertung vom 07.01.2015, Kassenpatient

#### Ein sehr menschlicher und kompetenter Arzt

Ich bin seit ca. 15 Jahren immer wieder Patientin.

Herr [REDACTED] nimmt sich Zeit, hat einen guten Blick auf das Ganze, nimmt einem Ängste und nimmt einen ernst. Für mich der optimale Hausarzt.

Dazu: sehr arbeitnehmerfreundliche Sprechzeiten und internationales Personal mit vielen Sprachen"

#### Notenbewertung dieses Patienten

Behandlung	1,0	Gesamtnote <b>1,0</b>
Aufklärung	1,0	
Vertrauensverhältnis	1,0	
Genommene Zeit	1,0	
Freundlichkeit	1,0	
Wartezeit Termin	1,0	
Wartezeit Praxis	1,0	
Sprechstundenzeiten	1,0	
Betreuung	1,0	
Entertainment	1,0	
Praxisausstattung	1,0	
Telefonische Erreichbarkeit	1,0	
Öffentliche Erreichbarkeit	1,0	

[Diesem Patienten anonym eine Frage stellen >](#)

[Bewertung kommentieren \(nur für Herr \[REDACTED\]\) >](#)

[Problem melden >](#)

Wie hilfreich fanden Sie diese Bewertung? ★★★★★

Diese Bewertung ist die subjektive Meinung eines Patienten und nicht die der jameda GmbH.

## 7.3.9 BEWERTUNG 9: „SEHR ZU EMPFEHLEN“

## 16 Weitergaben

Note

Bewertung vom 11.10.2014, Kassenpatient, Alter: unter 30

1,0

” sehr zu empfehlen

Ich bin seit 18 Jahren Patient und bin sehr sehr zufrieden Herr [REDACTED] und Her [REDACTED] sind sehr sehr nette Ärzte die sich sehr lange zeit für Ihre Patienten nehmen teilweise habe ich sogar ein schlechtes gewissen gegenüber anderen Patienten Herr [REDACTED] hat sogar das Leben meines Vaters durch die richtige Diagnose gerettet ich hoffe das die beiden sehr sehr lange weiter machen denn solche Ärzte findet man nicht einfach"

## Notenbewertung dieses Patienten

Behandlung	1,0	Gesamtnote <b>1,0</b>
Aufklärung	1,0	
Vertrauensverhältnis	1,0	
Genommene Zeit	1,0	
Freundlichkeit	1,0	
Wartezeit Termin	1,0	
Wartezeit Praxis	2,0	
Sprechstundenzeiten	1,0	
Betreuung	1,0	
Entertainment	2,0	
alternative Heilmethoden	2,0	
Kinderfreundlichkeit	1,0	
Barrierefreiheit	1,0	
Praxisausstattung	1,0	
Telefonische Erreichbarkeit	1,0	
Parkmöglichkeiten	3,0	
Öffentliche Erreichbarkeit	1,0	

[Diesem Patienten anonym eine Frage stellen >](#)
[Bewertung kommentieren \(nur für Herr \[REDACTED\]\) >](#)
[Problem melden >](#)

Wie hilfreich fanden Sie diese Bewertung? ★★★★★

Diese Bewertung ist die subjektive Meinung eines Patienten und nicht die der jameda GmbH.

### 7.3.10 BEWERTUNG 10: „SEHR KOMPETENTER UND NETTER ARZT“

15 Weitergaben

Note

1,0



Bewertung vom 05.03.2014, Kassenpatient, Alter: unter 30

„sehr kompetenter und netter Arzt

Dr. [REDACTED] ist ein sehr vertrauensvoller und kompetenter Arzt der sich Zeit für einen nimmt. außerdem sind seine sprechzeiten der arbeitnehmerfreundlich!“

#### Notenbewertung dieses Patienten

Behandlung	1,0	Gesamtnote <b>1,0</b>
Aufklärung	1,0	
Vertrauensverhältnis	1,0	
Genommene Zeit	1,0	
Freundlichkeit	1,0	
Wartezeit Termin	2,0	
Wartezeit Praxis	2,0	
Sprechstundenzeiten	1,0	
Betreuung	1,0	
Entertainment	3,0	
alternative Heilmethoden	3,0	
Praxisausstattung	1,0	
Telefonische Erreichbarkeit	1,0	

Dieser Patient möchte leider nicht anonym befragt werden.

[Bewertung kommentieren \(nur für Herr \[REDACTED\]\)](#) >

[Problem melden](#) >

Wie hilfreich fanden Sie diese Bewertung? ★★★★★

Diese Bewertung ist die subjektive Meinung eines Patienten und nicht die der jameda GmbH.

## 7.4 LEITFADEN FÜR PROBLEMZENTRIERTES INTERVIEW

### Problemzentriertes Interview

- Vielen Dank.
- Jetzt haben Sie die Möglichkeit ausführlich zu erzählen, was Sie zum Thema Arztbewertungen denken. Sie können also über alles sprechen. Es gibt keine richtigen oder falschen Themen. Alles was interessant für Sie ist, ist auch interessant für mich.

### Nutzung von ABP (*plus: Einstiegsfrage, um Narration generieren*)

- Jetzt haben Sie gerade viele Arztbewertungen gelesen. Haben Sie vorher auch schon einmal eine Arztbewertung gelesen? (*Ziel: Lockere Erzählung, Anknüpfungspunkte, ist trotz Think-Aloud nötig*)
- Haben Sie von diesen Arztbewertungsportalen schon einmal gehört: *Jameda, Weisse Liste, Docinsider*?
- Haben Sie schon einmal einen Arzt auf einem dieser Arztbewertungsportale gesucht?
- Welche Erfahrungen haben Sie bisher mit Bewertungen gemacht?
- Haben Sie schon einmal einen bestimmten Arzt besucht oder nicht besucht wegen einer Arztbewertung?
- Haben Sie schon einmal selbst eine Arztbewertung geschrieben?

### Bewertungen durchgehen

- *Notizen von SCA zu den einzelnen Arztbewertungen gemeinsam durchgehen*
- Hier haben Sie das gesagt/ gemacht. Was meinen Sie damit? Auf was bezieht sich das?
- *Wenn bei einer Bewertungen mehrere Gründe aufgetaucht sind:* Sie sagen X und Y, was ist für sie wichtiger? Verstärken sich die beiden Sachen, schwächen sie einander ab?
- Wann würden Sie eine Bewertung auf gar keinen Fall weitergeben?
- Wann würden Sie eine Bewertungen auf jeden Fall weitergeben?
- Worauf achten Sie noch wenn Sie eine Bewertung lesen?
- Bei abstrakte Begriffen nachfragen (z. B. „kompetent“): Was verstehen Sie darunter? Woran erkennen Sie, dass ein Bewerter „kompetent“ ist?
- Worauf achten Sie bei Bewertungen?
- *Antworten abgleichen mit dem, was Teilnehmer beim Think-Aloud gemacht hat*

### Probes/ Folgefragen zur Klärung

- Sagen Sie, was Sie mit XXX meinen.
- Können Sie mir ein Beispiel dafür sagen?
- Wie kommen Sie darauf? Woran erkennen Sie das?
- Warum ist das wichtig?
- Was ist mit Grund X? Ist das wichtig für Sie?

### Schluss

- Vielen Dank! Jetzt beende ich die Aufzeichnung.
- Haben Sie Interesse daran, die Ergebnisse der Studie zu erfahren?

- ggf. Informationen zur Studie geben

## 7.5 CODEBUCH FÜR QUALITATIVE INHALTSANALYSE

Alle Äußerungen der Teilnehmer zu den Bewertungen wurden codiert.

Kodiereinheit: Sinneinheit

Bezeichnung	Beschreibung	Ankerbeispiel
<b>Vorcodierungen</b>		
SCA	Aussagen von Interviewerin	<i>automatische Codierung des Transkripts</i>
T01, T02,...	Aussagen von Teilnehmer	<i>automatische Codierung des Transkripts</i>
Think-Aloud-AB01, Think-Aloud-AB02,...	Bewertung, die der Teilnehmer vorliest: „AB01“ für erste Bewertung, „AB02“ für zweite Bewertung usw.	„Hat immer ein offenes Ohr...“
Feedback-AB01, Feedback-AB02,...	Bewertung, auf die der Teilnehmer Bezug nimmt: „AB01“ für erste Bewertung, „AB02“ für zweite Bewertung usw.	„Bei der ersten Bewertung ist mir aufgefallen,...“
Feedback-Allgemein	Allgemeine Aussagen des Teilnehmers zu Bewertungen	„Ich achte bei Bewertungen immer darauf, dass...“
<b>Vertrauensmodell</b>		
Komplexität	Aussagen zur Komplexität der modernen Welt	
Vertrauensobjekt	Bewerter bzw. Bewertung	
<b>Ungenügendes Wissen</b>		
<i>Kompetenz des Lesers</i>	Beruf im Gesundheitswesen, neue Stadt, Abgleich eigener Meinung	
<i>Markeneinstellung des Teilnehmers</i>	Einstellung des Teilnehmers gegenüber Ärzten	„Ich mag keine Ärzte.“, „Alle Ärzte können das.“
<b>Risikobewusstsein</b>		
<i>Allgemeine Vertrauen des Lesers</i>		
<i>Einstellung des Lesers gegenüber Online-Bewertungen</i>	Positiv/negativ/abwägend, Grund: wegen Fälschung, Grund: wegen Subjektivität, Grund: nicht repräsentativ, Grund: kennt Bewerter nicht	

	Grund: kann nicht nachfragen	
<b>Strukturelle Voraussetzungen für Vertrauen</b>	Aussagen zur Freiwilligkeit, Erlernbarkeit, Adressierbarkeit, Kontrollierbarkeit, Sanktionierbarkeit	
<b>Dimension Faktenselektivität</b>		
<b>Allgemeines Vertrauen des Teilnehmers</b>	Teilnehmer äußert sich zu seinem allgemeinen Vertrauen	„Ich glaube alles, was man mir erzählt.“, „Ich bin eher der kritische Typ.“
<b>Ausführlichkeit der Bewertung</b>	Teilnehmer äußert sich zur Ausführlichkeit der Bewertung (Unterkategorien: Ausführlich ja/nein, welche Details (Krankheit, Sprechzeiten, Wartezeiten, Betreuung bzw. Behandlung, Ausstattung, Barrierefreiheit, beide Ärzte, Entertainment, Internationalität des Personal, Parkplatzsituation nur Informationen zu den beruflichen Leistungen des Arztes/nichts privates, Kommentare zur Länge)	„Da fehlen mir Details zu...“, „Hier steht alles, was ich wissen muss, wie...“
<b>Ausgewogenheit der Bewertung</b>	Teilnehmer äußert sich zur Ausgewogenheit der Bewertung (extreme Ausprägung der Bewertung (z. B. eine Note 1 oder eine Note 6) als auch die durchgehende Bewertung mit ein und derselben Note (z. B. nur die Note 3 in allem)	„Er sagt nur, was er alles schlecht fand.“, „Sie zeigt die guten und die schlechten Seiten.“
<b>Dimension Explizite Bewertung</b>		
<b>Argumentation der Bewertung</b>	Teilnehmer äußert sich zur Argumentation der Bewertung (Argumentation ja/nein, genannte Kriterien	„Mit... kann er das gut begründen.“, „... ist kein Argument.“

	(Nachvollziehbarkeit der Begründung: Kriterium „Entertainment“, Freundlichkeit eines Arztes, Umgang mit Peinlichkeit, Erfassen der Akutheit der Situation, Entfernung einer Zecke, Wartezeit), Maßstab: falsche Erwartungen an den Arzt)	
<b>Attribuierung</b>	Teilnehmer äußert sich zur Attribuierung der Bewertung zum Bewerteten (external) oder zum Bewerter (internal), Darstellung des Bewerter als „genervt“ etc.	„Der Bewerter meckert einfach gerne.“, „Da hat der Arzt gepfuscht.“
<b>Übereinstimmung der Zahlen- und Textbewertung</b>	Teilnehmer äußert sich zur Übereinstimmung der Bewertung im Kommentartext mit der Zahlenbewertung	„Die Note passt aber nicht zum Text.“
<b>Spezifischer Grund: Bestätigung durch mitgeteilte Erfahrungen anderer</b>		
<b>Bewertung der Bewertung</b>	Teilnehmer äußert sich zur Bewertung der Bewertung	„Diese Bewertung wurde als ‚hilfreich‘ bewertet.“
<b>Unspezifischer Grund: Rückschluss von Mitteilung auf Selektion</b>		
<b>Emotionalität der Bewertung</b>	Teilnehmer äußert sich zur Emotionalität der Bewertung	„Das ist sachlich.“, „Der Bewerter ist wütend.“
<b>Foto des Bewerteten</b>	Teilnehmer äußert sich zu Fotos des Bewerteten	
<b>Kompetenz des Bewerter</b>	Teilnehmer äußert sich zur Kompetenz des Bewerter in Bezug auf das Bewertete	„Er ist schon lange bei dem Arzt und kann ihn gut beurteilen.“, „Sie ist bestimmt selbst Ärztin.“
<b>Rechtschreibung und Grammatik der Bewertung</b>	Teilnehmer äußert sich zur Rechtschreibung und Grammatik der Bewertung (ja/nein, warum)	„Hier fehlt ein Komma.“, „Das ist grammatikalisch einwandfrei.“
<b>Valenz der Bewertung (negativ/ positiv)</b>	Teilnehmer äußert sich zur negativen/positiven Valenz der Bewertung	„Er benotet ihn sehr schlecht.“, „Sie lobt ihn in den Himmel.“
<b>Unspezifischer Grund: Rückschluss von Kontext der Mitteilung auf Selektion</b>		
<b>Ähnlichkeit des Bewerter</b>	Teilnehmer äußert sich zur Ähnlichkeit des Bewerter zu sich selbst, z. B.	„Sie ist über 50, wie ich.“, „Ich hätte an seiner Stelle nachgefragt.“

	bezüglich Geschlecht, Alter, Beruf, Geschmack	
<b>Alter der Bewertung</b>	Teilnehmer äußert sich zum Alter der Bewertund	„Die Bewertung ist sehr aktuell.“
<b>Alter des Bewerter</b>	Teilnehmer äußert sich zum Alter des Bewerter	„Das ist ein ganz junger Kerl.“
<b>Anzahl der Bewertungen</b>	Teilnehmer äußert sich zur Anzahl der Bewertungen, z. B. 357 statt 1 Bewertungen, 30 statt 3 Bewerter oder „many members“	„Das sind mir zuwenige Bewertungen.“, „Dieser Arzt hat viele Bewertungen.“
<b>Authentifizierung des Bewerter</b>	Teilnehmer äußert sich zur Authentifizierung des Bewerter	„Die Plattform hat bestätigt, dass der Bewerter das Produkt gekauft hat.“
<b>Ethnizität des Bewerter</b>	Teilnehmer äußert sich zur Ethnizität des Bewerter	„Klingt nach Migrationshintergrund.“
<b>Geschlecht des Bewerter</b>	Teilnehmer äußert sich zum Geschlecht des Bewerter	„Das hat eine Frau geschrieben.“
<b>Krankenversicherung des Bewerter</b>	Teilnehmer äußert sich zur Krankenversicherung des Bewerter	„Er ist bestimmt privat versichert.“
<b>Qualität der Plattform</b>	Teilnehmer äußert sich zur Qualität der Plattform	„Die Seite ist gut aufgebaut.“, „Mit Datenschutz nehmen sie es nicht so genau.“
<b>Soziale Nähe des Bewerter</b>	Teilnehmer äußert sich zur sozialen Nähe mit dem Bewerter	„Den finde ich sympathisch.“, „Mit so jemandem kann ich gar nichts anfangen“
<b>Übereinstimmung mit anderen Bewertungen</b>	Teilnehmer äußert sich zur Übereinstimmung der Bewertung mit anderen Bewertungen	„Dasselbe stand schon in der anderen Bewertung.“, „Er ist der erste, der das schreibt.“
<b>Unabhängigkeit des Bewerter</b>	Teilnehmer äußert sich zur Unabhängigkeit des Bewerter ggü. dem Arzt	„Klingt wie ein guter Freund es Arztes.“
<b>Vertrauenswürdigkeit der Plattform</b>	Teilnehmer äußert sich zur Vertrauenswürdigkeit der Plattform	„Jameda finde ich glaubwürdig.“ „Bei Jameda bin ich skeptisch.“
<b>Vertrautheit mit Plattform</b>	Teilnehmer äußert sich zur Vertrautheit mit Plattform	„Ich war noch nie auf dieser Seite.“, „Ich suche meine Ärzte immer auf Jameda.“

<b>Leser- und Bewertetes-bezogene Faktoren</b>		
<i>Alter des Teilnehmers</i>	[Kurzfragebogen]	
<i>Bildung des Teilnehmers</i>	[Kurzfragebogen]	
<i>Einstellung des Teilnehmers gegenüber Online-Bewertungen</i>	Teilnehmer äußert sich zu seiner gegenüber Online-Bewertungen	„Auf Bewertungen gebe ich nichts.“, „Bewertungen helfen mir sehr.“
<i>Erfahrung des Teilnehmers mit Ärzten</i>	Teilnehmer äußert sich zu seiner Erfahrung mit Ärzten	„Mir helfen Ärzte immer.“, „Ich hatte mal einen sehr schlechten Arzt.“
<i>Erfahrung des Teilnehmers mit Online-Bewertungen</i>	Teilnehmer äußert sich zu seiner Erfahrung mit dem Lesen oder Schreiben von Online-Bewertungen	„Ich habe noch nie eine Bewertung abgegeben.“, „Ich bewerte regelmäßig.“
<i>Ethnizität des Teilnehmers</i>	[Kurzfragebogen]	
<i>Fokus des Teilnehmers</i>	Teilnehmer äußert sich zu seinem Fokus beim Lesen der Bewertung	„Ich möchte einen schlechten Arzt vermeiden.“, „Ich will einen guten Arzt finden.“
<i>Geschlecht des Teilnehmers</i>	[siehe Kurzfragebogen]	
<i>Internetnutzung des Teilnehmers</i>	Teilnehmer äußert sich zu seiner Internetnutzung des Teilnehmers	„Ich bin ständig online.“, „Ich mache selten was im Netz.“
<i>Involviertheit des Teilnehmers</i>	Teilnehmer äußert sich zu seiner Involviertheit	„Ich will nur Bewertungen weitergeben, die ich gut finde.“
<i>Kompetenz des Teilnehmers</i>	Teilnehmer äußert sich zu seiner Kompetenz in Bezug auf Ärzte	„Ich arbeite in einem Krankenhaus.“, „Ich habe Medizin studiert.“
<i>Kontrollillusion des Teilnehmers</i>	Teilnehmer äußert sich zu seiner Kontrollillusion	„Ich erkenne, ob ein Bewerter die Wahrheit sagt.“, Ich kann solche Bewertung nicht einschätzen.“
<i>Motivation des Teilnehmers</i>	Teilnehmer äußert sich zu seiner Motivation beim Lesen der Bewertung	„Ich will die Bewertungen finden, die am meisten helfen.“
<i>Orientierung des Teilnehmers an anderen</i>	Teilnehmer äußert sich zu seiner Orientierung an anderen	„Wenn meine Freunde das sagen, übernehme ich das.“, „Die Meinung von anderen Menschen ist mir egal.“
<i>Personal Sense of Power des Teilnehmers</i>	Teilnehmer äußert sich zu seinem Personal Sense of Power r zur Teilnehmer	„Das kann man eh nicht ändern.“, „Ich verändere das.“

	nimmt seine eigene Macht, etwas zu gestalten, als gering oder groß wahr	
<b>Reputation der Plattform</b>	Teilnehmer äußert sich zur Reputation von <i>Jameda</i>	„Jameda ist sehr bekannt.“, „Jameda hat keinen guten Ruf.“
<b>Selbstoffenbarung des Teilnehmers</b>	Teilnehmer äußert sich zu seiner Selbstoffenbarung	„Ich betreibe einen Blog zu meinem Leben.“, „Bei Facebook bin ich unter falschen Namen angemeldet.“
<b>Trägheit des Teilnehmers</b>	Teilnehmer äußert sich zu seiner Trägheit, z. B. hoher Unwille des Teilnehmers, sein Verhalten zu ändern	„Ich werde mein Verhalten nicht ändern.“, „Ich passe mein Verhalten an meine Erfahrungen an.“
<b>Typ des Bewerteten</b>	Bewertetes ist ein Erfahrungs- oder Suchgut	
<b>Sonstige Faktoren</b>		
<b>Konsequenz des Vertrauens</b>	Teilnehmer äußert sich zur Konsequenz seines Vertrauens, z. B. Weitergabe der Information	„Das würde ich weitergeben.“, „Er rät davon ab, also würde ich nicht dahin gehen.“

## 7.6 TRANSKRIPTE

Bei Interesse stelle ich die anonymisierten Transkripte aller Teilnehmer gerne zur Verfügung, dafür genügt eine E-Mail an: [sevda.arslan@lmu.de](mailto:sevda.arslan@lmu.de)

## **8**      EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quellen gekennzeichnet. Insbesondere habe ich nicht die entgeltliche Hilfe von Vermittlungs- und Beratungsdiensten in Anspruch genommen.

Sevda Can Arslan