

Karlsruher Institut für Technologie
Institut für Technikzukünfte, Teilinstitut Wissenschaftskommunikation

Masterarbeit

Erstbetreuer: Prof. Dr. Markus Lehmkuhl

Zweitbetreuerin: Dr. Sarah Kohler

Nachrichtenauswahl im Wissenschaftsjournalismus - systematisch, aber unvorhersehbar?

Entwicklung eines automatisierten Verfahrens zur Prognose der
Berichterstattung über wissenschaftliche Studien

Nikolai Promies

Master Wissenschaft – Medien – Kommunikation

Abgabedatum: 17. September 2019

Inhaltsverzeichnis

1. Einleitung	1
2. Theoretischer Hintergrund	6
2.1 Theorien der Nachrichtenauswahl im allgemeinen Journalismus	6
2.1.1 Nachrichtenwerttheorie	7
2.1.2 Gatekeeper-Forschung	14
2.1.3 Weitere Ansätze	20
2.2 Forschung zur Nachrichtenauswahl im Wissenschaftsjournalismus.....	23
2.2.1 Inhaltliche Faktoren.....	24
2.2.2 Wissenschaftsbezogene Faktoren.....	28
2.2.3 Bedeutung wissenschaftlicher Zeitschriften.....	30
2.2.4 Bedeutung der Pressearbeit wissenschaftlicher Institutionen	33
2.2.5 Weitere Faktoren	36
2.2.6 Bisherige Systematisierungen	37
2.3 Synthese: Schwächen und Lücken vorhandener Ansätze	40
3. Nachrichtenauswahl als komplexe Entscheidungssituation	44
3.1 Modell soziologischer Erklärung und Handlungstheorie.....	45
3.2 Nachrichtenauswahl in der Theory of Planned Behavior.....	47
3.3 Rationalität und Komplexität von Entscheidungen.....	51
3.4 Synthese: Nachrichtenauswahl im Wissenschaftsjournalismus als Entscheidung unter Komplexität.....	59
4. Prognosen und datengestützte Verfahren in der Kommunikationswissenschaft	62
4.1 Prognosen als Instrument der Überprüfung von Theorien	62
4.2 Relevante bisherige Verwendungen von Prognosemodellen	64
4.2.1 Prognosen in der Kommunikationswissenschaft.....	64
4.2.2 Prognosemodelle der Popularität von Medieninhalten in der Informatik.....	66
4.2.3 Prognosen der Auswahl wissenschaftlicher Studien.....	70

4.4 Synthese: Datengestützte Prognosen der journalistischen Selektivität als Verfahren zur Prüfung von Theorien der Nachrichtenauswahl.....	74
5. Methode.....	75
5.1 Input-Materialien.....	76
5.2 Output-Materialien	79
5.3 Prädiktoren	83
5.3.1 Publikationsort	84
5.3.2 Verfügbarkeit von Pressemitteilungen	84
5.3.3 Sprachmodelle	85
5.4 Prognoseverfahren: Random Forests	90
5.4.1 Grundlage: Recursive Partitioning und Entscheidungsbäume	91
5.4.2 Ensemble-Verfahren und andere Erweiterungen	93
5.4.3 Konkrete Umsetzung.....	94
5.4.4 Genauigkeitsmaße von Machine Learning-Verfahren	96
5.5 Synthese: Kombination von Input-Output-Analyse und Machine Learning	98
6. Ergebnisse	99
6.1 Deskriptive Beschreibung	99
6.2 Ergebnisse verschiedener Prognosemodelle	103
6.2.1 Grundmodell.....	103
6.2.2 Erweiterung um Sprachmodelle	107
6.3 Synthese: hochselektive Auswahl anhand weniger Faktoren.....	109
7. Diskussion	110
7.1 Folgen der Ergebnisse für die Leistungsfähigkeit des Journalismus	113
7.2 Limitationen und mögliche Erweiterungen.....	116
8. Fazit und Ausblick	118
Anhang	123
Literatur.....	127

Abbildungsverzeichnis

Abbildung 1: Nachrichtenfaktoren im Wissenschaftsjournalismus nach Badenschier und Wormer.....	27
Abbildung 2: TPB-Modell der journalistischen Darstellung von Unsicherheit, entwickelt von Guenther und Kollegen.....	50

Tabellenverzeichnis

Tabelle 1: Beispieldarstellung einer Confusion Matrix.	97
Tabelle 2: Verteilung der Studien nach dem MSM-Score, unterteilt in Zehnerschritte.....	100
Tabelle 3: Kreuztabelle Vorkommen Pressemitteilung/Medienaufmerksamkeit	101
Tabelle 4: Liste der Zeitschriften mit der höchsten Auswahlwahrscheinlichkeit	102
Tabelle 5: Confusion Matrix des Vorhersagemodells ohne inhaltliche Informationen, Grenzwert bei MSM-Score ≥ 1	104
Tabelle 6: Permutation Importance der verschiedenen Prädiktoren bei der Vorhersage im Grundmodell.....	105
Tabelle 7: Confusion Matrix des Prognosemodells ohne inhaltliche Informationen, Grenzwert bei MSM-Score ≥ 10	106
Tabelle 8: Confusion Matrix des Vorhersagemodells ohne inhaltliche Informationen, Grenzwert bei MSM-Score ≥ 50	106
Tabelle 9: Confusion Matrix des Prognosemodells mit inhaltlichen Informationen, Grenzwert bei MSM-Score ≥ 1	107
Tabelle 10: Gegenüberstellung der Genauigkeitsmaße der verschiedenen Vorhersagemodelle.....	108
Tabelle 11: Inhaltsanalysen, in denen die Quellenauswahl von Wissenschaftsjournalisten untersucht wurde.	123
Tabelle 12: Ergebnisse von Befragungen zur Quellenauswahl von Wissenschaftsjournalisten	124
Tabelle 13: Confusion Matrix des Vorhersagemodells ohne inhaltliche Informationen, Grenzwert bei MSM-Score ≥ 10	125
Tabelle 14: Confusion Matrix des Vorhersagemodells ohne inhaltliche Informationen, Grenzwert bei MSM-Score ≥ 50 . Es wurden nur Studien verwendet, deren Score mindestens bei 1 liegt.....	125

Tabelle 15: Confusion Matrix des Vorhersagemodells mit inhaltlichen Informationen, Grenzwert bei MSM-Score ≥ 10	125
Tabelle 16: Confusion Matrix des Vorhersagemodells mit inhaltlichen Informationen, Grenzwert bei MSM-Score ≥ 50	125
Tabelle 17: Confusion Matrix des Vorhersagemodells mit inhaltlichen Informationen, Grenzwert bei MSM-Score ≥ 10 . Es wurden nur Studien verwendet, deren Score mindestens bei 1 liegt.	126

Abkürzungsverzeichnis

BMJ: British Medical Journal

CART: Classification and Regression Trees

DOI: Digital Object Identifier

FN: Falsch negative Vorhersagen

FP: Falsch positive Vorhersagen

JAMA: Journal of the American Medical Association

LDA: Latent Dirichlet Allocation

MSM-Score: Mainstream-Media-Score

NEJM: New England Journal of Medicine

PNAS: Proceedings of the National Academy of Sciences of the United States of America

RN: Richtig negative Vorhersagen

RP: Richtig positive Vorhersagen

SMC: Science Media Center

TPB: Theory of Planned Behavior

1. Einleitung

Es ist Mittwochmorgen, vor der Uni ist noch etwas Zeit und mir kommt der Gedanke, dass ich mich ja mal wieder darüber informieren könnte, was denn so los ist draußen in der Welt. Auch wenn sich durch das Internet mit sozialen Netzwerken und Blogs die Zahl der Quellen vervielfacht hat, die ich zu diesem Zweck heranziehen könnte, führt der klassische Weg noch über die Massenmedien, in meinem Fall die Zeitung (wenn auch digital). Angenommen, ich hätte keine Lust, mich mit Konflikten und Katastrophen, politischen Intrigen oder Sportergebnissen zu beschäftigen und wollte stattdessen herausfinden, vielleicht begünstigt durch das Studium der Wissenschaftskommunikation, was so in der Wissenschaft vor sich geht. Je nach Medium müsste ich in verschiedenen Ressorts suchen oder könnte mich auf einer spezialisierten Seite informieren, in beiden Fällen müsste ich wahrscheinlich nicht mehr als fünf Beiträge lesen und wüsste dann Bescheid darüber, was Forscher aus aller Welt so beschäftigt. Oder zumindest ungefähr 0,1 % von ihnen. Denn würde ich versuchen, mich nicht in Massenmedien, sondern in wissenschaftsinternen Quellen über die neuesten Forschungsergebnisse zu informieren, müsste ich etwas mehr Zeit mitbringen.

Nach groß angelegten bibliometrischen Untersuchungen verdoppelt sich die Menge wissenschaftlicher Publikationen zurzeit alle 24 Jahre, sie wächst jährlich um 3 % (vgl. Bornmann/Mutz 2015). In absoluten Zahlen ausgedrückt: in der wissenschaftlichen Literaturdatenbank Scopus, die der Verlag Elsevier betreibt, werden jährlich ungefähr 2 Millionen neue Articles und Reviews eingestellt, im Schnitt erscheinen damit täglich 5.500 neue Ergebnisse. Und hier ist nur ein Ausschnitt des wissenschaftlichen Outputs vertreten. Von dieser ungeheuren Menge neuer Erkenntnisse bekomme ich jedoch üblicherweise nichts mit, wenn ich mich außerhalb des Systems Wissenschaft bewege. Die Vielfalt der Erkenntnisse wird durch den Journalismus auf ein Kleinstmaß von Ergebnissen verringert, bei denen Medienrezipienten dann nur hoffen können, dass es die für sie relevanten, interessanten oder unterhaltsamen sind. Das öffentliche Bild von Wissenschaft und die Informationen, die ein Großteil der Bevölkerung über oder aus der Wissenschaft erhalten, sind beide vor allem durch Medien vermittelt (vgl. bspw. Brossard 2013).

Welches Bild sich in den Medien findet, hängt von der Auswahl ab, die Journalisten aus den möglichen Berichterstattungsanlässen treffen. Diese Auswahl jedoch stellt sie vor ein komplexes Problem. Wenn es durch die schiere Masse der Publikationen schon für Wissenschaftler so gut wie unmöglich ist, auch nur der Entwicklung in ihrem speziellen Feld vollständig zu folgen, wie sollen dann externe Beobachter komplette Disziplinen oder sogar die

gesamte Wissenschaft überblicken? Schon allein durch die begrenzten Ressourcen von Journalisten, was Platz, Zeit oder Verarbeitungskapazitäten betrifft, ist es klar, dass nur ein kleiner Teil neuer wissenschaftlicher Ergebnisse den Weg in die Öffentlichkeit finden kann und auch nur ein Bruchteil der Ergebnisse überhaupt für die Auswahl berücksichtigt werden kann. Zur Zahl der Studien kommt noch hinzu, dass es für Journalisten quasi unmöglich ist, die (wissenschaftliche) Relevanz und Qualität der Studien einzuschätzen. Hier stellt sich eine Frage, die Kommunikationswissenschaftler seit Jahrzehnten beschäftigt: wieso wird über manche Ereignisse, Themen oder Fragestellungen berichtet und über andere nicht? Oder, anders formuliert, nach welchen Kriterien entscheiden Journalisten, worüber sie berichten, was bestimmt ihre Nachrichtenauswahl? Für den Wissenschaftsjournalismus könnte man dann spezifisch fragen: wie wählen Wissenschaftsjournalisten in einer Situation Ereignisse für die Berichterstattung aus, in der sie weder alle möglichen Anlässe überblicken können, noch ihre Qualität oder Relevanz wirklich einschätzen können?

Für den allgemeinen beziehungsweise insbesondere politischen Journalismus haben sich im Laufe der Zeit verschiedene Ansätze etabliert, um die Nachrichtenauswahl zu erklären. Am weitesten verbreitet sind die Nachrichtenwerttheorie, der Gatekeeping-Ansatz und der News-Bias-Ansatz (vgl. Staab 1990a: 11). Während die Forschung zur Nachrichtenwerttheorie sich vor allem mit den Merkmalen von Ereignissen (oder der journalistischen Wahrnehmung oder Konstruktion dieser Ereignisse) beschäftigt hat, betrachten Gatekeeping und News Bias zur Erklärung der Auswahl vor allem Eigenschaften der Journalisten, ihrer Einstellungen und ihres Umfelds. Zu allen drei Ansätzen gibt es eine Vielzahl von Arbeiten, gerade die ersten beiden haben sich immer weiter ausdifferenziert. Für die Nachrichtenwertforschung liegen inzwischen lange Nachrichtenfaktorkataloge vor, die in diversen Studien getestet worden sind. Und Untersuchungen zum Gatekeeping haben den Fokus immer weiter weg von der Betrachtung des einzelnen Akteurs hin zu seinem organisationalen, gesellschaftlichen und systemischen Umfeld verschoben (vgl. Shoemaker et al. 2009). Sowohl Gatekeeping-Ansatz als auch Nachrichtenwerttheorie waren immer wieder in der Lage, entweder die empirisch beobachtete Berichterstattung oder Aussagen von Journalisten zu ihren Auswahlkriterien mindestens teilweise zu erklären.

Für den Wissenschaftsjournalismus dagegen gibt es wenige Untersuchungen, die versucht haben, die Nachrichtenauswahl systematisch zu erklären. Meist wurden einzelne spezifische Aspekte untersucht, so konnte gezeigt werden, dass Journalisten bevorzugt über Studien aus einer kleinen Menge renommierter wissenschaftlicher Zeitschriften berichten (vgl. van Trigt et al. 1994; Pahl 1998; Suleski/Ibaraki 2010; Granado 2011). Zusätzlich erhöht das Vorkommen

einer Pressemitteilung beträchtlich die Wahrscheinlichkeit, dass über eine Studie berichtet wird (vgl. Semir et al. 1998; Stryker 2002; Schwartz et al. 2012). Damit wurden einzelne Routinen identifiziert, die aber bisher kaum in einen größeren theoretischen Rahmen eingebunden wurden. Dabei ermöglicht es die besondere Situation des Wissenschaftsjournalismus eigentlich, Theorien zur Nachrichtenauswahl zumindest für einen begrenzten Teil der Berichterstattung in großem Umfang und in einem umfassenden Ansatz zu testen.

Eine Schwäche hat die bisherige Forschung zur Nachrichtenauswahl nämlich: es wurde selten die tatsächliche Auswahl untersucht, also die Menge der Ereignisse oder Anlässe, über die berichtet werden könnte, denjenigen Ereignissen und Anlässen gegenübergestellt, über die tatsächlich berichtet wurde. Die Nachrichtenwertforschung hat sich fast nur mit dem journalistischen Output beschäftigt, also mit den Inhalten der Berichterstattung und die Prominenz einzelner Ereignisse (gemessen über die Platzierung, den Raum oder Bebilderung) zur Untersuchung der Bedeutung von Nachrichtenfaktoren verwendet. Und die Gatekeeping-Forschung hat sich vor allem in Befragungen oder Beobachtungsstudien mit Abläufen in Redaktionen oder Einflüssen auf Journalisten beschäftigt. Aussagen zur Nachrichtenauswahl lassen eigentlich nur Input-Output-Analysen zu, die allerdings wegen des verbundenen Aufwands selten durchgeführt werden (vgl. Maier et al. 2018b: 51). In solchen Analysen wird ein Input für den Journalismus insgesamt oder einen speziellen Journalisten oder eine Redaktion (beispielsweise Pressemitteilungen oder Agenturmeldungen) dem Output, also der Berichterstattung, gegenübergestellt.

Mit der zunehmenden digitalen Verfügbarkeit wissenschaftlicher Ergebnisse in Literaturdatenbanken und neuen Methoden zum automatisierten Datenabruf ergibt sich die Möglichkeit, die Menge der möglichen Berichterstattungsanlässe für eine spezielle Form des Wissenschaftsjournalismus mindestens näherungsweise zu bestimmen. Der zeitlich aktuelle und wissenschaftsgenerierte Wissenschaftsjournalismus berichtet vor allem über neue wissenschaftliche Resultate, dadurch liegt sein Ereignishorizont in Form der wissenschaftlichen Publikationen vor, die in einem bestimmten Zeitraum veröffentlicht werden (vgl. Lehmkuhl 2008: 302–303). Auch für die Bestimmung des journalistischen Outputs liegt bei dieser Form des Wissenschaftsjournalismus eine Sondersituation vor. In den letzten Jahren haben altmetrics langsam, aber sicher an Bedeutung gewonnen, alternative Maßzahlen, die die Bedeutung einer Studie durch die Messung von Aufmerksamkeit oder Resonanz in verschiedenen Online-Quellen erheben. Manche Anbieter von altmetrics beobachten auch die Berichterstattung in Online-Medientiteln. Durch die freie Zugänglichkeit dieser Daten ist es möglich, auch den journalistischen Output und damit die Selektion der Studien systematisch zu bestimmen. Das

heißt, für den Wissenschaftsjournalismus können alle Studien eines Jahres, über die berichtet wurde, mit allen Studien aus dem selben Jahr verglichen werden, über die nicht berichtet wurde. Dabei kann nach Regelmäßigkeiten gesucht werden, die auf bestimmte Routinen oder Auswahlkriterien der Journalisten hindeuten. Oder aber es können in vorherigen Untersuchungen gefundene oder postulierte Auswahlkriterien in großflächigen Analysen geprüft werden.

Eine solche Analyse wird in dieser Arbeit präsentiert. Mit Daten aus der Literaturlatenbank Scopus und des altmetrics-Anbieters Altmetric wurde eine umfangreiche Input-Output-Analyse der Nachrichtenauswahl des Wissenschaftsjournalismus durchgeführt. Um die Auswahlentscheidungen zu erklären, die sich dabei zeigen, wurde ein innovatives methodisches Vorgehen verwendet. Mit einem Vorhersageverfahren, das auf einem Machine Learning-Ansatz basiert, wurde versucht, die beobachtete Struktur der Auswahl nachzustellen. Dafür wurden verschiedene Prädiktoren verwendet, die sich aus der Analyse des Forschungsstands zur Nachrichtenauswahl im Wissenschaftsjournalismus ergaben. Das Prognosetool wurde in Kooperation mit dem Science Media Center Germany (SMC) entwickelt. Meik Bittkowski war für die praktische Entwicklung und Programmierung des Verfahrens in der Programmiersprache Python verantwortlich, während der Autor dieser Arbeit die notwendigen Daten bereitstellte, die Ergebnisse des Tools auswertete und die theoretischen Grundlagen erarbeitete.

Mit dem methodischen Vorgehen sollen verschiedene Ziele erreicht werden: praktisch bietet ein solches Prognoseverfahren eine Hilfestellung für Journalisten oder aber auch für Institutionen wie das SMC, die versuchen, die Berichterstattung über neue Studien durch wissenschaftliche Einordnungen oder Experten-Statements zu unterstützen. Inhaltlich und theoretisch kann durch die Vorhersage von Auswahlentscheidungen geprüft werden, welche Faktoren in der Lage sind, die Nachrichtenauswahl zu erklären und damit, welche theoretischen Modelle Gültigkeit für den Wissenschaftsjournalismus haben. Denn die Genauigkeit der Vorhersage kann als Indiz dafür gewertet werden, wie gut das Modell, das der Prognose zugrunde liegt, die untersuchte Realität repräsentieren kann. Eine genaue Vorhersage spricht dann für ein verlässliches Modell, eine ungenaue Vorhersage für ein schlechtes oder unvollständiges Modell. Und methodisch kann gezeigt werden, welches Potenzial Verfahren zum Abruf großer Datenmengen sowie Prognoseverfahren für die Journalismusforschung haben.

Die Arbeit ist wie folgt aufgebaut: einführend werden allgemeine Theorien der Nachrichtenauswahl im Detail vorgestellt und ihre mögliche Anwendung auf den Wissenschaftsjournalismus diskutiert. Zudem werden Schwächen oder Lücken identifiziert, die sich bei der Verwendung für Prognosen ergeben. Um auf diese zu reagieren, wird ein Modell der Nachrichtenauswahl vorgestellt, das die Auswahl wissenschaftlicher Studien als Entscheidung unter Unsicherheit präsentiert. Bei dieser Entscheidung macht es zunehmende Komplexität in der Sach-, Zeit- und Sozialdimension erforderlich, dass Journalisten komplexitätsreduzierende Strategien verfolgen, die insbesondere in der Orientierung an wissenschaftsinternen Kriterien bestehen. Theoretische Grundlage sind zusätzlich handlungstheoretische Ansätze aus dem Umfeld des strukturell-individualistischen Ansatzes (vgl. Reinemann 2007; Reinemann/Baugut 2016). Getestet wird das Modell dann in einer großen Input-Output-Analyse und durch ein darauf basierendes Prognoseverfahren. Für alle Studien, die zwischen August 2016 und Juli 2017 in Scopus eingestellt wurden, wird anhand von Altmetric-Daten ermittelt, über welche Studien in über 2.000 Onlinenachrichtenquellen berichtet wurde.

Nach einer deskriptiven Darstellung des verwendeten Datensatzes wird ein Machine Learning-Verfahren verwendet, um die Nachrichtenauswahl vorherzusagen. Als Prädiktoren werden der Publikationsort, das Vorkommen einer Pressemitteilung und abstrakte Informationen zum Inhalt verwendet. Die dafür verwendete Methode sind sogenannte Random Forests, das sind Ensembles von Regressions- und Klassifikationsbäumen, die nach dem Prinzip des recursive partitioning einen Datensatz in immer kleinere Gruppen unterteilen, in denen der Wert einer Zielvariable möglichst homogen ist (vgl. Strobl et al. 2009: 325). Die Betrachtung der Genauigkeitswerte des Vorhersageverfahrens lässt Rückschlüsse auf die Bedeutung einzelner Merkmale der Studien für die Nachrichtenauswahl zu. Die Ergebnisse verschiedener Prognosemodelle werden beschrieben und gegenübergestellt, um dann in der Diskussion auf ihre Bedeutung für vorhandene Erkenntnisse zur journalistischen Auswahl wissenschaftlicher Ergebnisse einzugehen. Abschließend werden mögliche Erweiterungen der vorgestellten Methode diskutiert. Das betrifft beispielsweise die stärkere Einbindung inhaltsbezogener Merkmale.

2. Theoretischer Hintergrund

Um die Nachrichtenauswahl von Journalisten prognostizieren zu können, muss man sie erklären können.¹ Damit ist ein Kernbereich der Journalismusforschung adressiert. Seit Jahrzehnten beschäftigen sich Forscher aus verschiedenen Perspektiven damit, wieso bestimmte Ereignisse und Themen in die Nachrichten gelangen und andere nicht. Verschiedene theoretische Modelle, die die Nachrichtenauswahl erklären sollen, wurden im Laufe der Zeit entwickelt und immer weiter verfeinert. Zwei dieser allgemeinen Ansätze sowie verschiedene alternative Konzepte sollen vorgestellt werden und dann den empirischen Ergebnissen zur Nachrichtenauswahl im Wissenschaftsjournalismus gegenübergestellt werden. Aus dem spezifischen Untersuchungsbereich, dem Wissenschaftsjournalismus sowie dem spezifischen Interesse, der Prognose von Nachrichtenauswahl, ergeben sich einige Lücken bisheriger Konzepte, diese werden am Ende des Kapitels vorgestellt. Sie begründen die Notwendigkeit eines anderen Modells der Nachrichtenauswahl, aus dem konkrete Faktoren abgeleitet werden können, die eine Prognose von Auswahlentscheidungen ermöglichen.

2.1 Theorien der Nachrichtenauswahl im allgemeinen Journalismus

Es können drei große Forschungsrichtungen und –traditionen auf dem Gebiet der Nachrichtenauswahl in Massenmedien unterschieden werden. Diese unterscheiden sich vor allem dadurch, dass sie verschiedene Variablen im Selektionsprozess von Journalisten betrachten. Die Nachrichtenwertforschung untersucht vor allem den Einfluss abstrakter inhaltlicher Ereignis- oder Beitragsmerkmale auf die Auswahlentscheidungen. Der Gatekeeper-Ansatz beschäftigt sich mit Eigenschaften der Journalisten sowie der Organisationen und des Umfelds, in dem sie arbeiten. Und die News-Bias-Forschung sucht nach einem Einfluss der (politischen) Einstellungen von Journalisten auf ihre Auswahl und nach dadurch entstehenden Verzerrungen der Berichterstattung (vgl. Eilders 1996: 14–15). Diese drei Ansätze prägen seit Langem die Erforschung journalistischer Nachrichtenauswahl, dabei gibt es durchaus Überschneidungen und diverse Querverbindungen zwischen den Konzepten (vgl. Staab 1990a: 11). Relevant für diese Arbeit sind vor allem Nachrichtenwerttheorie und Gatekeeper-Ansatz, deswegen werden sie im Folgenden ausführlicher vorgestellt und die Ergebnisse verschiedener wichtiger Forschungsarbeiten präsentiert.

¹ Es sei denn, man stützt sich allein auf datengetriebene Verfahren, deren Gewinn für die Erforschung des Journalismus ist jedoch eher gering, wenn sie nicht mit theoretischen Überlegungen gestützt werden, siehe auch Kapitel 4.

Die News-Bias-Forschung dagegen soll hier nur kurz dargestellt werden, da ihre Bedeutung für den Wissenschaftsjournalismus eher fraglich ist. Sie beschäftigt sich damit, ob Medien und Journalisten Nachrichten nach einer bestimmten Einstellung oder politischen Linie auswählen (vgl. Maier et al. 2018b: 133–134). Dafür wird nach unausgewogener, tendenziöser oder einseitiger Berichterstattung gesucht und versucht, ein Zusammenhang zu den Einstellungen herzustellen, die Journalisten in Befragungen äußern (vgl. Staab 1990a: 27). Ein Beispiel ist eine Untersuchung von Patterson und Donsbach, in der diese Journalisten in verschiedenen europäischen Ländern und den USA zu ihrer politischen Einstellung und der ihrer Redaktion befragten und sie dann in einem Experiment bei vier hypothetischen Situationen verschiedene Nachrichtenentscheidungen treffen ließen. Dabei zeigte sich ein geringer Einfluss der politischen Einstellung der Journalisten auf ihre Entscheidungen (vgl. Patterson/Donsbach 1996). Der Ansatz wurde vor allem für den politischen Bereich entwickelt. Beim Wissenschaftsjournalismus gibt es zwar durchaus Situationen, in denen die politischen Einstellungen von Journalisten einen Einfluss haben könnten (beispielsweise die Berichterstattung über Forschungsergebnisse zum Klimawandel), jedoch stellen diese in der klassischen Berichterstattung über wissenschaftliche Studien wohl eher die Ausnahme dar. Zudem liegen in dieser Arbeit andere Einflussfaktoren stärker im Fokus.

Neben den drei großen Traditionen gibt es eine Reihe weiterer Ansätze, die sich nicht ohne weiteres einordnen lassen und geringere Resonanz in der Forschung erhalten haben. Bei der Betrachtung mancher dieser Ansätze zeigen sich Schwachstellen von Nachrichtenwerttheorie und Gatekeeping-Ansatz.

2.1.1 Nachrichtenwerttheorie

„Die Nachrichtenwerttheorie geht davon aus, dass die für die Publikation ausgewählten Ereignisse bestimmte Merkmale aufweisen, die ihren Nachrichtenwert und ihre Publikationswürdigkeit bestimmen. Diese Merkmale nennt man Nachrichtenfaktoren.“
(Maier et al. 2018b: 18)

Das heißt, die Grundannahme der Nachrichtenwerttheorie ist, dass man die Nachrichtenauswahl von Journalisten dadurch erklären kann, dass verschiedene Ereignisse aufgrund bestimmter Nachrichtenfaktoren mit unterschiedlicher Wahrscheinlichkeit von Journalisten ausgewählt werden. Damit fungieren die Nachrichtenfaktoren als eine Art Selektionskriterien der Journalisten (vgl. Eilders 2016: 431–432). Zwei Begriffe sind dabei zentral und müssen unterschieden werden, Nachrichtenfaktor und Nachrichtenwert. Nachrichtenfaktoren sind die Ereignismerkmale, die von Journalisten wahrgenommen werden

(vgl. Maier et al. 2018b: 18–19). Meistens werden sie als Liste von Eigenschaften dargestellt, beispielsweise Prominenz, geographische Nähe, Konflikt oder Überraschung. Da diese Merkmale von Journalisten als publikationswürdig betrachtet werden, verleihen sie in ihrer Kombination und unterschiedlichen Intensität einem Ereignis einen Nachrichtenwert. Ein höherer Nachrichtenwert wiederum sorgt für eine höhere Auswahlwahrscheinlichkeit (vgl. Eilders 2006: 6). In der grundlegendsten Beschreibung wird damit postuliert, dass Ereignisse umso eher publiziert, prominenter platziert und umfangreicher dargestellt werden, desto größer ihr Nachrichtenwert ist, der sich aus den Nachrichtenfaktoren ergibt (vgl. Kepplinger/Bastian 2000: 462).

Auf den ersten Blick erscheint das Konzept recht simpel, dabei aber plausibel. Das Konzept der Nachrichtenauswahl nach Nachrichtenfaktoren ist nach Meinung vieler Autoren in den Grundzügen auch empirisch bestätigt worden, wobei die genauen Listen der Nachrichtenfaktoren sich unterscheiden (vgl. Eilders 1996: 57; 2006: 9; Maier et al. 2018b: 83). Betrachtet man jedoch die verschiedenen Untersuchungen, die sich über Jahrzehnte mit der Nachrichtenwerttheorie auseinandergesetzt haben, zeigt sich zweierlei: erstens ist der Ansatz bei genauerer Betrachtung doch nicht wirklich trivial - wo genau soll man beispielsweise die Nachrichtenfaktoren verorten, sind es tatsächlich Eigenschaften der Ereignisse selbst, der journalistischen Wahrnehmung dieser Ereignisse oder aber der Medienbeiträge zu diesen Ereignissen? Als Folge gibt es zweitens sehr verschiedene Konzepte, die unter den Begriff Nachrichtenwerttheorie fallen und zum Teil unterschiedliche Vorstellungen von der Verortung und Bedeutung der Nachrichtenfaktoren für die Nachrichtenauswahl haben.

Erstmals verwendet wurde der Begriff news value 1922 von Walter Lippmann, damit war die Publikationswürdigkeit von Ereignissen gemeint. Dieser news value resultiert aus Kriterien, die späteren Nachrichtenfaktoren zum Teil sehr ähnlich sind. Das Grundkonzept news value wurde in der Folge vielfach in amerikanischen Journalismushandbüchern aufgegriffen, die Listen von Faktoren enthielten, die ein Ereignis publikationswürdig machen. Genannt wurden vor allem Direktheit, Nähe, Prominenz, Überraschung, Konflikt und Bedeutung (vgl. Staab 1990a: 40–41; 1990b: 424; Maier et al. 2018b: 29–31). In den 1960er Jahren begründeten dann verschiedene Arbeiten norwegischer Friedensforscher die europäische Forschungstradition der Nachrichtenwertforschung, die sich aus empirisch-sozialwissenschaftlicher Perspektive mit Nachrichtenfaktoren beschäftigte. Östgaard (1965) nannte drei der Nachrichtenentstehung inhärente Faktoren, die Nachrichten für den Leser interessant machen sollen: Vereinfachung, Identifikation und Sensationalismus (vgl. auch Eilders 1996: 21–22; Uhlemann 2012: 30–31).

Darauf aufbauend entwickelten Galtung und Ruge (1965) den ersten systematischeren und empirisch geprüften Nachrichtenfaktorenkatalog. Zudem verwendeten sie erstmals den Begriff „news factors“. Basierend auf einer Analyse der Berichterstattung über vier internationale Krisen und angelehnt an bestimmte wahrnehmungspsychologische Annahmen entwickelten sie eine Liste aus acht kulturunabhängigen und vier kulturabhängigen Faktoren, die die Nachrichtenauswahl beeinflussen sollen. Sie waren der Meinung, dass die Nachrichtenfaktoren allgemeine menschliche Wahrnehmungsprozesse widerspiegeln. Zudem stellten sie fünf Hypothesen zum Zusammenspiel und der Wirkung der Faktoren auf. Nach der Selektivitätshypothese gehen sie davon aus, dass ein Ereignis publikationswürdiger ist, je stärker ein Nachrichtenfaktor zutrifft, nach der Additivitätshypothese verstärken auch mehr Nachrichtenfaktoren die Publikationswürdigkeit. Nach der Komplementaritätshypothese können sich Nachrichtenfaktoren auch gegenseitig ergänzen. Zusätzlich werden nach der Verzerrungshypothese nachrichtenfaktorhaltige Aspekte in der Darstellung betont und nach der Replikationshypothese in jeder Stufe des Vermittlungsprozesses verstärkt (vgl. auch Uhlemann 2012: 31–34; Eilders 2016: 433; Maier et al. 2018b: 23, 36–38).

Während Galtung und Ruge ihre Kriterien und Hypothesen der Nachrichtenauswahl nur ansatzweise in ihrer Inhaltsanalyse testen konnten, war ihre Arbeit der Ausgangspunkt für eine Vielzahl von Studien zur Nachrichtenselektion. Diese Folgearbeiten lassen sich in vier Gruppen einteilen. Die erste Gruppe sind Inhaltsanalysen der medialen Berichterstattung ohne Vergleich mit anderen Datenquellen. Entweder haben Forscher Nachrichtenfaktoren in den untersuchten Medien beschrieben und von besonders häufigem oder intensivem Vorkommen auf die Bedeutung einzelner Faktoren geschlossen oder aber den mediale Beachtungsgrad von Ereignissen in Beziehung zu Nachrichtenfaktoren gesetzt. Eine zweite Gruppe bilden Vergleiche von Inhaltsanalysen der Medienberichterstattung mit Daten aus nicht-medialen Quellen, die dritte Gruppe sind Input-Output-Analysen, in denen Agenturmaterial mit den veröffentlichten Beiträgen verglichen wurde und die letzte Gruppe experimentelle Überprüfungen der Auswahlkriterien (vgl. Eilders 1996: 29–31). Das ideale Verfahren zur empirischen Überprüfung ist dabei eine Input-Output-Analyse, diese wurde jedoch wegen des verbundenen Aufwands so gut wie nie durchgeführt (vgl. Maier et al. 2018b: 22).

Gerade Studien aus der ersten Gruppe dominierten die Nachrichtenwertforschung. Sande (1971) prüfte für sechs der Nachrichtenfaktoren den Zusammenhang mit dem Beachtungsgrad in den Medien. Dafür entwickelte er einen „Prominence Index“, der die Bedeutung messen soll, die Journalisten einem Ereignis zuweisen. In dieser wie in anderen frühen Arbeiten wurden Nachrichtenfaktoren als Eigenschaften von Ereignissen definiert, die darüber entscheiden, ob

ein Ereignis zu einer Nachricht wird oder eine Nachricht zu einem Artikel (Uhlemann 2012: 36). Eine etwas andere Sichtweise präsentierte Schulz (1976), für ihn sind Nachrichtenfaktoren journalistische Hypothesen von Realität, keine objektiven Eigenschaften von Ereignissen. Demnach spiegeln Nachrichten wider, was Journalisten als Realität wahrnehmen, Nachrichtenfaktoren sind Merkmale von Beiträgen, nicht von Ereignissen. Dadurch ergibt sich nach Schulz eine neue Grundlage für die empirische Untersuchung von Nachrichtenfaktoren, deren Relevanz kann anhand des Beachtungsgrads bestimmt werden, der wiederum aus Platzierung und Umfang von Beiträgen gebildet wird. In einer umfassenden Untersuchung prüfte er einen überarbeiteten Katalog von 18 Nachrichtenfaktoren, die in vier verschiedenen Intensitätsstufen vorliegen können. Seine Ergebnisse deuten darauf hin, dass Nachrichtenfaktoren Platzierung und Umfang beeinflussten, aber in sehr unterschiedlichem Ausmaß und vor allem zusammengenommen. Zudem zeigten sich deutliche Unterschiede zwischen verschiedenen Medientiteln (vgl. auch Staab 1990b: 426–427; Eilders 1996: 34–36; Uhlemann 2012: 36–38).

Staab (1990a; 1990b) legte eine weitere Überarbeitung der Nachrichtenwerttheorie vor. Sie richtete sich vor allem gegen das, was er als „Kausalmodell“ der Nachrichtenfaktoren bezeichnete. Demnach wird in den meisten Analysen der Nachrichtenauswahl der Zusammenhang zwischen Nachrichtenfaktoren und Auswahlentscheidungen als Kausalzusammenhang dargestellt, Journalisten handeln in dieser Darstellung als passive Informationsvermittler, die reflexhaft auf Nachrichtenfaktoren reagieren. Das widerspricht für Staab den Ergebnissen der News Bias-Forschung, die zeigen, dass Journalisten durchaus auch bewusst Nachrichten auswählen, um bestimmte Ziele zu erreichen oder Ansichten zu unterstützen. Daher entwickelt er das Finalmodell der Nachrichtenfaktoren, in dem Selektionsentscheidungen zielgerichtete Handlungen von Journalisten sind und Nachrichtenfaktoren aktiv zugeschrieben werden, um die Beachtung von Beiträgen zu steuern oder Selektionsentscheidungen zu rechtfertigen (vgl. Staab 1990a: 93–100). Um das Finalmodell zu prüfen, führte Staab ebenfalls eine Inhaltsanalyse durch, in der er 22 verschiedene Nachrichtenfaktoren prüfte, die er in indizierbare und konsensbedingte Faktoren unterteilte. Indizierbare Faktoren wie Nähe können mehr oder weniger objektiv identifiziert werden, konsensbedingte dagegen eher subjektiv (vgl. ebd.: 120–122). Bei ihm hatte nur ein Teil der Nachrichtenfaktoren einen Einfluss auf Umfang und Platzierung von Meldungen, in verschiedenen Analysen konnte das Finalmodell die Berichterstattung mindestens genauso gut erklären wie das Kausalmodell (vgl. ebd.: 214).

Mit dem Zusammenhang zwischen Nachrichtenfaktoren und Nachrichtenwert beschäftigte sich Kepplinger in verschiedenen Untersuchungen und entwickelte das Zwei-Komponenten-Modell der Nachrichtenauswahl. Für ihn ergibt sich der Nachrichtenwert nicht einfach aus der Summe oder Kombination der Nachrichtenfaktoren, stattdessen spricht er vom „Nachrichtenwert der Nachrichtenfaktoren“. Demnach sind diese keine objektiven Auswahlkriterien, die für alle Journalisten gleich wichtig sind, sondern verschiedene Faktoren haben für unterschiedliche Redaktionen einen unterschiedlichen Nachrichtenwert. Auch er spricht sich damit aus einer anderen Perspektive dagegen aus, dass Journalisten einfach auf objektive Ereignismerkmale reagieren. Die Nachrichtenwürdigkeit eines Ereignisses ergibt sich stattdessen aus der Intensität der Nachrichtenfaktoren eines Ereignisses und dem relativen Einfluss dieser Faktoren auf die Nachrichtenauswahl, dem Nachrichtenwert (vgl. Kepplinger 1998: 19–22). Dieser kann beispielsweise in Befragungen von Journalisten erhoben werden, näherungsweise aber auch durch den Einfluss von Nachrichtenfaktoren auf Umfang und Platzierung. Kepplinger prüfte das Zwei-Komponenten-Modell in zwei Studien, in denen er versuchte, basierend auf Nachrichtenfaktoren und Nachrichtenwerten den Umfang und die Platzierung von Beiträgen zu prognostizieren. In einem Fall ermittelte er die Nachrichtenwerte aus Daten einer Inhaltsanalyse (Kepplinger/Bastian 2000), im zweiten Fall in einer Befragung von Studenten (Kepplinger/Ehmig 2006). Während im ersten Fall eine relativ moderate Korrelation zwischen berechnetem und gemessenem Umfang gefunden wurde, war diese in der zweiten Studie sehr hoch (0,8), dafür hatte die Untersuchung aber erhebliche methodische Schwächen.² Bestätigt werden konnte dafür, dass sich die Bedeutung von Nachrichtenfaktoren zwischen verschiedenen Medien sowie im Zeitverlauf unterscheidet.

Verschiedene Studien haben sich dann auch detaillierter mit der Gültigkeit der verschiedenen Nachrichtenfaktoren in unterschiedlichen Medientiteln, Ländern oder im Zeitverlauf beschäftigt. Ruhrmann und Kollegen (2003) untersuchten das Vorkommen von Nachrichtenfaktoren in Fernsehnachrichten und verglichen öffentlich-rechtliche und private Sender. Unter anderem fanden sie eine besondere Bedeutung des neuen Faktors Visualität. Der ursprüngliche Katalog von Galtung und Ruge wurde in zwei Studien 2001 und 2017 für Inhaltsanalysen der britischen Berichterstattung verwendet, dabei sollte untersucht werden, ob sich die Prozesse der Nachrichtenauswahl im Laufe der Zeit verändert haben. Schon 2001 kam ein Teil der Nachrichtenfaktoren sehr selten in der Berichterstattung vor, dafür fanden sich

² Beispielsweise wurde der Nachrichtenwert ermittelt, indem Studenten der Kommunikationswissenschaft befragt wurden, wie wichtig verschiedene Beiträge für drei Medientitel sein könnten. Bei der Prognose wiederum sollten wiederum Studenten die Publikationswürdigkeit von fünf Beiträgen in eine Rangfolge bringen, das heißt, es wurden nur Rangkorrelationen für fünf Datenpunkte berechnet.

Faktoren wie Unterhaltung oder Bezug zu etwas Positivem, die Galtung und Ruge nicht genannt hatten (vgl. Harcup/O'Neill 2001). 2017 wurden diese neuen Faktoren mit erhoben und es bestätigte sich wiederum die sich wandelnde Bedeutung von Nachrichtenfaktoren (vgl. dies. 2017). Ob Nachrichtenfaktoren auch in Medien nicht-europäischer und nicht-westlicher Länder eine Rolle spielen, untersuchte Schwarz (2006). Seinen Ergebnissen zufolge eignen sich Nachrichtenfaktoren dafür, die Nachrichtenentscheidungen von mexikanischen Qualitätszeitungen zu erklären.

Mit verschiedenen weiteren Faktoren, die den Einfluss von Nachrichtenfaktoren verändern könnten, hat sich Uhlemann (2012) beschäftigt. In einer der wenigen vorhandenen Input-Output-Analysen zur Nachrichtenwerttheorie analysierte sie die Auswahl von Agenturmeldungen in Lokalzeitungen. Als Einflussfaktoren untersuchte sie den Zeitpunkt, an dem eine Meldung in einer Redaktion eingeht, die Nachrichtenlage, also die Menge konkurrierender Meldungen und die Menge der Meldungen zu einem Thema, also die Betonung eines Themas durch die Nachrichtenagenturen. Für das Timing zeigte sich kein Einfluss, dafür gab es eine Korrelation zwischen der Häufigkeit von Agenturmeldungen und der Auswahl von Ereignissen, die über die Korrelation mit dem Nachrichtenwert hinausgeht. Gleiches gilt für zusätzliche Verarbeitungen durch die Nachrichtenagenturen. Der Einfluss der Konkurrenz, also der aktuellen Nachrichtenlage, auf den Beachtungsgrad dagegen war wiederum sehr gering. Dafür fand sie deutliche Hinweise darauf, dass bei der Wirkung des Nachrichtenwerts ein Schwellenwert vorliegt, ab dem dieser die Nachrichtenauswahl besonders beeinflusst (vgl. ebd.: 215–217).

Die Input-Output-Analyse von Uhlemann stellt eine Ausnahme in der Nachrichtenwertforschung dar, die von reinen Output-Analysen dominiert wird. Eine weitere Ausnahme ist das Experimental-Design, mit dem Helfer und van Aelst (2016) die Nachrichtenauswahl politischer Journalisten aus verschiedenen Ländern untersuchten. In einer Online-Umfrage wurden Journalisten verschiedene kurze fiktionale Pressemitteilungen präsentiert, in denen die Intensität von vier Nachrichtenfaktoren manipuliert wurde. Die Journalisten sollten dann bei sechs Pressemitteilungen angeben, ob sie sie für eine Nachricht verwenden würden. Hier zeigte sich ein Einfluss des politischen Systems eines Landes auf die Bedeutung von Nachrichtenfaktoren: je nach Land orientierten sich die Journalisten an unterschiedlichen Faktoren (vgl. ebd.: 67–70).

In den meisten Studien wurde die Liste der verwendeten Nachrichtenfaktoren nicht systematisch hergeleitet. Galtung und Ruge begründeten sie zwar mit einer Metapher zum

Abhören von Radiosignalen, trotzdem waren sie eigentlich nicht mehr als heuristische Ableitungen. Eilders versuchte 1996 erstmals, die Nachrichtenfaktoren theoretisch als kollektive menschliche Relevanzindikatoren zu begründen, die mit allgemeinen Wahrnehmungsprinzipien in Verbindung gebracht werden können. Die verschiedenen Faktoren leitet sie aus Überlegungen zur evolutionären Theorie, allgemein-menschlichen psychologischen Mechanismen und sozialer Relevanz her. Unter den ersten Aspekt fallen mehr oder weniger automatische Reaktionen auf bestimmte Stimuli, der zweite bezieht sich darauf, dass bekannten Objekten, Ereignissen und Aspekten mehr Aufmerksamkeit geschenkt wird und der letzte darauf, dass es bestimmte Überschneidungen in individuellen Relevanzstrukturen gibt, die sich aus gemeinsamer Sozialisation ergeben und notwendig für Interaktion und Kommunikation in der Gesellschaft sind (vgl. Eilders 1996: 96–105; 2006: 12–16).

Um zu prüfen, ob Nachrichtenfaktoren tatsächlich allgemeine Relevanzindikatoren sind, untersuchte Eilders (1996) die Bedeutung von Nachrichtenfaktoren für die Beitragsauswahl und Erinnerung von Beiträgen durch Rezipienten. Das heißt, sie übertrug das Nachrichtenwertkonzept auf die Medienwirkungsforschung. In ihrer Untersuchung verwendete sie ein Input-Output-Design, in dem Daten aus Inhaltsanalysen genutzter Medienbeiträge und aus Rezipientenbefragungen zu diesen Beiträgen gegenübergestellt wurden. Die Beitragsauswahl korrelierte mit dem Vorkommen und der Stärke der Nachrichtenfaktoren, allerdings kann dieser Zusammenhang auch dadurch entstehen, dass die Rezipienten vor allem Beiträge auswählten, die durch die Journalisten formal besonders hervorgehoben wurden, da diese auch mehr Nachrichtenfaktoren enthalten. Denn formale Variablen erklärten einen deutlich höheren Anteil der Varianz als die Nachrichtenfaktoren (vgl. ebd.: 206). Bei der Beitragsverarbeitung erinnerten die Rezipienten überproportional viele Inhalte, in denen keine Nachrichtenfaktoren vorkamen. Für die einzelnen Faktoren zeigten sich unterschiedliche Verarbeitungsmuster, die meisten zeigten jedoch einen Effekt, wenn auch erst ab einer relativ hohen Intensität der Nachrichtenfaktoren. Insgesamt konnte die Sichtweise von Nachrichtenfaktoren als allgemein-menschliche Selektionskriterien nicht wirklich bestätigt werden, höchstens für einzelne Faktoren (vgl. ebd.: 265–269).

Auch Fretwurst (2008) beschäftigte sich in seiner Arbeit zur Nachrichtenwerttheorie mit der Bedeutung von Nachrichtenfaktoren für die Rezeption und fand ähnlich wie Eilders Unterschiede zwischen der Selektion von Journalisten und Rezipienten sowie vor allem einen Einfluss der journalistischen Beachtung auf die Auswahl. Zusätzlich systematisierte er die Nachrichtenfaktoren als Faktorenbündel und unterscheidet zwischen Darstellungsfaktoren, die sich vor allem auf Eigenschaften der Medienbeiträge beziehen, Ereignisfaktoren, die sich auf

die Ereignisse beziehen, die den Beiträgen zugrunde liegen, und Indikatoren, beispielsweise Prominenz und Einfluss (vgl. ders. 2008: 112–139).

Die bisher vorgestellten Studien entstammen vor allem der deutschsprachigen Forschung zur Nachrichtenwerttheorie. International stark rezipiert wird dagegen vor allem das Nachrichtenwertkonzept von Shoemaker und Cohen (2006), die den Nachrichtenwert von Ereignissen aus verschiedenen Aspekten der beiden Dimensionen „Deviance“ und „Significance“ ableiten. Deviance bezieht sich auf unübliche Ereignisse, die in irgendeiner Form von Normen oder dem Normalzustand abweichen, Significance auf Auswirkungen auf verschiedene gesellschaftliche Bereiche (vgl. ebd.: 7–15). Aus dieser Konzeption ergibt sich direkt, dass in unterschiedlichen Kulturen andere spezifische Faktoren von Relevanz für die Nachrichtenauswahl sind, je nachdem, welche Ereignisse gesellschaftlich bedeutsam oder abweichend sind. Zuletzt haben Caple und Bednarek in den letzten Jahren ein Konzept von Nachrichtenfaktoren aus einer diskursiven Perspektive entwickelt. Demnach werden Nachrichtenfaktoren in Medienbeiträgen durch bestimmte sprachliche Ressourcen und Praktiken erzeugt. Das heißt, sie bestimmen nicht unbedingt die Auswahl von Ereignissen, sondern werden durch einen Diskurs als Aspekte von Akteuren, Ereignissen oder Themen konstruiert. Nachrichtenwürdigkeit ist dann eine Eigenschaft von Texten. Dabei unterscheiden sie „news values“ als nachrichtenwerte Aspekte von Objekten, die durch bestimmte gesellschaftliche Werte etabliert sind, von den eigentlichen Selektionskriterien (vgl. Caple/Bednarek 2016: 438–441).

2.1.2 Gatekeeper-Forschung

Während die Nachrichtenwertforschung versucht, die Nachrichtenauswahl anhand der Ereignisse zu erklären, aus denen ausgewählt wird, fokussiert die Gatekeeper-Forschung die Auswahl selbst und die Personen und Organisationen, die sie treffen. Gatekeeping ist “the process of selecting, writing, editing, positioning, scheduling, repeating and otherwise massaging information to become news.” (Shoemaker et al. 2009: 73) Gatekeeper sind ganz allgemein Instanzen, die kontrollieren, welche Informationen in welcher Form durch einen Kanal gelangen, Beispiele können neben Journalisten auch Verhaltenscodes oder Algorithmen sein (vgl. ebd.: 74). Die Gatekeeper-Forschung beschäftigt sich damit, welche Einflussfaktoren die Auswahl und Präsentation von Information bestimmen und wieso Journalisten als Gatekeeper so agieren, wie sie es tun (vgl. Engelmann 2016a: 11). In diesem Konzept kann man sich Journalisten als eine Art Filter vorstellen, die aus einer Vielzahl von Ereignissen diejenigen selektieren, die später den Weg in die Medien finden (vgl. Staab 1990b: 423). Gatekeeping

bezieht sich jedoch nicht nur auf die reine Auswahl von Informationen, sondern auch auf Entscheidungen über Platzierung, Umfang oder Menge von Meldungen (vgl. Shoemaker et al. 2009: 75). Es können drei verschiedene Formen von Gatekeeping unterschieden werden: selektives Gatekeeping bezeichnet die Auswahl von Informationen für die Berichterstattung; qualitatives Gatekeeping die Präsentation der ausgewählten Informationen und quantitatives Gatekeeping die Entscheidung, wie viel Raum den ausgewählten Informationen gegeben wird (vgl. Engelmann 2016a: 14–15). Zusätzlich kann zwischen primärem Gatekeeping durch die Journalisten und sekundärem Gatekeeping durch das Publikum unterschieden werden (vgl. ebd.).

Im Laufe der Zeit hat sich der Fokus und damit auch der Untersuchungsbereich der Gatekeeper-Forschung sehr verändert. Während die ersten Studien vor allem den individuellen Journalisten und dessen subjektive Einstellungen zur Erklärung der Nachrichtenauswahl untersuchten, wurde später auch zunehmend der Einfluss von Medienorganisationen berücksichtigt, sowohl vom direkten Redaktionsumfeld wie auch von Medien als wirtschaftliche Unternehmen. Journalisten wurden damit als Teil von Organisationen betrachtet, deren Hierarchien, Strukturen und Routinen die Nachrichtenproduktion beeinflussen. (vgl. Shoemaker et al. 2009: 76–78; Pürer 2014: 133). Die allgemeine Idee des Gatekeepers wurde 1947 von Kurt Lewin formuliert, der das Bild von verschiedenen Kanälen entwickelte, über die Informationen (oder in seiner ersten Untersuchung Lebensmittel) Menschen erreichen, diese Kanäle werden wiederum von Gatekeepern kontrolliert. White (1950) übertrug dieses Bild als erster auf den Journalismus. Redaktionen, Redakteure, Nachrichtenagenturen oder PR-Abteilungen können als Kanäle interpretiert werden, in denen es verschiedene Gates gibt, an denen über die Selektion und Verarbeitung von Informationen entschieden wird. Die Entscheidungen, die von Gatekeepern getroffen werden, werden durch Einflussfaktoren („forces“) mitbestimmt. Konkret untersuchte White in einer Art Beobachtungsstudie das Selektionsverhalten eines einzelnen Redakteurs, der dafür zuständig war, aus Agenturmeldungen diejenigen auszuwählen, über die berichtet wird. Ergebnis war, dass vor allem die Erfahrungen, Einstellungen und Erwartungen des Redakteurs die Auswahl prägten (vgl. White 1950: 383-384, 388-389; Engelmann 2016a: 26–27).

Die darauffolgende Weiterentwicklung des Gatekeeper-Ansatzes kann in drei Phasen unterteilt werden. Bis Ende der 1970er wurde der Ansatz in der Forschung etabliert, darauf folgte die Systematisierungsphase, in der die bis dahin identifizierten Einflussfaktoren in umfangreiche Modelle integriert wurden, diese Phase hält bis heute an. Die dritte Phase begann mit der zunehmenden Nutzung des Internets, vor diesem Hintergrund kam es zu einer Neuorientierung

der Gatekeeper-Forschung (vgl. Engelmann 2016a: 32). In der ersten Phase gab es wiederum drei verschiedene Ansätze. Der erste, zu dem auch die Gatekeeper-Studie von White gezählt werden kann, untersuchte vor allem einzelne Journalisten und individuelle psychologische Faktoren. Der zweite dagegen verstand Redakteure als Mitglieder von Medieninstitutionen, wichtige Einflüsse sind demnach institutionelle Rollen und äußere Faktoren wie Zeit- und Platzmangel oder Informationsquellen. Zuletzt betrachtet der kybernetische Ansatz sowohl die Prozesse der Nachrichtenverarbeitung innerhalb von Redaktionen sowie den Einfluss äußerer Variablen (vgl. ebd.: 32–34). Während in den verschiedenen Studien der ersten Phase bereits eine Vielzahl möglicher Wirkfaktoren auf den Prozess der Nachrichtenauswahl und -verarbeitung identifiziert wurden, standen sie meist unverbunden nebeneinander, es fehlte ein Rahmen zur Systematisierung. Daher wurden in der zweiten Phase verschiedene integrative Modelle entwickelt.

Meist wurden in diesen Modellen Klassifikationen verschiedener Ebenen oder Hierarchien von Einflussfaktoren entwickelt. In der deutschsprachigen Forschung findet vor allem das Zwiebelmodell von Weischenberg breite Anwendung. Er unterscheidet vier verschiedene hierarchisch angeordnete Kontexte journalistischer Arbeit. Im Inneren des Modells steht der Rollenkontext der individuellen Journalisten mit ihren Einstellungen und Merkmalen. Darüber liegen der Funktionskontext der Medienaussagen, der beispielsweise von Darstellungsformen oder Informationsquellen gebildet wird sowie der Strukturkontext der Medieninstitutionen. Zuletzt liegt auf der obersten (bzw. äußersten) Ebene der Normenkontext der Mediensysteme, der beispielsweise gesellschaftliche Bedingungen oder professionelle Standards beinhaltet. Die äußeren Ebenen schränken den individuellen Akteur in seiner subjektiven Auswahl von Medieninhalten ein, dadurch liegt der Fokus auf dem System Journalismus mit seinen Regeln und Normen (vgl. Hanitzsch 2009: 155; Engelmann 2016a: 18–19).

International wird meist das Hierarchy of Influences-Modell von Shoemaker und Reese (1996) verwendet. Sie unterscheiden fünf verschiedene Einflüssebenen. Auf der Ebene des „individual journalist“ liegen persönliche Merkmale und Erfahrungen der Journalisten genauso wie ihre Einstellungen, Überzeugungen und professionelle Rollenbilder. Die „routines of journalism“-Ebene beinhaltet neben Einschränkungen durch Zeit und Platz die Orientierung an anderen Medien und Kollegen, am Publikum sowie an den Quellen. Konkret können beispielsweise Nachrichtenfaktoren als Auswahlkriterien oder bestimmte Darstellungsformen auf dieser Ebene ins Modell integriert werden. Die dritte Ebene ist das „organizational level“, hier sind Organisationsziele, -ressourcen und -strukturen von Bedeutung. Auf der „extra media“-Ebene geht es vor allem um Einflüsse externer Akteure auf die Nachrichtenproduktion. Hier spielen

das Publikum, journalistische Quellen, insbesondere Pressearbeit, sowie die Märkte, auf denen Medienorganisationen agieren, eine Rolle. Zuletzt beinhaltet das „social system level“ die Makroebene vor allem gesellschaftlicher oder politischer Einflüsse (vgl. Shoemaker et al. 2009: 81–82; Hanitzsch 2009: 156; vgl. Engelmann 2016a: 60–77). Daneben gibt es weitere ähnliche Modelle, die ebenfalls verschiedene Ebenen identifizieren basieren und sich darin unterscheiden, wie viele Ebenen unterschieden werden und wo bestimmte Faktoren eingeordnet werden. Hanitzsch hat diese verschiedenen Modelle gegenübergestellt und fasst sie so zusammen, dass sie insgesamt auf eine fünfdimensionale Struktur hindeuten, aus Individuen, Medienroutinen, Organisationen, Medienstrukturen und Mediensystemen (vgl. Hanitzsch 2009: 156–157).

Die dritte Phase der Gatekeeping-Forschung setzte mit den technischen und dadurch ausgelösten institutionellen Veränderungen im Mediensystem ein, die in den letzten Jahren vor allem das Internet mit sich brachte. Die Aufmerksamkeit der Forschung verschob sich dabei vor allem auf die Online-Umgebung, in der Nachrichten konstruiert werden (vgl. Shoemaker et al. 2009: 78–79). Während manche Autoren sogar ein Ende des Gatekeeper-Ansatzes vermuten, da im Internet Auswahl-Gates umgangen werden können und weniger Selektionsdruck aufgrund von Zeit- und Platzmangel besteht, ist der Konsens eher, dass die bisherigen Gatekeeping-Modelle weiterhin gültig sind, aber angepasst werden müssen. Beispielsweise fungieren auch technische Strukturen wie Suchmaschinen als eine Art Gatekeeper. Zudem wird durch die Möglichkeit, die Nutzung von Beiträgen in Form von Aufrufstatistiken, Kommentaren oder Verbreitung in sozialen Netzwerken direkt und genau zu beobachten, der Einfluss des Publikums auf die Beitragsauswahl potenziell deutlich größer (vgl. Engelmann 2016a: 36–38). Dieser Einfluss ist auf verschiedenen Ebenen möglich. Indirekt könnten Rezipienten die Bedeutung von Nachrichtenfaktoren auf die Auswahl beeinflussen, sodass sie keine Rolle mehr für die Auswahl spielen und durch das Nutzerinteresse ersetzt werden oder aber andere Nachrichtenfaktoren an Bedeutung gewinnen. Und direkt würde der Einfluss, wenn sich Journalisten bei ihrer Auswahl rein daran orientieren, welche Beiträge oft aufgerufen wurden.

Untersucht wurde eine möglicherweise veränderte Bedeutung von Nachrichtenfaktoren unter anderem von Burggraaff und Trilling (2017) sowie Wendelin und Kollegen (2017). Erstere verglichen das Vorkommen bestimmter Nachrichtenfaktoren in über 700.000 Artikeln aus niederländischen Online- und Printmedien. Dafür entwickelten sie verschiedene elaborierte Verfahren zur automatisierten Codierung von Nachrichtenfaktoren. Sie untersuchten die These, dass sich die Routinen der Online-Nachrichtenproduktion von denen von Offline-Redaktionen

unterschieden, das sollte sich in den Nachrichtenfaktoren zeigen. Tatsächlich fanden sie verschiedene Unterschiede, allerdings nicht immer in der erwarteten Richtung. Beispielsweise waren in den Online-Artikeln weniger Nachrichten zu Berühmtheiten oder zur Unterhaltung (vgl. Burggraaff/Trilling 2017). In der zweiten Studie wurden dagegen Auswahlkriterien von Journalisten und Onlinepublikum verglichen. Dafür wurde die journalistische Auswahl über eine besonders prominente Platzierung und Auswahl des Publikums über Nutzerrankings operationalisiert. In 667 verschiedenen Artikeln wurden 19 Nachrichtenfaktoren (diese wurden nach dem Ansatz von Shoemaker und Cohen zu Deviance und Significance zusammengefasst) sowie das Thema codiert. Bezüglich der Nachrichtenfaktoren zeigten sich relativ wenige Unterschiede zwischen den Artikeln, die prominent platziert wurden und die weit oben in Nutzerrankings standen, größer waren die Unterschiede zwischen verschiedenen Medien. Bei den Themen dagegen zeigten sich unterschiedliche Präferenzen. Während die Journalisten politische Themen bevorzugten, wählte das Publikum vor allem Service-Themen aus. Insgesamt waren vor allem die Qualitätszeitungen in ihrer Platzierung am unabhängigsten von der Auswahl des Online-Publikums (vgl. Wendelin et al. 2017).

Andere Studien beschäftigten sich noch direkter mit dem Einfluss des Publikums auf Nachrichtenplatzierung oder Auswahl. Lee und Kollegen konnten einen zeitverzögerten Effekt der Aufrufzahlen von Artikeln auf deren Platzierung zeigen, aber keinen Effekt der Platzierung auf die Aufrufzahlen. Da der Effekt der Aufrufzahlen negativ war, also Artikel mit hohen Aufrufzahlen in der Folge weniger prominent platziert wurden, können die Ergebnisse nicht wirklich einen Einfluss des Publikums auf die Auswahl oder Platzierung zeigen (vgl. Lee et al. 2014). In einem aufwendigen Vergleich der Themen, die in verschiedenen Online-Medien sowie sozialen Netzwerken und Blogs diskutiert werden, konnten Russell Neuman und Kollegen dagegen einen umgekehrten Agenda Setting-Effekt beobachten. Die Medien gaben nicht immer, wie es die Agenda Setting-Theorie klassisch behauptet (vgl. bspw. Coleman et al. 2009), die Themenagenda des Publikums vor, es war in vielen Fällen eine umgekehrte Beeinflussung beobachtbar. Um das zu zeigen, wurden Beiträge von 100 Millionen Twitternutzern, aus 160 Millionen Blogs und 300.000 Foren sowie aus verschiedenen Medien automatisiert nach dem Vorkommen von 29 verschiedenen politischen Themen durchsucht. Für 18 der 29 Themen verursachte Aufmerksamkeit in den sozialen Medien im Zeitverlauf höhere Aufmerksamkeit in den Medien (vgl. Russell Neuman et al. 2014). Auch Welbers und Kollegen kamen in ihrer Untersuchung von fünf niederländischen Online- und fünf Printmedien zu dem Ergebnis, dass Nutzerstatistiken zumindest bei 7 der zehn Medien einen Einfluss auf die Themenwahl hatten. Dafür identifizierten sie bestimmte Storylines in der Berichterstattung und

bestimmten mit cross-lagged-Regressionen den Einfluss der Aufrufzahlen auf die Häufigkeit der Artikel zu diesen Storylines (vgl. Welbers et al. 2016).

Neben solchen durch automatisierte Verfahren gestützten Inhaltsanalysen konnte auch in qualitativen Untersuchungen und quantitativen Befragungen der Einfluss des Publikums über die Erhebung von Nutzerzahlen auf die journalistische Arbeit gezeigt werden. Während Journalisten fast immer angaben, Aufrufzahlen regelmäßig zu beobachten, wurde ihr direkter Einfluss auf die Nachrichtenauswahl meistens verneint, höchstens ein Einfluss auf die Platzierung von Beiträgen wurde eingeräumt. Allerdings unterschieden sich die Ergebnisse auch zwischen unterschiedlichen Medien (vgl. Anderson 2011; Vu 2014; Welbers et al. 2016).

Wie die kurze Vorstellung der Arbeiten aus den verschiedenen Phasen der Gatekeeper-Forschung schon im Ansatz zeigen konnte, ist das Spektrum verwendbarer Methoden hier deutlich größer als bei der Nachrichtenwertforschung, in der Output-Analysen dominieren. Hanitzsch unterscheidet zwei Untersuchungsanordnungen, entweder kann der direkte Einfluss von verschiedenen Faktoren auf Inhalte, Redaktionsprozesse oder Orientierungen untersucht werden, wofür Inhalte und Einflussfaktoren erhoben werden müssen. Die Bedeutung der Faktoren kann dann durch Regressionsanalysen bestimmt werden. Andererseits können Journalisten danach befragt werden, wie sie die Bedeutung verschiedener Einflüsse wahrnehmen. Beide Vorgehensweisen sind komplementär (vgl. Hanitzsch 2009: 153–154).

Geht man weiter ins Detail, können Beobachtungsstudien, Befragungen, quasi-experimentelle Designs sowie Inhaltsanalysen als Output- oder Input-Output-Analysen unterschieden werden. Beobachtungen untersuchen tatsächliche Selektionsentscheidungen, können aber für sich genommen nicht die Bedeutung der Handlungen oder die Absichten dahinter bestimmen. Experimentelle Designs haben eine ähnliche Logik wie Input-Output-Analysen, als Input fungieren kurze Texte, aus denen Journalisten auswählen können. Problematisch ist hier vor allem die künstliche Auswahlssituation. Nur mit Input-Output-Analysen kann man „die Gatekeeping-Prozesslogik tatsächlich und vollständig abbilden, da man Ausgangsinformationen und Selektionsergebnisse vergleichen kann, statt nur aus den publizierten Medieninhalten auf die Auswahlprozesse zu schließen.“ (Engelmann 2016a: 51) Der Nachteil ist dafür, dass sie keine Erkenntnisse über die Prozesse zwischen Input und Output liefern können (vgl. ebd.: 40–51). Mit den unterschiedlichen Designs können verschiedene Aspekte oder Ebenen von Gatekeeping untersucht werden. Daher bieten sich Mehr-Methoden-Designs an, vor allem Kombinationen von Beobachtung und Befragung oder auch aus Beobachtung und Input-Output-Analyse (vgl. ebd.: 51–55).

2.1.3 Weitere Ansätze

Neben diesen beiden dominierenden Konzepten (und der News Bias-Forschung) wurden im Laufe der Zeit viele weitere Ansätze entwickelt, die die Nachrichtenauswahl von Journalisten erklären sollen. Meist reagierten diese neuen Ansätze auf bestimmte Schwachstellen, die in bestehenden Theorien gefunden wurden, sehr häufig finden sich zudem Übertragungen von Konzepten aus anderen wissenschaftlichen Disziplinen und Gebieten, vor allem aus der Psychologie und Soziologie, aber auch aus den Wirtschaftswissenschaften oder der Chaostheorie.

Diese verwendete Frerichs als Inspiration für sein Modell der Nachrichtenauswahl. Sein Grundgedanke ist, dass die Auswahl und Gestaltung von Nachrichten keine linearen Vorgänge sind, sondern sich besser als dynamische Ordnungsbildung und Selbstorganisation verstehen lassen. Aus chaostheoretischer Sicht ist der Nachrichtenjournalismus nichtlinear und dynamisch, kleine Veränderungen der Rahmenbedingung der Nachrichtenauswahl, also beispielsweise der politischen Einstellung von Journalisten oder der redaktionellen Verhältnisse, führen zu großen Veränderungen der Kommunikationsprodukte. Das liegt auch daran, dass der Nachrichtenfluss von den Quellen bis zu den Nutzern durch mehrstufige Auswahlprozesse gekennzeichnet ist, die zu einer Unschärfe der Nachrichtenauswahl führen. Zudem beobachtete er bestimmte Phänomene aus der Chaostheorie im Journalismus, beispielsweise fraktale Strukturen von Nachrichten oder einen Wechsel von Ordnung und Chaos im Nachrichtenfluss (vgl. Frerichs 2016).

Fengler und Ruß-Mohl übertragen dagegen Konzepte aus der Wirtschaftstheorie auf den Journalismus. Ihrer Ansicht nach kann eine ökonomische Theorie des Journalismus, die vor allem den Rational Choice-Ansatz miteinbezieht, zu neuen Erkenntnissen über journalistisches Handeln führen. In dieser Perspektive können Interaktionen von Journalisten untereinander oder mit anderen Akteuren als Transaktion von Akteuren mit eingeschränkter Rationalität verstanden werden. Das heißt, Journalisten verfolgen in ihrem Handeln bestimmte Eigeninteressen und versuchen, diese möglichst umfänglich zu verfolgen, wenn sie Nachrichten auswählen, produzieren oder verbreiten. Das von Journalisten verfolgte Interesse ist meist öffentliche Aufmerksamkeit. Markttransaktionen finden beispielsweise zwischen Journalisten und ihren Quellen oder PR-Mitarbeitern statt. Der Journalist ist dabei auf der Suche nach exklusiven Informationen, er zahlt für die Informationen mit öffentlicher Aufmerksamkeit, die sein Medium bietet. Als ein Indiz für die Validität ihres Konzepts führen die Autoren verschiedene externe Effekte an, die durch das ökonomische Handeln und den Wettbewerb im

Journalismus entstehen. Ein Beispiel ist Herdenjournalismus, der auftritt, wenn Breaking News direkt nach der Veröffentlichung von anderen Medien aufgegriffen werden. Insgesamt erklären in diesem Konzept nicht nur Nachrichtenfaktoren, sondern auch persönliche Interessen der verschiedenen an der Nachrichtenproduktion teilnehmenden Agenten die spezifischen Ergebnisse, Verzerrungen, Über- und Unterrepräsentation von Ereignissen in den Nachrichten (vgl. Fengler/Ruß-Mohl 2008).

Die spezifischen Ziele und Wahrnehmungen von Journalisten als Akteure werden auch in Ansätzen betont, die sich an der Psychologie, insbesondere der kognitiven und Sozialpsychologie orientieren. Zhong und Newhagen benennen die Denkprozesse von Journalisten als wichtigste Einflussfaktoren für die Auswahl und Verarbeitung von Nachrichten, diesen sprechen sie mehr Bedeutung zu als dem, was tatsächlich in der Welt passiert. Um diese Denkprozesse zu untersuchen, verwenden sie einen Ansatz aus der Wahrnehmungspsychologie. Nachrichtenentscheidungen basieren demnach auf Nachrichtenelementen, das sind unverarbeitete Nachrichtenmaterialien oder Medienstimuli. Wenn diese allgemein akzeptierte Nachrichtenfaktoren beinhalten, werden sie als Nachrichtenelemente auf dem „cognitive system level“ bezeichnet. Reflektieren oder stimulieren sie kulturelle Werte der Journalisten, werden sie als Nachrichtenelemente auf dem „cultural system level“ bezeichnet. Zuletzt sind Nachrichtenelemente auf dem „rational system level“ solche, die bestimmte ideologische Werte beinhalten und Ansichten der Journalisten zur Gesellschaft und bestimmten Normen reflektieren (vgl. Zhong/Newhagen 2009: 590–591). Diese drei Systemebenen können hierarchisch angeordnet werden, kognitive Elemente sind die Grundlage der Nachrichtenentscheidungen, die beiden anderen Ebenen liegen darüber und werden nur relevant für eine Entscheidung, wenn vorhandene Elemente kulturelle oder gesellschaftliche Werte berühren. Um die Validität des Modells zu testen, wurde ein Experiment mit chinesischen und US-amerikanischen Journalisten durchgeführt, die Entscheidungen zu Nachrichtenelementen auf den verschiedenen Ebenen treffen sollten. Dabei wurde erwartet, dass sich bei den oberen beiden Ebenen deutliche Unterschiede zeigen, da die Journalisten ja sehr unterschiedliche kulturelle und ideologische Hintergründe haben. Jedoch wurden nur sehr geringe Unterschiede zwischen den Journalistengruppen gefunden, was die Autoren als Beleg für eine zunehmende Globalisierung journalistischer Standards interpretierten (vgl. ebd.: 599–603).

Auch Donsbach entwickelte ein Modell der Nachrichtenauswahl aus psychologischer Sicht. Er charakterisiert darin die Nachrichtenauswahl als komplexe Entscheidungssituation, in der Journalisten unter extremen Zeitdruck und Wettbewerb Entscheidungen treffen müssen, für die

sie selten objektive Kriterien haben. Damit befinden sie sich in unsicheren oder unbestimmten Situationen, in denen vor allem zwei allgemeine Bedürfnisse der Journalisten die Nachrichtenauswahl erklären können: das Bedürfnis, eigene Wahrnehmungen sozial zu validieren und das Bedürfnis, existierende Voreinstellungen zu bewahren. In unbestimmten Situationen kann eine Validierung von eigenen Ansichten und Erfahrungen vor allem dadurch erreicht werden, dass die wahrgenommene Realität mit anderen geteilt wird. Für Journalisten sind als Referenzgruppe dabei vor allem andere Journalisten relevant. Damit kann beispielsweise erklärt werden, wie es zu konsonanter (also übereinstimmender) Berichterstattung oder dem Einfluss von Schlüsselereignissen auf die Menge der Berichterstattung kommen kann. Nach Ereignissen mit hoher Sichtbarkeit wie Katastrophen nehmen Berichte über ähnliche Ereignisse signifikant zu. Neben anderen Journalisten kann auch der Einfluss organisationaler Rollen und von Public Relations als Teil des sozialpsychologischen Prozesses verstanden werden, in dem Nachrichten in unbestimmten Situationen definiert werden (vgl. Donsbach 2004: 136–145).

Der zweite wichtige psychologische Faktor sind bestehende Einstellungen und vorhandenes Wissen. Journalisten können aus einer potenziell unbegrenzten Menge von möglichen Nachrichten auswählen und müssen daher deutlich selektivere Entscheidungen treffen als beispielsweise ihr Publikum. Diese Selektivität kann unterteilt werden in selektive Aufmerksamkeit, die bestimmt, welchen Stimuli überhaupt Aufmerksamkeit geschenkt wird, selektive Wahrnehmung, die bestimmt, wie Informationen verarbeitet und gespeichert werden und selektive Erinnerung, die bestimmt, was langfristig ins Gedächtnis übernommen wird. In allen drei Stufen wird ein Einfluss davon ausgeübt, welches Wissen und welche Ansichten vor der Wahrnehmung vorhanden waren. Insgesamt ergibt sich aus beiden Bausteinen ein Modell der journalistischen Wahrnehmung aus drei Komponenten: Charakteristiken des Objekts, Voreinstellungen des Individuums und Kommunikation mit anderen (vgl. ebd.: 145–148).

Diese vorgestellten Ansätze stellen natürlich nur einen Bruchteil der Konzepte dar, mit denen versucht wurde, die Nachrichtenauswahl von Journalisten zu erklären. Sie zeigen aber bereits zwei Aspekte auf, die in den großen Forschungstraditionen nur am Rande vorkommen. Einerseits stellen sie Journalisten als Akteure in den Fokus, die nicht nur passiv auf bestimmte Einflüsse oder Nachrichtenfaktoren reagieren, sondern mit bestimmten Intentionen handeln, dabei aber gewissen Einschränkungen unterliegen, die vor allem mit der allgemeinen menschlichen Psychologie zusammenhängen. Die Gatekeeper-Forschung betrachtet zwar Journalisten als Individuen, vernachlässigt jedoch die Frage, wie die verschiedenen auf sie wirkenden Einflüsse dann konkret in Auswahlentscheidungen verarbeitet werden. Auswahl

wird immer schon auf einer aggregierten Ebene betrachtet. Und die Nachrichtenwerttheorie beschäftigt sich abgesehen vom Finalmodell auch nur am Rande mit den Einstellungen der Journalisten. Andererseits verweisen sie darauf, dass Nachrichtenauswahl gleichzeitig auch das eher zufällige Ergebnis eines komplexen Gefüges von Einflüssen sein kann, seien es die externen Effekte in der Wirtschaftstheorie oder die nichtlinearen Prozesse der Chaostheorie. Nach der Darstellung des Forschungsstands zur Nachrichtenauswahl im Wissenschaftsjournalismus werden diese und weitere Gedanken wieder aufgegriffen, um aufzuzeigen, welche Vorteile ein neues Modell der Nachrichtenauswahl im Wissenschaftsjournalismus haben kann, das Journalisten als Akteure in unsicheren und komplexen Situationen wahrnimmt.

2.2 Forschung zur Nachrichtenauswahl im Wissenschaftsjournalismus

Die verschiedenen Theorien und Modelle zur Erklärung der Nachrichtenauswahl im Journalismus haben eine Gemeinsamkeit: sie wurden vor allem anhand von Untersuchungen des politischen Journalismus entwickelt und überprüft. Während ihre Validität für diesen Bereich zumindest in Teilen gesichert ist, ist bei anderen, speziellen Bereichen des Journalismus nicht klar, inwiefern die allgemeinen Ansätze hier Geltung haben. Ein Beispiel für einen solchen Bereich ist der Wissenschaftsjournalismus. Macht beispielsweise ein Ansatz wie der des News Bias Sinn für eine Form des Journalismus, in der politische Einstellungen höchstens in seltenen Kontroversen einen signifikanten Einfluss auf die Berichterstattung haben?³ Oder wie steht es um Nachrichtenfaktoren wie geographische Nähe oder Prominenz, spielen sie im Wissenschaftsjournalismus auch eine wichtige Rolle? Manche Forscher haben versucht, Nachrichtenwerttheorie (bspw. Badenschier/Wormer 2012) oder Gatekeeping-Ansatz (bspw. Guenther/Ruhrmann 2013) auf den Wissenschaftsjournalismus zu übertragen und anzupassen, sie stellen damit jedoch eine Ausnahme dar. Es fehlt insgesamt an systematischen Analysen der Selektionskriterien von Wissenschaftsjournalisten, ob basierend auf bekannten Modellen und Theorien oder auf neu entwickelten. Schäfer (2011: 403) konstatierte in einem Review der Forschung zur Kommunikation über Wissenschaft in Massenmedien, dass die Kriterien, nach denen Wissenschaftsjournalisten Themen auswählen, bisher eher selten erforscht wurden.

³ Ein vergleichbarer Einfluss auf die Berichterstattung könnte allgemein das Vertrauen in oder die Einstellung gegenüber der Wissenschaft oder bestimmter Forschungsbereiche sein. Die meisten Untersuchungen finden hier ein relativ hohes Vertrauen bis hin zu einer Unterstützung der Wissenschaft, was dazu führen könnte, dass die Berichterstattung unkritischer ist, als es vielleicht angebracht wäre (vgl. Schäfer 2011: 403). Oder aber Themenpräferenzen der Journalisten, die dazu führen, dass bestimmte Disziplinen überrepräsentiert sind.

Die existierenden Studien haben sich meist auf verschiedene einzelne Aspekte der Nachrichtenauswahl konzentriert, zeigen allerdings zusammengenommen, dass es auch hier potenziell eine Vielzahl von Einflussfaktoren gibt, die darüber entscheiden, über welche Themen oder Ereignisse berichtet wird. Allan (2009: 153) spricht beispielsweise von einem „complex array of institutional imperatives“, das bestimmt, welche wissenschaftlichen Entwicklungen den Weg in die Medien finden, Hodgetts und Kollegen (2008: 61) nennen als mögliche Einflüsse auf Auswahlentscheidungen kulturelle Bedeutungen, professionelle Normen, redaktionelle Prozesse, Wahrnehmungen des Publikums und Vorstellungen des öffentlichen Interesses. Im Folgenden sollen die bisher eher verstreut vorliegenden Ergebnisse zur Nachrichtenauswahl im Wissenschaftsjournalismus nach verschiedenen Ebenen und Einflussquellen geordnet vorgestellt werden. Dabei liegt der Fokus auf den Faktoren, die für den in dieser Untersuchung betrachteten Typ des Wissenschaftsjournalismus relevant sind, der aktuell über wissenschaftsgenerierte Ereignisse berichtet.

2.2.1 Inhaltliche Faktoren

Ein erster wichtiger Einflussfaktor auf die Auswahlentscheidungen sind die Merkmale der Ereignisse, über die Journalisten berichten, das heißt konkret der Forschung, die den Anlass für die Berichterstattung liefert. Auf einer allgemeineren Ebene haben verschiedene Studien darauf verwiesen, dass über unterschiedliche Medientypen hinweg bestimmte Wissenschaftsbereiche bevorzugt werden (vgl. Lehmkuhl 2008: 304–305; Allan 2009: 153; Badenschier/Wormer 2012: 61; Blöbaum 2017: 231). Das heißt, Forschungsergebnisse einzelner Disziplinen werden von den Journalisten besonders häufig ausgewählt. Welche Disziplinen das sind, unterscheidet sich jedoch in verschiedenen Untersuchungen. Eine große Anzahl von Inhaltsanalysen stellten eine Dominanz von Medizin, Biologie und Verhaltenswissenschaften fest, in anderen Worten den sogenannten Life Sciences (bspw. Bucchi/Mazzolini 2003: 10, 16-17; Clark/Illman 2006: 505; Elmer et al. 2008: 885). In anderen Studien wiederum erzielten die Sozialwissenschaften sehr hohe Werte (bspw. Hijmans et al. 2003: 161; Summ/Volpers 2016: 784). Eine Ursache für die unterschiedlichen Ergebnisse ist das Verständnis von Wissenschaftsjournalismus, das in den Studien jeweils verwendet wird (vgl. Lehmkuhl 2008: 304–305; Blöbaum 2017: 231). Bei einem engen Verständnis von Wissenschaftsjournalismus, das vor allem die Berichterstattung in spezialisierten Wissenschaftsressorts oder –sendungen umfasst, dominieren die Life Sciences und Naturwissenschaften, bei einem weiteren Verständnis dagegen nehmen auch die Sozial- und Geisteswissenschaften viel Raum ein (vgl. Blöbaum 2017: 231). Das liegt vor allem daran, dass sich die Berichterstattung über diese Disziplinen vielfach außerhalb der

spezialisierten Ressorts findet, über Sozialwissenschaften wird beispielsweise auch im Politikteil oder Feuilleton berichtet (vgl. Summ/Volpers 2016: 783).

Gerade in der Berichterstattung über medizinische Forschung findet sich ein weiteres eher abstraktes Auswahlkriterium, es werden bestimmte Frames oder „media templates“ (Hodgetts et al. 2008: 50) verwendet, um vorgestellte Ergebnisse für den Leser relevant erscheinen zu lassen. Demnach werden eher Ergebnisse ausgewählt, die zu diesen Frames oder Templates passen. Ein Beispiel ist der Lifestyle-Frame, in dem die persönliche Verantwortung für die eigene Gesundheit betont wird (vgl. Hodgetts et al. 2008: 50; Amend/Secko 2012: 262). Eine Untersuchung konnte zeigen, dass in der Berichterstattung über 12 verschiedene Krankheiten solche Studien mit höherer Wahrscheinlichkeit ausgewählt wurden und mehr mediale Aufmerksamkeit erhielten, deren Ergebnisse in den Lifestyle-Frame passen (vgl. Dumas-Mallet et al. 2017: 5–6). Zu ähnlichen Ergebnissen kam eine Input-Output-Analyse, die Zeitungsartikel zu Studien in zwei medizinischen Zeitschriften untersuchte (vgl. Stryker 2002: 524). Auch in anderen Disziplinen ist es wahrscheinlich, dass bevorzugt Ergebnisse ausgewählt werden, die in bestimmte Muster oder Templates der Berichterstattung passen (vgl. auch Guenther/Ruhrmann 2013: 10; Rosen et al. 2016: 343).

Auch für den Wissenschaftsjournalismus haben verschiedene Autoren die Bedeutung bestimmter Nachrichtenfaktoren als Auswahlkriterien behauptet. Vielfach wird davon ausgegangen, dass hier die gleichen Nachrichtenfaktoren von Bedeutung sind wie im allgemeinen Journalismus (bspw. Amend/Secko 2012: 262; Blöbaum 2017: 230). Eine erste Gruppe von Untersuchungen hat den Einfluss einzelner Nachrichtenfaktoren genauer untersucht, ohne diese in eine größere Systematik einzuordnen. Genannt wurde besonders häufig „Human Interest“⁴ als wichtiges Auswahlkriterium (vgl. Hansen 1994: 114–115; Wilkie 1996: 1309; Allan 2009: 154). Gemeint ist damit, dass die Relevanz der Ergebnisse für das tägliche Leben der Leser dargestellt werden kann oder die Ergebnisse mit einem „human angle“ dargestellt werden können.⁵ Journalisten betonten die Wichtigkeit dieses Faktors in verschiedenen qualitativen Befragungen (vgl. Hansen 1994: 114–115; Rosen et al. 2016: 343) und in mindestens einer Input-Output-Analyse wurde ein Einfluss gefunden. Kiernan (2003) untersuchte die Berichterstattung von 25 US-Zeitungen sowie der Nachrichtenagentur Associated Press über Studien in vier großen wissenschaftlichen Zeitschriften (*JAMA*, *Nature*,

⁴ Entnommen wurde der Faktor dem Nachrichtenfaktorenkatalog von Shoemaker und Reese (vgl. Kiernan 2003: 908), in der deutschsprachigen Nachrichtenwertforschung gibt es keine direkte Entsprechung.

⁵ Das könnte auch eine Erklärung für den hohen Anteil der Medizin in der Berichterstattung sein (vgl. Hansen 1994: 115).

NEJM, und *Science*). Sowohl in den Pressemitteilungen der Zeitschriften wie auch den journalistischen Beiträgen codierte er Nachrichtenfaktoren, der häufigste war jeweils Human Interest (vgl. ders. 2003: 911–913).

Relativ viel Aufmerksamkeit wurde auch dem Faktor geographische Nähe geschenkt, allerdings mit widersprüchlichen Ergebnissen. Einerseits gaben Journalisten in Befragungen oft an, bevorzugt über Ergebnisse zu berichten, die von nationalen Forschern kommen (vgl. Hansen 1994: 116; Entwistle 1995: 921; Rosen et al. 2016: 344). Andererseits konnte in verschiedenen Inhaltsanalysen kein signifikanter Einfluss des Herkunftslands bzw. der nationalen Verortung der Forscher gezeigt werden (vgl. Einsiedel 1992: 95; Bucchi/Mazzolini 2003: 12; Kiernan 2003: 916). Für die Berichterstattung über biomedizinische Studien in Zeitungen aus acht verschiedenen Ländern fanden Dumas-Mallet et al. (2019) dagegen eine bevorzugte Auswahl von Forschungsergebnissen, an denen nationale Forscher beteiligt waren. Dafür setzten sie die Zahl der Zeitungsberichte ins Verhältnis zur Zahl der Studien, an denen Forscher der verschiedenen Länder beteiligt waren. Das heißt, in den absoluten Zahlen zeigte sich nicht unbedingt eine Präferenz, diese wurde jedoch deutlich, wenn berücksichtigt wird, wie groß das Angebot ist, aus dem in den verschiedenen Ländern jeweils gewählt werden kann. Demnach führen die Autoren die Ergebnisse der anderen Analysen, die keinen Einfluss finden konnten, auch darauf zurück, dass nicht beachtet wurde, dass die Wissenschaftssysteme verschiedener Länder unterschiedlich produktiv sind (vgl. ebd.: 197).

Neben diesen beiden häufig genannten und einigermaßen gut untersuchten Faktoren werden in der Forschungsliteratur verschiedene weitere inhaltliche Merkmale genannt, die die Auswahl beeinflussen könnten. Diese Faktoren sind Kontroverse (vgl. Entwistle 1995: 921; Kiernan 2003: 911–913; Hodgetts et al. 2008: 49; Rosen et al. 2016: 343), Bezug zu Anwendungen (vgl. Guenther/Ruhrmann 2013: 10), Überraschung (vgl. Rosen et al. 2016: 343; Blöbaum 2017: 230), Kuriosität (vgl. Entwistle 1995: 921; Kiernan 2003: 911–913; Rosen et al. 2016: 343) und Prominenz (vgl. Kiernan 2003: 911–913). Beim Einfluss von Aktualität oder auch Neuigkeit gibt es wiederum widersprüchliche Ergebnisse. Je nach Art des Wissenschaftsjournalismus oder der Medientypen, die betrachtet werden, zeigt sich eine unterschiedliche Bedeutung der beiden Faktoren (vgl. bspw. Rosen et al. 2016: 343).

Eine erste systematische Untersuchung zur Validität der Nachrichtenwerttheorie im Wissenschaftsjournalismus haben Badenschier und Wormer durchgeführt. Dafür entwickelten sie zuerst eine Art heuristisches Modell der Nachrichtenauswahl, das aus „Importance factor“, „Surprise factor“ und „Usability factor“ (Badenschier/Wormer 2012: 66) gebildet wird. Darauf

aufbauend prüften sie für eine Liste von Nachrichtenfaktoren, die aus einem Literaturreview insbesondere der deutschsprachigen Nachrichtenwertforschung abgeleitet wurde, inwiefern die Faktoren für den Wissenschaftsjournalismus Relevanz haben könnten und welche Faktoren verändert oder ergänzt werden müssten. Als neue Faktoren wurden beispielsweise wissenschaftliche Nähe, Überraschungsaspekt, Komposition der Berichterstattung in Bezug auf wissenschaftliche Gebiete oder Einfluss des Experten auf die Publikation eingeführt. Insgesamt ergab sich ein Katalog aus 29 verschiedenen Nachrichtenfaktoren (vgl. ebd.: 71–72). Dieser wurde im Rahmen einer quantitativen Inhaltsanalyse und durch Leitfaden-Interviews mit Wissenschaftsjournalisten geprüft. In der Inhaltsanalyse fanden sich die Faktoren Überraschung, Erstaunen und Komposition besonders häufig in den Artikeln. Die befragten Redakteure schätzten zusätzlich die Bedeutung von Reichweite sowie Relevanz für die Rezipienten oder die Gesellschaft insgesamt als sehr wichtig ein, das wurde durch die Inhaltsanalyse jedoch nicht bestätigt. Insgesamt reduzierten die Autoren den Katalog basierend auf der Überprüfung auf 14 Faktoren (vgl. ebd.: 74–78, s. Abbildung 1). Dieser adaptierte Katalog lässt sich nach Meinung von Badenschier und Wormer vor allem auf die Auswahl klassischer Wissenschaftsartikel anwenden, die durch wissenschaftliche Publikationen oder Konferenzen ausgelöst wurden (vgl. ebd.: 80).

14 News factors with the highest impact on the selection process of science news (alphabetic order)	
Astonishment	Political relevance
Composition	Range (number of affected people)
Controversy	Reference to elite persons
Economic relevance	Relevance to recipients/society
Graphical material	Scientific relevance*
Intention	Actuality (Trigger)
Personalisation	Unexpectedness

Abbildung 1: Nachrichtenfaktoren im Wissenschaftsjournalismus nach Badenschier und Wormer. Quelle: Badenschier/Wormer, 2012: 79.

Einen ähnlichen methodischen Ansatz für die Prüfung einer etwas andere Theorie hat Molek-Kozakowska angewendet. Sie hat den newsworthiness-framework von Bednarek und Caple (s. Kapitel 2.1.1) auf den Wissenschaftsjournalismus übertragen und dabei explorativ versucht, die sprachlichen Ressourcen zu finden, mit denen speziell im Wissenschaftsjournalismus Nachrichtenwert konstruiert wird. Einerseits prüfte sie dafür die Nachrichtenfaktoren Neuheit, Superlative, Aktualität, Einfluss, Negativität/Positivität sowie Prominenz und andererseits spezielle stilistische Aspekte, die besonders im Wissenschaftsjournalismus Anwendung finden

(direkte Ansprache bzw. Gesprächsstil und Rationalisierung, d.h. Darstellung der Relevanz von Ergebnissen, ohne deren Glaubwürdigkeit zu gefährden). Besonders häufig wurde in der untersuchten Stichprobe auf Neuheit und Aktualität sowie Negativität oder Positivität verwiesen, in etwas unter der Hälfte der Artikel auf Superlative der Forschung (vgl. Molek-Kozakowska 2017: 76–80).

Zuletzt hat sich eine Untersuchung auch mit inhaltlichen Faktoren beschäftigt, die dafür sorgen, dass Forschungsergebnisse nicht von Journalisten ausgewählt werden. Mellor argumentiert, dass Wissenschaftsjournalisten bei der Nachrichtenauswahl Ereignisse nicht nur mit einem Satz von Nachrichtenfaktoren filtern, sondern gleichzeitig „non-news values“ anwenden, um bestimmte Ergebnisse auszuschließen (vgl. Mellor 2015: 93). Damit sind Eigenschaften gemeint, die systematisch als nicht-nachrichtenwürdig eingeschätzt werden oder die Verarbeitung eines Ereignisses für die Berichterstattung erschweren. Diese Nicht-Nachrichtenfaktoren leitet sie aus einer qualitativen Inhaltsanalyse der Inhalte ab, die in der Berichterstattung über Wissenschaft fehlen, aber in den Quellen, also Studien oder Pressemitteilungen, vorhanden waren. Es handelt sich vor allem Faktoren, die die Aussagekraft wissenschaftlicher Ergebnisse einschränken oder in Frage stellen, beispielsweise die Angabe von Förderung oder Limitationen. Allgemeiner bezeichnet Mellor die non-news values als „provisionality“ (Vermutungen, Einschränkungen und Schwächen), „contingency“ (die Voraussetzungen oder Vorbedingungen der Forschung) und „dissonance“ (Verhalten, das vom Bild der objektiven Wissenschaft abweicht) (ebd.: 110). Die Autorin verbindet ihre Analyse mit der Kritik an einem ihrer Ansicht nach zu unkritischen Wissenschaftsjournalismus, der das Bild einer unangreifbaren und objektiven Wissenschaft reproduziert (vgl. ebd.: 111). Ob diese Kriterien aber wirklich in der Nachrichtenauswahl Einsatz finden und nicht nur bei der Darstellung, ist fraglich und wurde nicht weiter überprüft.⁶

2.2.2 Wissenschaftsbezogene Faktoren

Als Besonderheit des Wissenschaftsjournalismus spielt in der Forschungsliteratur noch eine weitere Gruppe von Ereignismerkmalen eine Rolle für die Auswahl der Berichterstattungsanlässe. Verschiedene Autoren sprechen davon, dass neben dem Thema und klassischen oder abgewandelten Nachrichtenfaktoren auch die Verlässlichkeit oder Qualität der wissenschaftlichen Informationen wichtig ist (vgl. Wilkie 1996: 1309; Allan 2009: 154). Das heißt, eigentlich wissenschaftsinterne Kriterien werden von den Journalisten bei ihrer Selektion

⁶ Das gilt allerdings zum Großteil auch für die anderen Nachrichtenfaktoren, es wird fast immer die Darstellung und nicht die eigentliche Auswahl untersucht. Dazu mehr in Kapitel 2.3.

angewendet. Schäfer begründet das mit dem hohen Vertrauen der Wissenschaftsjournalisten in die Wissenschaft sowie „personal and source pressure to conform to scientific values“ (Schäfer 2011: 403). Besondere Bedeutung hat die wahrgenommene wissenschaftliche Evidenz bei Themen, deren gesellschaftliche Relevanz besonders hoch erscheint (vgl. Blöbaum 2017: 230). In welchem Ausmaß die Orientierung an wissenschaftlicher Qualität tatsächlich einem Einfluss auf die Auswahl hat, wurde jedoch abgesehen von Befragungen (vgl. Amend/Secko 2012: 262) selten untersucht.

In ihrer oben bereits zitierten Studie zur Konstruktion von Nachrichtenwert im Wissenschaftsjournalismus widmete sich Molek-Kozakowska besonders „rationalization strategies“, mit denen die Glaubwürdigkeit der Artikel gestärkt werden soll. Diese Strategien werden aus der wissenschaftlichen Kommunikation übernommen, Beispiele sind die Verwendung von Statistiken oder quantitativen Daten aus Messungen, wissenschaftlicher Jargon oder auch Metasprache (Molek-Kozakowska 2017: 80–82). Sie finden sich in den von ihr untersuchten Artikeln sehr häufig. Während daraus noch nicht darauf geschlossen werden kann, dass wissenschaftliche Kriterien für die Auswahl eines Ereignisses wichtig sind, zeigt sich zumindest die Bedeutung bei der Verarbeitung von Forschungsergebnissen. Verschiedene andere Untersuchungen haben sich damit beschäftigt, ob bestimmte Studientypen bevorzugt ausgewählt werden. Gerade in der Medizin gibt es eine Art Hierarchie der Evidenz, die durch verschiedene Studiendesigns erreicht wird. Der Goldstandard sind dabei Randomised Control Trials (vgl. Guyatt et al. 2008). Demnach könnte man davon ausgehen, dass Wissenschaftsjournalisten, falls sie sich an der Verlässlichkeit von Ergebnissen orientieren, bevorzugt über solche Studien berichten. Gezeigt wurde jedoch eher das Gegenteil, dass diese Studien nämlich seltener ausgesucht wurden (vgl. Bartlett et al. 2002: 82)⁷ oder es keinen signifikanten Einfluss des Studiendesigns gab (vgl. Wang et al. 2015: 5). Auch für die Samplegröße konnte kein Einfluss nachgewiesen werden (vgl. ebd.: 3).

Gezeigt werden konnte dafür in der Berichterstattung über biomedizinische Studien eine Präferenz der Journalisten für Initialstudien, das heißt für Untersuchungen, in denen zum ersten Mal über einen Zusammenhang berichtet wird. Das wiederum widerspricht auch der Annahme, dass Journalisten nach wissenschaftlicher Evidenz auswählen, da die Ergebnisse von Initialstudien gerade in der Biomedizin oft nicht repliziert werden können und sie damit einen

⁷ In dieser Studie wurde zusätzlich untersucht, ob das Studiendesign (und andere Studieneigenschaften) einen Einfluss darauf haben, ob eine Pressemitteilung zu einer Studie erstellt wurde. Und hier war die Wahrscheinlichkeit bei Randomised Control Trials etwas höher. Das heißt, sie wurden seltener ausgewählt, auch wenn es etwas öfter Pressemitteilungen gab (vgl. Bartlett et al. 2002: 83).

eher geringen Evidenzgrad darstellen. Gonon und Kollegen untersuchten die Berichterstattung zur ADHS-Forschung und fanden deutlich weniger Berichterstattung über Folgestudien oder Meta-Analysen, selbst wenn diese Ergebnisse aus Studien abschwächten oder widerlegten, über die berichtet worden war (vgl. Gonon et al. 2012). Zu ähnlichen Ergebnissen kam auch eine umfassendere Analyse der Berichterstattung zu zehn verschiedenen Krankheiten, neben einer Präferenz für Initialstudien zeigte sich zusätzlich die bevorzugte Auswahl von Studien mit positivem Ergebnis, das heißt in diesem Fall, dass ein postulierter Zusammenhang bestätigt werden konnte.⁸ Außerdem wurde von den Initialstudien, über die berichtet wurde, nur die Hälfte in folgenden Meta-Analysen bestätigt (vgl. Dumas-Mallet et al. 2017).

Diese Ergebnisse sprechen also eher dagegen, dass die wissenschaftliche Verlässlichkeit oder Glaubwürdigkeit ein Kriterium ist, das Wissenschaftsjournalisten bei der Auswahl von Studienergebnissen für die Berichterstattung anwenden. Bei genauerer Überlegung erscheint es auch nicht wirklich plausibel, dass Journalisten die Qualität von Ergebnissen einschätzen oder gegeneinander abwägen sollten und daraus ihre Auswahl ableiten. Denn: wie sollten Journalisten, die häufig über eine große Bandbreite wissenschaftlicher Disziplinen berichten, in der Lage sein, für die verschiedensten Studien eine Bewertung der Qualität abzugeben?⁹ Das heißt dabei nicht, dass Überlegungen zur Qualität von Forschung keine Rolle für die Nachrichtenauswahl spielen. Diese Überlegungen betreffen jedoch meist nicht die Inhalte einzelner Studien, sondern stattdessen vor allem die Zeitschrift, in der eine Studie veröffentlicht wird, also die Quellenwahl der Journalisten.

2.2.3 Bedeutung wissenschaftlicher Zeitschriften

Im Gegensatz zur Nachrichtenauswahl von Wissenschaftsjournalisten wurde deren Quellensuche und –verwendung relativ gut erforscht. In einer Synthese der qualitativen Forschung zu Wissenschaftsjournalismus waren Praktiken der Quellensuche das am häufigsten vorkommende Thema in der analysierten Literatur (Amend/Secko 2012: 260–261). Ein Grund dafür ist die besondere Quellensituation, die sich im Wissenschaftsjournalismus präsentiert. Ausgewählt werden können Ergebnisse aus mehr als einer Million Studien jährlich, die in über 50.000 Zeitschriften veröffentlicht werden (vgl. Schäfer 2011: 404). Das heißt, es ist relativ gut bekannt, welches Quellenrepertoire vorhanden ist und angesichts der gewaltigen Menge möglicher Quellen klar, dass eine Art Vorauswahl getroffen werden muss, die dann eingehender

⁸ Ein negatives Ergebnis wäre stattdessen, dass der Zusammenhang widerlegt wurde und ein Null-Effekt, dass er weder belegt noch widerlegt werden konnte.

⁹ Gerade, wenn es schon Zweifel daran gibt, dass es der innerwissenschaftlichen Qualitätsüberprüfung, meist in Form von Peer Review, gelingt, die Qualität von Studien verlässlich zu bewerten (vgl. bspw. Smith 2006).

beobachtet wird. Denn neben den Studien gibt es für den allgemeinen Wissenschaftsjournalismus ja noch eine Vielzahl anderer möglicher Quellen, sowohl fachspezifisch (wissenschaftliche Konferenzen, direkter Kontakt mit Wissenschaftlern oder Wissenschafts-PR) sowie allgemeiner (andere Medien oder Nachrichtenagenturen) (vgl. Blöbaum et al. 2004: 37). Dadurch ergibt sich eine Flut von Informationsmaterialien und möglichen Berichterstattungsanlässen. Wilkie berichtet schon vor der großflächigen Verbreitung des Internets von 158 Informationsmaterialien, die er alleine per Post innerhalb von vier Tagen zugeschickt bekommt (vgl. Wilkie 1996: 1309). Inzwischen hat sich die Menge der Materialien sicherlich um ein Vielfaches erhöht, ein Beispiel ist die Zunahme von Pressemitteilungen, die Universitäten verschicken (vgl. Serong et al. 2017).

Die Bedeutung, die den verschiedenen Quellen zugewiesen wird, unterscheidet sich zwischen Medientiteln oder Journalisten. Die wichtige Rolle von Fachzeitschriften wurde jedoch in verschiedenen Studien gezeigt (vgl. Raupp 2008: 350–351; Blöbaum 2017: 228), nach einzelnen Wissenschaftlern sind sie meist die wichtigste Quelle. Das zeigt sich auch daran, dass Journalisten in verschiedenen Befragungen angeben, bestimmte Fachzeitschriften in der Redaktion abonniert zu haben und jede Ausgabe nach möglichen Anlässen zu durchsuchen (vgl. van Trigt et al. 1994: 640; Entwistle 1995: 920; Blöbaum et al. 2004: 45–46). Zudem korreliert die Themenagenda der Wissenschaftsberichterstattung, beispielsweise über Medikamente, mit der der Fachzeitschriften (vgl. van Trigt et al. 1995). Ein Grund für die wichtige Rolle als Quelle für den Wissenschaftsjournalismus besteht darin, dass Journalisten davon ausgehen, dass die Veröffentlichung in diesen Journals inklusive der vorherigen Prüfung durch das Peer Review-Verfahren die Relevanz und Verlässlichkeit der präsentierten Ergebnisse sicherstellt (vgl. van Trigt et al. 1994: 639; Blöbaum et al. 2004: 57, 62; Granado 2011: 802). Das trifft nicht für alle Zeitschriften gleichermaßen zu, sondern besonders für eine kleine Gruppe renommierter Journals, denen besondere Bedeutung zugesprochen wird. Durch die Konzentration auf einige wenige Quellen reagieren die Journalisten damit auch auf das Problem, nicht mehrere zehntausend Zeitschriften gleichzeitig beobachten zu können. Die Auswahl dieser Zeitschriften erfolgt tatsächlich vor allem nach wissenschaftsinternen Kriterien, was sich beispielsweise daran zeigt, dass bevorzugt Ergebnisse aus Zeitschriften mit hohem Impact Faktor ausgewählt werden (vgl. Pahl 1998: 250; Rosen et al. 2016: 346). Wormer stellt zudem die These auf, dass multidisziplinäre oder auf Medizin fokussierte Zeitschriften bevorzugt werden oder auch solche, die eine Disziplin in Gänze abdecken, also ein breites Themenspektrum aufweisen (vgl. Wormer 2008: 324).

In den verschiedensten Untersuchungen zeigt sich eine sehr konstante Gruppe weniger Zeitschriften, die einen herausragenden Einfluss auf die Auswahl von Forschungsergebnissen haben. Diese Gruppe besteht im Kern aus den multidisziplinären Journals *Nature*, *Science* und den *Proceedings of the National Academy of Sciences* (PNAS) sowie den medizinischen Zeitschriften *New England Journal of Medicine* (NEJM), *Journal of the American Medical Association* (JAMA), *Lancet*, *British Medical Journal* (BMJ) und *Nature Medicine*. Sehr häufig genannt wird zusätzlich der *New Scientist*, der aber eher populärwissenschaftlich orientiert ist. Diese Zeitschriften werden in eher anekdotischen Überblicksarbeiten zum Wissenschaftsjournalismus als wichtige Quellen bezeichnet (vgl. Wilkie 1996; Wormer 2008), in Befragungen von Wissenschaftsjournalisten genannt (vgl. Hansen 1994; Blöbaum et al. 2004; Guenther/Ruhrmann 2013) und stellen die meisten Berichterstattungsanlässe in Inhaltsanalysen (bspw. van Trigt et al. 1994; Pahl 1998; Suleski/Ibaraki 2010). Die Tabellen 11 und 12 im Anhang zeigen einen Überblick der verschiedenen Studien, die die Auswahl wissenschaftlicher Fachzeitschriften untersucht hat, mit Listen der identifizierten wichtigen Journals.

Diese Gruppe ist dabei weitestgehend unabhängig von verschiedenen Merkmalen der durchgeführten Analysen. In Deutschland (vgl. Pahl 1998), Spanien (vgl. Semir 1996), den Niederlanden (vgl. van Trigt et al. 1994), Großbritannien (vgl. Hansen 1994) oder den USA (vgl. Suleski/Ibaraki 2010) stimmten die genannten oder identifizierten Zeitschriften abgesehen von einigen lokalen Zeitschriften überein (die allerdings meist auch geringe Bedeutung hatten, beispielsweise die *Münchener Medizinische Wochenschrift*, vgl. Pahl 1998). Interessanterweise handelt es sich durchweg um naturwissenschaftliche oder medizinische Zeitschriften, es fehlt dagegen an dominanten geistes- oder sozialwissenschaftlichen Quellen. Evans konnte zeigen, dass sich in der Berichterstattung über sozialwissenschaftliche Ergebnisse keine Orientierung an einzelnen renommierten Zeitschriften zeigt. Keine Quelle aus diesem Bereich wurde in seiner Inhaltsanalyse besonders oft ausgewählt. Seiner Meinung nach ist das Fehlen von dominanten Zeitschriften auch ein Grund dafür, dass die Sozialwissenschaften öffentlich weniger sichtbar sind (vgl. Evans 1995). Dafür fanden sich auch, wenn man spezifischere Themengebiete in den Naturwissenschaften oder der Medizin betrachtet, insbesondere *Nature* und *Science* immer wieder, beispielsweise in der Krebsberichterstattung (vgl. Moriarty et al. 2010) oder Genforschung (vgl. Bubela/Caulfield 2004).

Im Laufe der Zeit scheint sich eine gewisse Veränderung vollzogen haben, gerade durch die Verbreitung des Internets und das damit verbundene Aufkommen neuer Publikationsformen oder -formate sowie die noch größere Verbreitung von Pressematerialien der

wissenschaftlichen Verlage. Kiernan untersuchte die Quellenauswahl in der Science Times-Sektion der New York Times und fand über die Zeit eine zunehmende Diversität in der Quellenwahl. Das betrifft sowohl den Anteil einiger großer Zeitschriften an allen Zitierungen in den Medien wie auch den Anteil, den bestimmte Disziplinen an der Berichterstattung einnehmen. Nach Kiernan führt der erweiterte Online-Zugang zu wissenschaftlichen Zeitschriften dazu, dass Journalisten eine größere Bandbreite wissenschaftlicher Zeitschriften zitieren können (vgl. Kiernan 2016: 203–204). Dafür gaben Wissenschaftsjournalisten in einer anderen Untersuchung an, dass die Verwendung des Internets für ihre Arbeit zu einer verstärkten Abhängigkeit von wissenschaftlichen Zeitschriften und deren Pressemitteilungen geführt hat (vgl. Granado 2011: 806, 809). Bestätigt wird das auch durch eine Zunahme der Zitierungen von wissenschaftlichen Zeitschriften insgesamt, die sich bei Kiernan (2016: 203) findet (vgl. auch Suleski/Ibaraki 2010: 120–121).

Insgesamt dominieren also einige wenige Zeitschriften die aktuelle Berichterstattung über wissenschaftliche Forschungsergebnisse. Begründet wird diese Dominanz von den Journalisten mit dem großen Renommee, das diese Zeitschriften genießen und der damit verbundenen Hoffnung, dass die in ihnen präsentierten Ergebnisse sowohl relevant als auch valide sind. In einzelnen Befragungen gaben Journalisten ganz offen an, dass sie sich nicht in der Lage fühlen, die Bedeutung oder Glaubwürdigkeit von Ergebnissen einzuschätzen, weswegen sie sich auf das Peer Review der Zeitschriften verlassen (vgl. Entwistle 1995: 922). Dadurch geben diese Zeitschriften den Großteil des Informationspools vor, aus dem Wissenschaftsjournalisten Ergebnisse auswählen können, und legen die Grundlage dafür, welche wissenschaftlichen Ergebnisse in die Öffentlichkeit gelangen. Eine Folge ist die bereits dargestellte Präferenz der Journalisten für Initialstudien (deren Ergebnisse oft nicht bestätigt werden), da diese Präferenz schon bei den großen Zeitschriften besteht (vgl. Dumas-Mallet et al. 2017; Gonon et al. 2011). Und der Einfluss geht sogar noch etwas weiter. Denn ein weiterer Grund für die enge Kopplung des Wissenschaftsjournalismus an bestimmte Zeitschriften ist, dass diese früh begonnen haben, einzelne Studienergebnisse in Pressemitteilungen oder anderen Vorabinformationen besonders hervorzuheben und etwas verständlicher darzustellen (vgl. Semir 1996: 1165; Pahl 1998: 252; Blöbaum et al. 2004: 46). Diese Vorverarbeitungen der Forschungsergebnisse grenzen die Menge der Studien, aus denen Journalisten auswählen, zusätzlich ein.

2.2.4 Bedeutung der Pressearbeit wissenschaftlicher Institutionen

Neben den Veröffentlichungen in wissenschaftlichen Zeitschriften selbst sind auch Pressemitteilungen oder andere Formen der Pressearbeit zu diesen Veröffentlichungen oder zur

Forschung insgesamt eine wichtige Quelle für Wissenschaftsjournalisten. Manche Autoren gehen sogar davon aus, dass die Abhängigkeit des Wissenschaftsjournalismus von vorverarbeitetem Material stärker ist als in anderen Journalismusformen (vgl. Mellor 2015: 97). Dabei erstellen verschiedene Institutionen solche Pressematerialien, einerseits Forschungseinrichtungen wie Universitäten oder auch forschende Unternehmen (vgl. dazu Raupp 2008), andererseits auch wissenschaftliche Verlage oder einzelne Zeitschriften in unterschiedlichem Umfang (vgl. Woloshin/Schwartz 2002). Wie viel Relevanz Wissenschaftsjournalisten den Pressemitteilungen oder Pressekonferenzen für ihre Themenfindung und ihre Beiträge zusprechen, unterscheidet sich je nach Befragung, Quelle der Pressematerialien und Medium. Gerade gegenüber den Pressematerialien aus der Wirtschaft, beispielsweise von Pharmaunternehmen, äußern sich Journalisten oft kritisch und geben an, diese in ihrer Arbeit eher selten zu nutzen (vgl. van Trigt et al. 1994: 640; Blöbaum et al. 2004: 58). Materialien von Universitäten oder Zeitschriften dagegen werden unter anderem als eine wichtige Quelle bei der Themensuche genannt (vgl. Entwistle 1995: 920; Guenther/Ruhrmann 2013: 10; Rosen et al. 2016: 344). In der Untersuchung von Blöbaum und Kollegen nannten die befragten Journalisten sehr unterschiedliche Verarbeitungsroutinen im Umgang mit Pressematerialien. Während manche angaben, alles sofort wegzuwerfen, nutzten sie andere mindestens als Information über wichtige Termine oder mögliche Themen, andere wiederum nutzen Pressemitteilungen direkt für ihre Beiträge, allerdings in unterschiedlicher Intensität (Blöbaum et al. 2004: 52, 59).

Die verschiedenen Befragungen von Journalisten erwecken also den Eindruck, dass Pressemitteilungen zwar genutzt werden, aber durchaus mit einer gewissen kritischen Distanz und in begrenztem Ausmaß. Die Ergebnisse von Inhaltsanalysen deuten dagegen auf einen deutlich größeren Einfluss hin. Das betrifft einerseits direkt die Auswahl von Studien selbst. Es gibt relativ eindeutige Belege dafür, dass gerade auch die Auswahl von Studien aus den renommierten Zeitschriften, die die Grundlage für einen Großteil der journalistischen Auswahlentscheidungen stellen, auf die Arbeiten konzentriert, zu denen es eine Pressemitteilung gab oder aber die in einem Editorial hervorgehoben wurden (vgl. Entwistle 1995; Pahl 1998; Semir et al. 1998; Bartlett et al. 2002; Stryker 2002). Dieser Zusammenhang geht aber noch weiter, auch die Studien, die in Pressemitteilungen besonders hervorgehoben wurden, also beispielsweise als erstes genannt wurden, wurden mit noch höherer Wahrscheinlichkeit ausgewählt (vgl. Semir et al. 1998: 295; Bartlett et al. 2002: 82). Das heißt, die Journalisten orientieren sich in ihrer Auswahl an den Studien, die durch Zeitschriften oder Universitäten besonders hervorgehoben werden. Dementsprechend werden Pressemitteilungen

auch relativ häufig als Quelle von Berichterstattung identifiziert (vgl. Blöbaum et al. 2004: 94–96).

Andererseits wird aber auch die Berichterstattung selbst, also auf Ebene der Aussagen, die sie enthält, durch Pressemitteilungen beeinflusst. In einer qualitativen Analyse der Berichterstattung zu 50 Studien über „Superfoods“ beispielsweise wurde für 86 % der 64 Zeitungsartikel, die über diese Studien berichteten, eine Pressemitteilung gefunden. In diesen Zeitungsartikeln war dann immer mindestens ein Teil der enthaltenen Zitate direkt aus der Pressemitteilung übernommen (vgl. Weitkamp/Eidsvaag 2014: 877). Ähnliche Ergebnisse wurden für verschiedene Formen von Ungenauigkeiten oder Übertreibungen gefunden, die Medienbeiträge zu wissenschaftlichen Ergebnissen enthalten können. Schwartz und Kollegen untersuchten das Vorkommen von 13 Qualitätsmaßen in 100 Studien aus fünf großen medizinischen Zeitschriften sowie den zugehörigen Pressemitteilungen und Medienbeiträgen. Das klare Ergebnis: für alle Qualitätsmaße war es wahrscheinlicher, dass sie in den Medienbeiträgen vorhanden waren, wenn sie auch in der Pressemitteilung enthalten waren. Der Einfluss war auch größer als der des Vorkommens der Information im Abstract einer Studie. Und noch interessanter: 12 der 13 Qualitätsmaße fanden sich mit höherer Wahrscheinlichkeit in der Berichterstattung, wenn es keine Pressemitteilung gab, als wenn es eine gab, in der sie nicht erwähnt wurden (vgl. Schwartz et al. 2012: 3-4). Ebenso finden sich verschiedene Formen von Übertreibungen wie kausale Aussagen, die aus Korrelationen abgeleitet wurden, mit deutlich höherer Wahrscheinlichkeit in den Medien, wenn sie auch in Pressemitteilungen enthalten waren. Das gilt für Pressemitteilungen sowohl von Universitäten (vgl. Sumner et al. 2014) als auch von wissenschaftlichen Zeitschriften (vgl. Sumner et al. 2016).¹⁰

Das heißt, nicht nur die Auswahl wissenschaftlicher Studien ist abhängig von Pressemitteilungen, sondern auch die konkrete Darstellung der Ergebnisse inklusive der Ungenauigkeiten oder Verfälschungen, die dem Wissenschaftsjournalismus oft vorgeworfen werden. Diese entstammen häufig den Pressemitteilungen, die aus der Wissenschaft stammen. Wissenschaftsjournalisten fügen sie meist nicht selbst hinzu, sind aber auch nicht in der Lage, sie zu erkennen (vgl. auch Brechman et al. 2009). Übertreibungen und Ungenauigkeiten in den Pressemitteilungen wiederum hängen häufig mit bereits vorhandenen Ungenauigkeiten in den Studien selbst zusammen (vgl. Yavchitz et al. 2012).

¹⁰ Die Rate der Übertreibungen war dabei in den Pressemitteilungen von Universitäten höher als in denen von wissenschaftlichen Zeitschriften (vgl. Sumner et al. 2016: 13).

Zumindest für die Berichterstattung über wissenschaftliche Institutionen gibt es allerdings auch Befunde, die den Einfluss der Pressearbeit auf die Berichterstattung etwas relativieren. In einer Studie, in der der Zusammenhang zwischen verschiedenen Eigenschaften von Hochschulen und der Menge der Aufmerksamkeit untersuchte, die sie in verschiedenen Medien erhielten, hatte keiner der Indikatoren, die mit der Pressearbeit der Institutionen zusammenhängen, einen signifikanten Einfluss. Abhängig war die Medienresonanz stattdessen vor allem von der Hochschulgröße, hier gemessen in der Zahl der Professoren einer Hochschule, sowie vom lokalen Medienumfeld (vgl. Friedrichsmeier et al. 2015). Zusammengenommen kann man die Ergebnisse dann vielleicht so interpretieren, dass die Pressearbeit wissenschaftlicher Institutionen auf die weitere Berichterstattung über Wissenschaft einen eher geringen Einfluss hat, während die engere Form des Wissenschaftsjournalismus in ihrer Auswahl und der Darstellung stark davon beeinflusst wird, welche Ergebnisse für Pressemitteilungen ausgewählt werden. Oder aber die Pressearbeit wissenschaftlicher Zeitschriften hat einfach einen deutlich größeren Einfluss als die von Universitäten.¹¹

2.2.5 Weitere Faktoren

Neben den genannten Einflussfaktoren, die in einer größeren Zahl von Studien untersucht wurden und sich bestimmten Faktorgruppen zuweisen lassen, wurden einige weitere Einflüsse auf die Nachrichtenauswahl genannt oder in Untersuchungen identifiziert. Zu nennen sind ökonomische oder organisationale Einflüsse auf Redaktionen, Ressorts oder Journalisten (vgl. Allan 2009: 155), diese wurden aber selten darauf hin untersucht, wie sie sich konkret auswirken könnten. Gleiches gilt für Ereignisse oder Stimmungslagen außerhalb der Wissenschaft, die dafür sorgen können, dass es bestimmte Trendthemen gibt (vgl. Semir 1996: 1164).¹² Zusätzlich könnten auch zeitabhängige Faktoren eine Rolle spielen, ähnlich, wie es auch Uhlemann (2012) für die Nachrichtenwerttheorie vermutet hat. So beeinflusst wahrscheinlich die jeweilige Nachrichtenlage an einem konkreten Tag, wie wahrscheinlich überhaupt über wissenschaftliche Ergebnisse berichtet wird. Hier zeigt sich dann auch der relativ geringe Status, den die Wissenschaft in vielen Redaktionen hat. So können einerseits politische Nachrichten selbst die interessanteste Wissenschaftsgeschichte verdrängen, da ihnen

¹¹ Ähnliche Überlegungen stellt zumindest auch Raupp an, die schreibt: „Insgesamt muss auf der Grundlage dieser Forschungsergebnisse die Thematisierungsleistung durch Wissenschafts-PR geringer veranschlagt werden als beispielsweise die Thematisierung durch PR im Bereich der politischen Kommunikation. Fachzeitschriften als spezifische Kommunikationsform innerhalb des Wissenschaftssystems bilden auch für die Kommunikation an den Grenzstellen des Systems Wissenschaft - also im Bereich des Wissenschaftsjournalismus und auch der Wissenschafts-PR - einen wesentlichen Referenzpunkt für Anschlusskommunikation, d.h. für die Aufbereitung von Wissenschaftsthemen für die Nicht-Fachöffentlichkeit.“ (Raupp 2008: 352).

¹² Das würde ungefähr einem Nachrichtenfaktor wie Thematisierung entsprechen.

höhere Relevanz zugewiesen wird, und andererseits, „[i]f there is a lack of interesting events on the news market [...], nearly any science story will be regarded as a timeless beauty.“ (Badenschier/Wormer 2012: 62)

Einen großen Einfluss auf die Nachrichtenauswahl im allgemeinen Journalismus haben die Nachrichtenagenturen durch ihre Ereignisauswahl. Im Wissenschaftsjournalismus ist nicht ganz klar, wie groß dieser Einfluss ist. Bei tagesaktuellen Medien werden sie wohl durchaus auch zur Themenfindung verwendet oder als Grundlage für kurze Meldungen. Allerdings schätzten befragte Journalisten sie als nicht besonders qualifiziert für Wissenschaftsnachrichten ein (vgl. Blöbaum et al. 2004: 39). Für die Wissenschaftsberichterstattung in US-Medien konnte dagegen ein großer Einfluss der Nachrichtenagenturen festgestellt werden, hier war die wichtigste Einflussgröße für die Auswahl einer Studie durch Zeitungen, ob die Associated Press, also eine US-Nachrichtenagentur, über die Studie berichtete (vgl. Kiernan 2003: 917).

Zuletzt wendete eine Studie die Theorie des „influence of presumed media influence“ auf die Berichterstattung über israelische Forscher an. Nach dieser Theorie sind die Interaktionen von Akteuren mit den Massenmedien davon abhängig, wie groß sie den Einfluss der Medienberichterstattung einschätzen. Demnach würden solche Forscher, die davon ausgehen, dass der Einfluss von Medien groß ist (beispielsweise auf ihre Karriere oder zukünftige Förderung), eher bereit sein, mit den Medien zu interagieren und dann auch öfter in ihnen genannt werden.¹³ In der Überprüfung dieses Modells fand sich zwar ein relativ starker Zusammenhang zwischen dem wahrgenommenen Medieneinfluss und der Motivation, in den Medien vertreten zu sein, allerdings war der Effekt dieser Motivation auf die tatsächliche Menge der Berichterstattung relativ gering (vgl. Tsfaty et al. 2011). Mit anderen Worten, nur, weil sich Forscher bemühten, in die Medien zu kommen, gelang ihnen das nicht unbedingt. Hier scheinen andere Faktoren wichtiger zu sein, nach den Ergebnissen, die bisher dargestellt wurden, wohl beispielsweise das Journal, in dem ein Forscher Ergebnisse veröffentlicht, oder der Themenbereich, zu dem er forscht.

2.2.6 Bisherige Systematisierungen

Die Ordnung, in die die verschiedenen bisher erkannten Einflussgrößen auf die Nachrichtenauswahl im Wissenschaftsjournalismus in den vorangegangenen Kapiteln gebracht

¹³ Dieser Annahme entsprechende Ergebnisse haben sich in verschiedenen Untersuchungen zu den Auswirkungen der Governance-Reformen an deutschen Hochschulen auf deren Streben nach öffentlicher Sichtbarkeit gezeigt. Demnach bemühen sich vor allem solche Hochschulen um mehr Medienresonanz, deren Entscheidungsträger davon ausgehen, dass die Medienberichterstattung einen großen Einfluss auf die Entscheidungen in der Wissenschaftspolitik hat (vgl. Friedrichsmeier et al. 2013: 14–17). Eingegangen wird auf Aspekte dieser Entwicklung später in der Diskussion.

wurden, stellt eine relativ einfache Systematisierung dar. Eine Rolle könnten demnach inhaltliche Aspekte von Forschungsergebnissen spielen, beispielsweise das Thema, die Disziplin, der sie zugeordnet werden können oder vielleicht auch bestimmte Nachrichtenfaktoren. Dazu kommen wissenschaftsspezifische Kriterien, die sich allerdings praktisch vor allem auf die Auswahl einer kleinen Gruppe von Zeitschriften beschränken, die als besonders renommiert oder verlässlich angesehen werden. Als weitere Orientierungsgröße werden Pressemitteilungen verwendet. Es gibt in der Forschungsliteratur einige wenige andere Ordnungs- oder Systematisierungsversuche, die als zusätzliche Alternativen vorgestellt werden sollen. Das beginnt mit Listen von Faktoren oder Kriterien, die komplizierte Auswahlroutinen auf einen einfacheren Nenner bringen sollen. Allan (2009: 154) beispielsweise nennt aus einer praktisch orientierten Perspektive Faszinationswert, Größe des betroffenen oder interessierten Publikums, Relevanz, Qualität oder Verlässlichkeit der Ergebnisse und Aktualität. Stryker (2002: 520) kommt dagegen in einem Forschungsüberblick zu ähnlichen Ergebnissen wie den hier vorgestellten und nennt drei Faktoren, die die Sichtbarkeit von medizinischen Forschungsergebnissen bestimmen sollen: Herkunftsquelle der Informationen, Vorkommen einer Pressemitteilung und Nachrichtenwert der Informationen.

Auf einer etwas höheren Abstraktionsebene liegt eine Untersuchung von Schäfer und Kollegen, in der diese ein analytisches Modell prüften, das die Medienaufmerksamkeit für bestimmte Themen anhand weniger Faktoren erklären soll, im konkreten Fall für den Klimawandel. Möglicherweise ließe sich ein solches Modell aber auch auf andere Wissenschaftsthemen oder die Wissenschaftsberichterstattung insgesamt anpassen. Faktorengruppen in diesem Modell sind „problem indicators“, „focussing events“ und „feedback“. Unter die letzte Faktorengruppe würden dann auch wissenschaftliche Publikationen fallen, deren Einfluss auf die Medienaufmerksamkeit für den Klimawandel jedoch relativ gering war (vgl. Schäfer et al. 2014). Während das Modell gut in der Lage war, die Medienaufmerksamkeit in drei Ländern im Zeitverlauf zu erklären, eignet es sich vermutlich nicht besonders gut dafür, konkrete Auswahlentscheidungen besser zu verstehen, da es eher die aggregierte Berichterstattung in den Blick nimmt.

Besser dafür geeignet könnte eine letzte Systematisierung sein, die in zwei Arbeiten entwickelt und geprüft wurde. So, wie Badenschier und Wormer die Nachrichtenwerttheorie auf den Wissenschaftsjournalismus übertragen haben, haben Guenther und Ruhrmann (2013) Annahmen aus der Gatekeeper-Forschung auf den Wissenschaftsjournalismus angewendet. Dafür leiteten sie zuerst aus dem Forschungsstand ab, welche Einflussfaktoren sich auf den unteren vier Ebenen des Hierarchy-of-Influences-Modells von Shoemaker und Reese finden.

Auf der individuellen Ebene waren das beispielsweise Rollenvorstellungen von Journalisten oder ihre persönlichen Interessen, auf der Ebene der Kommunikationsroutinen wiederum Nachrichtenfaktoren, auf der organisationalen Ebene unter anderem Hierarchien in Redaktionen oder Normen und Werte in einer Organisation sowie auf der Ebene sozialer Institutionen die Quellen der Journalisten, das Publikum und andere Medien. Die Ebene des sozialen Systems wurde in einer zweiten Studie genauer betrachtet, in der verschiedene Länder verglichen wurden (Rosen et al. 2016).

Um die Bedeutung der einzelnen Einflüsse zu prüfen, wurde eine qualitative Befragung von 21 deutschen Wissenschaftsjournalisten durchgeführt. Auf der individuellen Ebene sahen die meisten Journalisten ihre Rolle als die des „information provider“, ein Großteil gab einen Einfluss persönlicher Interessen an (vgl. Guenther/Ruhrmann 2013: 8). Auf der zweiten Ebene wurden verschiedene Nachrichtenfaktoren genannt, auf der dritten vor allem der Einfluss von Redaktionskonferenzen oder Gesprächen mit Kollegen betont. Auf der Ebene der sozialen Institutionen wurden PR und wissenschaftliche Zeitschriften als wichtige Quellen genannt, zudem sprach zumindest ein Teil der Journalisten der Berichterstattung anderer Medien einen Einfluss auf ihre Auswahlentscheidungen zu (vgl. ebd.: 8–10). Die zweite Studie, in der neben deutschen auch argentinische und französische Wissenschaftsjournalisten befragt wurden, bestätigte zum Großteil die Ergebnisse der ersten Untersuchung. Es fanden sich einzelne Unterschiede zwischen den Ländern, beispielsweise sahen sich deutsche und französische Journalisten häufiger als Kritiker, argentinische Journalisten legten einen stärkeren Fokus auf die Berichterstattung über lokale Forschung (vgl. Rosen et al. 2016: 341–347).

Damit gibt es auch einzelne umfassendere Modelle zur Nachrichtenauswahl im Wissenschaftsjournalismus. Diese eignen sich jedoch nur bedingt zur Erklärung tatsächlicher Auswahlentscheidungen von Wissenschaftsjournalisten. Speziell das Gatekeeping-Modell für den Wissenschaftsjournalismus beispielsweise nennt nur mögliche Einflussfaktoren und die Ebenen, auf denen diese verortet werden können, nicht aber, wie diese zusammenspielen oder welchen Einfluss sie denn überhaupt auf die Auswahl haben. Was sagt es beispielsweise konkret für die Nachrichtenauswahl aus, wenn ein Journalist sich als neutralen Vermittler von Informationen sieht? Die fehlende Erklärungskraft dieses Modells liegt jedoch nicht unbedingt an der konkreten Umsetzung von Guenther und Ruhrmann, sondern grundsätzlich an einer Schwäche des Gatekeeping-Ansatzes. Ähnliches gilt für die Arbeit von Badenschier und Wormer zur Nachrichtenwerttheorie. Im folgenden Abschnitt soll nach einer Zusammenfassung des Kapitels dargelegt werden, wo die Schwächen der (meisten) bisher entwickelten Theorien der Nachrichtenauswahl liegen, speziell für den Fall des

Wissenschaftsjournalismus. Daraus ergibt sich die Notwendigkeit einer anderen theoretischen Grundlage, die dann im dritten Kapitel detailliert vorgestellt wird.

2.3 Synthese: Schwächen und Lücken vorhandener Ansätze

Die Frage, nach welchen Kriterien Journalisten die Themen oder Ereignisse auswählen, über die sie dann berichten, beschäftigt Forscher seit Jahrzehnten und steht im Zentrum der Journalismusforschung. Bestimmende Forschungstraditionen sind vor allem die Nachrichtenwertforschung und Gatekeeper-Forschung. Erstere erklärt die Auswahl über Eigenschaften der Gegenstände der Berichterstattung, die Nachrichtenfaktoren. Deren Status hat sich im Laufe der Zeit von mehr oder weniger objektiv erkennbaren Ereignismerkmalen zu journalistischen Hypothesen der Realität gewandelt, deren Bedeutung sich zwischen Medien, Ländern und Zeitpunkten unterscheidet. Der Gatekeeper-Ansatz dagegen betrachtet die Eigenschaften der auswählenden Personen und Institutionen sowie der Einflüsse, die auf sie wirken. Dabei können Nachrichtenfaktoren als ein Einflussfaktor integriert werden. Auch diese Forschungsrichtung hat eine enorme Wandlung durchlebt, von einem engen Fokus auf individuellen Journalisten als Gatekeepern, deren Einstellungen und Erfahrungen die Auswahl bestimmen, bis zu komplexen Mehr-Ebenen-Modellen, die Einflüsse bis zur gesamtgesellschaftlichen Ebene beinhalten. In letzter Zeit ist zunehmend die Wirkung des Publikums in den Fokus geraten, da die Aufzeichnung von Nutzungsstatistiken die Beobachtung der Präferenzen von Rezipienten erlaubt.

Beide Ansätze sind im Laufe der Zeit immer weiter überarbeitet und geprüft worden, viele ihrer Annahmen gelten als bestätigt. Umso überraschender ist es, dass beide bisher von einzelnen Ausnahmen abgesehen kaum in der Analyse der Nachrichtenauswahl im Wissenschaftsjournalismus verwendet wurden. Stattdessen besteht der Forschungsstand aus einer verstreuten Menge von Einzelergebnissen, die sich mit verschiedenen allgemeinen und spezifischen Faktoren beschäftigt haben. Gezeigt haben diese Ergebnisse vor allem den enormen Einfluss einiger wissenschaftlicher Zeitschriften sowie der Pressearbeit wissenschaftlicher Institutionen auf die Auswahl von Wissenschaftsjournalisten. Es mangelt jedoch an Systematisierungen. Nun könnte man im Folgenden einfach versuchen, etwas detaillierter als in vorherigen Arbeiten den Gatekeeper-Ansatz oder die Nachrichtenwerttheorie auf den Wissenschaftsjournalismus zu übertragen. Der Einfluss von Pressearbeit und Zeitschriften lässt sich ganz gut auf der Extra-Media-Ebene von Shoemaker und Reese einordnen, von der Validität der Nachrichtenfaktoren im Wissenschaftsjournalismus sprechen bereits verschiedene Autoren. Im Folgenden soll jedoch gezeigt werden, dass beide Ansätze

sich nicht wirklich dafür eignen, die konkrete Nachrichtenauswahl zu erklären. Die vorgestellten alternativen Konzepte haben schon auf erste Schwachstellen hingewiesen, beispielsweise werden die Intentionen und Denkprozesse von Journalisten zu wenig berücksichtigt und die Auswahl-situation als solche vernachlässigt. Daneben gibt es noch weitere Schwächen, die sich besonders zeigen, wenn man versucht, Auswahlentscheidungen zu prognostizieren, was durch eine valide Theorie der Nachrichtenauswahl möglich sein sollte.

Gerade die Nachrichtenwerttheorie hat eine Vielzahl von Schwachstellen. Die meisten Autoren gestehen schon ein, dass die Nachrichtenwerttheorie die journalistische Selektion nicht alleine erklären kann, beispielsweise auch in einem Lehrbuch zur Nachrichtenwerttheorie (vgl. Maier et al. 2018b: 121). Man kann jedoch auch argumentieren, dass sie journalistische Selektion eigentlich überhaupt nicht erklären kann. Denn so wie sie bisher untersucht und angewendet wurde, beschäftigt sie sich nur damit, wie Nachrichten verarbeitet und präsentiert werden, nicht aber, wie sie ausgewählt wurden (vgl. O'Neill/Harcup 2009: 168). Das hängt einerseits damit zusammen, wie Nachrichtenfaktoren erhoben werden. Wie oben beschrieben wurde, besteht ein Großteil der Nachrichtenwertforschung aus Inhaltsanalysen der Medienberichterstattung. Das heißt, untersucht wird, wie viele Nachrichtenfaktoren die Berichterstattung in welcher Intensität enthält und vielleicht noch, ob es einen Zusammenhang zu Umfang und Platzierung gibt (und nicht einmal der kann immer eindeutig gezeigt werden). Die meisten Arbeiten setzen erst an, wenn die Auswahl schon getroffen wurde. Damit werden, jetzt in den Begriffen des Gatekeeping-Ansatzes ausgedrückt, nur qualitatives und quantitatives Gatekeeping untersucht, also die Rolle der Nachrichtenfaktoren in der Verarbeitung, nicht aber selektives Gatekeeping (vgl. Eilders 2006: 11–12). Denn um Aussagen darüber machen zu können, ob Nachrichtenfaktoren Auswahlkriterien sind, müsste man ihre Verteilung in den möglichen Anlässen der Berichterstattung kennen und dann nach einem Zusammenhang mit den ausgewählten Anlässen suchen. Mit anderen Worten, es bräuchte Input-Output-Analysen, die aber sehr selten sind (vgl. Staab 1990b: 433–436; Kepplinger 1998: 22; Maier et al. 2018b: 51).

Schulz versucht dieses Problem dadurch zu umgehen, dass er Nachrichtenfaktoren als journalistische Konventionen oder eben Hypothesen der Realität bezeichnet. Dadurch wird aber aus dem erst einmal rein methodischen Problem ein theoretisches. Denn folgt man diesem Gedanken, kann die Nachrichtenwerttheorie die Frage der eigentlichen Nachrichtenauswahl nicht beantworten, sie ist nur ein Modell zur Beschreibung und Analyse der Medienrealität (vgl. Staab 1990a: 109–110). Es sprechen auch weitere Gründe dafür, dass Nachrichtenfaktoren zumindest in ihrer bisherigen Form die Auswahl von Berichterstattungsanlässen nicht erklären können. Einerseits ist da der unklare und unsystematische Status der Nachrichtenfaktoren. Ein

Teil von ihnen sind objektiv messbare Eigenschaften von Ereignissen (bspw. Nähe), weitere sind subjektiv wahrgenommene Eigenschaften der Ereignisse (bspw. Überraschung oder Kontroverse) und wieder andere sind eigentlich Eigenschaften der Berichterstattung (Komposition oder Frequenz). Verschiedene Autoren wie Staab oder Fretwurst haben auf dieses Problem reagiert, indem sie Nachrichtenfaktoren in verschiedene Kategorien unterteilt haben, von denen nur ein Teil wirklich geeignet wäre, Auswahlentscheidungen zu beeinflussen (vgl. Staab 1990a: 121–122; Fretwurst 2008: 112–113). Selbst wenn man nur die Faktoren betrachtet, die sich auf Ereigniseigenschaften beziehen, hat man das Problem, wie Ereignisse definiert werden. Analysen aus Disziplinen wie der Wissenssoziologie oder Sozialpsychologie verweisen darauf, dass Ereignisse eigentlich nur in ihrer Wahrnehmung durch Subjekte, die sie erfassen und definieren, Wirkung zeigen können. Im Wissenschaftsjournalismus liegen mit der Publikation von Studien zwar relativ gut intersubjektiv abgrenzbare Ereignisse vor, deren Wahrnehmung kann sich jedoch von Journalist zu Journalist erheblich unterscheiden, wodurch eine allgemeine Wirkung objektiver Eigenschaften der Ereignisse nicht wirklich plausibel ist (vgl. Staab 1990b: 430–433; Uhlemann 2012: 93–95).

Hinzu kommt noch, dass selbst wenn Nachrichtenfaktoren Auswahlkriterien wären, sie Auswahlentscheidungen kaum erklären könnten. Das lässt sich gut am Beispiel des Wissenschaftsjournalismus veranschaulichen. Wenn täglich über 5000 Studien veröffentlicht werden, ist die Wahrscheinlichkeit sehr hoch, dass viele von ihnen einen ähnlichen Nachrichtenwert besitzen. Auswahlprognosen sind so nur dort möglich, wo man es mit überschaubaren Auswahlmöglichkeiten zu tun hat (vgl. Lehmkuhl/Promies 2019). Kepplinger formuliert es so, dass

*„Nachrichtenfaktoren zwar eine *Conditio sine qua non* für positive Selektionsentscheidungen sind, sie aber nicht erklären, weil es hinreichend viele Meldungen gibt, die nicht publiziert werden, obwohl sie die gleichen Eigenschaften besitzen.“* (Kepplinger 1998: 22)

Speziell für die Prognose von Nachrichtenauswahl ergibt sich ein Problem dadurch, dass die Nachrichtenwerttheorie zwar sehr allgemein formuliert ist, aber nicht allgemein Gültigkeit hat. Unter anderem Kepplinger (1998: 23–27) hat darauf verwiesen, dass die Bedeutung der Nachrichtenfaktoren weder themen-, zeit- oder kulturunabhängig, noch in allen Situationen der Berichterstattung gleich ist. Beispielsweise wird nicht zwischen Routineberichterstattung und Ausnahmesituationen unterschieden. Darauf aufbauend diagnostiziert er einen Mangel an prognostischem Gehalt der Theorie (vgl. Kepplinger/Bastian 2000: 463). Diesen versucht er zu

beheben, indem er das oben beschriebene Zwei-Komponenten-Modell einführt, das allerdings auch nicht viel besser zur Prognose von Auswahlentscheidungen geeignet ist. Denn das hier unterstellte Modell der Nachrichtenauswahl ist, wie Uhlemann (2012: 126–127) bemerkt, so komplex, dass es nicht plausibel ist, dass es auch nur annähernd die Überlegungen von Journalisten in der Praxis widerspiegelt. Diese müssten ja mehr oder weniger implizit erst den Nachrichtenwert verschiedener Artikel berechnen und dann gegenüberstellen und dabei vielleicht noch abwägen, wie mit ähnlichen Werten umgegangen werden soll. Plausibler erscheint ihr die Verwendung von einfacheren Urteilsheuristiken.

Einen letzten Einwand gegen die Geltung der Nachrichtenwerttheorie für die Nachrichtenselektion bringt Shoemaker (2006) vor. Sie verweist darauf, dass in der bisherigen Forschung eine gewisse begriffliche Unschärfe bei den Begriffen „news“ und „newsworthiness“ vorliegt. Während oft angenommen wird, dass die Prominenz und Darstellung von Nachrichten nur ein Ergebnis von deren Nachrichtenwürdigkeit ist, muss man beide Begriffe klarer trennen. Denn empirisch zeigt sich oft ein deutlicher Unterschied zwischen dem, was Menschen nachrichtenswürdig finden und was tatsächlich in den Nachrichten ist. Das heißt, nur, weil Journalisten bestimmte Ereignismerkmale publikationswürdig finden, muss das nicht bedeuten, dass diese auch wirklich dafür sorgen, dass etwas publiziert wird. Veranschaulicht haben diese Unterscheidung Strombäck und Kollegen (2012), die Journalisten nach der normativen und tatsächlichen Bedeutung verschiedener Ereignismerkmale für die Nachrichtenauswahl befragten. Das heißt, die Journalisten sollten einmal angeben, wie wichtig die Eigenschaft sein sollte und einmal, wie wichtig sie tatsächlich ist. In der Befragung zeigte sich, dass gerade die Ereignismerkmale in der Wahrnehmung der Journalisten weniger wichtig sind, als sie es vielleicht sein sollten, wichtiger sind dafür Produktionsroutinen oder ökonomische Überlegungen (vgl. ebd.: 725–726).

Während sich also viele Einwände dagegen finden, die Nachrichtenwerttheorie als Theorie der Nachrichtenauswahl zu bezeichnen, ist das beim Gatekeeper-Ansatz schwieriger. Auch hier kann kritisiert werden, dass selten die tatsächliche Auswahl in Form von Input-Output-Analysen untersucht wird. Diese sind insgesamt selten, Ausnahmen stellen vor allem Untersuchungen zu einzelnen Bereichen des Journalismus dar, in denen wie im Wissenschaftsjournalismus bekannt ist, aus welchen Ereignissen insgesamt ausgewählt werden kann. Ein Beispiel ist die Berichterstattung über schwere Verbrechen (vgl. bspw. Chermak 1998; Lundman 2003; Chermak/Chapman 2007). Zudem wurden Input-Output-Analysen verwendet, um den Einfluss von Pressearbeit oder Nachrichtenagenturen auf die

Berichterstattung zu untersuchen (bspw. Baerns 1991; Fröhlich/Rüdiger 2006; Riesmeyer 2007). Hier wurde jedoch nur ein Teil möglicher Berichterstattungsanlässe untersucht, die zusätzlich schon eine erste Auswahlstufe durchlaufen sind. Abgesehen von dieser methodischen Kritik lässt sich wenig gegen die verschiedenen Gatekeeping-Modelle vorbringen. Das liegt jedoch vor allem daran, dass diese so allgemein formuliert sind, dass sie kaum Aussagen über die tatsächliche Auswahl zulassen und damit auch keine Kritik an diesen Aussagen. Denn was soll dagegen vorgebracht werden, dass eine Vielzahl von Einflüssen auf Journalisten wirken, wenn sie Nachrichten auswählen?

Die Schwäche dieser sehr allgemeinen Formulierung zeigt sich wiederum, wenn man versucht, das Modell für Prognosen der Nachrichtenauswahl zu verwenden. Wie soll man aus einem Modell mit fünf Ebenen, die zudem mehr oder weniger unverbunden übereinanderstehen, Prognosen über die Nachrichtenauswahl ableiten? Man könnte sagen, welche Faktoren überhaupt eine Rolle spielen könnten, wie diese aber tatsächlich aussieht und wie dann das konkrete Zusammenspiel der Faktoren bei einer konkreten Entscheidung aussieht, darüber lassen sich keine Prognosen treffen. Durch die sehr abstrakte Modellierung gerät der Prozess der Nachrichtenauswahl einzelner Journalisten aus dem Blick. Die Nachrichtenwertforschung behauptet zumindest noch, die Nachrichtenauswahl im Einzelnen erklären zu können, die Gatekeeper-Forschung betrachtet dagegen eine sehr allgemeine Ebene, auch wenn sie mit der Beobachtung von individuellen Redakteuren begonnen hat.

3. Nachrichtenauswahl als komplexe Entscheidungssituation

Es hat sich also gezeigt, dass die vorhandenen Konzepte der Nachrichtenauswahl nicht die Grundlage für eine Prognose konkreter Auswahlentscheidungen bieten können. Daher soll im Folgenden ein Modell für die Nachrichtenauswahl von Wissenschaftsjournalisten entwickelt werden, das von der konkreten Entscheidungssituation ausgeht und darauf aufbauend Faktoren identifiziert, die die Auswahl erklären können. Bevor das Modell im Einzelnen hergeleitet wird, soll eine weitere Schwäche der vorhandenen Nachrichtenauswahltheorien adressiert werden, die bisher nicht angesprochen wurde. Es fehlt ihnen an einer theoretischen Fundierung, beispielsweise an einem Handlungsmodell (vgl. Engelmann 2016b: 457). Eine mögliche Grundlage, die die Nachrichtenauswahl in einen größeren theoretischen Rahmen einbettet, ist das Modell soziologischer Erklärung von Esser, das gesellschaftliche Prozesse und Strukturen über das Handeln individueller Akteure erklärt. Engelmann (2010; 2011; 2012) hat in verschiedenen Arbeiten ein Modell der Nachrichtenauswahl entwickelt, das sich auf diesen theoretischen Hintergrund stützt und für das Erklärung des Handelns von Journalisten die

Theory of Planned Behavior verwendet, eine bestimmte Art von Handlungstheorie. Engelmanns Ansatz und seine theoretischen Grundlagen werden in dieser Arbeit als Basis für das Modell der Nachrichtenauswahl von Wissenschaftsjournalisten verwendet, allerdings um eine zusätzliche Komponente erweitert. Denn nicht berücksichtigt wird darin bisher die Komplexität von journalistischen Entscheidungen, auf die unter anderem Donsbach (2004) verweist und die sich im Wissenschaftsjournalismus in besonderer Form zeigt. Um den Einfluss dieser Komplexität auf das Handeln von Journalisten darzulegen, wird vor allem auf Überlegungen von Schimank (2005) zu verschiedenen Dimensionen der Komplexität von Entscheidungen zurückgegriffen.

3.1 Modell soziologischer Erklärung und Handlungstheorie

Grundgedanke des Modells soziologischer Erklärung von Esser beziehungsweise des „strukturell-individualistischen Ansatzes“ (Reinemann/Baugut 2016: 310) ist, dass gesellschaftliche Phänomene, Prozesse und Strukturen nur über das individuelle Handeln der Menschen erklärt werden können. Das heißt, auch wenn Forscher an Makrophänomenen interessiert sind, müssen sie im Sinne einer echten Erklärung das individuelle Handeln berücksichtigen (vgl. Esser 2007: 27; Reinemann/Baugut 2016: 310). Dieses Handeln ist immer in eine soziale Situation eingebunden und hat darüber hinausgehende Folgen, die neu entstehende Situationen und damit das folgende Handeln prägen. Dadurch verläuft die Erklärung gesellschaftlicher Vorgänge über drei Schritte: die Untersuchung der Logik der Situation, die Akteure wahrnehmen, die Erklärung des Handelns der Akteure über eine Logik der Selektion und dann die Ableitung der gesellschaftlichen Folgen des Handelns vieler Akteure über eine Logik der Aggregation (vgl. Esser 2007: 27).

Für die subjektive Definition der Situation, also die Logik der Situation, sind zwei Faktoren wichtig: einerseits die äußeren Bedingungen der Situation und andererseits die inneren Bedingungen der Akteure. Die äußeren Bedingungen sind beispielsweise Einflüsse sozialer Strukturen, externer Ereignisse oder Informationen. Welche Aspekte einer Situation handlungsrelevant sind, wird im Modell mithilfe sogenannter Brückenannahmen beschrieben. Diese können beispielsweise theoretisch hergeleitet werden oder empirisch unter anderem durch Inhaltsanalysen oder Befragungen ermittelt werden. Im Falle des Journalismus könnten die relevanten äußeren Bedingungen beispielsweise aus Gatekeeper-Modellen abgeleitet werden, die die Einflussfaktoren auf Journalisten systematisieren (vgl. Reinemann/Baugut 2016: 312–313). Engelmann nennt Redaktions- oder Ressortumfeld als wichtige äußere Bedingungen für die Logik der Situation (vgl. Engelmann 2012: 42). Die relevanten äußeren

Bedingungen werden dann vor dem Hintergrund der inneren Bedingungen der Akteure wahrgenommen und interpretiert. Damit ist es für das Erklären des Handelns essenziell, die Situationsdefinitionen der Akteure nachzuvollziehen, da sie spätere Entscheidungen determinieren (vgl. Reinemann 2007: 54–56).

Im nächsten Schritt wird in der Logik der Selektion untersucht, wie Akteure vor dem Hintergrund der Situationsdefinition aus verschiedenen Handlungsalternativen wählen. Um das Handeln zu erklären, braucht es hier eine Handlungstheorie. Im Modell soziologischer Erklärung wird Handeln als spezielle Form des Verhaltens betrachtet, die mit der Bildung von Intentionen und der Kalkulation zukünftiger Situationen verbunden ist. Durch diesen Grundgedanken steht das Modell Theorien des rationalen Handelns nahe. Handeln beruht demnach auf einer mehr oder weniger bewussten Entscheidungen zwischen Handlungsalternativen, die sich an deren erwarteten Folgen und der Bewertung dieser Folgen orientiert, auf einem subjektiven Sinn, der mit dem Handeln verbunden wird (vgl. Esser 2007: 28–29). Akteure versuchen dann in Entscheidungssituationen, unter den jeweils gegebenen Restriktionen ihre Präferenzen möglichst gut zu realisieren. Folge ist, dass die Selektion einer Handlungsalternative nur dadurch erklärt werden kann, dass Forscher die subjektive Perspektive der Akteure inklusive ihrer Präferenzen, wahrgenommenen Handlungsalternativen und Bewertungen von Handlungsfolgen einnehmen (vgl. Reinemann/Baugut 2016: 311).

Für die konkrete Erklärung der Auswahl einer Handlungsalternative wird oft die Wert-Erwartungstheorie verwendet. Nach diesem Ansatz wählen Akteure die Alternative, „bei der das *Produkt* aus der *Bewertung* gewisser Konsequenzen ihres Tuns und der *Erwartung*, dass die jeweilige Alternative zu den verschiedenen Konsequenzen führt, [...] maximal ist.“ (Esser 2007: 31, Hervorhebung im Original). Das heißt, es wird abgewogen, welche Folgen eine Handlung haben könnte, ob diese positiv oder negativ bewertet werden und wie wahrscheinlich die Folgen eintreten. Hier wird eine extreme Form von intentionalem Handeln unterstellt, in der Akteure alle Handlungsalternativen kennen und auch über deren Wirksamkeit gut informiert sind. Da eine solche Form des Handelns in den meisten Situationen nicht plausibel ist, wurden verschiedene Erweiterungen der Wert-Erwartungstheorie entwickelt, von denen mit der Theory of Planned-Behavior im nächsten Kapitel eine genauer dargestellt wird.

Hat man über eine Handlungstheorie die Selektion einer Handlungsalternative vor dem Hintergrund der Situationsdefinition von Akteuren erklärt, fehlt noch ein Schritt, um vom Handeln in konkreten Situationen auf die Entwicklung gesellschaftlicher Strukturen zu schließen. Dafür braucht es in der Logik der Aggregation bestimmte Aggregationsregeln, nach

denen die Ergebnisse individueller Handlungswahlen aufeinander zu beziehen sind. Hier fehlt es bisher an allgemeinen methodischen Regeln, wie diese zu bilden sind. Auf der Mesoebene kann beispielsweise das Handeln einer Medienorganisation als aggregiertes Ergebnis der Interaktionen aller Mitarbeiter betrachtet werden. Auf der Makroebene zeigt sich die Bedeutung der Aggregation besonders in Resultaten, die von den Handelnden nicht beabsichtigt waren, den externen Effekten, von denen auch in der Ökonomie (unter anderem auch bei Fengler und Ruß-Mohl (2008: 678)) gesprochen wird (vgl. Reinemann/Baugut 2016: 317).

Der strukturell-individualistische Ansatz lässt sich auf eine Fülle von Fragestellungen der Journalismusforschung anwenden. Für die konkrete Ausformulierung ist das Hintergrundwissen bedeutsam, das die an sich leere Handlungstheorie füllt und die Erklärung, die für die Auswahl von Handeln geliefert wird. Das heißt, man muss jeweils plausibel erklären, wie Akteure die Handlungssituation wahrnehmen und warum sie bestimmte Handlungsalternativen wählen (vgl. ebd.: 318–319). Der Vorteil für die Erklärung der Nachrichtenauswahl besteht vor allem darin, dass die individuelle Perspektive der Journalisten in der Auswahl-situation mit der sich daraus ergebenden Berichterstattung verbunden wird. Die Auswahlmuster der Berichterstattung, die sich in der Input-Output-Analyse zeigen, können damit mit dem Handeln einzelner Akteure in Verbindung gebracht werden. Dieses wiederum wird vor allem theoretisch aus dem vorhandenen Forschungsstand hergeleitet. Damit ist auch die Prognose konkreter Auswahlentscheidungen möglich. Als mögliche handlungstheoretische Grundlage hat Engelmann die Theory of Planned Behavior bereits auf den Journalismus übertragen.

3.2 Nachrichtenauswahl in der Theory of Planned Behavior

Die Theory of Planned Behavior (TPB) kann als eine weite Version einer Theorie des rationalen Handelns bezeichnet werden. Das heißt, sie geht davon aus, dass Akteure in ihrem Handeln versuchen, bestimmte Präferenzen zu verfolgen. Diese müssen anders als in der klassischen Theorie rationalen Handelns aber nicht egoistisch sein. Zudem haben die Akteure keine vollständigen Informationen und auch subjektiv wahrgenommene Restriktionen beeinflussen das Handeln (vgl. dazu auch Reinemann 2007: 51–52). Entwickelt wurde die TPB von Fishbein und Ajzen (2010). Der Ansatz hat vor allem in der Sozialpsychologie und Soziologie breite Anwendung gefunden und ist eine der am besten belegten Theorien zur Erklärung menschlichen sozialen Handelns (vgl. Ajzen 2012; Engelmann 2012: 43; Guenther et al. 2015: 203). In der Journalismusforschung wird er erst in den letzten Jahren aufgegriffen.

Die TPB ist dreistufig konzipiert. Auf der ersten Stufe werden realisierte Handlungen von den Handlungsintentionen eines Akteurs determiniert. Das heißt, es besteht eine Korrelation zwischen dem, was Akteure tun wollen und dem, was sie tatsächlich tun. Die Handlungsintentionen wiederum werden auf der zweiten Stufe von drei verschiedenen Einstellungen der Akteure („Attitudes“) bestimmt. Die erste Einstellung betrifft die verschiedenen Handlungsalternativen, also deren Bewertung, die zweite die subjektiv wahrgenommene Norm, das heißt, wie bestimmte Bezugsgruppen verschiedenes Handeln bewerten könnten und die letzte ist die subjektiv wahrgenommene Handlungskontrolle, also wie der eigene Einfluss auf das Handeln eingeschätzt wird. Auf der dritten Stufe wiederum sind verschiedene Vorstellungen („Beliefs“) die Grundlage der Einstellungen, sie werden in Form einer Wert-Erwartungstheorie bestimmt. Die handlungsbezogenen Vorstellungen entstehen aus der wahrgenommenen Wahrscheinlichkeit, dass ein Handeln bestimmte Folgen hat (Erwartung) und der Bewertung dieser Folgen (Wert). Zusammengenommen ergeben sie eine positive oder negative Einstellung zu einem Handeln. Die Bezugsgruppen- oder Normvorstellungen entstehen aus den antizipierten Reaktionen verschiedener Referenzgruppen (Wert) und der Motivation, entsprechend diesen Reaktionen zu handeln, in anderen Worten, der Bedeutung dieser Referenzgruppen für das eigene Handeln (Erwartung). Zusammengenommen ergeben sie die subjektive Norm oder wahrgenommenen sozialen Druck. Und die Kontroll- oder Ressourcenvorstellungen entstehen daraus, wie stark die eigene Kontrolle, zum Beispiel in Form von Ressourcen oder Gelegenheiten wahrgenommen wird (Erwartung) und inwiefern diese Ressourcen oder Gelegenheiten Handeln erleichtern oder erschweren (Wert). Zusammengenommen ergeben sie die wahrgenommene Kontrolle über das eigene Handeln (vgl. Ajzen 2012: 17–18; Engelmann 2016b: 458–460)

Um das Handeln von Akteuren erklären zu können, muss man also untersuchen, wie sie zu verschiedenen Handlungen und deren Folgen stehen, welche Normen ihr Handeln beeinflussen könnten und wie viel Kontrolle über ihr Handeln sie wahrnehmen. Im Kern sind immer subjektive Bewertungen und Erwartungen von Bedeutung, aus denen sich subjektive Einstellungen und dann Handlungsintentionen ergeben. Engelmann (2010; 2011; 2012) hat das Konzept auf den Journalismus übertragen, um speziell die Nachrichtenwerttheorie theoretisch besser zu stützen und sie mit der News Bias-Forschung verbinden zu können. Demnach können Nachrichtenfaktoren als Handlungsvorstellungen bezeichnet werden, die einerseits in unterschiedlicher Stärke wahrgenommen werden (das wäre die Erwartungskomponente) und andererseits unterschiedlich wichtig sein können (Wert-Komponente). So könnte erklärt werden, warum Nachrichtenfaktoren medienspezifisch unterschiedlich verwendet werden,

zudem wird über die Wahrnehmungen von Nachrichtenfaktoren das Problem gelöst, diese an einem Ereignis festmachen zu müssen, stattdessen sind sie in diesem Modell nur in der Berichterstattung sichtbar. Normvorstellungen ergeben sich aus redaktionellen oder außerredaktionellen Bezugsgruppen. Die Erwartungen von Kollegen zum Beispiel können die Handlungsvorstellungen von Nachrichtenfaktoren verstärken. Als relevant für die Kontrollvorstellungen bezeichnet Engelmann vor allem die Berichterstattung in eigenen oder anderen Medien. So könnten Nachrichtenfaktoren dadurch verstärkt oder abgeschwächt wahrnehmen, dass andere Medien schon über einen Sachverhalt berichtet haben. Insgesamt gibt es damit drei Einflussbündel: die wahrgenommene Stärke von Nachrichtenfaktoren, journalistische Bezugsgruppen und die Medienorientierung. Gerade letztere bezeichnet sie als wichtig für Selektionsentscheidungen (vgl. Engelmann 2012: 43–46; 2016b: 462–463).

Um das tatsächliche Handeln erklären zu können, fehlt dann noch eine weitere Komponente, die die Auswahl einer Handlung erklärt. Dafür erweitert Engelmann das Modell entscheidungstheoretisch, die Entscheidung wird dadurch getroffen, dass für jede gegebene Handlungsoption die Wert-Erwartungsprodukte der einzelnen Vorstellungen berechnet werden und die Option gewählt wird, in der die Summe dieser Produkte am höchsten ist. Das heißt, je höher der wahrgenommene Nachrichtenwert im Vergleich zu anderen Ereignissen ist, desto höher die Selektionswahrscheinlichkeit (vgl. Engelmann 2010: 527). Für die Bildung des Nachrichtenwerts schlägt sie verschiedene Modelle vor. In einer nutzenrationalen Variante werden bei jeder Selektionsoption alle objektiv gegebenen Nachrichtenfaktoren mehr oder weniger stark wahrgenommen und nach ihrer Wichtigkeit bewertet. Aus allen Alternativen wird dann diejenige ausgewählt, in der die Produktschme die Summe dieser Faktoren am höchsten ist. Deutlich plausibler sind alltagsrationale Modelle, in denen nur wenige Nachrichtenfaktoren die Auswahl steuern, diese wird als Prozess zielgerichteter Aufmerksamkeitsfokussierung auf wenige Nachrichtenfaktoren definiert (vgl. dies. 2016b: 463–464).

Geprüft wurde das Modell in einem quasi-experimentellen Design mit Befragungen von Journalisten, in denen diese eines von zwei Themen sowie verschiedenen Quellen und Statements hypothetisch für die Berichterstattung auswählen sollten. Als unabhängige Variablen wurde die Beurteilung der Intensität verschiedener Nachrichtenfaktoren in den Input-Materialien sowie die Beurteilung von deren Wichtigkeit abgefragt. Bei der Themenwahl hatte nur die wahrgenommene Intensität der Nachrichtenfaktoren einen Einfluss, nicht aber die Bewertung oder das Produkt der Faktoren. Damit konnte die Bedeutung von Nachrichtenfaktoren als Handlungsvorstellungen nicht wirklich bestätigt werden (dies. 2010: 533–539). In einer weiteren Untersuchung wurde mit ähnlichem Design der Einfluss

verschiedener Bezugsgruppen untersucht, einen Einfluss hatte vor allem die Orientierung an der Berichterstattung anderer Medien, weniger die redaktionelle Kontrolle (vgl. dies. 2012: 57). Das von Engelmann entwickelte Modell erscheint gerade durch den Fokus auf Nachrichtenfaktoren nicht besonders überzeugend (und kann auch durch ihre eigenen Untersuchungen nicht wirklich bestätigt werden). Der Ansatz an sich ist dagegen mit seiner Betonung der Perspektive individueller Journalisten und ihrer Wahrnehmungen verschiedener Handlungsfolgen und normativer Einflüsse auf das Handeln vielversprechend. Für die Anwendung auf den Wissenschaftsjournalismus müsste er noch angepasst werden.

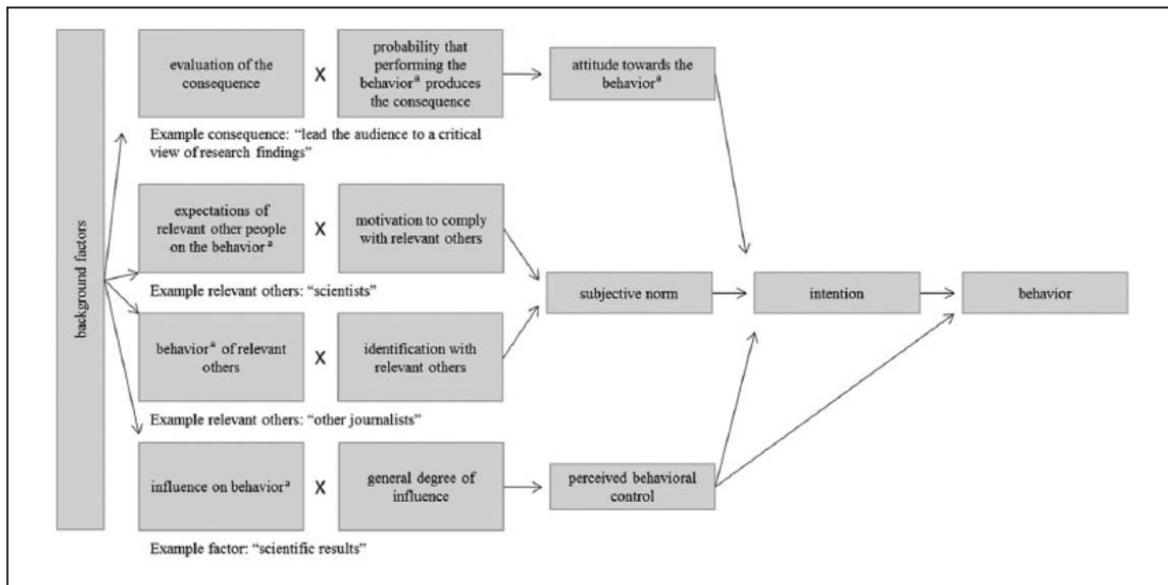


Abbildung 2: TPB-Modell der journalistischen Darstellung von Unsicherheit, entwickelt von Guenther und Kollegen. Erkennbar sind die drei verschiedenen Komponenten, die die Handlungsintention bestimmen, wobei hier die subjektive Norm in zwei Bestandteile aufgeteilt wurde.

Erste Ansätze in diese Richtung gibt es bereits, gerade in Bezug auf die Darstellung wissenschaftlicher Unsicherheit wurden Modelle für das Handeln von Journalisten (vgl. Guenther et al. 2015; Guenther/Ruhrmann 2016) und Stakeholdern (vgl. Post/Maier 2016) entwickelt, die auf der TPB basieren. Als mögliche Folgen, die Journalisten mit der Darstellung von Unsicherheit verbinden könnten, wurden in qualitativen Befragungen beispielsweise eine kritischere Sicht des Publikums zu bestimmten Themen, eine Verbesserung der Qualität ihrer Arbeit, aber auch eine negativere Einstellung zu Technologie bei den Rezipienten genannt. Als relevante Bezugsgruppen wurden Wissenschaftler, das Publikum und Kollegen genannt, seltener PR-Mitarbeiter oder Politiker. Als Kontrollvorstellungen, die die Darstellung von Unsicherheit erleichtern oder erschweren, wurden die wissenschaftlichen Ergebnisse selbst, die Qualität der Quellen und der Vergleich zu vorherigen Ergebnissen angegeben (vgl. Guenther et al. 2015: 206–210). Das so erstellte Modell wurde zusätzlich in einer quantitativen Befragung

geprüft, in der jeweils getrennt Wert und Erwartung der verschiedenen Vorstellungen abgefragt wurden. Korrelationen mit der Intention, Unsicherheit darzustellen, fanden sich vor allem bei den Bezugsgruppenvorstellungen sowie den Kontrollvorstellungen. Wichtig waren insbesondere die Darstellung von Unsicherheit in anderen Medien, Erwartungen des Publikums und die Wahrnehmung der Journalisten, dass Unsicherheit in der Forschung vorliegt. Die handlungsbezogenen Vorstellungen hatten dagegen wenig Einfluss (vgl. Guenther/Ruhrmann 2016: 935–938).

Diese Ergebnisse sind immerhin ein erster Hinweis darauf, dass sich die die TPB auch auf den Wissenschaftsjournalismus übertragen lässt. Die wichtige Bedeutung insbesondere von Bezugsgruppenvorstellungen und Kontrollansichten lässt sich auch gut mit bisherigen empirischen Ergebnissen verbinden. Die Bedeutung wissenschaftlicher Zeitschriften könnte man beispielsweise bei den Kontrollvorstellungen einordnen, die wahrgenommene Qualität oder Reputation einer Quelle wäre demnach ein Faktor, der die Auswahl deutlich erleichtert, ähnliches gilt für Pressemitteilungen zu einer Studie. Gleichzeitig könnte die Orientierung an anderen Medien als wichtige Bezugsgruppe eine weitere Erklärung für die Dominanz weniger, in den Medien akzeptierter Quellen sein. Ein wichtiger Aspekt fehlt aber noch zur Erklärung der Nachrichtenauswahl, beziehungsweise erscheint dieser in bisherigen Modellen nicht plausibel. Unterstellt wird ja eine mehr oder weniger rationale Entscheidung zwischen den gegebenen Handlungsalternativen. Gerade vor dem Hintergrund der enormen Menge wissenschaftlicher Ergebnisse und ihrer sachlichen Komplexität ist eine rationale Wahl aber kaum möglich, nur ein Bruchteil aller möglichen Handlungsalternativen wird überhaupt in Betracht gezogen. Um das zu erklären, kann auf vorhandene Modelle zur Entscheidung unter Komplexität zurückgegriffen werden.

3.3 Rationalität und Komplexität von Entscheidungen

Der Begriff der Rationalität beziehungsweise des rationalen Handelns ist die größte Schwachstelle im Modell der Nachrichtenauswahl als TPB. Denn es sind erhebliche Zweifel daran angebracht, dass das journalistische Handeln bei der Nachrichtenauswahl tatsächlich als (vollständig) rational betrachtet werden kann. Begründet werden diese Zweifel im Zusammenspiel des Handelns mit einem weiteren Phänomen, der Komplexität von Entscheidungssituationen, das insbesondere von Schimank (2005) ausführlich betrachtet wird.

Doch zuerst muss genauer betrachtet werden, was im Zusammenhang mit Entscheidungshandeln als rational bezeichnet werden kann. Nach Schimank ist kennzeichnend für Entscheidungen, dass diese eine bewusste, kalkulierte Auswahl zwischen

Handlungsalternativen darstellen, wodurch sie sich von traditionalem oder auch gefühlsgelitetem Handeln unterscheiden (vgl. ders. 2005: 44–45). In dieser bewussten Auswahl wird Rationalität angestrebt. Diese liegt „in dem Maße vor, wie eine Entscheidung der jeweiligen Situation und den von ihr gestellten Problemen gerecht wird.“ (ebd.: 65) Eine Entscheidung ist damit umso rationaler, je mehr Situationsdimensionen berücksichtigt werden. Entscheidungen sind die einzige Handlungsform, deren Ziel Rationalität ist. Unterschieden werden kann zwischen Ergebnis- und prozeduraler Rationalität. Erstere bezieht sich darauf, dass ein Akteur die Intentionen erfüllen kann, die er im Moment des Handelns hatte, letztere meint, dass ein Akteur bei der Wahl und Umsetzung einer Entscheidungsalternative bestimmte Regeln des Vorgehens befolgt. Nur um diese Form der Rationalität kann sich ein Akteur tatsächlich bemühen (ebd.: 55–57). Spezifizieren kann man rationale Entscheidungen durch unterschiedliche Entscheidungskomponenten, in denen möglichst rational gehandelt werden soll. Diese Komponenten sind Problemdiagnose, Kriterienformulierung, Alternativensuche, Alternativenbewertung, Implementation und Evaluation (vgl. ebd.: 174–175). Durch die Notwendigkeit, möglichst viele Dimensionen einer Situation erfassen zu müssen, ergibt sich eine weitere Eigenschaft von Entscheidungen: diese sind deutlich aufwendiger als andere Handlungsformen, da viel mehr Informationen bewusst verarbeitet werden müssen. Der erhöhte Aufwand führt dazu, dass Entscheidungen nur in bestimmten Situationen angewendet werden können, in denen die weniger aufwendigen Handlungsformen nicht mehr angebracht sind. Das trifft vor allem auf unklare oder schlecht definierte Situationen zu (vgl. ebd.: 68–73).

Hier kommt dann der zweite Begriff ins Spiel, die Komplexität von Situationen. Denn ein Kennzeichen solcher schlecht definierten Situationen ist ihre Komplexität, rationales Handeln reagiert also auf Komplexität, wird aber auch durch sie eingeschränkt. Worum es sich genau bei Komplexität handelt, ist schwierig anzugeben. Der Begriff hat sich in den letzten Jahren beinahe zu einem Modewort der Forschung entwickelt; in der Sozialwissenschaft, aber gerade auch in der Kommunikationswissenschaft wird geradezu inflationär von Komplexität gesprochen.¹⁴ Manchmal werden dabei explizite Bezüge zu Komplexitätstheorien hergestellt (diese werden hier nicht genauer betrachtet, vgl. für einen Überblick Urry 2005; Page 2015), in vielen Fällen nicht. Allgemein ist mit Komplexität die Einsicht bezeichnet, dass „das Verständnis des Beobachters angesichts der Beschaffenheit des Beobachteten überfordert ist“ (Adolf 2017: 501). Im Speziellen liegen dann verschiedene Konzepte oder Unterteilungen von

¹⁴ Beispiele dafür sind ein kürzlich erschienener Sammelband zur Komplexität im Journalismus (Dernbach et al. 2019) oder auch ein eigenes Themenheft der deutschen Zeitschrift „Medien & Kommunikationswissenschaft“ (bspw. Katzenbach/Pentzold 2017).

Komplexität vor. Scholl und Loosen beispielsweise sehen Komplexität immer dort, wo eine Vielzahl von Optionen vorliegt, die die Übersicht und Entscheidungsfindung erschweren und zu Unsicherheit führen (vgl. Scholl/Loosen 2019: 15–17). Gerade in der Kommunikationswissenschaft bezieht sich die Rede von Komplexität darauf, dass die empirischen Gegenstände beispielsweise durch Digitalisierung und Vernetzung auf vielerlei Weise komplexer geworden sind (vgl. Katzenbach/Pentzold 2017: 484). Das trifft dann auch auf den Journalismus zu, der auch in der Nachrichtenselektion mit zunehmender Komplexität umgehen muss.

Diese Arbeit folgt dem Komplexitätsbegriff von Schimank, der sich besonders für die Analyse einzelner Entscheidungssituationen eignet. Er unterscheidet drei analytische Dimensionen von Komplexität, die Sach-, Sozial- und Zeitdimension. Die Sozialdimension besteht darin, dass das Entscheidungshandeln eines Akteurs immer mit dem Handeln anderer Akteure zusammenwirkt. Hier zeigt sich Komplexität in Form von Konflikten und mangelnder Erwartungssicherheit, die entsteht, wenn ein Akteur unsicher ist, wie andere Akteure handeln werden (vgl. Schimank 2005: 122–128). Die Sachdimension ist der Informationsgehalt einer Entscheidungssituation, das ist die Menge der Informationen, die benötigt wird, um sich ein umfassendes Bild einer Entscheidungssituation zu machen. Innerhalb der Sachdimension ergibt sich Komplexität, wenn die Menge bedeutsamer Informationen die Informationsverarbeitungskapazität des Akteurs überschreitet (vgl. ebd.: 155–157). In der Zeitdimension zeigt sich Komplexität vor allem als Zeitknappheit, die durch Termine und Fristen verstärkt werden kann. Letztendlich sind Sach- und Sozialdimensionen nur Erscheinungsformen zeitlicher Komplexität, da diese dazu führt, dass nicht unbegrenzt Informationen verarbeitet werden können oder Erwartungssicherheit aufgebaut werden kann (vgl. ebd.: 166–170).

Im Umgang mit dieser Komplexität gibt es grundsätzlich zwei Möglichkeiten: Komplexität kann intern oder extern reduziert werden. Im ersten Fall wird die Informationsverarbeitungskapazität eines Systems oder Akteurs verbessert, im zweiten Fall wird die Umwelt vereinfacht (vgl. Adolf 2017: 509). Schimank spricht hier aufbauend auf Orrin Klapp von „opening“ oder „closing“, beides liegt in einer funktionalen und dysfunktionalen Form vor. Good Opening meint eine Öffnung für mehr und vielfältigere Informationen, die zu einer umfassenderen Informationsverarbeitung führt, bad opening eine Überflutung mit Informationen, die nicht mehr verarbeitet werden können. Closing meint dagegen ein Abschließen gegenüber weiteren externen Informationen, good closing vermeidet ein bad opening, indem im richtigen Moment keine weiteren Informationen verarbeitet werden, bad

closing meint dagegen ein stumpfes Entscheiden nach Schema F, das zu wenig Informationen über die konkrete Situation berücksichtigt. Ausgedrückt in diesen Begriffen meint rationales Entscheiden, dass zuerst ein good opening stattfindet, das dann durch good closing geschlossen wird (vgl. Schimank 2005: 52–54). Bezogen auf die Nachrichtenauswahl könnte man sagen, rationales Handeln wäre zuerst eine Öffnung gegenüber den neuen wissenschaftlichen Ergebnissen, die in einem Zeitraum veröffentlicht wurden. Diese Öffnung würde dann aber ausreichend begrenzt, sodass die schiere Menge der Ergebnisse den Journalisten nicht überfordert.

Genauso, wie Rationalität ein Mittel zur Verarbeitung von Komplexität ist, ist Komplexität auch eine Herausforderung für rationales Handeln. Untersucht wurde das Zusammenspiel beider Komponenten vor allem vor der Diagnose zunehmender Komplexität in modernen Gesellschaften. So diagnostiziert beispielsweise die Theorie reflexiver Modernisierung, dass die heutige Gesellschaft zunehmend durch Ungewissheit, Uneindeutigkeit und Unsicherheit gekennzeichnet ist, die rationales Handeln zunehmend erschweren (vgl. Böhle/Wehrich 2009: 9–12). Gleichzeitig wächst der Handlungs- und Entscheidungsdruck auf die Akteure quantitativ und zeitlich, was ebenfalls eine Erschwerung für rationales Handeln ist (vgl. Hofer 2009). Schimank geht sogar so weit, die moderne Gesellschaft als Entscheidungsgesellschaft zu bezeichnen, in der immer mehr Handeln entscheidungsförmig durchgeführt wird und werden soll. Dem Druck, immer möglichst rational handeln zu sollen, steht eine zunehmende Überforderung der Akteure aufgrund der Komplexität der Entscheidungen entgegen. „Im Spannungsverhältnis von Komplexität und Rationalität drückt sich die Problematik des Entscheidens in der modernen Gesellschaft aus.“ (Schimank 2005: 121)

Denn in allen Dimensionen der Komplexität findet eine Steigerung von Komplexität statt. In der Sachdimension erschweren Ungewissheit, Uneindeutigkeit und Unsicherheit die Erfassung der Informationen zu einer Situation, die für eine Entscheidung notwendig sind. Beispielsweise kann Wissen fehlen über die zu bedenkenden Entscheidungskriterien, die möglichen Entscheidungsalternativen und über die erwartbaren Folgen einer Entscheidung (ders. 2009: 82–84). Auf der Sozialdimension führt gerade die funktionale Differenzierung der Gesellschaft zu größerer Erwartungsunsicherheit und mehr Konflikten (vgl. ders. 2005: 148–149). Je größer aber die Komplexität einer Entscheidungssituation ist, desto schwieriger wird es für Akteure, die Anforderungen an eine rationale Entscheidung zu erfüllen, praktisch wird damit überall nur ein deutlich reduziertes Rationalitätsniveau erreicht. Dadurch gibt es einen deutlichen Unterschied zwischen dem tatsächlichen Entscheiden und dem Ideal rationalen Entscheidens, der durch die sachliche, soziale und zeitliche Komplexität von Entscheidungssituationen

entsteht. Je höher die Komplexität, desto geringer das erreichbare Rationalitätsniveau (vgl. ders. 2005: 223).

Damit wird es in Situationen mit hoher Komplexität unwahrscheinlich, dass Akteure so handeln, wie es in vielen Handlungstheorien vorgesehen wird, unter anderem in der Theory of Planned Behavior. Die meisten Forscher verwenden bereits abgeschwächte Versionen von Theorien rationalen Handelns, in denen nicht mehr unterstellt wird, dass Akteure vollständige Informationen über eine Situation haben. Gesprochen wird dann beispielsweise von „subjektiver Rationalität“ (vgl. Reinemann 2007: 51). Engelmann unterscheidet zwischen der Rationalität der Wahl und der Rationalität des Wissens, die sie auch als Alltagsrationalität bezeichnet. Im ersten Fall meint rational eine logische oder geplante Wahl, bei der die Handlungsalternative den größten zu erwartenden Nutzen bringt. Im zweiten Fall meint rational dagegen den intersubjektiven Sinn einer Wahl, Handeln ist für andere verständlich oder eine kluge Wahl. Ihrer Meinung nach ist die zweite Variante deutlich häufiger (vgl. Engelmann 2016b: 460–462). Diese leichten Abschwächungen des Begriffs der Rationalität reichen jedoch noch nicht aus, um gerade das Handeln in sehr komplexen Situationen noch vollständig erklären zu können. Und solche liegen in der Nachrichtenauswahl des Wissenschaftsjournalismus oft vor.

In der Sachdimension sorgen hier vor allem zwei Entwicklungen für eine Zunahme der Komplexität. Das ist einerseits die schon beschriebene Zunahme wissenschaftlicher Publikationen und die daraus entstehende Überfülle neuer wissenschaftlicher Ergebnisse, aus denen ausgewählt werden kann. Hier stellt sich die Frage, wie im Sinne eines good closing die relevanten Ergebnisse identifiziert werden sollen. Die Antwort scheint nach den empirischen Beobachtungen eine Konzentration auf einige wenige Quellen zu sein (vgl. Lehmkuhl 2019: 205–207). Hinzu kommt dann noch die Spezialität der Befunde, die dazu führt, dass die Ergebnisse für sich genommen meist wenig Relevanz haben und Journalisten diese Relevanz meist nicht einschätzen können. Um damit umzugehen, vertrauen sie Bewertungen aus der Wissenschaft, was sich unter anderem in der beschriebenen Übernahme von Inhalten aus Pressemitteilungen zeigt (vgl. ebd.: 208–209). Zuletzt könnten auch Zweifel an der Integrität wissenschaftlicher Ergebnisse komplexitätssteigernd wirken, die durch die zunehmende Zahl von Betrugsfällen in der Wissenschaft entstehen könnten (vgl. Lehmkuhl/Promies 2019).

In zeitlicher Hinsicht entsteht Komplexität dadurch, dass die Auswahlentscheidungen in den meisten Medien unter Zeitdruck getroffen werden müssen. Dieser wird natürlich durch die Menge der Ergebnisse verstärkt, die täglich publiziert werden. Für den

Wissenschaftsjournalismus wurde die Bedeutung dieses Faktors bisher selten erforscht, es gibt aber Hinweise darauf, dass gerade das Aufkommen des Internets den Zeitdruck verstärkt hat, da mehr Publikationen wahrgenommen werden können (vgl. Granado 2011: 805–807). In der Sozialdimension sind drei Interaktionskontexte von Bedeutung, die Interaktion mit anderen Journalisten, mit den Quellen und mit dem Publikum. Im ersten Kontext sorgt vor allem das gleichzeitige Streben nach Exklusivität und die gleichzeitige Orientierung an anderen Medien für Komplexität. Für den Wissenschaftsjournalismus liegen bisher wenige Studien vor, die die Relevanz der Kollegenorientierung untersucht hätten. Es gibt aber Hinweise darauf, dass sich auch im Wissenschaftsjournalismus Redaktionen gegenseitig dauerhaft beobachten, beispielsweise sind Leitmedien wichtige Quellen für Themen. (vgl. ebd.: 801). Im zweiten Kontext ist neben der enormen Bedeutung einzelner Quellen für die Auswahl umgekehrt auch die Bedeutung des Journalismus für die Wissenschaft beziehungsweise speziell für wissenschaftliche Institutionen angestiegen, wodurch es ein verstärktes Bestreben gibt, Medienaufmerksamkeit zu erhalten. Das zeigt sich beispielsweise in einer Zunahme der Zahl der Pressemitteilungen, die Universitäten versenden (vgl. Serong et al. 2017). Wirksam ist hier vor allem eine Entwicklung, die als neue Governance der Wissenschaft bezeichnet wird und auf die in der Diskussion noch einmal vertiefend eingegangen wird (Kapitel 7.1). Im Kontext des Publikums könnte auch für den Wissenschaftsjournalismus die zunehmende Nutzung von Aufrufstatistiken eine Rolle für die Auswahl spielen (s. Kapitel 2.1.2), hierzu gibt es aber bisher keine Forschungsergebnisse.

Zusammengenommen kann auf Grundlage der empirischen Forschung zum Wissenschaftsjournalismus also davon ausgegangen werden, dass wir es bei der Nachrichtenauswahl mit einer Situation hoher Komplexität zu tun haben, in der rationales Handeln kaum möglich ist. Beispielhaft gezeigt werden kann das an der Entscheidungskomponente der Alternativensuche. Wegen der hohen Zahl der wissenschaftlichen Ergebnisse, aus denen sie potenziell wählen könnten, sind Wissenschaftsjournalisten nicht in der Lage, eine vollständige Alternativenliste aufzustellen (vgl. Schimank 2005: 209). Das heißt, sie können nicht alle möglichen Alternativen überblicken, um dann rational diejenige auszuwählen, die am ehesten ihren Entscheidungskriterien entsprechen würde. Stattdessen richten sie ihr Augenmerk auf einen Bruchteil der möglichen Alternativen, wobei diese Auswahl auf der Sozialdimension von der Pressearbeit wissenschaftlicher Institutionen beeinflusst wird. Wenn perfekt rationales Entscheiden nicht möglich ist, welche Entscheidungsstrategie verfolgen Wissenschaftsjournalisten dann? Anders ausgedrückt, wie ist zumindest begrenzt rationales

Handeln unter Komplexität möglich? Schimank identifiziert hier verschiedene Möglichkeiten, die sich jeweils bei unterschiedlichen Komplexitätsgraden anbieten. Zwei davon lassen sich gut auf den Wissenschaftsjournalismus übertragen.

Die erste Möglichkeit ist der sogenannte Inkrementalismus, der sich in Situationen mittlerer Komplexität anbietet. Mit Komplexität wird hier dadurch umgegangen, dass meist viele kleine Entscheidungen statt weniger großer getroffen werden, weswegen das Konzept auch als die Wissenschaft des „Sich-Durchwurstelns“ bezeichnet wird (vgl. ders. 2005: 237). Relevant für den Wissenschaftsjournalismus sind hier vor allem drei Strategien: bei der Alternativensuche wird der Verarbeitungsaufwand durch eine sogenannte „simple-minded search“ reduziert. Betrachtet werden dann statt des kompletten Alternativenspektrums nur die Möglichkeiten im engsten Umkreis eines Problems (vgl. ebd.: 244–246). Beim Wissenschaftsjournalismus könnten das dann nur die Arbeiten sein, zu denen beispielsweise bei EurekaAlert Pressemitteilungen eingestellt wurden. Zudem kann eine Vereinfachung durch eine Typisierung von Entscheidungssituationen getroffen werden, wodurch sich schneller bestimmte Auswahlstrategien wählen lassen (vgl. ebd.: 249–250). Ein Beispiel für eine solche Typisierung könnten die Muster der Berichterstattung sein, beispielsweise der Lifestyle-Ansatz, die besonders gerne ausgewählt werden. Zuletzt wird die Auswahl einer Handlungsalternative nach dem Konzept des „Satisficing“ getroffen. Statt der Handlungsalternative, die Entscheidungskriterien maximal erfüllt, wird die erste Alternative gewählt, die sie auf einem bestimmten Niveau erfüllt. Dieses Anspruchsniveau definiert ein Minimum an sachlicher und sozialer Adäquanz (vgl. ebd.: 274–279). Im Wissenschaftsjournalismus könnte das vereinfacht gesagt dazu führen, dass die erste Studie aus Nature gewählt wird, die etwas mit Medizin zu tun hat.

In Situationen sehr hoher Komplexität, in denen nicht einmal mehr Inkrementalismus möglich ist, benennt Schimank verschiedene Muster sub-inkrementalistischen Entscheidens, von denen vor allem die Rationalitätsfiktionen relevant sind. Damit ist Handeln bezeichnet, das zwar als rationale Entscheidung deklariert ist, aber eigentlich nicht rational ist. Das geht so weit, dass selbst die handelnden Akteure davon überzeugt sind, dass sie rational handeln. Rationalitätsfiktionen sind also intersubjektiv geteilte Routinen, die sich darstellen, als ob es sich um Entscheidungen handele. Schimank erklärt das am Beispiel einer Person, die Sport treibt, um etwas für die eigene Gesundheit zu tun. Grundlage der Entscheidung ist kein rationales Abwägen verschiedener möglicher Alternativen oder überhaupt die Erkenntnis des Akteurs, dass ihm die eigene Gesundheit besonders wichtig ist, sondern der Eindruck, dass es als rationale Entscheidung gilt, Sport zu treiben (vgl. Schimank 2005: 373–374). Übertragen

auf den Wissenschaftsjournalismus könnte man sagen: Journalisten wählen bevorzugt Studien aus großen Zeitschriften aus, weil das als rationale Entscheidung gilt, aber nicht unbedingt, weil sie die Studien aus diesen Zeitschriften mit denen anderer möglicher Quellen verglichen haben.

In einem solchen Handeln wird keine Entscheidung im eigentlichen Sinne getroffen, weswegen sie auch keine prozedurale Rationalität hat. Es könnte ja sein, dass es eine viel relevantere Studie in einer anderen Zeitschrift gibt. Aber auch der Akteur selbst glaubt, sich rational entschieden zu haben, was daran liegt, dass er überzeugt ist, seine Wahl auf einer gut begründeten Grundlage getroffen zu haben. Wiederum im Beispiel des Wissenschaftsjournalismus wäre diese Grundlage die Annahme, dass das Peer Review von *Nature* und Co dafür sorgt, dass nur Studien hoher Qualität in ihnen erscheinen oder dass der hohe Impact Factor dafür spricht, dass in einer Zeitschrift sehr relevante Studien erscheinen. Nützlich sind Rationalitätsfiktionen, weil sie in zeitlicher Hinsicht große Zeitgewinne verschaffen, in sachlicher Hinsicht angesichts unvollständiger Informationen Unsicherheit verringern und in sozialer Hinsicht einer Handlung Legitimität verleihen. Rationalitätsfiktionen entstehen durch bestimmte Akteurskonstellationen, vor allem in Beobachtungskonstellationen. Hier beobachten Akteure gegenseitig ihr Handeln und passen sich daran an, wie andere agieren. Wenn ein Akteur in einer solchen Konstellation eine Entscheidung zu treffen hat, orientiert er sich an Akteuren, die ähnliche Entscheidungen getroffen haben. Im Journalismus sorgt demnach die Orientierung an anderen Medien dafür, dass bestimmte Quellen bevorzugt ausgewählt werden. Dazu kommt als zweite Komponente noch die Beeinflussung durch Expertenwissen (vgl. ebd.: 374–382).

„Legitimation bekommt eine Rationalitätsfiktion [...] erst dadurch, dass Experten auf eine solide, wenn möglich wissenschaftlich abgesicherte Empirie verweisen und theoretische Begründungen dafür darzulegen vermögen, warum das von der Rationalitätsfiktion empfohlene Entscheidungshandeln Erfolg versprechend und anderen Alternativen überlegen ist.“ (ebd.: 385)

Für die Orientierung an bestimmten Quellen könnte dieses Expertenwissen in den verschiedenen Maßzahlen bestehen, mit denen in der Wissenschaft die Qualität oder Relevanz von Zeitschriften oder Studien gemessen werden soll. Insgesamt entsteht die (begrenzte) Rationalität von Entscheidungen, die auf Rationalitätsfiktionen basieren, dadurch, dass sich Akteure in Situationen auf ein sozial verbreitetes Urteil verlassen, in denen sie sonst mehr oder weniger zufällig wählen müssten.

3.4 Synthese: Nachrichtenauswahl im Wissenschaftsjournalismus als Entscheidung unter Komplexität

Als Grundlage für ein Modell der Nachrichtenauswahl des Wissenschaftsjournalismus wurden zwei verschiedene theoretische Ansätze vorgestellt. Das Modell soziologischer Erklärung von Esser bietet eine Möglichkeit, gesellschaftliche Strukturen und Vorgänge durch das individuelle Handeln einzelner Akteure in sozialen Situationen zu erklären. Die Erklärung ergibt sich aus der Logik der Situation, die Akteure vor ihren inneren und den äußeren Bedingungen wahrnehmen und der Logik der Selektion, die vor diesem Hintergrund erklärt, welche Handlung Akteure in der Situation auswählen. Zuletzt werden gesellschaftliche Phänomene auf höheren Ebenen durch die Logik der Aggregation erklärt. Durch dieses Modell liegt der Fokus stark auf den individuellen Wahrnehmungen und Intentionen der Akteure, in diesem Beispiel also der Wissenschaftsjournalisten sowie auf einzelnen Situationen, in denen sie handeln, in diesem Fall die Situation der Nachrichtenauswahl. Demnach wird die Selektion von wissenschaftlichen Ergebnissen theoretisch als Entscheidungssituation betrachtet, in der zwischen einer Vielzahl möglicher Alternativen gewählt werden kann. Dadurch, dass viele Wissenschaftsjournalisten in dieser Situation ähnlich handeln, entsteht eine Struktur, in diesem Fall die Auswahlmuster der Berichterstattung, die in dieser Arbeit untersucht werden. Um zu erklären, wie es zu diesen Mustern kommt, braucht es eine Handlungstheorie, mit der die Logik der Situation und Selektion bestimmt werden. Als Handlungstheorie wird hier die Theory of Planned Behavior verwendet, in der das Handeln von Akteuren über ihre Handlungsintentionen erklärt wird. Diese wiederum ergeben sich aus Einstellungen zu den verschiedenen Handlungsoptionen, dem wahrgenommenen sozialen Druck auf ihr Handeln und der wahrgenommenen Kontrolle über ihr Handeln. Engelmann hat dieses Konzept auf den Journalismus übertragen und versucht, insbesondere die Nachrichtenwerttheorie so besser theoretisch zu fundieren. Zudem liegen TPB-Modelle zur Darstellung von Unsicherheit von Wissenschaftsjournalisten vor.

Eine Modell der Nachrichtenauswahl basierend auf der Theory of Planned Behavior könnte dann so aussehen, dass auf der Ebene der Handlungsvorstellungen bestimmte inhaltliche Präferenzen verortet sind, also die Bevorzugung bestimmter Disziplinen oder Themengebiete oder auch Nachrichtenfaktoren, die in der Berichterstattung hervorgehoben werden sollen. Hier könnte auch die wahrgenommene wissenschaftliche Qualität der Studien verortet sein. Auf der Ebene der Normvorstellungen sind dann neben den Erwartungen von Kollegen und des Publikums auch Erwartungen von Wissenschaftlern verortet. Besondere Relevanz dürften die Erwartungen anderer Journalisten sowie der eigenen Redaktion haben, wodurch sich eine

Orientierung an anderen Medien ergibt, die beispielsweise ein Faktor für die Auswahl weniger Zeitschriften ist, die von allen Medien beobachtet werden. Zentral ist für die Erklärung der Struktur der Berichterstattung jedoch die Ebene der Kontrollvorstellungen. Hier sind Faktoren eingeordnet, die ein Handeln erleichtern oder erschweren. Beispiele wären das Vorkommen einer Pressemitteilung oder die Publikation in einem renommierten Journal, die einerseits die Sichtbarkeit eines Ergebnisses erhöhen und andererseits dessen wahrgenommene Relevanz. Mit dieser handlungstheoretischen Fundierung ist ein erster Baustein dafür gelegt, die Nachrichtenauswahl aus Sicht der Journalisten in konkreten Situationen zu erklären.

Was noch fehlt, ist eine Erklärung dafür, wie zur Verwirklichung der Intentionen eine konkrete Entscheidung getroffen werden kann. Die Modelle rationaler Entscheidungen, die meistens für solche Modelle verwendet werden und davon ausgehen, dass die Alternative ausgewählt wird, bei der die Summe der verschiedenen Einstellungen maximal ist, sind hier wenig plausibel. Grund dafür ist die Komplexität der Situation, in der Wissenschaftsjournalisten eine Auswahl treffen. Basierend auf Schimank kann diese Komplexität in den drei Dimensionen soziale Komplexität, zeitliche Komplexität und sachliche Komplexität spezifiziert werden. Auf allen drei Dimensionen, aber insbesondere auf der sachlichen Dimension müssen Wissenschaftsjournalisten Entscheidungen unter hoher Komplexität treffen. Diese hohe Komplexität macht ein rationales Entscheiden unmöglich. Möglich sind stattdessen alternative Entscheidungsstrategien, je nach Komplexitätsniveau Inkrementalismus oder Rationalitätsfiktionen. Ersterer läuft vor allem auf eine deutlich reduzierte Alternativensuche hinaus, die sich beispielsweise an den eigenen Erfahrungen orientiert, sowie auf eine Auswahl, die nur ein Mindestanspruchsniveau erfüllen muss. Rationalitätsfiktionen wiederum beinhalten eigentlich keine wirkliche Entscheidung mehr, sondern ein Handeln, das sich an dem orientiert, was gesellschaftlich als rationale Entscheidung betrachtet wird. Beide Strategien laufen darauf hinaus, dass nur ein kleines Spektrum möglicher Berichterstattungsanlässe in Betracht gezogen werden. Gerade das Handeln nach Rationalitätsfiktionen könnte die besondere Bedeutung einiger weniger Zeitschriften sowie auch der Pressearbeit von wissenschaftlichen Institutionen erklären. So würde es als rationale Entscheidung betrachtet, eine Arbeit aus einer Zeitschrift mit hohem Renommee auszuwählen oder aber über eine Studie zu schreiben, die durch wissenschaftliche Institutionen besonders hervorgehoben wird. Eine eigentliche Entscheidung würde dann nicht stattfinden, weil ja nicht verglichen werden kann, ob eine andere Studie die Entscheidungskriterien eines Journalisten besser erfüllen würde.

Kombiniert man beide Ansätze, kommt man damit zu folgendem Modell: Wissenschaftsjournalisten handeln als subjektiv rationale Akteure in Situationen, in denen sie

Auswahlentscheidungen treffen müssen. Ihr Handeln basiert auf bestimmten Intentionen, die unter anderem aus inhaltlichen Präferenzen, der Orientierung an anderen Medien und den Ressourcen gebildet werden, die ihr Handeln erleichtern. Sowohl die Identifizierung als auch die Auswahl einer Handlungsalternative, also die eigentliche Entscheidung, wird durch die Komplexität beeinträchtigt, die in der Auswahl-situation vorliegt. Die Intentionen können nur in begrenztem Rahmen erfüllt werden, weswegen auch nur begrenzt rationales Handeln möglich ist. Stattdessen orientiert sich das eigentliche Handeln an Rationalitätsfiktionen oder Inkrementalismus. Mit anderen Worten: Die Journalisten wollen in ihrem Handeln vielleicht bestimmte Themen präsentieren, Nachrichtenwerte hervorheben oder wissenschaftlich valide Ergebnisse präsentieren und dabei bestimmte normative Erwartungen erfüllen, das können sie aber nur sehr begrenzt umsetzen. Denn getroffen wird die Auswahl nur für eine sehr kleine Menge an Handlungsalternativen und beeinflusst wird sie durch die Wahrnehmung dessen, was insgesamt, ob gesellschaftlich, im Journalismus oder in der Wissenschaft als rationale Auswahlentscheidung betrachtet wird.

Folge eines solchen Modells der Nachrichtenauswahl wäre auf der Ebene der Berichterstattung dann die klare Dominanz weniger Zeitschriften sowie die Orientierung an Pressemitteilungen. Beide bilden eine Art Heuristik, mit der rationale Entscheidungen simuliert werden. Zugleich müsste sich aber auch ein Einfluss der Intentionen der Journalisten zeigen, der vor allem in bestimmten thematischen Präferenzen bestehen dürfte. Geprüft werden soll im Folgenden vor allem der erste Aspekt, also die Folgen der Komplexität der Auswahl-situation auf die Struktur der Berichterstattung. Wenn Journalisten Rationalitätsfiktionen folgen, sollte das zwei Folgen haben: einerseits die Dominanz weniger Zeitschriften, da die Auswahl von Studien dieser Zeitschriften als rational betrachtet wird, andererseits die Orientierung an Pressemitteilungen, weswegen vor allem solche Studien ausgewählt werden dürften, zu denen es eine Pressemitteilung gab. Noch stärker wird vor allem das Gegenteil ausfallen, Studien, zu denen es keine Pressemitteilung gab und die in der falschen Zeitschrift erschienen sind, dürften so gut wie nie ausgewählt werden. Würden Journalisten in ihrer Auswahl rational handeln, dürfte das nicht unbedingt der Fall sein, da sie einerseits nach exklusiver Berichterstattung streben und es andererseits nicht plausibel ist, dass keine solche Studie nicht den Intentionen der Journalisten bei ihrer Auswahl entspricht. Geprüft werden diese Annahmen des Modells in einem innovativen Verfahren, in dem in großem Umfang die Auswahl von Studien für die Berichterstattung vorhergesagt wird. Als erklärende Faktoren werden der Publikationsort und das Vorkommen einer Pressemitteilung verwendet, zusätzlich noch abstrakte Informationen zum Inhalt. Wenn bereits mit den ersten beiden Komponenten die Auswahl recht genau

vorhergesagt werden kann, spricht das für ihre Bedeutung als Rationalitätsfiktionen. Wenn zusätzlich durch den Inhalt eine Verbesserung der Genauigkeit erzielt wird, wird die Bedeutung weiterer Faktoren gestützt, die durch die TPB erklärt werden können.

4. Prognosen und datengestützte Verfahren in der Kommunikationswissenschaft

Bevor im nächsten Kapitel das genaue methodische Vorgehen präsentiert wird, soll noch ein kurzer Blick darauf geworfen werden, inwiefern sich Prognoseverfahren, insbesondere solche, die auf automatisierten Methoden basieren, für die Prüfung von Theorien in der Sozialwissenschaft eignen. Zudem werden Arbeiten vorgestellt, die bereits Vorhersagemodelle im Umfeld der Erforschung des Journalismus erstellt haben. Damit sollen die Validität des verwendeten Verfahrens dargestellt und einige konkrete methodische Entscheidungen begründet werden.

4.1 Prognosen als Instrument der Überprüfung von Theorien

In der Einleitung wurde das in dieser Arbeit verwendete Vorgehen zur Erklärung der Nachrichtenauswahl als „methodisch innovativ“ beschrieben. Abgesehen von der konkreten Methode, die automatisierte Erhebungsverfahren und Machine Learning-Methoden miteinschließt, bezog sich das auch ganz grundlegend auf den Ansatz. Denn es ist zumindest in der Kommunikationswissenschaft selten, dass versucht wird, etwas vorherzusagen und dann die Vorhersage mit der Realität verglichen wird. Das ist eigentlich überraschend, wenn man betrachtet, welche Rolle Prognosen in der Wissenschaftstheorie spielen. Kepplinger hat in einer Kritik der Nachrichtenwerttheorie darauf verwiesen, dass diese keine Prognosen darüber zulässt, welchen Umfang eine Nachricht besitzen wird, die bestimmte Nachrichtenfaktoren aufweist. „Genau dies sind aber die Anwendungsfälle, in denen sich eine wissenschaftliche Theorie bewährt und genau darin liegt ihr praktischer Nutzen.“ (Kepplinger/Bastian 2000: 463) Damit bezieht er sich auf den Umstand, dass nach dem Falsifikationsprinzip wissenschaftliche Theorien dadurch geprüft werden können, dass aus ihnen abgeleitete Annahmen mit der Realität verglichen werden (vgl. Chalmers/Bergemann 2007: 51–52). Praktisch bedeutet das meist, dass bestimmte Hypothesen auf Grundlage einer Theorie abgeleitet werden und dann geprüft werden. Bei diesen Hypothesen handelt es sich im Kern ja um nichts anderes als Prognosen über die Realität. Das heißt, Theorien müssen sich daran messen lassen, ob aus ihnen Prognosen ableitbar sind, die bestätigt werden können.

Gleichzeitig kann davon gesprochen werden, dass eine wissenschaftliche Erklärung und eine wissenschaftliche Vorhersage strukturell gleich sind (vgl. Häder 2015: 59; Brühl 2015: 273–274). In beiden Fällen wird aus einer wissenschaftlichen Theorie und speziellen Anfangsbedingungen eine Aussage abgeleitet, im einen Fall eine Erklärung, im anderen eine Vorhersage. Erst bei genauerer Betrachtung können einige Unterschiede zwischen Erklärungen und Prognosen ausgemacht werden. So beziehen sich Erklärungen auf vergangene und Prognosen auf zukünftige Ereignisse, zudem werden jeweils etwas andere Dinge gesucht. Während bei Erklärungen der zu erklärende Sachverhalt bekannt ist und die erklärenden Faktoren gesucht werden, also insbesondere die Theorie, sind bei Prognosen Theorie und Anfangsbedingungen bekannt und es wird gesucht, welcher Sachverhalt sich daraus ergibt. Zuletzt sind die Anforderungen an eine Prognose etwas weniger restriktiv als an eine Erklärung. Bei dieser sollen Ursachen angegeben werden, also Kausalzusammenhänge hergestellt werden, bei Prognosen dagegen geht es vor allem um die korrekte Vorhersage eines Ergebnisses (vgl. Brühl 2015: 275–277).

Gerade dieser letzte Unterschied spiegelt sich dann in unterschiedlichen Methoden wider, die für Erklärungen oder Prognosen verwendet werden. Gezeigt werden kann das speziell am Beispiel von Regressionen. Nach Hothorn und Kollegen erfüllen Regressionen im Kern immer die beiden Funktionen von Prognose und Erklärung und verschiedene Regressionsmodelle können danach unterschieden werden, welchem Zweck sie sich eher widmen. Regressionen, die sich auf Erklärung fokussieren, verwenden eine gut lesbare und interpretierbare Darstellung. Regressionen, die sich dagegen auf die Vorhersage konzentrieren, können eine beliebig komplexe Form annehmen, die eine möglichst gute Prognose ermöglichen soll (vgl. Hothorn et al. 2006: 651–654). In dieses Spektrum kann man auch die verschiedenen Vorhersagemodelle einordnen, die in den folgenden Kapiteln dargestellt werden. Die wenigen Prognosen aus der Kommunikationswissenschaft sind eher darauf ausgerichtet, gut nachvollziehbar zu sein und Theorien oder Erklärungen zu prüfen, Prognosen in der Informatik sind dagegen vor allem an möglichst hoher Genauigkeit orientiert. Insgesamt sollten Prognosen eine Balance zwischen beiden Zielen erreichen. Denn Prognosen sollten einerseits gut begründet sein und andererseits genaue Vorhersagen erreichen (vgl. Brühl 2015: 287–288). Wenn das erste Kriterium nicht erfüllt werden kann, ist der Erkenntnisgewinn der Theorie gering, wenn das zweite Kriterium nicht erfüllt wird, ist einerseits der praktische Nutzen beschränkt und andererseits die theoretische Grundlage wohl wenig valide.

4.2 Relevante bisherige Verwendungen von Prognosemodellen

Drei verschiedene Gruppen von Prognosemodellen sollen im Folgenden kurz vorgestellt werden, um die Möglichkeiten und Herausforderungen dieser Methode zu veranschaulichen. Zuerst werden einige wenige Beispiele für Prognosemodelle in der Kommunikationswissenschaft vorgestellt, danach die deutlich vielfältigere Forschung zur Vorhersage der Popularität von Onlinenachrichten, die vor allem aus der Informatik kommt. Und zuletzt werden die wenigen Vorgänger dieser Arbeit aufgeführt, das heißt Studien, in denen versucht wurde, die Nutzung oder Verbreitung von wissenschaftlichen Studien vorherzusagen.

4.2.1 Prognosen in der Kommunikationswissenschaft

Wenn man davon absieht, dass jede Regression im Kern eine Prognose des Verhaltens einer abhängigen Variable ist (aber selten tatsächlich in dem Sinne verwendet wird), sind Prognosen in der Kommunikationswissenschaft insgesamt und in der Forschung zum Journalismus speziell sehr selten. Theorien werden fast nicht dadurch getestet, dass zukünftige Ereignisse vorhergesagt werden und dann die Vorhersage mit der Realität verglichen wird. Durch die Verwendung computergestützter Verfahren und des Entstehens einer „computational communication science“ (vgl. van Atteveldt/Peng 2018) verändert sich das erst langsam.

Eines der wenigen Beispiele, in der Prognosen verwendet wurden, um eine Theorie zu prüfen, stellen die bereits erwähnten Arbeiten von Kepplinger dar (vgl. Kepplinger/Bastian 2000; Kepplinger/Ehmig 2006). Dieser versuchte sein Zwei-Komponenten-Modell der Nachrichtenwerttheorie durch eine Prognose des Umfangs und der Platzierung von Beiträgen sowie von deren eingeschätzter Publikationswürdigkeit zu prüfen. Dafür verwendete er eine einfache lineare Regression, in der sich die verschiedenen abhängigen Variablen aus der Intensität von Nachrichtenfaktoren multipliziert mit dem Nachrichtenwert ergeben. Während die Nachrichtenfaktoren in einer Inhaltsanalyse erhoben wurden, wurden die Nachrichtenwerte in einem Fall über eine Inhaltsanalyse, im anderen durch eine Befragung erhoben. Schon dieses Vorgehen ist suboptimal, wie Kepplinger selbst bemerkt, da die Nachrichtenwerte eigentlich unabhängig von den Nachrichtenfaktoren erhoben werden sollten (vgl. Kepplinger/Bastian 2000: 472). Die Genauigkeit seiner Prognosen prüft er nur mit dem Korrelationskoeffizienten r^2 . Im ersten Fall liegt dieser mit 0,54 nicht besonders hoch, im zweiten Fall erreicht er eine höhere Korrelation, wobei nur eine Rangkorrelation mit wenigen Rängen gerechnet wird. Insgesamt unterscheidet sich sein Prognoseverfahren kaum von üblichen Regressionsanalysen, nur dass er es explizit als Prognose betitelt.

Komplexere Methoden wurden in zwei Studien verwendet, die den Einfluss verschiedener Faktoren auf die Nutzung und Verbreitung untersuchten. Berger und Milkman (2012) versuchten bei 7.000 New York Times-Artikeln vorherzusagen, welche es auf die Liste der am häufigsten per Mail versandten schafften. Dafür untersuchten sie das Vorkommen bestimmter Emotionen in den Artikeln und kontrollierten verschiedene andere Faktoren. Zur Vorhersage wurde eine logistische Regression verwendet. In der Analyse zeigte sich vor allem ein Einfluss des Erregungsgrads der Emotionen, die ein Artikel auslöst. Zusätzlich bestätigten sie die Zusammenhänge, die sich in der Inhaltsanalyse zeigten, in einem Experiment, in dem in den Stimulusmaterialien die Emotionen gezielt beeinflusst wurden (vgl. ebd.). Trilling und Kollegen untersuchten die Verbreitung von Nachrichtenartikeln in sozialen Netzwerken, ihr Interesse galt vor allem dem Einfluss von Nachrichtenfaktoren. Verschiedene Nachrichtenfaktoren wurden im Rahmen einer automatisierten Inhaltsanalyse erhoben, diese wurden dann neben mehreren Kontrollvariablen für ein Regressionsmodell verwendet. Es zeigte sich vor allem ein Einfluss der Faktoren Nähe und Positivität, während für andere Faktoren wie Konflikt kein Einfluss gezeigt werden konnte (vgl. Trilling et al. 2017). Diese beiden Studien ähnelten im Kern in ihrem Vorgehen den Studien aus der Informatik, die später vorgestellt werden. Verschiedene Variablen werden verwendet, um eine Zielvariable wie die Nutzung eines Artikels vorherzusagen und die Bedeutung der einzelnen Faktoren wird anhand von deren Einfluss auf die Vorhersage bestimmt.

Einen völlig anderen methodischen Ansatz, der aber auch einem Prognoseverfahren entspricht, wendete Waldherr (2012; 2014) an. Ihr Ziel war es, die fundamentalen Mechanismen zu verstehen, die für das Auftreten einer bestimmten Struktur in der Berichterstattung über einzelne Themen sorgen, sogenannte Nachrichtenwellen. Dafür entwickelte sie ein komplexes Simulationsmodell, das basierend auf der empirischen Forschung zur Entwicklung und Entstehung von Medienaufmerksamkeit verschiedene Akteure und Prozesse integriert. So enthält es Journalisten und Themensponsoren, die miteinander interagieren können und über Themen berichten oder diese bewerben. Waldherr bezeichnet ihr Simulationsmodell als „agentenbasiertes Modell der Medienarena“. Sie führt dann verschiedene Experimente mit diesem Modell durch, in denen sie bestimmte Grundannahmen variiert und prüft, welche ausreichend sind, um die empirisch beobachteten Muster der Medienaufmerksamkeit zu reproduzieren. Insgesamt gelingt es ihr mit dem Modell, typische Nachrichtenwellen-Muster zu erzeugen und sie kann den Einfluss einzelner Faktoren auf diese Muster klarstellen, beispielsweise scheinen Themensponsoren nicht unbedingt notwendig zu sein (vgl. Waldherr 2014; 2012). Der Nachteil eines solchen Vorgehens besteht vor allem in der Abstraktion der

realen Verhältnisse, durch die man Gefahr läuft, wichtige Faktoren zu vernachlässigen oder eine zu simple Darstellung zu entwickeln. Der Vorteil ist, dass man leicht verschiedene Szenarien durchspielen kann, um das Zusammenspiel der Faktoren genauer zu betrachten. Sinnvoll ist das natürlich nur, wenn man eine Repräsentation erstellt hat, die die Realität verlässlich nachstellen kann.

4.2.2 Prognosemodelle der Popularität von Medieninhalten in der Informatik

Während Prognoseverfahren, die diesen Namen auch verdienen, in der Kommunikationswissenschaft eher selten sind, sind sie in Disziplinen wie der Informatik ein selbstverständliches Untersuchungsinstrument. Eine erhebliche Zahl von Arbeiten hat sich auch dem Bereich der (Online-)Medien gewidmet. Vorhergesagt wurde dabei meist die Popularität von Medienbeiträgen mit unterschiedlichen Auswertungs- oder Machine Learning-Verfahren. Eingeordnet werden können diese Untersuchungen in das größere Forschungsfeld der Vorhersage der Popularität von Online-Inhalten. Ziel des Forschungsfeldes ist es vor allem, die Inhalte zu erkennen, die besonders populär sein werden, um diese Information für verschiedene Zwecke zu nutzen, beispielsweise die Optimierung von Empfehlungssystemen, die möglichst effiziente Platzierung von Werbung oder die Steuerung der Online-Infrastruktur (vgl. Tatar et al. 2014a: 1).

Popularität wird in dieser Forschung verstanden als Beziehung zwischen einem individuellen Item auf einer Website und den Online-Nutzern, die es konsumieren. Konkret kann diese Beziehung mit verschiedenen Metriken quantitativ erfasst werden. Bei Onlinenachrichten sind das neben der Zahl der Aufrufe auch die Zahl der Kommentare, die ein Artikel erhält (bspw. Tsagkias et al. 2009; Lee et al. 2012; Tatar et al. 2014b) oder aber die Erwähnung und Verbreitung in Sozialen Netzwerken (bspw. Bandari et al. 2012; Moniz et al. 2016; Arapakis et al. 2017). Das allgemeine Vorgehen bei der Vorhersage sieht dann so aus: Es werden zwei Zeitpunkte unterschieden, t_i und t_r . Ersteres ist der Zeitpunkt, an dem eine Vorhersage getroffen wird und letzteres der Zeitpunkt, für den die Popularität vorhergesagt wird. An t_i werden dann verschiedene Merkmale eines Webinhalts erhoben, je nach Vorgehen auch dessen Popularität zu diesem Zeitpunkt. Mit einem Machine Learning-Verfahren wird aus diesen Merkmalen dann eine Vorhersage der Popularität zum Zeitpunkt t_r erstellt. Unterschieden werden kann dabei eine numerische Vorhersage, in der ein genauer Wert der Popularität vorhergesagt werden soll, beispielsweise die genaue Zahl der Aufrufe eines Medienbeitrags oder aber eine Klassifikation, in der ein Popularitätsbereich vorhergesagt werden soll, in den ein Item wahrscheinlich fällt. Im einfachsten Fall sind die möglichen Klassen keine Aufmerksamkeit, also keine Aufrufe, und

Aufmerksamkeit, also mindestens ein Aufruf. Zuletzt gibt es auch noch die Möglichkeit, die Popularität verschiedener Artikel gegenüberzustellen, sie also nach ihrer Popularität zu sortieren. Dabei kann manchmal eine höhere Genauigkeit erreicht werden als bei der Klassifikation (vgl. Hensinger et al. 2013). Durch einen Vergleich der vorhergesagten Popularität zum Zeitpunkt t_r mit der tatsächlichen Popularität kann bestimmt werden, wie gut das Vorhersagemodell ist (vgl. Tatar et al. 2014a: 4).

Dieses allgemeine Vorgehen kann nach verschiedenen Kriterien in Untergruppen unterteilt werden. Es gibt Methoden, die nur Informationen von der Website verwenden, aus der die untersuchten Items stammen und Methoden, die Informationen verschiedener Websites extrahieren und kombinieren, beispielsweise noch aus Sozialen Netzwerken. Zweitens kann zwischen Vorhersagen vor oder nach der Veröffentlichung eines Inhalts unterschieden werden. Im ersten Fall können nur inhaltliche und Metainformationen verwendet werden (ein sogenannter „cold start“), im zweiten Fall kann auch aus der Popularität kurz nach der Veröffentlichung auf spätere Popularität extrapoliert werden. Die dritte Unterscheidung bezieht sich darauf, ob das aggregierte Nutzerverhalten für die Vorhersage verwendet wird oder aber das individuelle Nutzerverhalten untersucht wird, um die Vorhersage zu verfeinern (vgl. ebd.: 6, ein Beispiel für ein Modellierung auf Basis des Verhaltens einzelner Nutzer ist die Arbeit von Lerman und Hogg (2010)). Charakteristisch für Onlinenachrichten ist die kurze Lebensdauer der Popularität, das heißt, das Interesse lässt hier schnell nach, weswegen sich vor allem eine Vorhersage vor oder kurz nach der Veröffentlichung anbietet. Zudem ist die Popularität von Onlinenachrichten sehr ungleich verteilt, sie kann durch Power Law- oder Lognormal-Verteilungen beschrieben werden (vgl. Tatar et al. 2014a: 3). Das heißt, einem Großteil von Beiträgen mit geringer Popularität stehen wenige sehr populäre Beiträge gegenüber, weswegen sich viele Studien darauf konzentrieren, die wenigen sehr populären Fälle vorherzusagen (bspw. Moniz et al. 2016).

Auch wenn es diese beiden Schwerpunkte gibt, werden insgesamt die verschiedensten Ansätze verwendet, um die Popularität von Onlinenachrichten vorauszusagen. Prognosen vor der Veröffentlichung sind etwas schwieriger, da hier noch überhaupt keine Informationen zur möglichen Popularität vorhanden sind. Ausgeglichen wird das vor allem dadurch, dass eine Vielzahl von Merkmalen der Artikel extrahiert und zur Prognose verwendet wird. Im Kontext von Machine Learning werden diese Merkmale als Features bezeichnet. Die möglichen Features reichen von einfachen Metadaten wie dem Veröffentlichungsdatum oder der Quelle eines Artikels zu inhaltlichen Informationen auf verschiedenen Abstraktionsstufen. Verschiedene Verfahren des sogenannten Natural Language Processing werden verwendet, um

Themen, Personen oder Emotionen in Beiträgen automatisiert erfassen zu können. Zudem werden manchmal Informationen zur früheren Popularität von bestimmten Schlagworten (Fernandes et al. 2015), Artikelkategorien oder Quellen (Bandari et al. 2012), Überschriften (Kim et al. 2016), Autoren (Kong et al. 2018) oder erwähnten Personen (Arapakis et al. 2014) verwendet.

Ein Beispiel für die Vielfalt der Features, die dann in ein Prognosemodell eingehen können, ist die Arbeit von Fernandes und Kollegen (2015). Sie verwenden insgesamt 47 verschiedene Features, um die Popularität von Artikeln der Plattform Mashable vorherzusagen. Darunter fallen unter anderem das Vorkommen von Bildern oder Videos, Popularität von Schlagworten und im Beitrag verlinkten Artikeln, automatisiert erhobene Themen und die Subjektivität der Beiträge. Das mit diesen Features erstellte Prognosemodell wird zusätzlich noch verwendet, um ein „Intelligent Decision Support System“ zu entwickeln, das vor der Publikation eines Beitrags vorschlägt, wie dessen Popularität optimiert werden könnte (vgl. ebd.). In dieser Arbeit zeigt sich dann auch der Nachteil des Ansatzes, einfach willkürlich alle Merkmale für die Vorhersage zu verwenden, die man bekommen kann. Die Vorschläge des Support Systems sehen dann nämlich beispielsweise so aus, dass man doch noch sechs Videos und drei Bilder zum Beitrag hinzufügen könnte oder ihn statt am Donnerstag am Montag veröffentlichen sollte (vgl. ebd.: 544). Der inhaltliche Gewinn für das Verständnis von Popularität ist minimal.

Gleichzeitig führen mehr Features auch nicht unbedingt zu einer besseren Vorhersageleistung. Die Informatik reagiert darauf vor allem mit statistisch motivierten Verfahren zur Feature-Auswahl. Verschiedene dieser Verfahren wurden am Datensatz von Fernandes und Kollegen durchgespielt, sie basieren beispielsweise auf der Korrelation von Features untereinander und mit der Zielvariable (Kamel/Nour 2018), auf Faktorenanalysen (Khan et al. 2018) oder aber auf dem Beitrag der einzelnen Features zur Verbesserung der Vorhersage (Rezaeenour et al. 2018). Meistens führen sie zu einer verbesserten Genauigkeit der Vorhersage. Ein weiterer mehr oder weniger theoretisch begründeter Ansatz besteht darin, nur solche Features zu verwenden, die auch für Nutzer vor der Nutzung eines Artikels sichtbar wären, also vor allem Titel, Autoren, Zusammenfassung und Veröffentlichungszeitpunkt (vgl. Hensinger et al. 2013; Liu et al. 2017). Auch bei den verwendeten Verfahren finden sich verschiedenste Ansätze. Die meisten Studien verwenden mehrere Machine Learning-Ansätze zur Vorhersage und vergleichen deren Genauigkeit, um das beste Verfahren auszuwählen. Zum Teil werden auch die Ergebnisse mehrerer Verfahren kombiniert, wodurch sich die Vorhersage aber nicht immer verbessert (vgl. Long et al. 2018).

Insgesamt sind in der Literatur verschiedene Positionen dazu vertreten, ob es überhaupt möglich ist, die Popularität eines Artikels vor der Veröffentlichung alleine aufgrund des Inhalts verlässlich vorherzusagen. Die meisten vorgestellten Studien erreichen akzeptable Werte bei der Vorhersage und gehen demnach auch davon aus, dass es möglich ist. Arapakis und Kollegen haben sich dagegen in zwei Studien explizit mit der Umsetzbarkeit einer Prognose „at cold start“ auseinandergesetzt. Dafür replizieren sie die vielzitierte Arbeit von Bandari et al. (2012) und werten die Ergebnisse der Prognose neu aus. In den Ergebnissen zeigt sich, dass die hohen Genauigkeitswerte vor allem dadurch zustande kommen, dass geringe Popularität sehr gut vorhergesagt werden kann. Bei den relevanteren Fällen, in denen ein Artikel später sehr hohe Popularität erreicht, ist die Klassifikation dagegen sehr schlecht. Daraus folgern sie, dass „predicting news popularity at cold start is not yet a solved problem. This is mainly a consequence of the highly skewed distribution of popularity metrics.“ (Arapakis et al. 2017: 1163) Als eine Lösung für dieses Problem wurde vorgeschlagen, beim Training von Vorhersagemodellen überdurchschnittlich viele populäre Fälle zu verwenden und die Vorhersageleistung bei diesen Fällen stärker zu gewichten, wenn die Genauigkeit des Verfahrens geprüft wird (vgl. Moniz et al. 2014).

Einfacher scheint dagegen die Vorhersage der Popularität mit Daten kurz nach der Veröffentlichung zu sein. Das hängt vor allem damit zusammen, dass eine sehr hohe Korrelation zwischen frühen und späten Popularitätswerten besteht (vgl. Szabo/Huberman 2010: 83–84). Diesen Umstand machen sich viele Arbeiten zunutze, die schon mit einfachen Regressionsmodellen eine relativ hohe Genauigkeit bei der numerischen Vorhersage von Aufrufzahlen oder Kommentaren erreichen (vgl. Szabo/Huberman 2010; Tatar et al. 2011; Tatar et al. 2012; Tatar et al. 2014b). Neben linearen Regressionen werden hier auch komplexere Modelle verwendet, beispielsweise Verfahren aus der Ereigniszeitanalyse wie die Cox-Regression (vgl. Lee et al. 2010; Lee et al. 2012). Verbessert wird die Vorhersage natürlich noch durch die Kombination zeitlicher Daten zur Entwicklung der Aufrufzahlen mit Metadaten sowie inhaltlichen Informationen. Solche Verfahren übertreffen die Komplexität der oben beschriebenen Vorhersagemodelle zum Teil nochmal beträchtlich, Keneshloo und Kollegen (2016a) beispielsweise verwenden 105 verschiedene Features, um die Seitenaufrufe von Artikeln der Washington Post mit unterschiedlichen Regressionsverfahren vorherzusagen. Abgesehen von der gleichwertigen Verwendung der verschiedenen Arten von Features gibt es auch die Möglichkeit, im Rahmen eines hybriden Verfahrens bis zur Veröffentlichung eine Vorhersage nur auf Grundlage von Metadaten und inhaltlichen Eigenschaften zu treffen und

die Prognose dann nach der Veröffentlichung sukzessive basierend auf der Entwicklung der Popularität zu aktualisieren (vgl. Moniz et al. 2016).

Verschiedene Arbeiten haben untersucht, wie viel inhaltliche Features zur Verbesserung der Vorhersage beitragen können, meist durch einen Vergleich eines „Baseline“-Verfahrens, das nur frühe Popularitätswerte verwendet, mit umfangreicheren Modellen. Meist wurde mit den umfangreicheren Modellen eine Verbesserung im Vergleich zur Baseline erzielt (vgl. Tsagkias et al. 2009; van Canneyt et al. 2018), mindestens eine Studie fand jedoch eine erhebliche Dominanz der Features, die auf der zeitlichen Entwicklung der Popularität beruhen. Dafür wurde eine etwas andere Art des Vergleichs verwendet, Modelle ohne solche zeitlichen Informationen wurden Vorhersagen gegenübergestellt, die verschiedene Informationen zur frühen Popularität verwendeten. Dabei konnte selbst eine einzige zeitliche Eigenschaft bessere Vorhersagen liefern als Modelle, die keine zeitlichen Eigenschaften nutzen (vgl. Shulman et al. 2016).

Neben dem Standardvorgehen, basierend auf einer Vielzahl von Features einen bestimmten Popularitätswert oder -bereich vorherzusagen, gibt es bei Verfahren, die auch Daten nach der Veröffentlichung nutzen, eine größere Vielfalt von methodischen Ansätzen. Verschiedene Studien haben sich beispielsweise genauer mit der zeitlichen Entwicklung der Aufrufzahlen beschäftigt und Verteilungskurven zur Vorhersage verwendet. Althoff und Kollegen (2013) suchten unter semantisch ähnlichen Themen nach ähnlichen Entwicklungskurven und extrapolierten auf deren Grundlage auf die zukünftige Popularität eines Themas. Andere Arbeiten identifizierten verschiedene Cluster oder Kategorien der zeitlichen Entwicklung und entwickelten je nach Cluster eigene Vorhersageverfahren (vgl. Keneshloo et al. 2016b; Ouyang et al. 2016) oder unterschieden verschiedene Komponenten der Aufrufzahlen je nach Quelle, aus der diese stammten (beispielsweise externe Websites oder soziale Netzwerke), um eine feinere Vorhersage zu ermöglichen (vgl. Castillo et al. 2014; van Canneyt et al. 2018).

4.2.3 Prognosen der Auswahl wissenschaftlicher Studien

Einzelne Arbeiten haben auch bereits versucht, die Aufmerksamkeit für und Nutzung von wissenschaftlichen Ergebnissen zu prognostizieren und dadurch zu erklären, welche Faktoren eine wichtige Rolle für ihre Nutzung spielen. Dabei machen sie sich den praktischen Umstand zunutze, auf dem auch diese Arbeit basiert: die gute Zugänglichkeit von Daten zu wissenschaftlichen Publikationen und ihrer Erwähnung in verschiedenen Online-Quellen. Die meisten Arbeiten nutzen dabei die verschiedenen Scores des Anbieters Altmetric (s. Kapitel 5.2). Der Fokus lag dabei oft, aber nicht immer auf der Auswahl von Studien für die

Medienberichterstattung. In einer ersten Studie wurden Verbindungen zwischen der Aufmerksamkeit in den verschiedenen von Altmetric beobachteten Quellen untersucht. Für je 50.000 Studien, die keine Aufmerksamkeit in den Medien erhalten haben und 50.000 Studien, die Aufmerksamkeit erhalten haben, wurden Daten zu Erwähnungen in vier verschiedenen sozialen Netzwerken, auf Wikipedia, in Blogs und einem Peer-Review-Portal abgerufen. Drei verschiedene Machine Learning-Verfahren wurden verwendet, um aus diesen Erwähnungen vorherzusagen, ob eine Studie Medienaufmerksamkeit erhalten hat. Mit einem Random Forest-Verfahren (s. Kapitel 5.4) wurde eine Genauigkeit von fast 93 % erreicht, das heißt für 9 von 10 Studien konnte aus den Erwähnungen in den Onlinequellen richtig vorhergesagt werden, ob über sie in den Medien berichtet wurde (vgl. Siravuri/Alhoori 2017).¹⁵

Sinnvollere, weil besser interpretierbare Prädiktoren der Medienberichterstattung untersuchten Chapa und Kollegen (2017). Sie durchsuchten alle Ausgaben von fünf großen medizinischen Zeitschriften eines Jahres nach randomised control trials. Gefunden wurden 242 Studien, aus denen Informationen wie der Publikationsort, Förderquelle, Land der Durchführung und verschiedene methodische Details extrahiert wurden. Gesucht wurde zusätzlich nach Zitationen in Google Scholar, dem Attention Score von Altmetric sowie Einzelscores für Erwähnungen in Medien, Blogs, Twitter und Facebook. Insgesamt korrelierten die Erwähnungen, die von Altmetric gemessen wurden, vor allem mit der Größe der Bevölkerung, die von den Ergebnissen der Untersuchung betroffen sein könnte und bestimmten Krankheitstypen. Die methodischen Details hatten dafür kaum einen Einfluss, was auch zu den Ergebnissen zur Nachrichtenauswahl im Wissenschaftsjournalismus passt, die in Kapitel 2.2.2 vorgestellt wurden (vgl. ebd.).

Eine weitere Arbeit beschäftigte sich mit dem Einfluss verschiedener textbezogener Eigenschaften auf die Verbreitung von Studien, gemessen an der Zahl der Downloads, der Zitationen sowie der Bookmarks in einem Literaturverwaltungsprogramm. Untersucht wurde also eine eher innerwissenschaftliche Aufmerksamkeit, verwendet wurden dafür der linguistische Stil sowie die Lesbarkeit von Abstracts von Studien aus der Physik und Astronomie. Beide wurden mit automatisierten Verfahren erhoben. Es zeigte sich

¹⁵ Bei genauerer Betrachtung ist dieses Vorgehen aus verschiedenen Gründen wenig sinnvoll. Einerseits berücksichtigt der Ansatz nicht, dass Medienaufmerksamkeit ja zu Erwähnungen in sozialen Netzwerken etc. führen wird, weswegen die hohe Genauigkeit bei der Vorhersage wenig überraschend ist. Schließlich ist es sehr plausibel, dass die Studien, die von Medien aufgegriffen wurden, insgesamt stärker online diskutiert werden. Hier fehlt eine Korrektur für die Reihenfolge der Erwähnung. Das führt auch dazu, dass der Sinn des Verfahrens unklar ist, denn die Daten, die hier zur Vorhersage verwendet werden, sind ja erst nach der Veröffentlichung der Studien vorhanden, also zu einem Zeitpunkt, an dem die meiste Berichterstattung der Studien bereits passiert sein dürfte.

beispielsweise, dass in den untersuchten Abstracts eher persönliche Sprache verwendet wird sowie zukunftsbezogene Verben. Die Lesbarkeit unterscheidet sich zwischen den verschiedenen Quellen, in denen Aufmerksamkeit gemessen wurde, so sind die oft heruntergeladenen Studien leichter lesbar als die, die oft in Literaturverwaltungsprogrammen gespeichert werden (vgl. Guerini et al. 2012). Während die erhobenen Variablen interessant und relativ gut zu erheben sind, ist das Design eher fraglich, weil kein Vergleich der Studien mit viel Aufmerksamkeit mit Studien ohne Aufmerksamkeit durchgeführt wurde, weshalb unklar ist, ob die identifizierten Merkmale überhaupt charakteristisch für diese Studien sind.

Während der Nutzen dieser drei Studien für diese Arbeit eher begrenzt ist und sich darauf beschränkt, das breite Spektrum möglicher Herangehensweisen bei der datengestützten Untersuchung der Auswahl wissenschaftlicher Studien aufzuzeigen, sind drei weitere Untersuchungen von größerer Relevanz. Das betrifft erstens vor allem die Studie von MacLaughlin und Kollegen, die in den Grundzügen auf dem gleichen Untersuchungsdesign wie diese Arbeit basiert. Sie verwendeten einen Datensatz von knapp über 90.000 Studien, die 2016 veröffentlicht wurden und zu denen bei Altmetric Daten vorlagen. Für jede Studie wurden verschiedene Metadaten abgerufen sowie Titel und Abstract. Zusätzlich wurden Erwähnungen in drei verschiedenen Pressemitteilungsportalen erhoben. Basierend auf diesen Daten sollte dann die Medienaufmerksamkeit in Form einer Sortieraufgabe vorhergesagt werden. Das heißt, die Studien sollten in der Reihenfolge ihrer Medienaufmerksamkeit angeordnet werden, gemessen in der Zahl der Erwähnungen in Massenmedien. Die Autoren verglichen dann die Vorhersageleistung mit verschiedenen Variablen, am genauesten war die Vorhersage, wenn alle Informationen verwendet wurden. Von den einzelnen Variablen hatten vor allem das Vorkommen einer Pressemitteilung sowie bestimmte populäre Themen einen Einfluss. Ermittelt wurden diese Themen durch die Themenkategorien, die Altmetric zuordnete, und durch das Vorkommen von bestimmten Begriffen in Abstracts und Titeln. Das wichtigste Ergebnis war, dass inhaltliche Merkmale die Vorhersage deutlich verbessern konnten, anders als es oft bei der Vorhersage der Popularität von Online-Medienbeiträgen der Fall war. Zudem wurde die Bedeutung von Pressemitteilungen bestätigt, wenn nur Studien mit Pressemitteilungen sortiert werden sollten, wurde eine deutlich geringere Genauigkeit erreicht (vgl. MacLaughlin et al. 2018).

Zuletzt untersuchten zwei zusammengehörige Arbeiten, welche Faktoren die Berichterstattung über medizinische Studien in Medien sowie die Auswahl für Pressemitteilungen beeinflussen. Interessant ist an diesen Arbeiten natürlich vor allem, dass untersucht wird, wie wissenschaftliche Zeitschriften oder Institutionen die Studien auswählen, zu denen sie

Pressemitteilungen veröffentlichen. Dieser Aspekt wurde bisher kaum erforscht. Wenn sich jedoch die Nachrichtenauswahl von Wissenschaftsjournalisten stark daran orientiert, zu welchen Studien es Pressemitteilungen gibt, wird dieser Auswahlsschritt (insbesondere in den großen Zeitschriften) fast noch relevanter dafür, welche Studien öffentlich diskutiert werden. Zusätzlich wurde aber auch noch ein sehr durchdachtes Verfahren zur Erstellung einer Stichprobe von Studien sowie der zugehörigen Auswahlentscheidungen verwendet. In der ersten Arbeit (Wallace et al. 2015) wurde die Auswahl für Pressemitteilungen anhand von Studien aus der Zeitschrift JAMA untersucht und die Auswahl für die Berichterstattung anhand von Artikeln der Nachrichtenagentur Reuters zu biomedizinischen und Gesundheitsstudien, in denen diese direkt verlinkt wurden. Über die Links wurden die DOIs der Studien identifiziert und damit Metadaten zu den Studien aus PubMed abgerufen, einer Datenbank für biomedizinische Studien. Um zusätzlich einen Datensatz mit Studien zu erhalten, über die nicht berichtet wurde, wendeten die Autoren einen „matched sampling“-Ansatz an. Für jedes Positivbeispiel suchten sie bis zu 20 Negativbeispiele aus der gleichen Zeitschrift und dem gleichen Jahr, über die nicht berichtet wurde. So wurden bei den Pressemitteilungen fast 900 Positiv- und 9.914 Negativbeispiele ermittelt, für die Nachrichtenagentur 1343 Positiv- und knapp 27.500 Negativbeispiele (vgl. ebd.: 13).

Als Vorhersageverfahren wurde logistische Regression verwendet, als Variablen der Name des Journals, die Institution des Erstautors und Begriffe aus dem Titel, dem Abstract und den MeSH-Terms.¹⁶ In beiden Fällen konnten relativ gute Vorhersagen erreicht werden, bei Pressemitteilungen noch etwas bessere als beim Reuters-Korpus. Besonders aussagekräftig waren bestimmte Begriffe, beispielsweise „exercise“, „smoking“, „pregnancy“ oder „cancer“ (vgl. ebd.: 14). In einer zweiten Arbeit wurde der Ansatz noch einmal erweitert. Zum einen wurde ein weiterer Datensatz verwendet, der aus der Untersuchung von Sumner und Kollegen (2014) entnommen wurde, die oben bereits erwähnt wurde. Dieser enthält 462 Pressemitteilungen zu Studien aus biomedizinischen Zeitschriften von 20 Universitäten aus Großbritannien sowie die Berichterstattung, die sich daraus ergeben hat. Damit konnte der Datensatz einerseits zur Vorhersage von Berichterstattung verwendet werden, da nur über knapp die Hälfte der Studien in den Medien berichtet wurde, andererseits wurde die Auswahl für Pressemitteilungen untersucht, indem im „matched sampling“-Verfahren wiederum bis zu 10 Negativbeispiele für jede Auswahl für eine Pressemitteilung gesucht wurden. Zusätzlich wurde zur Vorhersage der Auswahl neben den oben genannten Faktoren auch eine automatisiert

¹⁶ MeSH oder Medical Subject Headings sind die Schlagworte, die in PubMed für die Indizierung von Studien verwendet werden.

erhobene Themenvariable verwendet. Die Genauigkeit der Vorhersage war beim neuen Datensatz deutlich geringer als zuvor, dabei war die Vorhersage, ob eine Pressemitteilung erstellt wird, noch etwas genauer. Positiven Einfluss hatten beispielsweise Themen, die mit Hirnforschung, Genforschung oder Diabetes verbunden waren (vgl. Zhang et al. 2016).

Insgesamt zeigen die wenigen vorhandenen Untersuchungen, in denen öffentliche Aufmerksamkeit oder Auswahl von Studienergebnissen vorhergesagt werden sollte, zweierlei: einerseits ist die Entwicklung solcher Vorhersageverfahren relativ gut möglich und es lässt sich je nach Verfahren eine gute Genauigkeit der Vorhersage erreichen. Andererseits können aufgrund der vielfältigen Daten, die frei zu wissenschaftlichen Studien verfügbar sind, relativ einfach die verschiedensten Faktoren in Vorhersagemodelle eingebaut werden. Wenn einfach alle Informationen oder die falschen in Vorhersagemodelle eingebaut werden, führt das schnell dazu, dass die Vorhersage entweder nicht mehr interpretiert werden kann oder aber die Genauigkeit zu wünschen übrig lässt. Ersteres tritt beispielsweise auf, wenn einfach alle möglichen Begriffe aus Abstracts und Titeln von Studien zur Vorhersage verwendet werden, wodurch dann Begriffe wie „people“ oder „fish“ im Vorhersagemodell damit zusammenhängen, dass eine Studie mehr Medienaufmerksamkeit erhält (vgl. MacLaughlin et al. 2018: 8). Zweiteres dagegen zeigt sich vor allem in der Arbeit von Chapa und Kollegen, in der der Fokus zu stark auf methodischen Details der Studien liegt, über die berichtet werden könnte. Beides spricht dafür, sich bei der Entwicklung solcher Vorhersagemodelle an den theoretischen Modellen und empirischen Ergebnissen der Forschung zur Nachrichtenauswahl zu orientieren.

4.4 Synthese: Datengestützte Prognosen der journalistischen Selektivität als Verfahren zur Prüfung von Theorien der Nachrichtenauswahl

Prognosen sind eigentlich ein Kernbestandteil der Prüfung wissenschaftlicher Theorien. Indem Vorhersagen aus einer Theorie abgeleitet und mit der Realität verglichen werden, kann getestet werden, wie gut die Realität abgebildet wird. Vorhersagen ermöglichen die Falsifizierung von Theorien. Damit können sie auch in der Kommunikationswissenschaft eingesetzt werden, um beispielsweise Annahmen über die Auswahlkriterien von Journalisten zu testen. Bisher ist das jedoch nur in seltenen Fällen passiert, nur wenige Studien haben versucht, beispielsweise die Verbreitung von Nachrichtenbeiträgen vorherzusagen. In der Informatik sind Prognosen dagegen wichtiger Bestandteil des methodischen Arsenal, oft werden dafür Machine Learning-Verfahren verwendet. Auch für die Popularität von Medieninhalten wurden diverse Vorhersagemodelle entwickelt.

Im Vergleich der in dieser Disziplin verwendeten Methoden mit den wenigen Verfahren aus der Kommunikationswissenschaft zeigt sich das Spannungsverhältnis, in dem Prognosen in der Wissenschaft immer stehen. Einerseits sollen sie möglichst genaue Vorhersagen liefern, andererseits soll noch interpretierbar sein, wie sie zu ihrer Vorhersage gekommen sind. Verfahren, die über 100 Features verwenden, um die Popularität eines Artikels vorherzusagen, mögen sehr genaue Ergebnisse erreichen, ihr inhaltlicher Nutzen für das Verständnis davon, wie Popularität zustande kommt, ist jedoch begrenzt. In dieser Arbeit soll die Balance dadurch erreicht werden, dass einerseits ein Machine Learning-Verfahren aus der Informatik verwendet wird, das sich gut für Vorhersagen eignet und andererseits wenige theoretisch begründete Faktoren in das Modell eingebaut werden. Es gibt bereits einige Arbeiten, in denen die Aufmerksamkeit für wissenschaftliche Studien vorhergesagt werden sollte. Gerade das Modell von MacLaughlin und Kollegen kann methodisch als Vorbild für diese Arbeit gesehen werden, allerdings legen sie zu wenig Wert auf die theoretische Auswahl der Prädiktoren, weswegen manche ihrer Ergebnisse wenig Erkenntniswert haben.

5. Methode

Im Überblick über die Forschungsliteratur hat sich eine bedeutende Lücke der Forschung zur Nachrichtenauswahl von Journalisten gezeigt: die wenigsten Analysen haben sich mit Auswahlentscheidungen im eigentlichen Sinne beschäftigt. In Inhaltsanalysen wurde oft nur die Berichterstattung untersucht und aus der Gewichtung von Ereignissen Rückschlüsse auf Auswahlkriterien gezogen. In der Gatekeeper-Forschung wurden zusätzlich vor allem Befragungen durchgeführt, die allerdings auch nur Erkenntnisse darüber zulassen, welche Einflüsse Journalisten nennen (wollen), nicht aber, wie welche Einflüsse die Auswahl prägen. Um das zu untersuchen, braucht es entweder Beobachtungen der Auswahl in Redaktionen oder aber Input-Output-Analysen, die die Inhalte, aus denen Journalisten auswählen könnten, denen gegenüberstellen, die ausgewählt wurden. Auch wenn beide Methoden ihre Schwächen haben (beide Untersuchungsdesigns erlauben es beispielsweise nicht direkt, die Überlegungen nachzuvollziehen, die zu einer Auswahl geführt haben), können sie doch eine Grundlage für die Untersuchung von Merkmalen sein, die mit der Auswahl zusammenhängen. In dieser Untersuchung wurde die besondere Situation des Journalismus ausgenutzt, dass die Menge möglicher Anlässe, aus denen ausgewählt wird, relativ gut bekannt ist. Das ermöglicht die Durchführung einer Input-Output-Analyse, in der für ein Jahr aus einem Großteil aller veröffentlichten Studien diejenigen identifiziert wurden, die mediale Aufmerksamkeit erhielten. Kombiniert wurde die Input-Output-Analyse mit einem Machine-Learning-

Verfahren, das einerseits basierend auf der gefundenen Auswahl eine Prognose zukünftiger Auswahlentscheidungen ermöglicht und andererseits zeigt, welchen Einfluss bestimmte Merkmale der Studien auf die Auswahl haben.

5.1 Input-Materialien

Das Interesse der Untersuchung gilt speziell dem Typ des Wissenschaftsjournalismus, der aktuell über Ereignisse berichtet, die in der Wissenschaft generiert wurden. Meistens handelt es sich bei diesen Ereignissen um die Publikation neuer Forschungsergebnisse. Durch die digitale Veröffentlichung der überwiegenden Mehrheit wissenschaftlicher Zeitschriften sowie die Archivierung und Indizierung in wissenschaftlichen Literaturdatenbanken sind diese Ereignisse leicht zugänglich und nachvollziehbar. Damit kann ein Großteil des Ereignishorizonts erfasst werden, aus dem Wissenschaftsjournalisten auswählen können, wenn sie über neue Forschungsergebnisse berichten. Vollständig untersucht werden kann er jedoch nicht, da einerseits neben Veröffentlichungen in wissenschaftlichen Zeitschriften auch andere Ereignisse Berichterstattung auslösen können, vor allem wissenschaftliche Konferenzen, Monografien und Sammelbände oder weniger standardisierte Veröffentlichungsformen wie Projektberichte oder Preprints (bzw. insgesamt die sogenannte graue Literatur). Dazu kommt, dass auch die großen Literaturdatenbanken wie Web of Science oder Scopus natürlich nicht alle wissenschaftlichen Zeitschriften abdecken, die weltweit veröffentlicht werden. Gerade bei nicht-englischsprachigen (und damit häufiger geistes- oder sozialwissenschaftlichen Zeitschriften) ist die Abdeckung meist eher schlecht. Trotzdem bieten sie insbesondere für die Naturwissenschaften ein relativ vollständiges Bild der *einflussreichen* internationalen Zeitschriften.

Verschiedene Quellen bieten sich an, um den Input für den Wissenschaftsjournalismus zu erheben. Die älteste und vermutlich am meisten genutzte Literaturdatenbank ist das Web of Science, inzwischen betrieben von Clarivate Analytics.¹⁷ 2004 kamen zwei große Konkurrenten hinzu, Scopus¹⁸, das durch den Verleger Elsevier betrieben wird sowie Google Scholar¹⁹. Seit 1996 gibt es zudem für die Biomedizin und Life Sciences mit PubMed²⁰ eine frei zugängliche Alternative, die Zugriff auf die MEDLINE-Datenbank der United States National Library of Medicine bietet. Während der freie Zugriff und eine kostenlose API für automatisierte Abfragen PubMed eigentlich zu einer guten und mit geringerem Aufwand durchsuchbaren Quelle für

¹⁷ <http://apps.webofknowledge.com>

¹⁸ <https://www.scopus.com/>

¹⁹ <https://scholar.google.de/>

²⁰ <https://www.ncbi.nlm.nih.gov/pubmed/>

solche Untersuchungen machen, fällt die Beschränkung auf einen Teilbereich der Wissenschaft schwer ins Gewicht. Auch wenn in der Forschungsliteratur häufig eine Dominanz medizinischer Studien in der Berichterstattung konstatiert wird, wird auch, und in manchen Studien in hohen Anteilen, über andere Wissenschaftsbereiche berichtet, die über PubMed nicht erfasst werden können. Da der Ereignishorizont des Wissenschaftsjournalismus möglichst umfangreich in die Analyse eingehen soll, muss PubMed daher als Datenquelle ausgeschlossen werden. Google Scholar bietet im Vergleich zu den anderen beiden Datenbanken eine größere Abdeckung auch von nicht-englischsprachigen Zeitschriften, Konferenzberichten oder Preprints (vgl. Bakkalbasi et al. 2006; Meho/Yang 2007) und ist ebenfalls frei zugänglich. Dafür ist unklar, welche Quellen insgesamt in die Datenbank mit einfließen und es gibt einen beträchtlichen Anteil an nicht seriösen Treffern, beispielsweise Websites, auf denen Haus- oder Abschlussarbeiten hochgeladen werden können oder Zeitschriften, die den Qualitätsstandards anderer Datenbanken nicht entsprechen. Hinzu kommt noch, dass es bei Google Scholar keine Möglichkeit des Exports von Informationen zu den Ergebnissen einer Suchanfrage gibt, es werden beispielsweise auch keine Abstracts verzeichnet.

Damit verbleiben Web of Science und Scopus als mögliche Quellen. Scopus deckt insgesamt eine etwas größere Zahl von Zeitschriften ab, dafür reichen die Daten von Web of Science weiter zurück (vgl. Bakkalbasi et al. 2006: 2). Da es in der Untersuchung um die Berichterstattung über relativ aktuelle Studien geht, ist dieser Vorteil des Web of Science nicht relevant. Wichtiger für die Auswahl ist aber ein anderer Vorteil von Scopus. Während es beide Datenbanken erlauben, die Ergebnisse von Suchen zu exportieren, ist die Exportfunktion beim Web of Science stark eingeschränkt, einerseits können pro Export nur 500 Einträge ausgewählt werden, andererseits ist die Auswahl der Datenformate vor allem auf verschiedene Literaturverwaltungsprogramme sowie tab-getrennte Textdateien beschränkt. Scopus erlaubt dagegen je nach Menge der Informationen, die zu einer Studie exportiert werden, zwischen 2000 und 20.000 Einträge pro Export,²¹ diese können im CSV-Format ausgegeben werden, das sich gut für eine Weiterverarbeitung in den verschiedensten Auswertungsprogrammen eignet. Das stellt eine deutliche Reduzierung des Aufwands bei ungefähr zwei Millionen zu exportierenden Einträgen dar. Daher wurde Scopus als Quelle für die Erzeugung eines Datensatzes gewählt, der den Output des Wissenschaftssystems in einem Jahr und damit den Input für den Wissenschaftsjournalismus repräsentiert.

²¹ Für die reinen Zitationsinformationen, also Titel, Autoren, Quelle etc. können 20.000 Treffer exportiert werden, für darüber hinausgehende Informationen ist der Export auf 2000 Studien beschränkt.

Als Untersuchungszeitraum wurde ein Jahr gewählt, vom 1. August 2016 bis zum 31. Juli 2017. Dieser Zeitpunkt liegt etwas weiter zurück, wodurch vermieden wird, dass die Scopus-Abdeckung durch Nachzügler verringert wird, die bereits online veröffentlicht wurden, aber noch nicht im Print und daher nicht in Scopus enthalten sind. Gleichzeitig sind für 2016 auch schon relativ umfangreiche Daten von Altmetric vorhanden, die später für die Output-Seite der Analyse verwendet werden. In diesem Zeitraum sind deutlich über 1,5 Millionen Studien bei Scopus eingestellt worden.²² Dazu, wie viele es genau sind, erhält man unterschiedliche Angaben bei Scopus. Sucht man nach einem Veröffentlichungsjahr, in der Syntax der Scopus-Suche „PUBYEAR = 2016“, werden etwas über 2,1 Millionen Treffer angezeigt. Teilt man die Suche dagegen in die einzelnen Monate auf, also „PUBDATETXT((august 2016) OR (september 2016) OR ... OR (july 2017))“, erhält man knapp 1,7 Millionen Treffer.²³ Für eine vollständigere Abdeckung wäre es demnach besser gewesen, die erste Variante für die Suche zu verwenden. Hier gibt es jedoch das Problem, dass maximal 20.000 Studien pro Export ausgewählt werden können. Und da zusätzlich immer nur die 20.000 Treffer ausgewählt werden können, die zuerst angezeigt werden können, längere Trefferlisten also nicht Schritt für Schritt abgearbeitet werden können, muss der gesamte Studienkorpus über mehrere Suchen erschlossen werden, die jeweils maximal 40.000 Treffer ergeben.²⁴

Eine erste Möglichkeit zur Unterteilung aller Studien besteht darin, die Subject Areas zu nutzen, die Scopus vorgibt. Studien werden dabei einzelnen Themengebieten wie Medizin oder Sozialwissenschaften zugeordnet, insgesamt 27. Während die Trefferzahl für Bereiche wie „Dentistry“ klein genug ist, um ein Jahr in einem Export vollständig zu erfassen, ist insbesondere die Medizin mit ca. 500.000 Studien, die in einem Jahr eingestellt wurden, zu groß, als dass allein diese Unterteilung genügen würde. Daher wird eine zweite Unterteilung benötigt, wofür sich eigentlich nur die einzelnen Monate eines Jahres anbieten. Das heißt, aufgebaut wurde die Suche dann insgesamt aus einzelnen Suchschritten, die folgende Syntax verwenden: „SUBJAREA(Themenbereich) AND DOCTYPE(ar OR re) AND PUBDATETXT((august 2016) OR (september 2016) OR (october 2016) ... OR (june 2017) OR (july 2017))“. Wenn die Trefferzahl zu groß war, wurde die Suche wiederum in einzelne Monate oder Monatsgruppen unterteilt. In Einzelfällen mussten bei Studien aus dem Bereich Medizin zusätzlich Article und Reviews getrennt exportiert werden, da die Zahl der Studien in

²² Da es vor allem um Forschungsergebnisse im engeren Sinne geht, wurde die Suche auf Articles und Reviews beschränkt, das heißt, die Suchstrings wurden um „DOCTYPE(ar OR re)“ ergänzt.

²³ Das könnte beispielsweise daran liegen, dass in manchen Studien kein oder ein fehlerhafter Monat eingetragen wurde, letztendlich lässt sich die Differenz jedoch nicht aufklären.

²⁴ 40.000 Studien können exportiert werden, indem die Treffer umsortiert werden, also einmal die Autoren von A-Z, einmal von Z-A, die beiden Exporte können dann zusammengefügt werden und Duplikate entfernt werden.

einem Monat knapp oberhalb von 40.000 lag. Mit diesem Verfahren wurden alle Studien aus dem genannten Zeitraum im Rahmen eines Projektseminars von Studentengruppen exportiert. Ein Vorteil der Unterteilung in kleinere Dateien, die durch Scopus vorgegeben wird, ist, dass sich diese besser für den späteren automatisierten Abruf eignen als eine große Datei, die alle Studien enthält. Bei einem Abruf der Daten zu 1,7 Millionen Studien würde einerseits viel Zeit benötigt und es besteht andererseits die Gefahr, dass der Abruf abbricht und von vorne gestartet werden muss. Zusammengefasst wurde der Input durch eine Reihe von Dateien mit Listen von bis zu 40.000 Studien gestellt, unterteilt in verschiedene Themenbereiche und Zeiträume. Exportiert wurden: der Studientitel; der Digital Object Identifier (DOI)²⁵; die Autoren der Studie; das Journal, in dem die Studie veröffentlicht wurde sowie Seitenzahlen, Jahrgang und Ausgabe; das Veröffentlichungsjahr; die Zahl der Zitationen, die Scopus für die Studie verzeichnet und der Dokument- und Quellentyp.

5.2 Output-Materialien

Während es Literaturdatenbanken schon länger erlauben, den Input für eine Analyse wie diese relativ einfach zu erheben, hat sich erst in den letzten Jahren durch andere Datenanbieter die Möglichkeit ergeben, auch den Output umfänglich zu erfassen. Klassisch hätte es zwei Möglichkeiten gegeben, in der Berichterstattung die Menge der Studien zu erfassen, über die tatsächlich berichtet wurde. Einerseits kann einfach generell in einer bestimmten Zahl von Medien nach der Berichterstattung über Wissenschaft gesucht werden, also zuerst der Output bestimmt werden. In allen Artikeln, in denen Forschungsergebnisse erwähnt werden, kann dann nach Referenzen auf Studien gesucht werden und versucht werden, diese Studien in Literaturdatenbanken zu identifizieren. Diese Methodik wurde beispielsweise von Suleski und Ibaraki (2010) verwendet, um anhand der Berichterstattung in zwei Medientiteln den Anteil aller Studien abzuschätzen, die in die Medien gelangen. Andererseits besteht eine zweite Möglichkeit darin, von einer begrenzten Menge von Studien auszugehen, beispielsweise zu einem bestimmten Forschungsthema, bei dem versucht wird, die Forschung möglichst umfänglich abzubilden. Ausgehend von diesem Input kann dann in Pressedatenbanken wie LexisNexis in einem größeren Sample von Medientiteln nach Berichterstattung zu den Studien gesucht werden. Dafür müssen Suchstrings entwickelt werden, mit denen man möglichst viele Artikel zu den Studien findet. So gingen unter anderem Dumas-Mallet et al. (2017) in ihrer

²⁵ DOIs sind Zeichenfolgen, die durch die International DOI Foundation für verschiedenste Objekte zugewiesen werden, beispielsweise für Studien, aber auch andere Dokumente, Datensätze, Programme... Die Zeichenfolgen bieten die Möglichkeit, die Objekte eindeutig zu identifizieren, so gut wie jede Studie hat eine DOI, mit der sie über verschiedene Plattformen hinweg erkannt wird.

Untersuchung der Berichterstattung über Studien zu 12 verschiedenen Krankheiten vor. Die Schwäche beider Ansätze ist vor allem die beschränkte Reichweite, die sich durch die notwendigen händischen Schritte ergeben. Im ersten Fall muss die Berichterstattung nach Referenzen auf Studien durchsucht und diese Referenzen dann richtig zugeordnet werden, im zweiten Fall müssen Suchstrings entwickelt und auf die einzelnen Studien angepasst werden. In dieser Untersuchung wurden stattdessen Daten verwendet, die auf einer automatisierten Suche nach Referenzen auf wissenschaftliche Ergebnissen in über 2.000 internationalen Medientiteln basieren. Diese Daten wurden vom altmetrics-Anbieter Altmetric²⁶ abgerufen.

Altmetrics haben sich in den letzten Jahren als alternative Methode zur Messung des Impacts von Forschung etabliert, die sich im Gegensatz zu Zitationsstatistiken wie dem Impact Factor oder H-Index vor allem auf kunden- und nutzerorientierte Daten aus den verschiedensten Quellen stützt (vgl. Franzen 2015: 225). Impact wird hier auf der Ebene des einzelnen Beitrags in verschiedenen Dimensionen gemessen, die vor allem auf den gesellschaftlichen Einfluss von Forschung abzielen. Damit sind altmetrics eine Antwort auf den zunehmenden Bedarf nach Indikatoren, die in der Lage sind, den Impact von Forschung auf bestimmte Gesellschaftsteile zu messen, vor allem in der Form von Interaktion oder Kommunikation mit gesellschaftlichen Stakeholdern (vgl. Bornmann 2014: 896–897). Gerade Daten aus sozialen Netzwerken werden verwendet, um den Gebrauch von Forschungsressourcen messen zu können (vgl. Haustein et al. 2014: 1146). Daneben werden oft auch Online-Leserzahlen, Erwähnungen in Blogs oder eben auch in Online-Medien gemessen. Durch den „Gebrauch“ von Studien in diesen verschiedenen Plattformen soll dann das öffentliche Engagement für Forschung gemessen werden (vgl. Bornmann 2014: 896). Für diese Untersuchung bieten sich altmetrics besonders an, da eine der Quellen, in der verschiedene Anbieter die Aufmerksamkeit für wissenschaftliche Ergebnisse messen, Online-Nachrichtenmedien sind.

Vorteile von altmetrics gegenüber Zitationsstatistiken sind unter anderem die schnellere Messung von Impact, die Erfassung eines breiten Spektrums an Forschungsergebnissen, nicht nur von Studien sowie, für diese Untersuchung besonders relevant, die leichte Zugänglichkeit. Ein Nachteil ist dagegen die Abhängigkeit von den beobachteten Datenquellen und Suchverfahren der altmetrics-Anbieter, die dazu führen kann, dass technische Veränderungen auch die Nutzungsstatistiken ändern, ohne dass sich die Nutzung geändert hat. Das erschwert die Replikation der altmetrics-Daten (vgl. ebd.: 898–900). Hinzu kommt noch, dass nicht bei allen Metriken klar ist, was eigentlich gemessen wird und ob damit ein relevanter Einfluss

²⁶ <https://www.altmetric.com/>

erfasst wird, der in Verbindung zur wissenschaftlichen Bedeutung steht (vgl. Sud/Thelwall 2014: 1132–1133). Altmetrics unterscheiden nicht zwischen positiver oder negativer Resonanz oder der Bedeutung der Referenz, was es für den Fall der Medien schwierig macht, die Erwähnungen zu vergleichen, schließlich ist unklar, ob eine Studie der Anlass für einen Artikel war oder nur am Rande oder als eine Art Quellenverweis erwähnt wurde.

Einer der ältesten Anbieter von altmetrics ist der Open-Access-Verleger PLOS, der sie in Form sogenannter article-level metrics 2009 für alle seine Journals einführte (beispielsweise PLOS One oder PLOS Biology). Erhoben werden seitdem Aufrufzahlen, Erwähnungen in bestimmten Blogs, auf Wikipedia und in sozialen Netzwerken, Zitationen in verschiedenen Datenbanken sowie Verwendung in Literaturverwaltungsprogrammen (vgl. Fenner 2013). Inzwischen sind drei weitere große Anbieter hinzugekommen, ImpactStory, Altmetric und PlumAnalytics, deren Indikatoren von verschiedenen Verlegern und Institutionen genutzt werden (vgl. Franzen 2015: 228). Die einzelnen Anbieter unterscheiden sich etwas in den Quellen, die sie jeweils beobachten und den Zielgruppen, die sie bedienen. Plum Analytics beispielsweise richtet sich vor allem an institutionelle Nutzer, beispielsweise an Zeitschriften oder Universitäten. Für diese Arbeit bot sich vor allem der Anbieter Altmetric an, weil er einerseits Erwähnungen in Massenmedien für eine große Bandbreite an Studien erhebt und andererseits über eine Programmierschnittstelle, eine API, automatisiert auf die Daten zugegriffen werden kann, die Altmetric erhoben hat (vgl. Adie/Roe 2013). Beim Zugriff auf die API können beispielsweise DOIs verwendet werden, wodurch die Abfrage gut mit den aus Scopus exportierten Listen kompatibel ist. Für eine gegebene DOI gibt die API die Daten zurück, die bei Altmetric dazu vorliegen, das beinhaltet vor allem die Scores für die verschiedenen Quellen, die Altmetric trackt, unter anderem auch soziale Netzwerke wie Facebook oder Twitter, Blogs oder auch Wikipedia.

Die Forschung zu altmetrics hat sich bisher vor allem mit der Abdeckung in verschiedenen Wissenschaftsbereichen (vgl. bspw. Hammarfelt 2014; Mohammadi/Thelwall 2014; Zahedi et al. 2014) oder der Korrelation mit üblichen Zitationsstatistiken (bspw. Thelwall et al. 2013) beschäftigt. Die Eignung für die Untersuchung medialer Resonanz wurde dagegen bisher kaum untersucht, abgesehen von den Studien, die im 4. Kapitel erwähnt wurden (Chapa et al. 2017; Siravuri/Alhoori 2017; MacLaughlin et al. 2018). Von diesen haben sich aber nur MacLaughlin und Kollegen näher mit der Qualität der Daten beschäftigt und vor allem auf das Vorkommen von Quellen verwiesen, die keine Massenmedien im klassischen Sinne darstellen, beispielsweise Portale für Pressemitteilungen oder Aggregatoren, die Pressemitteilungen oder Abstracts einfach kopieren. Die Liste der Medientitel, die Altmetric nach Referenzen auf

Studien durchsucht, wurde von Altmetric händisch und unsystematisch erstellt und ist auch eher schlecht einsehbar. Das heißt, man hat keinen Überblick darüber, wie verlässlich der Mainstream-Media-Score (MSM-Score) tatsächlich die Resonanz von Studien in der medialen Öffentlichkeit widerspiegelt. Eine Konsequenz ist, dass ein quantitativer Vergleich verschiedener Scores zwischen Studien wenig Sinn macht, man kann nicht begründet sagen, dass eine Studie mit MSM-Score = 50 doppelt so viel Aufmerksamkeit erhalten hat wie eine Studie mit MSM-Score = 100.

Um die Reliabilität der Daten von Altmetric vor der eigentlichen Analyse zu prüfen, wurden verschiedene Tests durchgeführt. Einerseits wurden Abfragen mit der gleichen Liste von DOIs zu verschiedenen Tageszeiten durchgeführt, um die Stabilität der API zu testen. Die Ergebnisse des Tests waren eindeutig, die Abfragen ergaben zu jedem Zeitpunkt gleich viele Resultate. Zusätzlich war bei ersten explorativen Abfragen aufgefallen, dass der Anteil der Studien, zu denen keine Daten zurückgeliefert werden, je nach Bereich relativ hoch ist, zum Teil bei über 70 %. Um sicherzustellen, dass hier keine Studien übersehen wurden, die tatsächlich Medienresonanz erhalten hatten, wurden verschiedene dieser „Missing Values“ händisch in LexisNexis sowie über Google gesucht. Zu keiner Studie aus der Teststichprobe konnten Medienberichte gefunden werden. Das spricht dafür, dass einerseits bei den DOIs, für die die Abfrage nichts zurückliefert, der Altmetric-Score einfach insgesamt bei 0 liegt (also in allen Quellen zusammengenommen) und damit keine Daten vorliegen und dass andererseits ein MSM-Score von 0 auch ein relativ verlässliches Indiz dafür ist, dass über eine Studie nicht berichtet wurde. Bei der bereits dargestellten Menge wissenschaftlicher Studien, die jedes Jahr veröffentlicht wird, erscheint es auch plausibel, dass die überwiegende Mehrheit keine Resonanz innerhalb der Wissenschaft erhält, geschweige denn außerhalb des Wissenschaftssystems.

Insgesamt wurde auf Grundlage dieser Tests entschieden, die MSM-Scores als Indikator für mediale Resonanz zu verwenden. Um diese von Altmetric abzurufen, wurde ein Python-Script entwickelt, das eine csv-Datei mit DOIs als Input verwendet, diese DOIs ausliest und für jede einzelne eine Anfrage an die API schickt. Die von Altmetric zurückgeschickten Scores werden dann mit den DOIs in eine neue csv-Datei geschrieben. Konkret wurde das Script dann für die verschiedenen Dateien ausgeführt, die im ersten Schritt aus Scopus exportiert wurden. Die Einzeldateien aus der Abfrage wurden erst auf der Ebene der Subject Areas zusammengeführt und mögliche Doppler entfernt, dann wurden die übergreifenden Dateien zu einem Masterdatensatz zusammengeführt. Auch hier mussten wieder Doppler entfernt werden, da viele Studien gleichzeitig verschiedenen Subject Areas zugewiesen werden. Eine später

entwickelte Version des Scripts kann zusätzlich von den zurückgelieferten Daten nur den MSM-Score auswählen und an die Zeile in der csv-Datei aus Scopus anhängen, aus der die DOI stammt. Das heißt, das Ergebnis ist dann eine Datei mit den Informationen aus Scopus sowie dem MSM-Score. Diese Variante des Scriptes wurde jedoch nur für Studien ausgeführt, deren Score oberhalb von 50 liegt, bei denen man also mit relativer Sicherheit von einer fokussierten Aufmerksamkeit ausgehen kann. Als Ergebnisse liegen dann zwei Masterdateien für den Zeitraum vor, eine Datei enthält Scores für alle Studien, zu denen Altmetric Daten hatte, die zweite die MSM-Scores und Informationen aus Scopus für alle Studien mit einem MSM-Score größer gleich 50. Dieser Datensatz wurde vor allem für die Identifizierung einflussreicher Journals sowie der Untersuchung der Verteilung journalistischer Aufmerksamkeit verwendet.

Bei der Verwendung des ersten Datensatzes für die Untersuchung des Auswahlverhaltens von Wissenschaftsjournalisten muss berücksichtigt werden, dass der genaue MSM-Score nicht wirklich interpretiert werden kann. In dieser Untersuchung geht es ja in erster Linie aber auch darum, ob eine Studie überhaupt von Journalisten ausgewählt wurde. Um diese Auswahl nachzuvollziehen, gibt es nun verschiedene Möglichkeiten. Es können einerseits alle Studien mit $\text{MSM-Score} \geq 1$ als Studien mit Medienaufmerksamkeit definiert werden, wobei bekannte Pressemitteilungsportale wie Alpha Galileo oder EurekAlert aus dem Score entfernt werden. Mit diesem intuitiven Grenzwert hat man vermutlich eine relativ große Gruppe ausgewählter Studien, dafür ist das Risiko relativ hoch, auch Studien miteinzuschließen, die keine Medienaufmerksamkeit erhalten haben, sondern nur von Aggregatoren oder ähnlichen Quellen kopiert wurden. Um dieses Risiko zu verringern, kann man höhere, mehr oder weniger willkürliche Grenzwerte setzen, beispielsweise bei einem Score von 10 oder 50. Dadurch wird allerdings die Gruppe ausgewählter Studien immer kleiner, wodurch die Vorhersageaufgabe schwieriger wird. In der Auswertung wurden Vorhersagemodelle für alle Grenzwerte erstellt, später werden die Ergebnisse verglichen.

5.3 Prädiktoren

Mit den Daten von Scopus und Altmetric kann ein Großteil aller möglichen Ereignisse, aus denen Wissenschaftsjournalisten auswählen konnten, der tatsächlichen Auswahl gegenübergestellt werden. Was dann noch fehlt, sind Merkmale, mit denen versucht werden kann, die beobachtete Auswahl zu erklären. Aus der Übersicht über den Forschungsstand zur Nachrichtenauswahl sowie insbesondere dem in Kapitel 3 entwickelten theoretischen Modell lassen sich besonders zwei mögliche Faktoren ableiten: das Journal, in dem eine Studie veröffentlicht wurde, sowie das Vorkommen einer Pressemitteilung. Zusätzlich kann auch von

einem Einfluss des Inhalts der Studien auf die Auswahl ausgegangen werden, wobei sich gerade für dieses Merkmal bei der großen Studienanzahl das Problem stellt, wie das Thema erfasst werden soll. Daher wurde aus Verfahren aus der Informatik beziehungsweise speziell aus dem Bereich des Natural Language Processing zurückgegriffen. Die Auswahl und Erstellung der einzelnen Prädiktoren orientierte sich auch daran, welche Informationen einerseits den auswählenden Journalisten und andererseits auch später im praktischen Betrieb eines Prognoseverfahrens zur Verfügung stehen. Damit fallen Informationen weg, die nur auf Scopus verfügbar sind, da Studien hier erst nach ihrer Veröffentlichung eingestellt werden, während das Prognoseverfahren kurz vor oder bei der Veröffentlichung eingesetzt werden soll.

5.3.1 Publikationsort

Der Publikationsort kann relativ einfach erhoben werden, da er ja als eine Information in den Scopus-Exporten mitgeliefert wird. Die Zeitschriftentitel wurden normalisiert, das heißt in Kleinbuchstaben umgewandelt und Formatierungen entfernt, für die Analyse wurden sie in Kategorien umgewandelt. Um neben dem Titel der Zeitschrift bereits eine Information über die Wahrscheinlichkeit einfließen lassen zu können, dass Studien aus einer Zeitschrift ausgewählt werden, wurde für alle Zeitschriften, die zwischen August 2015 und Juli 2018 eine Studie mit einem MSM-Score größer gleich 50 veröffentlicht haben, der Anteil an allen Studien mit einem MSM-Score größer gleich 50 berechnet. Die Liste dieser Zeitschriften wurde in einer Altmetric-Abfrage erstellt, die für eine andere Untersuchung verwendet wurde, allerdings dem gleichen Vorgehen folgt und das oben beschriebene zweite Script verwendet hat. Das heißt, für alle Zeitschriften liegen dann zwei zusätzliche Informationen vor, einerseits die absolute Zahl an relevanten Studien von Mitte 2015 bis Mitte 2018 und andererseits der Anteil an den relevanten Studien, der ein Maß für die Bedeutung dieser Zeitschrift für den Wissenschaftsjournalismus darstellen soll. Bei Zeitschriften, die keine relevante Studie veröffentlicht haben, liegt dieser Wert bei 0, die Summe aller Auswahlwahrscheinlichkeiten ergibt 1.

5.3.2 Verfügbarkeit von Pressemitteilungen

Das Vorkommen einer Pressemitteilung ist etwas schwieriger zu erheben als der Publikationsort, jedoch kann man sich hier eine Besonderheit des MSM-Scores zunutze machen. Wie oben bereits erwähnt wurde, werden als Teil dieses Scores auch EurekaAlert und AlphaGalileo, also die beiden größten internationalen Portale für wissenschaftliche Pressemitteilungen, nach Erwähnungen von Studien durchsucht. Während über die API kein direkter Zugriff auf die Daten möglich ist, aus denen sich die verschiedenen Scores ergeben, also auf die Erwähnungen selbst, bietet Altmetric mit dem Explorer eine weitere

Benutzeroberfläche zum Datenzugriff. Diese eignet sich weniger gut für den großflächigen Datenabruf, der mit der API möglich ist, dafür kann man hier für Listen von DOIs nach Erwähnungen in speziellen Quellen suchen. Für alle Studien, für die in der Abfrage ein MSM-Score von mindestens 1 ermittelt wurde, wurde im Explorer geprüft, ob eine Erwähnung in Alpha Galileo oder EurekAlert vorhanden war. Diese Informationen konnten für zwei Zwecke genutzt werden. Einerseits kann so aus dem Datensatz der Einfluss der beiden Portale auf den MSM-Score entfernt werden, indem dieser um eins oder zwei verringert wird, wenn eine Studie auf einer der beiden oder beiden Seiten erwähnt wurde. Und andererseits kann damit eine Variable „Pressemitteilung“ erstellt werden, deren Wert auf True gesetzt wird, wenn eine Erwähnung in einer der Quellen vorlag und auf False, wenn keine Erwähnung vorlag. Studien mit einem MSM-Score von 0 können auch keine Pressemitteilung erhalten haben, daher wurde die Variable Pressemitteilung für sie auf False gesetzt.

Damit ist für alle Studien aus dem Masterdatensatz die Information gegeben, ob es eine Pressemitteilung gab, die auf EurekAlert oder Alpha Galileo veröffentlicht wurde. Natürlich kann es auch Fälle geben, in denen es eine Pressemitteilung gab, diese aber nicht in diesen Portalen eingestellt wurde. Dafür, dass dies eher die Ausnahme ist, spricht die hohe Bedeutung, die Journalisten diesen Portalen zuweisen (vgl. bspw. Granada 2011: 801). Zudem finden sich hier auch die Pressemitteilungen der großen Zeitschriften. Insgesamt erscheint diese Möglichkeit der Erfassung von Pressemitteilungen die beste Alternative, da beide Portale die größten zentralen Quellen für internationale Pressemitteilungen darstellen und es im Rahmen dieser Arbeit keine Möglichkeit gab, eigene Verfahren zur Auswertung von EurekAlert und Alpha Galileo (bspw. über eigene Webscraper) zu entwickeln.

5.3.3 Sprachmodelle

Ein weiterer Faktor, der die Nachrichtenauswahl von Wissenschaftsjournalisten beeinflusst, ist der Inhalt der Studien. Sowohl die Disziplin, aus der eine Studie stammt, als auch das übergreifende Thema sowie verschiedene Nachrichtenfaktoren wurden als potenziell einflussreich für die Selektion identifiziert. Anders als beim Publikationsort oder dem Vorkommen einer Pressemitteilung ist es bei diesen Faktoren deutlich schwieriger, in großem Umfang die Bedeutung zu bestimmen, da es sich um komplexe Variablen handelt, bei denen mitunter (wie beim Thema) die Anzahl möglicher Ausprägungen vorher nicht bekannt ist. Eine erste, weniger aufwendige Lösung, um zumindest über Themen Aussagen machen zu können, könnte darin bestehen, die Themenbereiche oder Keywords zu nutzen, die Scopus vergibt. Hier gibt es jedoch einerseits das Problem, dass diese Informationen zum

Veröffentlichungszeitpunkt noch nicht verfügbar sind, weil sie ja von Scopus vergeben werden und andererseits sind sie inhaltlich vermutlich nicht besonders hilfreich. Die Subject Areas von Scopus beispielsweise sind sehr breit angelegt und eine Studie kann mehreren Bereichen zugeordnet sein. Eine direkte Codierung des Themas oder von Nachrichtenfaktoren im Rahmen einer Inhaltsanalyse der Studien(-abstracts) ist allein schon aufgrund der Zahl der Studien nicht umsetzbar. Eine Stichprobenbildung ist auch nicht wirklich sinnvoll, da die Gefahr besteht, beim geringen Vorkommen von Medienaufmerksamkeit relevante Fälle auszuschließen. Zudem müsste im Fall von Nachrichtenfaktoren ein eigenes Instrument entwickelt werden, mit dem diese in Studien codiert werden können, beim Thema müsste man zuvor eine Liste möglicher Ausprägungen explorativ identifizieren oder theoretisch herleiten. Beides wäre relativ aufwendig und angesichts der Heterogenität des Studienkorpus auch wenig erfolgversprechend. Auch eine automatisierte Codierung dieser Variablen erscheint daher nicht umsetzbar. Es gibt zwar Arbeiten zur automatisierten Identifizierung mancher Nachrichtenfaktoren (vgl. bspw. Burggraaff/Trilling 2017; Trilling et al. 2017), diese beziehen sich jedoch auf Medienbeiträge und benötigen vorher manuelle Codierung zum Training oder andere Vorarbeit.

Aussichtsreicher und etwas einfacher umsetzbar sind dagegen Verfahren, die keine Vorcodierung brauchen und alleine auf Grundlage der gegebenen Daten Themen oder ähnliche Strukturen in einem Textkorpus identifizieren. Bei Verfahren der automatisierten Inhaltsanalyse kann man grob drei Gruppen unterscheiden. Dictionary-Ansätze basieren auf genauen Codiervorgaben, beispielsweise Wortlisten, die automatisiert in einem Datensatz gesucht werden. Ein Beispiel dafür sind sogenannte Sentiment-Analysen, mit denen die Emotion von Texten bestimmt werden soll, basierend auf Listen von Worten, die eher positiv oder negativ sind. Supervised Machine Learning-Verfahren brauchen gewisse Vorgaben, beispielsweise einen vorcodierten Trainingsdatensatz, aus dem sie Ausprägungen der Zielvariable und deren Zusammenhang mit den Texten lernen können. Hierunter könnte man auch die oben genannten Arbeiten zur Identifizierung von Nachrichtenfaktoren zählen. Unsupervised Machine Learning-Verfahren dagegen erhalten vorher keinen inhaltlichen Input, hier müssen höchstens bestimmte Parameter vorher festgelegt werden wie die Themenzahl, die gefunden werden soll (vgl. zur Einteilung und zum Überblick Grimmer/Stewart 2013; Boumans/Trilling 2016; Günther/Quandt 2016). Das in der Kommunikationswissenschaft wohl bekannteste Beispiel ist Latent Dirichlet Allocation (LDA), eine Methode des Topic Modeling, die basierend auf dem gemeinsamen Vorkommen von Begriffen eine vorher festgelegte Zahl von Topics in einem Textkorpus findet, repräsentiert als Listen der Worte, die in den einzelnen

Topics besonders häufig vorkommen (vgl. Maier et al. 2018a). Man kann diese Gruppen auf einem Spektrum von deduktiv nach induktiv verorten, wobei Dictionary-Methoden deduktiv sind und unsupervised Machine Learning induktiv (vgl. Boumans/Trilling 2016: 10). Induktive Verfahren bieten sich besonders an, wenn die zu ermittelnden Kategorien (wie in diesem Fall die thematische Unterteilung) vorher nicht bekannt sind.

Grundlage von Verfahren, die zu unsupervised Machine Learning gezählt werden können, ist, dass der Text in einer Form modelliert wird, die die Anwendung von statistischen oder Machine Learning-Verfahren erlaubt. Das einfachste Verfahren sind sogenannte bag-of-words, in denen ein Dokument als Vektor reeller Zahlen dargestellt wird. Diese stehen dafür, wie oft ein bestimmtes Wort aus dem Vokabular aller Dokumente in einem Textkorpus in dem speziellen Dokument vorkommt. Das heißt: Dokumente werden reduziert auf eine Liste von Worthäufigkeiten. Meist durchlaufen Dokumente vorher verschiedene Vorverarbeitungsschritte, beispielsweise werden Worte normalisiert, also auf den Wortstamm reduziert, oder sehr häufige Wörter (Stopwords wie „und“, „der“ oder ähnliches) und sehr seltene Worte entfernt. Zudem gibt es eine populäre Erweiterung, tf-idf, in der die Häufigkeit eines Begriffs in einem Dokument ins Verhältnis zur Häufigkeit des Begriffes in allen Dokumenten gesetzt wird. So können Wörter gefunden werden, die ein Dokument besonders kennzeichnen, da sie vergleichsweise häufig vorkommen. Die erstellten Dokumentvektoren können auch zur Klassifizierung verwendet werden, indem beispielsweise Cluster gebildet werden oder die Ähnlichkeit zwischen Vektoren errechnet wird (vgl. Manning et al. 2009: 117–125). Ein Thema könnte dann dadurch erkannt werden, dass bestimmte Begriffe in einem Dokument besonders häufig vorkommen.

Ein Nachteil dieser Modellierung ist, dass sehr viele Informationen verloren gehen, einerseits wird die Wortreihenfolge vollkommen ignoriert, andererseits können Zusammenhänge zwischen Dokumenten oder Begriffen nicht richtig dargestellt werden, man kann ähnliche Begriffe nur indirekt identifizieren. Zudem haben die erstellten Vektoren sehr viele Dimensionen (so viele, wie Worte im Vokabular sind) und sind meist sehr dünn besetzt, das heißt viele Stellen des Vektors sind Null, da in einem Dokument meist nur ein Bruchteil des gesamten Vokabulars vorkommt (vgl. Le/Mikolov 2014: 2931). Daher wurden verschiedene Verfahren entwickelt, die auch die Strukturen in oder zwischen Dokumenten besser darstellen sollen. LDA beispielsweise modelliert Textsammlungen als Ergebnis latenter Inhaltsvariablen, der Topics. Die Grundidee ist, dass ein Dokument eine Mischung verschiedener Topics ist, ein Topic wiederum eine Wahrscheinlichkeitsverteilung über das Vokabular des gesamten Korpus. Die einzelnen Topics sind also dadurch gekennzeichnet, dass bestimmte Wörter besonders

häufig vorkommen, Dokumente dadurch, dass bestimmte Topics besonders wichtig sind. Das Verfahren versucht dann, aus den vorhandenen Worten in den Dokumenten auf die Themenverteilungen zu schließen, die diese Wörter am wahrscheinlichsten erzeugt haben (vgl. Blei et al. 2003: 995–997). LDA hat den Vorteil, dass das Ergebnis des Algorithmus, also die Wortlisten der einzelnen Topics sowie die Bedeutung eines Topics für ein einzelnes Dokument relativ gut interpretiert werden können, wenn sich aus den Wortlisten sinnvolle Themen ableiten lassen. Problematisch ist dafür, dass die Zahl der Themen vor der Anwendung festgelegt werden muss und man die richtige Anzahl an Themen nur anhand der Qualität der Ergebnisse mehr oder weniger valide bestimmen kann.

Eine andere Herangehensweise basiert ebenfalls auf einer Vektor-Darstellung, allerdings werden hier nicht die Dokumente als Vektoren der Begriffshäufigkeit dargestellt, sondern jeder Begriff aus dem Vokabular eines Textkorpus als Vektor über eine festgelegte Zahl von Dimensionen. Dadurch ergibt sich ein geteilter Vektorraum, in dem alle Wortvektoren angeordnet werden können und die Distanz zwischen Worten gemessen werden kann (vgl. Bengio et al. 2003: 1139). Ein solches Modell wird als „distributed representations of words“ bezeichnet. Die Repräsentation der Worte als Vektoren wird mit einem neuronalen Netzwerk, also einem Machine Learning-Verfahren erstellt, dabei werden die Darstellungen so gewählt, dass durch die Kombination von Vektoren der Worte aus einer Abfolge das nächste Wort besonders gut vorhergesagt werden kann (vgl. ebd.: 1141–1142). Das Ergebnis des Lernvorgangs ist, dass die Position der Wortvektoren im geteilten Vektorraum bestimmte semantische Zusammenhänge erfassen kann, so liegen dort ähnliche Worte näher beieinander und durch Vektoraddition können Zusammenhänge zwischen Worten nachvollzogen werden. Beispielsweise kann man die Differenz der Vektoren von „größter“ und „groß“ zum Vektor von „klein“ addieren, um „kleinster“ zu finden (vgl. Mikolov et al. 2013a: 5). Eine populäre Implementierung eines solchen Modells ist Word2Vec (vgl. Mikolov et al. 2013b).

Für diese Arbeit relevant ist eine Erweiterung des Word2Vec-Modells, Doc2Vec (vgl. Le/Mikolov 2014). Neben der Vektorrepräsentation von Worten kann dieses Verfahren auch Vektoren für ganze Sätze, Absätze oder Dokumente lernen. Das Vorgehen ist ähnlich wie bei Word2Vec, nur wird zusätzlich zu den Wortvektoren auch ein Paragraph-Vektor für Wortfolgen beliebiger Länge erstellt. Dieser wird beim Training ebenfalls dafür verwendet, das nächste Wort für eine gewisse Wortfolge vorherzusagen, dabei wird er wie ein weiteres Wort behandelt. Das heißt, bei der Vorhersage werden nicht mehr nur die Vektoren vorangegangener Wörter kombiniert, sondern zusätzlich noch eine Information über den Text verwendet, in dem sich das Wort befindet. Man kann diesen Vektor dann als eine Art abstrakte Information über

das Thema eines Texts oder Textabschnitts verstehen (vgl. dies. 2014: 2933). Die erstellten Vektoren können beispielsweise für eine Sentiment-Analyse verwendet werden und erzielen hier sehr gute Ergebnisse (vgl. ebd.: 2934–2936). Da mit diesem Verfahren ganze Dokumente oder in unserem Fall Abstracts auf eine einzigartige Darstellung in Form eines hochdimensionalen Vektors reduziert werden, die auch den Inhalt des Dokuments im Vergleich zu anderen Dokumenten erfasst, eignet es sich für eine Modellierung des Themas von Studien. Der Nachteil ist dabei, dass die Vektoren nicht direkt interpretiert werden können, das heißt, es sind später keine unmittelbaren Aussagen darüber möglich, welche Themen besonders viel Aufmerksamkeit erhalten. Denn es liegt für jedes Dokument ja nur ein hochdimensionaler Vektor vor, dessen einzelne Dimensionen nichts mit bestimmten Themenaspekten zu tun haben müssen. Dafür ist es relativ unaufwendig möglich, den Inhalt von Millionen von Abstracts in einer Form zu ermitteln, die von Prognoseverfahren gut verarbeitet werden kann. Das Verfahren eignet sich praktisch für die Erstellung eines Prognosemodells, lässt aber wenig Aussagen über die Präferenzen der Journalisten bei der Auswahl zu. Der Gewinn für die Erforschung der Auswahlpraxis liegt vor allem darin, dass es Einblicke darin ermöglicht, ob der Inhalt überhaupt relevant ist und wie groß diese Bedeutung im Vergleich zu den anderen beiden Merkmalen ist. Dafür werden später verschiedene Vorhersagemodelle mit und ohne Informationen aus den Sprachmodellen erstellt und die Ergebnisse verglichen.

Für die Erstellung des Doc2Vec-Modells wurde eine große Anzahl von Abstracts benötigt, sowohl von ausgewählten wie auch von nicht ausgewählten Studien. Da der Export von Scopus auf 2000 Studien pro Export beschränkt ist, wenn auch Abstracts ausgewählt werden, konnten nicht für alle Studien, die im untersuchten Zeitraum in Scopus eingestellt wurden, Abstracts exportiert werden. Um eine möglichst große Zahl relevanter Abstracts zu erhalten, wurde eine umfängliche Stichprobe gebildet. Dafür wurde auf Informationen aus dem oben beschriebenen zweiten Masterdatensatz zurückgegriffen, der nur Informationen zu Studien mit einem MSM-Score größer gleich 50 enthält und in einer anderen Untersuchung für die Analyse fokussierter journalistischer Aufmerksamkeit verwendet wurde (vgl. Lehmkuhl/Promies 2019). Eine Liste der Zeitschriften, von denen eine Studie in diesem Datensatz enthalten ist, wurde für den Export von Abstracts verwendet. Für alle Zeitschriften, in denen mindestens zwei Studien mit einem MSM-Score größer gleich 50 erschienen sind, wurden die Abstracts aller Articles und Reviews exportiert, die von August 2016 bis Juli 2017 erschienen sind. Für 215.990 Studien aus 540 Zeitschriften wurden Abstracts gespeichert.²⁷ So können zumindest für einen Großteil

²⁷ Die Syntax der Scopus-Suche ist dann ähnlich wie oben beschrieben, nur wird statt „SUBJAREA(Themenbereich)“ „SRCTITLE(Zeitschrift)“ verwendet.

relevanter Studien Inhaltsmodelle erstellt werden, zudem kann argumentiert werden, dass nur Studien ausgeschlossen werden, die wenig Relevanz für den Journalismus haben, wodurch die Leistung des Vorhersageverfahrens besser eingeschätzt werden kann. Denn es gibt ohnehin mehr Studien, die keine Aufmerksamkeit erhalten haben, wodurch dies leichter vorhersagbar ist. Aus den Abstracts wurden mit der Python-Implementierung von Doc2Vec im Package Gensim²⁸ Vektorrepräsentationen erstellt. Diese wurden nicht direkt verwendet, sondern für jeden Abstract die Kosinus-Ähnlichkeit des Vektors zum Mittelpunkt der Vektoren aller Abstracts berechnet, die Aufmerksamkeit erhalten bzw. nicht erhalten haben, je nach Prognosemodell MSM-Score ≥ 1 oder ≥ 10 (vgl. zu diesem Vorgehen auch Manning et al. 2009: 121). Das heißt es lag dann einmal die Ähnlichkeit zum durchschnittlichen Abstract ohne Medienaufmerksamkeit und einmal zum durchschnittlichen Abstract mit Medienaufmerksamkeit vor. So soll die thematische Ähnlichkeit zu den ausgewählten oder nicht ausgewählten Studien erfasst werden.

5.4 Prognoseverfahren: Random Forests

Mit den Informationen zum Veröffentlichungsort, dem Vorkommen einer Pressemitteilung und den Sprachmodellen aus dem Abstract ist die Grundlage für ein Prognosemodell gegeben, das auf Grundlage der Auswahlentscheidungen, die sich aus dem Vergleich von Input und Output ableiten lassen, versucht, zukünftige Entscheidungen vorherzusagen und dabei die Bedeutung der unterschiedlichen Merkmale aufzeigt. Das dafür verwendete Verfahren ähnelt den beschriebenen zur Vorhersage der Popularität von Onlineinhalten. Eine Reihe von Features wird einem Machine Learning-Verfahren übergeben, das den Wert einer Zielvariable vorhersagen soll. Dafür wird es zuerst mit einem Teil des Datensatzes trainiert, indem es bestimmte Zusammenhänge zwischen Features und Zielvariable „lernt“ und in der Folge wird mit einem weiteren Teil des Datensatzes geprüft, wie gut mit den gelernten Zusammenhängen die Zielvariable in Fällen vorhergesagt werden kann, die nicht zum Training verwendet wurden. Als Verfahren bieten sich aus verschiedenen Gründen sogenannte Random Forests an. Im Folgenden soll zuerst die Grundlage des Verfahrens, sogenannte Entscheidungsbäume, beschrieben werden, bevor die Vorzüge von Ensemble-Methoden wie Random Forests erläutert werden. Abschließend werden die konkrete Umsetzung in dieser Arbeit sowie die verwendeten Maße zur Darstellung der Genauigkeit der Vorhersage beschrieben.

²⁸ <https://radimrehurek.com/gensim/index.html>

5.4.1 Grundlage: Recursive Partitioning und Entscheidungsbäume

Recursive Partitioning oder Entscheidungsbäume sind ein einfaches nichtparametrisches Regressionsverfahren, das seit Jahren in Disziplinen wie der Biologie oder Informatik verwendet wird und dessen Vorteile langsam auch in den Sozialwissenschaften erkannt werden. Anders als klassische Regressionsverfahren wie eine logistische Regression ermöglicht es das Verfahren, auch Interaktionseffekte zu finden, die sonst übersehen würden. Es eignet sich auch für die Modellierung einer Vielzahl von Zusammenhängen, auch von nichtlinearen oder nicht monotonen Verbindungsregeln, die vorher nicht spezifiziert werden müssen (vgl. Berk 2006: 272; Strobl et al. 2009: 325). Zudem sind die Ergebnisse anders als bei anderen Machine-Learning-Verfahren relativ gut interpretierbar (vgl. Segal/Xiao 2011: 80). Der erste Algorithmus, der auf einem Regressionsbaum basiert, Automatic Interaction Detection, wurde bereits 1963 publiziert, mit der Erweiterung Theta Automatic Interaction Detection konnte das Verfahren auch zur Klassifikation verwendet werden (vgl. Loh 2014: 329–330). Die beiden populärsten Varianten sind CART (Classification und Regression Trees) und C4.5.

Grundlage von Recursive Partitioning ist ein zweistufiger Algorithmus: zuerst werden die Beobachtungen in einem Datensatz rekursiv immer weiter unterteilt und dann wird der Wert einer Ergebnisvariable für die entstandenen Untergruppen konstant vorhergesagt (vgl. Hothorn et al. 2006: 654–655). Die rekursive Unterteilung wird dabei so durchgeführt, dass die neu entstehenden Gruppen immer homogener im Wert der Zielvariable sind, die später vorhergesagt werden soll. Das Verfahren startet im Wurzelknoten, der alle Beobachtungen enthält und sucht nach der Variable und dem Unterteilungskriterium, die zu den homogensten Untergruppen führen (vgl. Kern et al. 2019: 75). Für diese Auswahl wird in den meisten Anwendungen die Reduzierung der „Impurity“ der Daten verwendet, gemessen über Bayes Error, Entropie-Funktion oder Gini-Index. Das gleiche wird für die entstandenen Untergruppen wieder durchgeführt und so weiter, bis ein Stop-Kriterium erfüllt ist. Ein solches Stop-Kriterium kann eine Mindestmenge an Fällen in einer Untergruppe sein oder ein bestimmter Wert der Reduktion der Impurity, der durch eine Unterteilung erreicht werden muss. (vgl. Strobl et al. 2009: 326–327). Das Ergebnis ist eine Top-Down-Baumstruktur, die sich ausgehend vom Wurzelknoten immer weiter verzweigt, bis in die sogenannten Blätter oder Endknoten, die nicht weiter unterteilt werden. In jedem dieser Endknoten wird dann ein Wert der Zielvariable vorhergesagt, indem bei einer Regression der Durchschnitt aller Werte gebildet wird oder bei einer Klassifikation die häufigste Klasse im Knoten (vgl. ebd.: 328). So ergibt sich eine gut interpretierbare Struktur, in der man anhand der einzelnen Unterteilungen nachvollziehen kann, welche Variablen einen Einfluss auf die Einteilung hatten und damit mit der Zielvariable

zusammenhängen könnten. Diese anhand eines Trainingsdatensatzes erstellte Struktur kann dann verwendet werden, um die Werte der Zielvariable bei neuen Fällen vorherzusagen. Dafür „wandern“ sie entlang der verschiedenen Unterteilungen durch den Baum, bis sie in einem Endknoten landen, dessen Wert dann für die Vorhersage benutzt wird.

Ein Problem bei der Verwendung von Entscheidungsbäumen für die Vorhersage ist das sogenannte Overfitting. Damit ist die Tatsache bezeichnet, dass ein Klassifikationsverfahren, das sich zu sehr an einen Datensatz angepasst hat, um diesen möglichst genau wiederzugeben, nicht nur die systematischen Komponenten der Struktur entdeckt, sondern auch die zufällige Variation der Struktur, die nur in diesem speziellen Datensatz vorliegt (vgl. Strobl et al. 2009: 327). Je genauer die Unterteilung wird, das heißt, je kleiner die Endknoten und je komplexer die Baumstruktur, desto größer ist das Risiko des Overfittings. Dadurch ergeben sich zwei Nachteile: einerseits ist die Genauigkeit bei der Erstellung des Modells zu optimistisch und andererseits passt das Modell schlecht auf neue Datensätze (vgl. Berk 2006: 277). Die Beschränkung der Größe von Bäumen durch Stop-Kriterien ist eine Möglichkeit, um Overfitting möglichst zu vermeiden, eine andere ist das sogenannte Pruning. Dabei werden Baummodelle, nachdem sie entwickelt wurden, verkleinert, indem Verzweigungen eliminiert werden, die die Vorhersagegenauigkeit nicht erhöhen (vgl. Strobl et al. 2009: 328). Der Zweck von Pruning ist es also, den besten Unterbaum eines gesättigten Baums zu finden, sodass die Qualität des Baumes möglichst hoch ist (vgl. Berk 2006: 272).

Auch das Pruning kann Overfitting jedoch nur begrenzt verhindern. Daher wurden verschiedene Erweiterungen der klassischen Entscheidungsbäume entwickelt, die dieses und weitere Probleme lösen sollen. Eine Variante sind conditional inference trees, die Variablenauswahl und Grenzwertsetzung beim Unterteilen in zwei verschiedene Schritte unterteilen. Bei der Variablenauswahl wird immer erst mit statistischen Tests geprüft, ob es einen Zusammenhang zwischen einer der unabhängigen Variablen und der Zielvariable gibt. Wenn das nicht der Fall ist, wird die Unterteilung gestoppt. Dadurch enthält das Verfahren auch direkt ein statistisch motiviertes Stopkriterium. Gibt es mindestens einen Zusammenhang, wird die Variable für die Unterteilung ausgewählt, die den stärksten Zusammenhang zeigt. Die Methode verringert das Problem des Overfittings, da nur Unterteilungen ins Modell eingehen, die einen signifikanten Zusammenhang mit der Zielvariable zeigen, gleichzeitig ist es für Daten auf beliebigen Skalen anwendbar und vermeidet eine systematische Tendenz zu Variablen mit vielen möglichen Splits oder vielen fehlenden Werten, die in klassischen Verfahren besteht (vgl. Hothorn et al. 2006).

Die unterschiedlichen Varianten von Entscheidungsbäumen bieten sich für die Auswertung einer Vielzahl von Datenformen und Fragestellungen in den Sozialwissenschaften an. Kern et al. beispielsweise sehen solche und weitere Machine Learning-Verfahren in der Umfrageforschung als flexiblen Ersatz für parametrische Regressionen und gleichzeitig als mächtige Vorhersagemethode (vgl. Kern et al. 2019: 87). Chiotta und Manzetti (2015) verwendeten conditional inference trees, um in Daten aus einer Befragung den Zusammenhang zwischen der Einschätzung der ökonomischen Situation eines Landes und der Bewertung der politischen Führung auf verschiedenen Aggregationsebenen zu prüfen. Dafür variierten sie die Zahl der Fälle pro Endknoten. Und Bubela und Caulfield (2004) untersuchten die Auswirkungen verschiedener Eigenschaften von Studien aus der Genforschung auf die Genauigkeit von Zeitungsberichten zu diesen Studien mit einem Classification Tree.

5.4.2 Ensemble-Verfahren

Eine weitere Möglichkeit zur Verbesserung der Vorhersageleistung von Entscheidungsbäumen und der Verringerung von Overfitting besteht in sogenannten Ensemble-Methoden. Damit ist die Kombination der Ergebnisse vieler verschiedener Entscheidungsbäume gemeint, die jeweils zufällige Stichproben aus dem Lerndatensatz verwenden (vgl. Strobl et al. 2009: 330). Die Idee ist, dass ein schwächeres Verfahren gute Ergebnisse erzielt, wenn es „by committee“ arbeitet (vgl. Berk 2006: 264). Zuerst werden dafür Bäume für eine Untermenge des Datensatzes gebildet und dann deren Vorhersagen kombiniert. Bei einer Regression erfolgt das über einen gewichteten oder ungewichteten Durchschnitt, bei einer Klassifizierung durch ein Mehrheitsvotum über alle Bäume. Durch die Verwendung vieler Bäume soll vor allem vermieden werden, dass kleine Veränderungen in den Lerndaten durch die hierarchische Struktur einen großen Einfluss auf das erstellte Modell haben. Zusätzlich sind Ensembles flexibler bei der Annäherung funktionaler Formen als einzelne Bäume, die Funktionen ja immer in gewissen Sprüngen abbilden (vgl. Strobl et al. 2009: 330–331). Als Grundlage können unterschiedliche Baumvarianten verwendet werden, oft wird CART benutzt.

Das älteste Ensemble-Verfahren ist das sogenannte Bagging. Dabei wird eine Vielzahl von Bäumen aus kleinen Bootstrapping-Stichproben eines Datensatzes ohne Pruning erzeugt. Die einzelnen Bäume zeigen dadurch Overfitting für die Stichprobe, mit der sie trainiert wurden, durch die Kombination der verschiedenen Bäume wird das jedoch ausgeglichen. Ein Nachteil ist es, dass keine einzelne Baumstruktur mehr vorliegt, die interpretiert werden kann, stattdessen gibt es ja viele Bäume mit verschiedenen Strukturen. Dadurch ist es auch nicht mehr ohne weiteres möglich, die Bedeutung einzelner Variablen zu bestimmen (vgl. Berk 2006: 278–

279). Ein weiterer Ensemble-Ansatz ist Gradient Boosting. Hier wird eine Abfolge von Bäumen verwendet, in der jede Komponente eine Verbesserung des vorherigen Baums darstellt. Das heißt, die Bäume stehen nicht mehr nebeneinander, sondern hintereinander, zudem verwendet jeder Baum den kompletten Datensatz. Die Bäume in Gradient Boosting sind meist schwache Prädiktoren, das heißt sehr kurz (vgl. ebd.: 286–288).

Als letzte große Alternative gibt es Random Forests (vgl. Breiman 2001). Dieses Verfahren arbeitet nach einem ähnlichen Prinzip wie Bagging, auch hier werden viele Bäume, die basierend auf zufälligen Stichproben aus dem Datensatz ohne Pruning entwickelt wurden, für die Vorhersage kombiniert. Allerdings enthalten Random Forests ein weiteres Zufallselement, das Overfitting zusätzlich verringert und die Vorhersageleistung weiter verbessert. Jeder Baum hat bei jeder Unterteilung nur ein zufälliges Sample von Prädiktoren zur Verfügung, um die Unterteilung auszuwählen (vgl. Berk 2006: 280–281). Das heißt, es wird immer in unterschiedlichen Mengen von Variablen nach dem optimalen Split gesucht, wodurch die verschiedenen Bäume noch diverser werden. So erzielen Random Forests eine deutliche Verbesserung der Vorhersageleistung im Vergleich zu einfachen Entscheidungsbäumen (vgl. Kern et al. 2019: 77). Dafür besteht hier wie bei Bagging der Nachteil, dass man keine Baumstruktur hat, die einfach interpretiert werden kann.

Insgesamt bieten sich gerade Ensemble-Methoden in den Sozialwissenschaften für Klassifikation und Vorhersagen an, unter anderem zur Prüfung der Beziehung zwischen erklärenden Variablen und Ergebnissen in Modellen. Die Forschung an Klassifikations- und Regressionsbäumen hat in den letzten Jahren stark zugenommen, Anwendungen kommen sogar noch schneller dazu. Ein Grund dafür ist die Interpretierbarkeit der Baumstrukturen, genauso aber auch die vernünftige Genauigkeit der Vorhersagen, die schnelle Berechnung und die weite Verfügbarkeit von Software. (vgl. Loh 2014: 343). Während der Black-Box-Charakter der Ensemble-Modelle ein Nachteil bei ihrer Anwendung für die Erklärung von Zusammenhängen ist, eignen sie sich doch dafür, Variablen zu identifizieren, die einen Zusammenhang mit einer Zielvariable aufweisen. Damit können Ensemble-Methoden von Entscheidungsbäumen, insbesondere Random Forests sowohl für die Entwicklung eines Prognoseverfahrens mit guter Vorhersageleistung verwendet werden wie auch für die Prüfung des Einflusses bestimmter Variablen auf die Auswahlentscheidungen von Journalisten.

5.4.3 Konkrete Umsetzung

Konkret wurden auf Grundlage der erhobenen Informationen zum Publikationsort, dem Vorkommen einer Pressemitteilung und den Informationen zum Inhalt der Abstracts

verschiedene Random Forest-Modelle aus CARTs berechnet. Variiert wurden sowohl die Zahl verwendeter Merkmale (mit oder ohne inhaltliche Informationen) als auch der Grenzwert, ab dem von Medienaufmerksamkeit gesprochen wird (ab 1, 10 oder 50). Damit ergeben sich sechs verschiedene Vorhersagemodelle. Durch den Vergleich der Ergebnisse mit oder ohne Informationen zum Inhalt soll dessen Einfluss auf die Selektionsentscheidungen zumindest ansatzweise bestimmt werden, durch die verschiedenen Grenzwerte für mediale Aufmerksamkeit soll einerseits die Leistung des Verfahrens bei der Identifizierung seltenerer Fälle geprüft werden und andererseits robustere Modelle entwickelt werden, die vermeiden, dass die Vorhersage durch Fälle verzerrt wird, in denen der MSM-Score keine richtige Medienaufmerksamkeit misst. Als Datengrundlage für die Modelle wurden die Altmetric-Daten sowie die Informationen zu den verschiedenen Prädiktoren über die DOI in einem Datensatz zusammengeführt. Da nicht für alle Studien Abstracts vorlagen, konnten von den 600.000 Studien, bei denen Daten von Altmetric vorlagen, 152.459 verwendet werden. Bei diesen Studien lagen dann der MSM-Score, der Publikationsort sowie die Auswahlwahrscheinlichkeit dieser Zeitschrift, die Information zum Vorkommen einer Pressemitteilung sowie der Paragraph-Vektor aus dem Doc2Vec-Modell vor. Der MSM-Score wurde dabei in zwei Gruppen umcodiert, Media Attention und No Media Attention, je nachdem, ob er oberhalb des Grenzwerts lag.

In jedem der Modelle wurden 80 % der Daten zum Training verwendet, das heißt, diese Studien wurden vollständig, also mit der Einordnung des MSM-Scores an das Verfahren übergeben. Dieses lernt dann wie oben beschrieben in einer Vielzahl von Bäumen, wie der Datensatz so unterteilt werden kann, dass der Wert der Zielvariable in den Untergruppen möglichst homogen ist. Dabei können verschiedene Parameter variiert werden, die Zahl der Bäume, die Zahl der Fälle pro Endknoten sowie das Kriterium zum Auswählen der optimalen Unterteilung. Bei einem Vergleich verschiedener Werte für die ersten beiden Parameter zeigten sich nur geringe Unterschiede in der Genauigkeit, daher wurden hier mit zehn Bäumen und mindestens drei Fällen pro Blatt übliche Werte gewählt. Als Splitkriterium wurde der Gini-Index verwendet. Das im Training erstellte Modell wurde dann verwendet, um bei den verbliebenen 20 % der Daten, die nicht zum Training verwendet wurden, vorherzusagen, ob diese Aufmerksamkeit erhalten haben. Aus den richtigen und falschen Vorhersagen des Modells kann dann bestimmt werden, wie gut es in der Lage ist, die Nachrichtenauswahl von Journalisten nachzuvollziehen und vielleicht sogar zu erklären.

5.4.4 Genauigkeitsmaße von Machine Learning-Verfahren

Bei Machine Learning-Verfahren gibt es verschiedene Maßzahlen, die es erlauben, die Genauigkeit der Klassifikation und Vorhersage einzuschätzen. Bei numerischen Vorhersagen sind das beispielsweise der Durchschnitt des quadrierten Fehlers oder dessen Wurzel (MSE oder RMSE) oder auch der absolute Fehler (MAE). Zuletzt gibt es den Mean Relative Error, der einen Fehler immer ins Verhältnis zur Größe des tatsächlichen Werts setzt. Zuletzt können hier auch Korrelationskoeffizienten wie R^2 verwendet werden, wobei dieser nur den Grad der linearen Verbindung zwischen Variablen misst (vgl. Tatar et al. 2014a: 4–5). Bei Klassifikationsaufgaben wie der in dieser Untersuchung durchgeführten werden verschiedene Maße verwendet, die auf dem Anteil der korrekt klassifizierten Fälle an allen Fällen (einer Klasse) basieren. Hier kann zwischen verschiedenen Arten von Fehlern unterschieden werden. Eine falsch negative Vorhersage liegt vor, wenn ein Fall nicht in eine Klasse eingeordnet wurde, obwohl er dort eigentlich hingehört. Bezogen auf unsere Untersuchung wäre eine falsch negative Vorhersage (FN), dass eine Studie als No Media Attention eingeordnet wird, tatsächlich aber Medienaufmerksamkeit erhalten hat. Eine falsch positive Vorhersage (FP) liegt dann umgekehrt vor, wenn für eine Studie prognostiziert wird, dass sie Aufmerksamkeit erhalten hat, es tatsächlich aber nicht der Fall war. Die Gegenstücke sind die richtig negative (RN) und richtig positive Vorhersage (RP). Häufig verwendet wird die sogenannte Accuracy, die den Anteil der korrekt klassifizierten Instanzen angibt. In unserem Fall wäre das der Anteil der Studien, bei denen richtig vorhergesagt wurde, ob sie Medienaufmerksamkeit erhalten haben, an allen Studien.

$$Accuracy = \frac{(RN+RP)}{(RN+RP+FN+FP)}$$

Die Accuracy ist nicht angebracht, wenn man es mit ungleich verteilten Klassen zu tun hat. Bei der Vorhersage von Medienaufmerksamkeit beispielsweise ist die Gruppe der Studien ohne Aufmerksamkeit deutlich größer als die Gruppe mit Medienaufmerksamkeit. In so einem Fall wird die Genauigkeit oft besser dargestellt, als sie ist, da in der großen Gruppe leichter gute Werte erreicht werden. Hier bieten sich dann Precision, Recall und F-Score an. Die Precision misst für eine Klasse den Anteil der richtigen Vorhersagen an allen Vorhersagen dieser Klasse. Also beispielsweise den Anteil aller Studien, denen richtig Medienaufmerksamkeit vorhergesagt wurde, an allen Studien, denen Aufmerksamkeit vorhergesagt wurde.

$$Precision = \frac{RN}{(RN+FN)} \text{ oder } Precision = \frac{RP}{(RP+FP)}$$

Damit ist die Precision ein Maß dafür, wie sehr man sich darauf verlassen kann, dass eine Vorhersage den richtigen Wert anzeigt, also beispielsweise wie sicher die Vorhersage ist, dass eine Studie Medienaufmerksamkeit erhält. Der Recall gibt dagegen an, welchen Anteil aller Fälle einer Klasse man mit der Vorhersage entdeckt. Für die Vorhersage von Medienaufmerksamkeit bedeutet das, dass der Recall beispielsweise angibt, wie viele Studien mit Medienaufmerksamkeit das Verfahren übersieht. Dafür werden die richtigen Vorhersagen einer Klasse ins Verhältnis zu allen Fällen in dieser Klasse gesetzt.

$$Recall = \frac{RN}{(RN+FP)} \text{ oder } Recall = \frac{RP}{(RP+FN)}$$

Der F-Score ist das harmonische Mittel von Precision und Recall. Will man die Genauigkeit über mehrere Klassen hinweg angeben, kann man Macro-Precision, Macro-Recall und Macro-Average verwenden (vgl. dies. 2014a: 5). Eine gängige Darstellung der Ergebnisse von Prognoseverfahren ist die sogenannte Confusion Matrix. Hier werden die verschiedenen richtigen und falschen Einordnungen gegenübergestellt, sodass sich die verschiedenen Genauigkeitsmaße leicht berechnen lassen. Tabelle 1 zeigt beispielhaft den Aufbau einer solchen Confusion Matrix.

Tabelle 1: Beispieldarstellung einer Confusion Matrix. Gegenübergestellt werden die tatsächlichen Klassen der Beobachtungen den vorhergesagten. In den inneren Zellen sind dann die verschiedenen richtigen und falschen Vorhersagen enthalten. Links oben ist sowohl die echte wie auch die vorhergesagte Klasse negativ, rechts unten sind beide positiv. Dadurch sind diese Zellen die richtigen Vorhersagen. Die beiden anderen sind falsche Vorhersagen. Bei einer genauen Vorhersage sind damit die Fallzahlen in der linken oberen und rechten unteren Zelle deutlich größer als in den beiden anderen.

Gesamtzahl vorhergesagter Fälle	Vorhergesagte Klasse: keine Medienaufmerksamkeit	Vorhergesagte Klasse: Medienaufmerksamkeit
Echte Klasse: keine Medienaufmerksamkeit	Zahl der richtig negativen Vorhersagen (RN)	Zahl der falsch positiven Vorhersagen (FP)
Echte Klasse: Medienaufmerksamkeit	Zahl der falsch negativen Vorhersagen (FN)	Zahl der richtig positiven Vorhersagen (RP)

Neben den Genauigkeitsmaßen, die eine Einschätzung der Leistung des Vorhersagemodells erlauben, gibt es speziell für Random Forests verschiedene Maßzahlen, mit denen die Relevanz einzelner Variablen für die Vorhersage des Ensembles abgeschätzt werden kann. Ein Beispiel ist die Gini Importance, die angibt, wie sehr eine Variable im Durchschnitt über alle Bäume hinweg die Impurity, gemessen als Gini Index, reduziert. Eine bessere Methode besteht darin, die Werte einer Variable über alle Bäume hinweg zufällig zu permutieren (also zufällig zu verändern) und zu prüfen, wie sehr sich die Vorhersagegenauigkeit verändert. Wenn sie sich kaum verändert, muss die Bedeutung der Variable gering sein, verändert sie sich sehr, hat die

Variable eine hohe Bedeutung für die Vorhersage. Daraus ergibt sich die permutation importance. Deren absolute Werte sollten nicht direkt interpretiert werden, stattdessen erlaubt es ein Vergleich der permutation importance verschiedener Variablen, diejenigen zu identifizieren, die für die Vorhersage tatsächlich von Bedeutung sind (vgl. Strobl et al. 2009: 335–336).

5.5 Synthese: Kombination von Input-Output-Analyse und Machine Learning

Insgesamt liegt damit ein Verfahren vor, das es später ermöglicht, großflächig Auswahlentscheidungen von Wissenschaftsjournalisten zu prognostizieren. In einer Input-Output-Analyse, die auf dem automatisierten Abruf von Publikations- und Mediendaten beruht, können zehntausende Auswahlentscheidungen nachvollzogen werden. Dafür werden mit Daten von Altmetric zur Referenz auf Studien in über 2000 Online-Medientiteln unter den Studien, die von August 2016 bis Juli 2017 in der Literaturdatenbank Scopus eingestellt wurden, diejenigen identifiziert, über die Medien berichtet haben. Um die so gefundene Auswahl zu erklären, werden drei Variablen für ein Machine-Learning-Verfahren verwendet. Der Publikationsort wird aus den Scopus-Daten extrahiert, für die Verfügbarkeit einer Pressemitteilung wird der MSM-Score von Altmetric verwendet, der auch Erwähnungen in zwei Pressemitteilungsportalen erfasst. Zuletzt wird die inhaltliche Verortung auf einer abstrakten Ebene mit einem Natural Language Processing-Verfahren ermittelt. Dafür werden die Studienabstracts als mehrdimensionale Vektoren repräsentiert, deren Position in einem gemeinsamen Vektorraum Aussagen über die Ähnlichkeit verschiedener Abstracts zulässt. Die Distanz eines Abstracts zum Mittelpunkt aller Abstracts, die Aufmerksamkeit oder keine Aufmerksamkeit erhalten haben, wird als letztes Merkmal verwendet.

Diese Informationen werden in einem Random Forest-Verfahren genutzt, um Medienaufmerksamkeit vorherzusagen. Random Forests basieren auf einem Ensemble von Entscheidungsbäumen, deren Vorhersage darauf basiert, dass sie einen Datensatz rekursiv in Untergruppen unterteilt, in denen der Wert einer Zielvariable immer homogener ist. Der Vorteil eines Ensembles besteht darin, dass die Vorhersageleistung besser ist als bei einzelnen Bäumen und das Verfahren besser auf neue Datensätze angewendet werden kann, die nicht zum Training benutzt wurden. Um zu testen, wie gut die verschiedenen Merkmale die Auswahl erklären können, wurden Modelle mit und ohne inhaltliche Informationen gerechnet. Zusätzlich wurden verschiedene Modelle berechnet, in denen unterschiedliche Mindestwerte des MSM-Scores verwendet wurden, um Medienaufmerksamkeit zu operationalisieren. Im einfachsten Fall ist Aufmerksamkeit ein MSM-Score von mindestens eins, zusätzlich wurden zehn und 50 als

Grenzwerte festgelegt, um zu prüfen, wie gut das Modell fokussierte Aufmerksamkeit vorhersagen kann und um robustere Modelle zu entwickeln. So wurden insgesamt sechs verschiedene Random Forest-Modelle erstellt und die Vorhersage mit einem Teil des Datensatzes geprüft. Die Aufbereitung der Datensätze und die Entwicklung der Vorhersagemodelle wurde am SMC von Meik Bittkowski in der Programmiersprache Python durchgeführt.

Ergebnis sind einerseits Angaben zur Genauigkeit der Vorhersage und andererseits zur Bedeutung verschiedener Variablen. Der Vergleich verschiedener Genauigkeitswerte zeigt, wie gut das Verfahren mit verschiedenen Variablenmengen und bei unterschiedlichen Grenzwerten die Auswahl erklären kann. Mit Precision und Recall sind Aussagen darüber möglich, wo die Schwächen des Verfahrens liegen. Mit der Permutation Importance kann bestimmt werden, welche Variablen besonders wichtig für die Vorhersage sind. Dadurch kann einerseits eingeschätzt werden, wie gut sich das Verfahren zukünftig für die praktische Prognose von Medienaufmerksamkeit eignet und wie gut es theoretisch erklären kann, welche Kriterien Wissenschaftsjournalisten bei der Nachrichtenauswahl anwenden. Und damit, wie gut das entwickelte Modell der Nachrichtenauswahl auf die tatsächliche Auswahlpraxis passt.

6. Ergebnisse

Bevor die Ergebnisse der verschiedenen Varianten des Prognosemodells vorgestellt werden, soll ein deskriptiver Blick auf einige Eigenschaften und Besonderheiten des Datensatzes geworfen werden. Das betrifft einerseits die Verteilung der Zielvariable, also der Medienaufmerksamkeit. Andererseits ist auch die Verteilung der verschiedenen Prädiktoren bedeutsam für die spätere Interpretation der Genauigkeit der Vorhersage.

6.1 Deskriptive Beschreibung

Knapp 1,7 Millionen Studien wurden aus Scopus exportiert, für fast 600.000 dieser Studien lagen bei Altmetric Daten vor. Für die Entwicklung der Prognoseverfahren wiederum konnte etwas mehr als ein Viertel dieser Studien verwendet werden, da nur für 152.459 Studien auch die Abstracts vorlagen, die für Entwicklung des Doc2Vec-Modells für die inhaltliche Modellierung benötigt werden. Von diesen Studien hatten knapp ein Viertel (24,2 %) einen MSM-Score von mindestens 1. Drei Viertel der Studien wurden also in keiner der über 2000 Onlinequellen erwähnt, die Altmetric nach Referenzen durchsucht.

Schaut man sich die Verteilung des MSM-Scores genauer an, zeigt sich eine sehr ungleiche Verteilung, in der sehr viele Studien in keinen bis wenigen Medientiteln erwähnt wurden und

sehr wenige in sehr vielen. Während immerhin noch 18,3 % der Studien einen MSM-Score zwischen 1 und 10 haben, liegt der Anteil der Studien mit einem Score zwischen 11 und 20 nur noch bei 2,7 %. Nur 1,5 % der Studien liegen über einem Score von 50 und nur noch 0,3 % aller wissenschaftlichen Ergebnisse wurden in mehr als 100 verschiedenen Quellen erwähnt (s. Tabelle 2). Die Menge der Studien, die einen bestimmten MSM-Score erhalten haben, fällt mit der Höhe dieses Scores exponentiell ab. Diese regelmäßige Verteilung journalistischer Aufmerksamkeit wurde für die fokussierte Berichterstattung über wissenschaftliche Studien, das heißt ab einem MSM-Score größer gleich 50, bereits genauer beschrieben und als sogenannte Power Law-Verteilung identifiziert (vgl. Lehmkuhl 2018; Lehmkuhl/Promies 2019). Charakteristisches Merkmal dieser Verteilung ist, dass sich bei der Darstellung auf logarithmischen Achsen eine Gerade zeigt. Damit folgt die Auswahl von Studien für die Berichterstattung einem ähnlichen Muster wie die Nutzung von Online-Nachrichten, bei der ebenfalls Power Law- oder die sehr ähnlichen Lognormal-Verteilungen gefunden wurden (vgl. Tatar et al. 2012; Tatar et al. 2014b; Moniz et al. 2016; van Canneyt et al. 2018). Für die Vorhersage der Aufmerksamkeit hat das zur Folge, dass die Prognose der seltenen Fälle sehr hoher Aufmerksamkeit besonders schwierig ist, da ja die Anzahl der Fälle exponentiell mit der Aufmerksamkeit abnimmt.

Tabelle 2: Verteilung der Studien nach dem MSM-Score, unterteilt in Zehnerschritte.

Bereich des MSM-Scores	Anteil dieses Bereichs an allen Studien
0	75,8 %
1-10	18,3 %
11-20	2,7 %
21-30	0,9 %
31-40	0,5 %
41-50	0,3 %
51-60	0,5 %
61-70	0,3 %
71-80	0,2 %
81-90	0,1 %
91-100	0,1 %
>100	0,3 %

Nur bei 7,7 % der Studien wurde eine Pressemitteilung gefunden, es zeigte sich jedoch schon in der ersten explorativen Untersuchung ein enger Zusammenhang zwischen dem Vorkommen einer Pressemitteilung und der Medienaufmerksamkeit, die eine Studie erhält. Besonders deutlich wird das in einer Kreuztabelle, in der beide Variablen gegenübergestellt werden (s. Tabelle 3, Medienaufmerksamkeit wurde definiert als MSM-Score von mindestens 1). Auffällig ist vor allem die sehr geringe Zahl der Studien, zu denen es eine Pressemitteilung gab, die aber keine Medienaufmerksamkeit erhielten. Ursache könnte der gewaltige Einfluss von Pressemitteilungen auf die Aufmerksamkeit sein oder aber die von Altmetric verwendete Quellenliste, in der auch Aggregatoren enthalten sind, die Pressemitteilungen kopieren. Der Chi-Quadrat-Test ergab entsprechend einen signifikanten Zusammenhang zwischen beiden Variablen, mit einem Cramer's V beziehungsweise Phi von 0,508. Zudem zeigten sich in einem t-Test signifikante Mittelwert-Unterschiede des MSM-Scores von Studien mit oder ohne Pressemitteilung. Bei Studien mit Pressemitteilung lag der MSM-Score im Schnitt bei knapp unter 22, bei Studien ohne Pressemitteilung bei knapp über 1. Allerdings müssen diese Werte mit Vorsicht interpretiert werden, da die ungleiche Verteilung des MSM-Scores eigentlich keine sinnvolle Berechnung des Mittelwerts zulässt. Aber auch nicht-parametrische Tests wie der Mann-Whitney-U-Test zeigen einen signifikanten Unterschied der Verteilung des Scores zwischen den beiden Gruppen.

Tabelle 3: Kreuztabelle Vorkommen Pressemitteilung/Medienaufmerksamkeit in absoluten Zahlen ($\chi^2=39315$, $df=1$, $p<0,001$)

	Medienaufmerksamkeit	Keine Medienaufmerksamkeit	Gesamt
Pressemitteilung nicht vorhanden	115.516	25.154	140.670
Pressemitteilung vorhanden	85	11.704	11789
Gesamt	115.601	36.858	152.459

478 Zeitschriften sind im Datensatz vertreten, mit Fallzahlen von 1 (bspw. *Oncoimmunology*) bis zu 16.225 (*Plos One*). Gerade die großen Open Access-Journals sind mit vielen Studien vertreten (neben *Plos One* noch *Scientific Reports* und *Nature Communications*), was vor allem an der großen Menge an Studien liegt, die sie im Jahr veröffentlichen. Neben den Zeitschriften selbst wurde auch die Zahl der Studien, die aus einer Zeitschrift fokussierte Aufmerksamkeit erhalten haben, für die Prognose verwendet sowie der Anteil einer Zeitschrift an allen Studien mit fokussierter Aufmerksamkeit. In der Verteilung dieser Werte zeigt sich ebenfalls eine Power Law-Verteilung, die an anderer Stelle bereits im Detail beschrieben wurde (vgl.

Lehmkuhl 2018; Lehmkuhl/Promies 2019). Die Zeitschriften mit der höchsten Auswahlwahrscheinlichkeit sind *Nature*, *Science*, *NEJM*, *PNAS*, *Plos One*, *Lancet*, *JAMA*, *Nature Communications*, *Pediatrics* und *Scientific Reports* (s. Tabelle 4). Bis auf *BMJ* findet sich hier also wieder die Gruppe renommierter Zeitschriften, die in vielen vorherigen Analysen identifiziert wurden. Neu hinzugekommen sind zudem die multidisziplinären Open Access-Journals, die auch mit der größten Fallzahl im Datensatz vertreten sind. Diese Zeitschriften wurden erst in den letzten Jahren gegründet und haben es anscheinend geschafft, sich als bedeutende Quellen für Journalisten zu etablieren.

Tabelle 4: Liste der Zeitschriften mit der höchsten Auswahlwahrscheinlichkeit

Journal	Zahl von Studien mit fokussierter Aufmerksamkeit von 2015 bis 2018 (übernommen aus Lehmkuhl/Promies 2019)
Nature	236
Science	217
NEJM	180
PNAS	174
Plos One	133
Lancet	121
JAMA	112
Nature Communications	111
Pediatrics	106
Scientific Reports	105

Um mehr über den Einfluss der Pressemitteilungen zu erfahren, wurde für die einzelnen Zeitschriften der Anteil der Studien berechnet, zu denen es eine Pressemitteilung gab, sowie der Anteil der Studien, die Medienaufmerksamkeit erhielten. Zwischen beiden Werten zeigte sich eine signifikante Korrelation mit einem relativ hohen Korrelationskoeffizienten von 0,729 (R^2), die wiederum den Einfluss von Pressemitteilungen auf die Nachrichtenauswahl zeigt. Zusätzlich wurde eine Korrelation berechnet zwischen der Wahrscheinlichkeit der Auswahl fokussierter Studien, die aus einem anderen Datensatz übernommen wurde (s. Kapitel 5.3.1) und dem Anteil der Studien, die im hier erstellten Datensatz Medienaufmerksamkeit erhielten. Auch hier zeigte sich eine signifikante Korrelation mit einem Koeffizienten von 0,738. Insgesamt sprechen diese Daten also bereits für einen Einfluss der ausgewählten Prädiktoren

auf die Nachrichtenauswahl im Wissenschaftsjournalismus. Ob dieser Einfluss bereits ausreicht, um die Auswahl auch vorherzusagen, wird in den nächsten Kapiteln erörtert.

6.2 Ergebnisse verschiedener Prognosemodelle

Sechs verschiedene Vorhersagemodelle wurden insgesamt berechnet, drei davon ohne und drei mit inhaltlichen Informationen aus den Abstracts, zudem wurden verschiedene Grenzwerte zur Klassifizierung der Studien verwendet. Zuerst sollen jetzt die Ergebnisse der einfacheren Prognosemodelle vorgestellt werden, die als Prädiktoren nur den Publikationsort und das Vorkommen einer Pressemitteilung verwendeten. Die Darstellung der Ergebnisse folgt dabei immer dem gleichen Muster, der Fokus liegt einerseits auf der Genauigkeit der Vorhersage, repräsentiert über eine Confusion Matrix, aus der abgelesen werden kann, welche Klassen gut und schlecht vorhergesagt werden. Andererseits bietet die Permutation Importance ein Maß für die Bedeutung der einzelnen Variablen für das Ergebnis der Vorhersage.

6.2.1 Grundmodell

Das einfachste Grundmodell verwendet nur Publikationsort, absolute und relative Auswahlwahrscheinlichkeit und das Vorkommen einer Pressemitteilung für die Vorhersage. Vorhergesagt wurde zuerst, ob eine Studie überhaupt Medienaufmerksamkeit erhält, der MSM-Score also bei mindestens 1 liegt. Wie in der Methode beschrieben, wurden 80 % des Datensatzes zum Training des Random Forest-Modells verwendet und für die verbliebenen 20 % wurde dann der Wert von Medienaufmerksamkeit vorhergesagt. Daher beziehen sich die folgenden Tabellen und Statistiken immer auf eine Fallzahl von 30.492 Fällen, also ein Fünftel der 152.459 Studien, die insgesamt im Datensatz vertreten waren. Wie die Confusion Matrix zeigt (Tabelle 5), waren deutlich mehr Fälle in Testset, bei denen keine Aufmerksamkeit vorlag, was ja aufgrund der ungleichen Verteilung der Medienaufmerksamkeit auch zu erwarten war. Die größte Gruppe findet sich in der linken oberen Zelle, das heißt, bei ihnen wurde richtig vorhergesagt, dass sie keine Medienaufmerksamkeit erhalten haben (RN). Besonders klein sind die Fallzahlen in der linken unteren und rechten oberen Zelle, also in den Zellen, in denen falsche Vorhersagen vorlagen. Die Zahl der falsch negativen Vorhersagen (es wurde keine Aufmerksamkeit vorhergesagt, es gab aber welche) war dabei noch etwas niedriger als die der falsch positiven (es wurde Aufmerksamkeit vorhergesagt, gab aber keine).

In den verschiedenen Genauigkeitsmaßen ausgedrückt ergibt sich dadurch folgendes Bild: Besonders hoch sind Recall und Precision bei der Vorhersage, dass eine Studie *keine* Aufmerksamkeit erhält. Bei fast 85 % der Studien, die wirklich keine Aufmerksamkeit erhalten

haben, war die Vorhersage richtig (Recall) und die Vorhersage, dass eine Studie keine Aufmerksamkeit erhält, stimmte in knapp über 89 % der Fälle. Das heißt, in fast 9 von 10 Fällen konnte man sich darauf verlassen, dass es stimmt, wenn einer Studie keine Aufmerksamkeit prognostiziert wird. Schlechter sind dagegen die Genauigkeitswerte bei der Vorhersage, dass eine Studie Medienaufmerksamkeit erhält. Hier lag der Recall bei etwa zwei Dritteln (67,61 %) und die Precision bei knapp unter 60 %. Inhaltlich bedeutet das, dass immerhin zwei von drei Studien, die von Journalisten für die Berichterstattung ausgewählt werden, durch das Prognoseverfahren entdeckt werden. Dafür ist unter den positiven Vorhersagen ein relativ hoher Anteil von Studien dabei, die eigentlich keine Aufmerksamkeit erhalten, die Vorhersage ist also nicht besonders präzise. Nimmt man die Werte zusammen, liegt die Genauigkeit dieses Modells bei 80,76 %. Acht von zehn Vorhersagen waren also insgesamt richtig, wobei die Genauigkeit durch die hohen Werte bei keiner Medienaufmerksamkeit etwas beschönigt wird.

Tabelle 5: Confusion Matrix des Vorhersagemodells ohne inhaltliche Informationen, Grenzwert bei MSM-Score ≥ 1 . Hoch ist besonders die Zahl richtig negativer Vorhersagen (links oben), der Anteil falsch positiver Vorhersagen ist noch relativ hoch.

Gesamtzahl vorhergesagter Fälle: 30492	Vorhergesagte Klasse: keine Medienaufmerksamkeit	Vorhergesagte Klasse: Medienaufmerksamkeit
Echte Klasse: keine Medienaufmerksamkeit	RN: 19601	FP: 3462
Echte Klasse: Medienaufmerksamkeit	FN: 2406	RP: 5023

Auch wenn die Permutation Importance bei so wenigen verwendeten Prädiktoren noch nicht wirklich aussagekräftig ist, soll sie kurz erwähnt werden. Wie beschrieben, sind die absoluten Werte nicht aussagekräftig, es geht allein darum, welche Variablen im Verhältnis besondere Bedeutung haben. Wie Tabelle 6 zeigt, haben sowohl das Vorkommen einer Pressemitteilung als auch der Publikationsort große Bedeutung für die Vorhersage, dafür jedoch die beiden Maße der Auswahlwahrscheinlichkeit nicht. Das könnte daran liegen, dass diese Variablen keine zusätzlichen Informationen enthalten, die nicht schon durch den Publikationsort selbst gegeben wären.

Tabelle 6: Permutation Importance der verschiedenen Prädiktoren bei der Vorhersage im Grundmodell. Die absoluten Zahlen können nicht interpretiert werden, wichtig ist allein, dass höhere Werte für eine höhere Bedeutung stehen.

Feature	Importance
Pressemitteilung	0,0644
Journal	0,0184
Auswahlwahrscheinlichkeit absolute Zahlen	-0,0210
Auswahlwahrscheinlichkeit relative Zahlen	-0,0210

Die Vorhersage, ob eine Studie überhaupt Medienaufmerksamkeit erhält, ist vergleichsweise einfach, da hier noch eine ausreichend große Fallzahl vorliegt. Schwieriger wird es dagegen bei der Prognose der seltenen Fälle, die mehr Aufmerksamkeit erhalten haben, wie auch die in Kapitel 4.2.2 vorgestellten Arbeiten gezeigt haben. Da jedoch die Daten von Altmetric gerade bei den niedrigen Scores wenig verlässlich sind, könnte diese Prognose noch mehr Wert haben. Daher wurden auch Prognosemodelle erstellt, die vorhersagen sollen, ob eine Studie einen MSM-Score von mindestens 10 oder mindestens 50 hat. Wie die Confusion Matrix für die Vorhersage eines Scores von mindestens 10 zeigt (Tabelle 7), erhöht sich hier vor allem die Zahl der richtig negativen Vorhersagen noch einmal erheblich. 85,6 % aller Fälle liegen in dieser Zelle. Dadurch ist insbesondere die Präzision, aber auch der Recall sehr hoch. In 98,57 % der Fälle stimmt die Vorhersage, dass eine Studie einen MSM-Score unter 10 hat.

Auch der Recall bei der Vorhersage eines Scores größer gleich 10 ist relativ hoch, er liegt bei 80 %. Sehr schlecht ist dagegen die Präzision dieser Vorhersage. Das liegt an der vergleichsweise hohen Fallzahl in der rechten oberen Zelle, das heißt den falsch positiven Vorhersagen. Während diese im Vergleich zur großen Zahl richtig negativer Vorhersagen kaum ins Gewicht fallen, sorgen sie dafür, dass man sich in weniger der Hälfte der Fälle auf die Vorhersage verlassen kann, dass eine Studie einen MSM-Score von mindestens 10 erhalten wird. Die Genauigkeit des Verfahrens liegt insgesamt mit 90,7 % höher als im vorherigen Verfahren, bedingt allein durch die hohe Zahl richtig negativer Vorhersagen. Ursache für die unterschiedlichen Genauigkeitswerte ist die ungleiche Verteilung der Studien auf die beiden Gruppen.

Tabelle 7: Confusion Matrix des Prognosemodells ohne inhaltliche Informationen, Grenzwert bei MSM-Score ≥ 10 . Im Vergleich zur vorherigen Tabelle ist die Zahl richtig negativer Vorhersagen noch größer geworden. Auffällig ist zusätzlich, dass die Zahl falsch positiver Vorhersagen höher ist als die richtig positiver.

Gesamtzahl vorhergesagter Fälle: 30492	Vorhergesagte Klasse: keine Medienaufmerksamkeit	Vorhergesagte Klasse: Medienaufmerksamkeit
Echte Klasse: keine Medienaufmerksamkeit	RN: 26112	FP: 2448
Echte Klasse: Medienaufmerksamkeit	FN: 379	RP: 1553

Das zeigt sich noch extremer, wenn versucht wird, einen Score von mindestens 50 vorherzusagen (Tabelle 8). Die Zahl der richtig negativen Vorhersagen wird nur unwesentlich höher, dafür werden noch mehr Studien mit geringer Medienaufmerksamkeit falsch eingeordnet. In den beiden unteren Zellen sind sehr wenige Fälle. Dadurch ist die Precision der Vorhersage, dass eine Studie einen Score unter 50 erhält, natürlich sehr hoch, sie liegt bei 99,5 %, dafür ist die Präzision der Vorhersage, dass eine Studie viel Aufmerksamkeit erhält, mit 7 % völlig unverlässlich. Interessante Unterschiede zeigen sich in der Betrachtung der Permutation Importance bei diesen beiden Modellen, während für die Vorhersage eines Scores größer oder kleiner 10 vor allem das Vorkommen einer Pressemitteilung größere Bedeutung hat, ist bei der zweiten Vorhersage der Einfluss des Journals wichtiger. Insgesamt erscheint das aber plausibel, da gerade für die seltenen Fälle sehr hoher Aufmerksamkeit eher die Publikation in einem prestigeträchtigen Journal ausschlaggebend sein dürfte, allein die Verfügbarkeit einer Pressemitteilung reicht dafür nicht aus.

Tabelle 8: Confusion Matrix des Vorhersagemodells ohne inhaltliche Informationen, Grenzwert bei MSM-Score ≥ 50 . Die Zahl richtig positiver Vorhersagen ist im Vergleich zu den falsch positiven sehr gering, wodurch die Vorhersage von Medienaufmerksamkeit hier deutlich öfter falsch als richtig ist.

Gesamtzahl vorhergesagter Fälle: 30492	Vorhergesagte Klasse: keine Medienaufmerksamkeit	Vorhergesagte Klasse: Medienaufmerksamkeit
Echte Klasse: keine Medienaufmerksamkeit	RN: 26201	FP: 3863
Echte Klasse: Medienaufmerksamkeit	FN: 139	RP: 289

Um eine Einschätzung der Genauigkeit der Vorhersagen bei diesen Klassen zu erhalten, wurden zwei weitere Modelle berechnet, für die nur Studien verwendet wurden, die einen Score von mindestens 1 haben (die zugehörigen Tabellen finden sich im Anhang, Tabelle 13 und 14). Die Prognoseaufgabe verändert sich also leicht, es soll jetzt nur noch zwischen geringer und höherer

Aufmerksamkeit unterschieden werden. Dafür blieben dann noch knapp 37.000 Fälle im Datensatz übrig, wovon 7389 für die Testvorhersage verwendet wurden. Für die Vorhersage, ob der MSM-Score größer oder kleiner 10 ist, war wiederum die Genauigkeit bei der Vorhersage niedriger Scores höher, der Recall lag bei 78,78 % und die Precision bei 88,47 %. Dafür ist gerade die Precision bei der positiven Vorhersage nicht mehr so schlecht wie vorher, da die beiden Klassen etwas ausgeglichener verteilt sind. Mit 56 % ist sie zwar immer noch nicht besonders gut, aber schon deutlich besser als vorher. In immerhin mehr als der Hälfte der Fälle ist die Vorhersage richtig, dass eine Studie einen Score von mindestens 10 erhält. Ähnlich ist es bei der Vorhersage eines Scores von mindestens 50, wobei die Precision der positiven Vorhersage hier mit 11,6 % zwar schon erhöht ist, aber immer noch sehr schlecht.

Insgesamt zeigen die Ergebnisse des Grundmodells also, dass bei der Vorhersage, ob eine Studie überhaupt Medienaufmerksamkeit erhält, schon mit den beiden verwendeten Variablen eine relativ genaue Vorhersage möglich ist. Besonders keine Medienaufmerksamkeit wird sehr gut erkannt. Das ist auch noch der Fall, wenn ein höherer Grenzwert für Medienaufmerksamkeit angesetzt wird, dafür ist die Vorhersage höherer Medienaufmerksamkeit dann kaum noch zu verwenden, insbesondere bei einem Grenzwert von 50.

6.2.2 Erweiterung um Sprachmodelle

Im zweiten Schritt wurden die Grundmodelle um die Informationen aus dem Doc2Vec-Modell ergänzt, die auf Grundlage der Abstracts erstellt wurden. Dadurch kamen als zusätzliche Variablen die inhaltliche Ähnlichkeit einer Studie zum Durchschnitt sowohl der ausgewählten wie der nicht ausgewählten Studien hinzu. Diese Ähnlichkeit wurde für die verschiedenen Grenzwerte jeweils neu berechnet. Die Ergebnispräsentation aus diesen Modellen konzentriert sich vor allem auf den Vergleich mit den Grundmodellen, um nach einem Einfluss des Inhalts auf die Auswahl zu suchen.

Tabelle 9: Confusion Matrix des Prognosemodells mit inhaltlichen Informationen, Grenzwert bei MSM-Score ≥ 1 . Im Vergleich zum Modell ohne inhaltliche Informationen konnte die Zahl der falsch positiven Vorhersagen verringert werden, von 3462 auf 2306.

Gesamtzahl vorhergesagter Fälle: 30492	Vorhergesagte Klasse: keine Medienaufmerksamkeit	Vorhergesagte Klasse: Medienaufmerksamkeit
Echte Klasse: keine Medienaufmerksamkeit	RN: 20757	FP: 2306
Echte Klasse: Medienaufmerksamkeit	FN: 2595	RP: 4834

Bei der einfachsten Aufgabe konnte die Verwendung der inhaltlichen Informationen die ohnehin schon akzeptable Genauigkeit der Vorhersage noch weiter verbessern (Tabelle 9). Während die Precision bei keiner Aufmerksamkeit und der Recall bei Aufmerksamkeit sehr ähnlich waren, verbesserten sich vor allem die beiden anderen Werte. Der Recall bei den negativen Vorhersagen wurde um 5 %, die Precision bei positiven Vorhersagen sogar um 8 % besser. Beides liegt daran, dass die Zahl der falsch positiven Vorhersagen verringert wurde. Das bedeutet, dass durch die inhaltlichen Informationen weniger Studien, die keine Aufmerksamkeit erhalten haben, fälschlich als „mit Aufmerksamkeit“ eingeordnet wurden. Mit diesem Prognosemodell liegt dann insgesamt ein Verfahren vor, das einerseits zwei Drittel aller Studien, die Aufmerksamkeit erhalten, im Datensatz findet und das andererseits auch in zwei Drittel der positiven Vorhersagen richtig liegt. Die Genauigkeit liegt zusammengenommen bei 84 %. Auch die Permutation Importance bestätigt die Bedeutung der inhaltlichen Informationen, die beiden daraus abgeleiteten Variablen haben auf Anhieb die höchste Bedeutung für die Vorhersage, insbesondere die Ähnlichkeit zu den ausgewählten Studien.

Ähnliche Ergebnisse zeigten sich auch bei den Vorhersagen mit anderen Grenzwerten (ein Überblick über die verschiedenen Genauigkeitswerte findet sich in Tabelle 10, hier ist auch der Vergleich mit den Modellen ohne inhaltliche Informationen möglich). Für einen Grenzwert von 10 und bei der Verwendung aller Studien wurde wiederum die Zahl falsch positiver Vorhersagen verringert, wodurch vor allem die Precision der Vorhersage erhöht wurde, dass eine Studie einen Score von mindestens 10 erhält, sie lag in diesem Modell bei 53 %. Dafür nahm der Recall dieser Vorhersage leicht ab. Die Ergebnisse bei der Prognose, ob der MSM-Score einer Studie bei mindestens 50 liegt, waren weiterhin kaum zu gebrauchen, aber immerhin konnte die Precision der positiven Vorhersage vergleichsweise deutlich auf 21 % erhöht werden. Interessant ist bei diesen beiden Vorhersagen, dass die Permutation Importance der neu eingeführten Variablen geringer ist als bei der ersten Vorhersage. Woran das genau liegt, kann nicht aufgelöst werden, ein Grund könnte sein, dass bei der ungleichen Verteilung der Studien auf die verschiedenen Klassen schon die beiden einfacheren Prädiktoren ausreichen, um einen Großteil der Studien auszuschließen, also negativ einzuordnen.

Tabelle 10: Gegenüberstellung der Genauigkeitsmaße der verschiedenen Vorhersagemodelle. Aus Gründen der Übersichtlichkeit wurden die Modelle mit einem Grenzwert von 50 weggelassen. Auffällig ist die Verbesserung insbesondere der Precision der positiven Vorhersage, die sich durch die Verwendung der inhaltlichen Informationen ergibt. In den Vorhersagen mit einem höheren Grenzwert

wird die Genauigkeit der negativen Vorhersage besser, dafür nimmt gerade die Precision der positiven Vorhersage ab.

Modell	Precision negative Prognose	Recall negative Prognose	Precision positive Prognose	Recall positive Prognose
MSM \geq 1, ohne Sprachmodell	89,07 %	84,99 %	59,2 %	67,61 %
MSM \geq 1, mit Sprachmodell	88,89 %	90 %	67,7 %	65,07 %
MSM \geq 10, ohne Sprachmodell	98,57 %	91,43 %	38,82 %	80,38 %
MSM \geq 10, mit Sprachmodell	98,25 %	95,51 %	53,02 %	74,9 %

Auch mit den inhaltlichen Informationen wurden für die höheren Grenzwerte Modelle erstellt, die nur die Studien verwendeten, deren MSM-Score mindestens bei 1 lag. Hier zeigten sich ähnliche Veränderungen wie in den anderen Modellen, verringert werden konnte meist die Zahl der falsch positiven Vorhersagen. Für einen Grenzwert von 10 wurde so eine Precision der positiven Vorhersage von 63 % erreicht, während die Ergebnisse bei einem Grenzwert von 50 weiterhin kaum zu nutzen waren (die zugehörigen Tabellen finden sich im Anhang, Tabelle 15-17).

6.3 Synthese: hochselektive Auswahl anhand weniger Faktoren

Bevor in der Diskussion genauer auf ihre Bedeutung eingegangen werden wird, sollen die Ergebnisse der verschiedenen Prognosemodelle kurz zusammengefasst werden. Es lag ein Datensatz vor, in dem die mediale Aufmerksamkeit sehr ungleich verteilt war. Ein Viertel der Studien erhielt überhaupt Medienresonanz, mit zunehmendem MSM-Score nahm die Zahl der Studien immer weiter ab. Diese Verteilung folgt einem genau zu beschreibenden Muster, einer sogenannten Power Law-Verteilung, in der die Aufmerksamkeit exponentiell abfällt. Das hat Konsequenzen für die Genauigkeit der Vorhersage. Während bei der Prognose, ob eine Studie überhaupt Aufmerksamkeit erhält, bereits mit dem Grundmodell eine recht hohe Genauigkeit erreicht, nimmt diese mit der Höhe des Grenzwerts ab, bei dem Medienaufmerksamkeit festgelegt wird. Das gilt im Speziellen für die Precision der positiven Vorhersage, bei einem Grenzwert von 50 liegt diese im Grundmodell nur bei 7 %, wenn alle Studien in der Vorhersage verwendet werden. Dadurch, dass die Fälle hoher Aufmerksamkeit sehr selten sind, wird es

immer schwieriger, sie präzise vorherzusagen. Da die Gruppe der Studien mit keiner oder geringer Aufmerksamkeit immer größer wird, reicht schon ein kleiner Anteil falsch positiver Vorhersagen aus, um dafür zu sorgen, dass die Vorhersage wenig verlässlich ist, dass eine Studie Medienaufmerksamkeit erhält. Etwas besser wird die Genauigkeit, wenn nur Studien berücksichtigt werden, deren MSM-Score mindestens 1 ist. Gerade für die praktische Anwendung ist ein solches Verfahren jedoch wenig hilfreich.

Durch die Verwendung der Informationen aus dem Inhalt der Abstracts konnte die Genauigkeit bei allen Vorhersagen verbessert werden. Insbesondere die Zahl falsch positiver Vorhersagen wurde reduziert, wodurch die Präzision der positiven Vorhersagen verbessert wurde. Die Bedeutung der inhaltlichen Variablen zeigte sich zumindest bei der Vorhersage eines Scores von mindestens 1 auch in der Permutation Importance, die angibt, welche Bedeutung einzelne Prädiktoren für das Ergebnis der Vorhersage haben. Die Werte der inhaltsbezogenen Variablen lagen hier sehr hoch. Bei den beiden anderen Modellen mit höheren Grenzwerten dagegen war die Bedeutung interessanterweise geringer, obwohl die Vorhersage verbessert wurde. Das könnte daran liegen, dass ein Großteil der Vorhersagen, die ja meistens negativ sind, durch die beiden anderen Prädiktoren getroffen wurde. Das genaueste Vorhersagemodell, das mit allen Prädiktoren vorhersagen sollte, ob eine Studie überhaupt Aufmerksamkeit erhalten hat, erreichte zusammengenommen eine Genauigkeit von 84 % und bei der positiven Vorhersage Recall- und Precision-Werte, die jeweils bei zwei Drittel lagen.

7. Diskussion

Insgesamt können die Ergebnisse der verschiedenen Vorhersagemodelle somit nicht vollständig überzeugen. Während keine Aufmerksamkeit sehr präzise vorhergesagt werden kann und der Anteil der Studien mit Aufmerksamkeit, die als solche entdeckt werden, noch akzeptabel ist, ist die Präzision bei der Prognose von Medienaufmerksamkeit gerade bei höheren Grenzwerten zum Teil sehr bescheiden. Was bedeuten diese Ergebnisse nun für unser Verständnis der Nachrichtenauswahl im Wissenschaftsjournalismus, was sagen sie aus über die Validität des entwickelten Modells der Nachrichtenauswahl als Entscheidungssituation? Da Studien wie diese in der Kommunikationswissenschaft bisher sehr selten sind, gibt es wenig Anhaltspunkte dafür, ab welchen Genauigkeitswerten man davon sprechen kann, dass ein Vorhersagemodell eine gute Repräsentation darstellt. Vergleicht man die Werte mit denen aus komplexeren Prognoseverfahren, die in der Informatik entwickelt werden, ist die Vorhersage nicht besonders gut. Gerade der praktische Nutzen der Prognose ist begrenzt, wenn die besonders interessanten Aussagen, dass nämlich eine Studie Aufmerksamkeit erhalten wird, nicht verlässlich sind.

Nimmt man dagegen die wenigen eher theoretisch fundierten Vorhersagemodelle als Maßstab, die bisher in der Journalismusforschung verwendet wurden, wirken die Ergebnisse deutlich besser. Und vergleicht man sie gar mit den Regressionsmodellen, mit denen der Einfluss von Nachrichtenfaktoren auf Umfang oder Platzierung gezeigt werden sollte, haben wir es hier mit einem sehr guten Ergebnis zu tun. Es kommt also auf die Perspektive an, wie man die erreichten Genauigkeitswerte interpretiert.

Drei Aspekte sprechen aber dafür, dass die Ergebnisse des Prognoseverfahrens durchaus vielversprechend sind. Erstens muss berücksichtigt werden, dass hier nur zwei beziehungsweise drei Prädiktoren für die Vorhersage verwendet wurden. In der Informatik ohnehin, aber auch in den Sozialwissenschaften werden meist deutlich komplexere Modelle berechnet, in die eine Vielzahl verschiedener Variablen eingeht. In dieser Arbeit ist es mit einem sehr simplen und dadurch gut interpretierbaren Modell gelungen, eine akzeptable Vorhersage eines komplexen Phänomens zu erreichen. Bedenkt man beispielsweise die vielen Faktoren der Mehrebenen-Modelle, mit der im Gatekeeping-Ansatz Nachrichtenauswahl erklärt werden soll, ist es erstaunlich, welche Präzision bereits nur mit dem Publikationsort und der Verfügbarkeit einer Pressemitteilung möglich war. Die niedrigeren Werte bei hohen Grenzwerten treten auch bei hochkomplexen Vorhersagemodellen auf und sind vor allem dadurch bedingt, dass bei der immer geringeren Fallzahl die Faktoren abnehmen, die die ausgewählten Studien gegenüber den nicht ausgewählten charakterisieren.

Zweitens sind die hohen Genauigkeitswerte bei der Vorhersage keiner oder geringer Aufmerksamkeit nicht allein das Ergebnis der ungleichen Verteilung der Studien auf diese Klassen, sondern eine Leistung des Modells. Dadurch haben sie auch einen Wert für das Verständnis des Wissenschaftsjournalismus. Um das zu verdeutlichen, muss man nur den Blickwinkel etwas verändern, von der Auswahl zur Nicht-Auswahl. Genauigkeitswerte von 90 % und mehr bei der Vorhersage, dass eine Studie keine Aufmerksamkeit erhalten wird, zeigen einmal mehr, wie hoch selektiv Wissenschaftsjournalisten bei der Auswahl von wissenschaftlichen Ergebnissen vorgehen. Denn: wenn es zu einer Studie keine Pressemitteilung gab oder sie in der falschen Zeitschrift erschienen ist, war die Wahrscheinlichkeit verschwindend gering, dass über diese Studie berichtet wird. Mit Kepplinger (1998: 22) könnte man sagen, dass im Wissenschaftsjournalismus eine Pressemitteilung und die Veröffentlichung im richtigen, renommierten Journal die „*Conditio sine qua non* für positive Selektionsentscheidungen“ sind. Ein Großteil der wissenschaftlichen Studien wird demnach nie für die Berichterstattung in Betracht gezogen.

Drittens zeigt die Verbesserung, die durch die Verwendung der inhaltlichen Informationen erzielt wurde, das Potenzial des Verfahrens, wenn weitere Faktoren integriert werden. Durch die Vektorrepräsentationen der Abstracts konnte eine abstrakte Darstellung ihrer Themen erstellt werden, die bereits eine Verbesserung der Vorhersage ermöglichte. Im Sinne eines praktischen Einsatzes des Verfahrens ist eine Vielzahl von Faktoren möglich, die in das Modell integriert werden könnten und die zum Teil in Kapitel 7.2 vorgestellt werden. Und bezüglich des theoretischen Nutzens des Ansatzes können die Ergebnisse dieser Arbeit mit denen von Modellen verglichen werden, die auf anderen theoretisch denkbaren Auswahl Faktoren basieren. Hier sind auch verschiedene Möglichkeiten denkbar. Beim Vergleich ist es entweder möglich, neue Faktoren in die bestehenden Modelle zu integrieren und zu prüfen, ob sich die Genauigkeit der Vorhersage verbessert und welche Permutation Importance die einzelnen Faktoren haben oder aber komplett neue Modelle zu entwickeln, um den Einfluss der Faktoren besser trennen zu können. Auch wenn sie aufgrund der geringen Zahl der verwendeten Features nicht richtig sinnvoll eingesetzt werden konnte, hat die Permutation auch in dieser Arbeit einige interessante Zusammenhänge gezeigt. Einer davon ist die geringere Bedeutung der Verfügbarkeit einer Pressemitteilung bei hohen MSM-Scores. Das könnte darauf hindeuten, dass es für die wenigen Studien, die von sehr vielen Medien ausgewählt werden, nicht mehr erheblich ist, ob es eine Pressemitteilung gab oder nicht (schließlich gibt es sehr viele Studien mit Pressemitteilung, die keine solche Aufmerksamkeit erzielen), die Zeitschrift dagegen hatte weiterhin einen Einfluss. Das könnte man wieder so interpretieren, dass es nur Studien aus sehr wenigen Zeitschriften schaffen, so eine hohe Aufmerksamkeit auf sich zu ziehen.

Welche Bedeutung die Ergebnisse nun für die Validität des in dieser Arbeit entwickelten Modells der Nachrichtenauswahl haben, darüber lassen sich vorsichtige Aussagen treffen. In erster Linie sind die Ergebnisse eine Bestätigung der vorgestellten Forschungsliteratur, in der der Einfluss des Publikationsorts und der Pressearbeit wissenschaftlicher Institutionen in diversen Arbeiten gezeigt werden konnte. Diese Ergebnisse werden durch die Input-Output-Analyse auf eine breitere Basis gestellt. Gleichzeitig spricht die Bestätigung dieser Faktoren auch für ein Modell der Nachrichtenauswahl, das sie integriert und ihre Bedeutung betont. Dafür bietet sich die Kombination aus TPB und Entscheidung unter Komplexität an, die im dritten Kapitel erstellt wurde. Die Struktur der Berichterstattung, die sich aus der Input-Output-Analyse ergeben hat, entspricht dem, was auf Grundlage des Handelns einzelner Journalisten zu erwarten wäre, die eine Auswahlentscheidung unter Unsicherheit treffen müssen. Die Beschränkung auf wenige Zeitschriften und die Orientierung an Pressemitteilungen können als Symptome einer simple-minded search oder aber eines Handelns auf Grundlage von

Rationalitätsfiktionen verstanden werden. Wenn man nicht unterstellt, dass alle Studien, die in den nicht beachteten Zeitschriften veröffentlicht wurden oder keine Pressemitteilung erhalten haben, aus anderen Gründen nicht den Auswahlkriterien der Journalisten entsprechen, liegt keine wirklich rationale Entscheidung vor.

Wenn man das Vorhersagemodell als sehr simple Repräsentation des Auswahlhandelns von Journalisten versteht, handeln sie wie folgt: es erfolgt einerseits eine Beobachtung weniger Zeitschriften und andererseits von Pressemitteilungen, beides vielleicht sogar in Kombination. Da auch die Zahl der Studien, die diesen Kriterien entsprechen, immer noch zu groß ist, als dass allein dadurch eine Auswahl möglich wäre, wird in dieser Vorauswahl nach weiteren Überlegungen ausgewählt, beispielsweise dem Thema einer Studie. Während der erste Schritt durch die Überlegungen Schimanks zu Entscheidungen unter Unsicherheit erklärt werden kann, könnte man für die genauere Untersuchung des zweiten Schritts Modelle auf Basis der Theory of Planned Behavior verfeinern. Unabhängig von der theoretischen Grundlage, mit der man das empirisch beobachtete Auswahlhandeln erklären möchte, hat es verschiedene Konsequenzen für die Leistungsfähigkeit des Journalismus, die im nächsten Kapitel angerissen werden.

7.1 Folgen der Ergebnisse für die Leistungsfähigkeit des Journalismus

Das Bild des Wissenschaftsjournalismus, das sich aus den vorgestellten Ergebnissen ergibt, ist das eines Beobachters zweiter Ordnung. Die große Zahl der Auswahlentscheidungen basiert auf verschiedenen vorher getroffenen Entscheidungen. Die Studien, unter denen Journalisten wählen, haben bereits mehrere Filterstufen durchlaufen. Einerseits die Auswahl und das Peer Review verschiedener wissenschaftlicher Zeitschriften und andererseits die Kommunikationsabteilungen oder Redaktionen, die einzelne Studien für Pressemitteilungen auswählen. Damit gibt es verschiedene andere Akteure, die abgesehen von Wissenschaftsjournalisten entscheiden, welches Bild von Wissenschaft die Öffentlichkeit hat. Zu nennen sind hier vor allem die Redakteure von Zeitschriften wie Nature und Science, die auswählen, welche Studien überhaupt ins Peer Review gelangen. Solange die Rationalitätsfiktionen gut begründet sind, auf denen die Orientierung an diesen anderen Akteuren und ihren Auswahlentscheidungen beruht, ist dagegen nichts zu sagen, es ist für den Journalismus sogar äußerst funktional. Wenn die großen Zeitschriften wegen ihrer Beliebtheit unter Wissenschaftlern nur die besten, verlässlichsten und relevantesten Arbeiten auswählen und wissenschaftliche Institutionen bei ihrer Pressearbeit darauf achten, solide Forschungsergebnisse auszuwählen, wird es dem Journalisten erspart, eine Vorauswahl zu treffen, die er aus Zeitmangel und fehlender Expertise gar nicht treffen könnte. Grundlage für

die Leistungsfähigkeit eines Journalismus, der sich so stark an wissenschaftlichen Selbstbeschreibungen orientiert, ist also, dass diese verlässlich sind (vgl. Lehmkuhl/Promies 2019).

Verschiedene Entwicklungen in den letzten Jahren haben jedoch dazu geführt, dass die Verlässlichkeit der Beschreibungen abgenommen hat. Zu nennen ist dabei eine Zunahme an Betrugsfällen oder fragwürdigen Forschungspraktiken (vgl. bspw. Anderson et al. 2007a; Fanelli 2009; Anderson et al. 2010; Steen et al. 2013), das Aufkommen sogenannter Predatory Publisher, die Zeitschriften herausgeben, die höchstens vorgeben, wissenschaftlichen Qualitätsansprüchen zu genügen (vgl. bspw. Shen/Björk 2015; Moher et al. 2017; Shamseer et al. 2017; Cobey et al. 2018) oder die bereits angesprochene Entwicklung, dass vor allem in Disziplinen wie der Medizin oder Biologie die Ergebnisse eines hohen Anteils von Studien nicht bestätigt werden können (vgl. bspw. Ioannidis 2005; Grimes et al. 2018). Man kann diese Entwicklungen mit Veränderungen im Wissenschaftssystem in Verbindung bringen, die zu einer erhöhten Konkurrenz zwischen Forschern und Institutionen geführt haben, die vor allem über Publikationsstatistiken oder Drittmittelinwerbung ausgetragen wird (vgl. Anderson et al. 2007b; Fanelli 2010; Grimes et al. 2018). In Deutschland sind hier vor allem die Veränderungen des Hochschulsystems im Rahmen des New Public Management zu nennen.

Neben einer Schwächung der staatlichen Regulierung und akademischen Selbstverwaltung kam es vor allem zu einer Stärkung der hierarchischen Selbststeuerung durch Rektorate oder Präsidien, einer Zunahme der Außensteuerung durch den Staat oder andere Akteure, beispielsweise durch Zielvereinbarungen sowie einer Stärkung des Wettbewerbs innerhalb und zwischen den Hochschulen (vgl. Grande et al. 2013: 25–26). Gerade Außensteuerung und zunehmender Wettbewerb befördern dabei die Verwendung von Instrumenten zur Leistungsmessung wissenschaftlicher Arbeit, die unter anderem dazu führen, dass Forscher versuchen, die Menge ihrer Publikationen zu optimieren (vgl. Schimank 2010; Grande et al. 2013: 28). Schimank stellt einen direkten Zusammenhang her zwischen diesem Streben nach mehr und hochrangigeren Publikationen und der Zunahme dysfunktionaler Entwicklungen in der Wissenschaft. Auf Grund ihrer besonderen innerwissenschaftlichen Bedeutung sind die renommierten Zeitschriften zusätzlich von diesem Problem betroffen, die Rate nicht-replizierbarer Ergebnisse und von zurückgezogenen Ergebnissen ist hier besonders hoch (vgl. Grieneisen/Zhang 2012; Hesselmann et al. 2017). Das könnte daran liegen, dass Wissenschaftler bei hochrangigen Zeitschriften eher auf fragwürdige Methoden zurückgreifen, um möglichst spektakuläre Ergebnisse präsentieren zu können.

Dazu kommt noch eine weitere Entwicklung, die mit dem New Public Management verbunden ist. Der Staat fällt zunehmend als Mittler zwischen Hochschulen und gesellschaftlichen Anspruchsgruppen weg. Die Hochschulen müssen den Erwartungen und Forderungen dieser unterschiedlichen Gruppen also direkt und über die Medien gegenüberreten (vgl. Friedrichsmeier et al. 2013: 8, Grande et al. 2013: 35f.). Gleichzeitig wird im Wettbewerb zwischen den Hochschulen um Fördergelder, Studierende und Mitarbeiter öffentliche Sichtbarkeit zu einem wichtigen Faktor. Es hat sich ein Gefüge wechselseitiger Beobachtung zwischen Hochschulen, Wissenschaftspolitik, Wissenschaftlern und gesellschaftlichen Anspruchsgruppen entwickelt (vgl. ebd.: 36., Friedrichsmeier/Fürst 2012: 53). Massenmedien ermöglichen diese Beobachtungen, Sichtbarkeit in den Medien entwickelt sich daher zu einem Indikator für die Aufmerksamkeit von Politik und Gesellschaft (vgl. ebd.: 55). Durch ihre Selbstdarstellung in Medien können wissenschaftliche Organisationen versuchen, die Wissenschaftsbeobachtung anderer gesellschaftlicher Akteure zu beeinflussen (vgl. Peters et al. 2013: 311). Die Bedeutung der Medienberichterstattung für die Hochschulen zeigt sich beispielsweise in Befragungen von Entscheidungsträgern (vgl. Heinrichs/Petersen 2009; Blöbaum et al. 2013; Friedrichsmeier et al. 2013). Diese sehen einen Druck zu mehr medialer Sichtbarkeit und erhoffen sich einen Bedeutungsgewinn in den Augen von politischen Entscheidern, denen sie eine hohe Mediennutzung und -orientierung unterstellen (vgl. Friedrichsmeier/Fürst 2012: 50–52; Friedrichsmeier et al. 2013: 16). Praktische Folge dieser Medienorientierung ist beispielsweise eine Zunahme der strategischen Medienkommunikation von wissenschaftlichen Institutionen, die sich in erhöhten Mitteln und Kapazitäten von Presseabteilungen oder verstärkter Aussendung von Pressemitteilungen zeigt (vgl. Friedrichsmeier et al. 2013: 23).

Beide Entwicklungen zusammen stellen eine Herausforderung für die Leistungsfähigkeit des Wissenschaftsjournalismus dar. Einerseits ist das Renommee und Peer Review wissenschaftlicher Zeitschriften nicht mehr unbedingt ein Garant für die Verlässlichkeit der Ergebnisse, die in ihnen publiziert werden. Zunehmender Wettbewerb in der Wissenschaft befördert die Verwendung fragwürdiger Forschungs- und Publikationspraktiken, die sowohl das wissenschaftliche Publikationswesen wie auch in der Folge den Wissenschaftsjournalismus je nach Blickwinkel vor ernsthafte Herausforderungen stellen oder gar überfordern. Hinzu kommt, dass das Streben wissenschaftlicher Institutionen nach Mediensichtbarkeit dazu führt, dass diese in ihrer Pressearbeit zunehmend strategisch kommunizieren. Eine Folge sind die bereits angesprochenen Übertreibungen und Falschdarstellungen in Pressemitteilungen, die der Journalismus nicht korrigieren kann (vgl. Woloshin/Schwartz 2002; Woloshin et al. 2009; Sumner et al. 2014; Sumner et al. 2016). Die Folgen dieser Entwicklungen fasst Lehmkuhl (2019: 209) so zusammen:

„Die Reputation des aktuellen Wissenschaftsjournalismus ist komplexitätsbedingt relativ eng gekoppelt an die Reputation der Wissenschaft. Wenn sich die wissenschaftlichen Selbstbeschreibungen aber zunehmend als unzuverlässig und strategisch motiviert erweisen, dann wirkt das unmittelbar auf die Leistungsfähigkeit dieses Journalismus zurück, dessen redaktionelle Selektionsprogramme auf die Redlichkeit wissenschaftlichen Kommunizierens angewiesen sind.“

7.2 Limitationen und mögliche Erweiterungen

Gerade aus der verwendeten Methode ergeben sich einige Limitationen für die Aussagekraft der Ergebnisse dieser Arbeit. Dabei kann zwischen solchen Limitationen unterschieden werden, die sich aus dem methodischen Ansatz an sich ergeben, und solchen, die sich aus dessen konkreter Umsetzung ergeben. Betrachtet werden sollen zuerst die Einschränkungen aus der zweiten Gruppe, die gleichzeitig auch Möglichkeiten für Erweiterungen oder Verbesserungen des Ansatzes aufzeigen. Die größte Schwäche der Untersuchung muss in den verwendeten Datenquellen verortet werden. Wie bereits diskutiert wurde, sind insbesondere die Altmetric-Daten, mit denen der journalistische Output erhoben wurde, ein wenig verlässlicher Indikator für tatsächliche journalistische Aufmerksamkeit. Das betrifft sowohl die konkrete Höhe des Scores, die eigentlich nicht interpretiert werden kann, als auch die Unterscheidung zwischen Studien mit einem Score von 0 oder 1. Gerade die niedrigen Werte können sich auch dadurch ergeben, dass eine Pressemitteilung auf einer anderen Website kopiert wurde, ohne dass wirklich von einer Auswahlentscheidung gesprochen werden kann.

Es gibt zwei Möglichkeiten, wie damit umgegangen werden kann. Entweder können die Altmetric-Scores im Nachhinein bereinigt werden, um nur „echte“ Medienaufmerksamkeit zu identifizieren. MacLaughlin und Kollegen (2018) identifizieren dafür Aggregatoren über verschiedene Strategien wie die Publikationshäufigkeit oder die Ähnlichkeit von Artikeln zu Pressemitteilungen. Dafür wird jedoch ein Zugriff auf die Medienbeiträge benötigt, in denen die Erwähnung vorliegt. Dieser war in dieser Arbeit nicht gegeben. Gleichzeitig stellt sich die Frage, ob bei dem verbundenen Aufwand nicht auch die zweite Möglichkeit gewählt werden könnte. Diese besteht darin, den journalistischen Output selbst zu erheben. Benötigt würde dafür einerseits ein Zugriff auf Volltexte journalistischer Beiträge und ein Verfahren, das automatisiert Referenzen auf Studien in diesen Beiträgen erkennt und zuordnen kann. Ein Vorteil dieser Methode wäre es, dass eine systematische Stichprobe von Medientiteln selbst ausgewählt werden kann und Vergleiche der Auswahlpraxis verschiedener Medien oder Länder mit geringem Aufwand möglich wäre.

Auf der Seite der Input-Daten konnte durch verschiedene Beschränkungen nur ein Zehntel der im Untersuchungszeitraum veröffentlichten Studien für das Vorhersageverfahren verwendet werden. Das ist zwar immer noch eine umfassende Abdeckung der Studienergebnisse eines Jahres, jedoch keine vollständige Input-Output-Analyse. Bedingt war diese Einschränkung vor allem durch den Export von Abstracts aus Scopus. Eine Lösung wäre hier wohl nur durch einen Zugriff auf die vollständige Scopus-Datenbank möglich, der nicht auf 2000 Studien pro Export beschränkt ist. Dieser war hier jedoch nicht gegeben. Es kann aber auch davon ausgegangen werden, dass die verringerte Studienzahl die Aussagekraft der Untersuchung nicht beeinträchtigt. Eine weitere Einschränkung gerade für die praktische Verwendung des Vorhersageverfahrens ist die geringe Genauigkeit bei der positiven Vorhersage. Diese könnte vielleicht durch die Einbindung weiterer Variablen verbessert werden, wodurch die Ergebnisse des Verfahrens dann aber inhaltlich schwieriger interpretierbar wären. Mögliche zusätzliche Faktoren wären die Affiliations der Studienautoren oder zu deren Renommee, sowohl wissenschaftlich in Form von Zitationen oder auch öffentlich in Form von früheren Medienauftritten. Zudem könnte die Einbindung inhaltlicher Aspekte noch verbessert werden. Hier wäre eine Anpassung von automatisierten Verfahren denkbar, mit denen Nachrichtenfaktoren erfasst werden (wobei deren Einfluss erst geprüft werden müsste) oder die händische Identifizierung bestimmter Trendthemen, die dann über Schlagworte gesucht werden könnten.

Neben diesen spezifischeren methodischen Einschränkungen gibt es auch einige grundlegende Schwachstellen des methodischen Vorgehens. Diese basieren auf der Verwendung eines Machine Learning-Verfahrens sowie insbesondere auf der Erhebung der thematischen Variablen. Auch wenn Random Forests im Vergleich zu anderen Ansätzen aus der Informatik wie neuronalen Netzen relativ gut nachvollziehbar zu ihren Prognosen kommen, ist die Interpretierbarkeit ihrer Ergebnisse doch eingeschränkt. Das zeigt sich beispielsweise in der Bedeutung einzelner Ausprägungen der Prädiktoren. Ohne die deskriptive Analyse beispielsweise hätte nicht beurteilt werden können, welche Zeitschriften eine hohe Auswahlwahrscheinlichkeit haben. In Random Forests ist es anders als bei Entscheidungsbäumen nicht mehr nachvollziehbar, wie konkret die Untergruppen gebildet werden, aus denen die Prognosen ermittelt werden. Rein aus den Ergebnissen der Vorhersagemodelle könnte nur abgeleitet werden, dass der Publikationsort einen Einfluss hat, aber nicht, wie dieser genau aussieht. Gleiches gilt für das Zusammenspiel der Prädiktoren. Wie genau aus diesen die Vorhersage erstellt wird, kann nicht nachvollzogen werden. Um die Auswirkungen dieser Einschränkung möglichst gering zu halten, wurden wenige Prädiktoren

verwendet, deren Einfluss vor dem Hintergrund des empirischen Forschungsstands plausibel ist.

Noch stärker gilt diese Einschränkung für die Näherung des Themas, die mit dem Doc2Vec-Modell erstellt wurde. Hier liegt eine Black Box vor, bei der nicht beurteilt werden kann, wie die Vektorrepräsentationen der Abstracts erstellt wurden oder was sie aussagen. So könnten keinerlei Aussagen darüber getroffen werden, welche Themen denn besonders oft ausgewählt werden. Möglich ist mit dieser Darstellung erst einmal nur die Aussage, dass der Inhalt einen Einfluss hat. Erweiterungen oder Verbesserungen des hier verwendeten Verfahrens könnten versuchen, besser interpretierbare Darstellungen des Inhalts zu verwenden. Bei LDA-Topics beispielsweise kann in einem gewissen Rahmen nachvollzogen werden, was sie inhaltlich erfassen. Problematisch ist hier, dass diese je nach Zahl verwendeter Themen zu ungenau sein könnten, um eine gute Vorhersage zu ermöglichen. Insgesamt muss hier genauso wie bei der Wahl des Prognoseverfahrens abgewogen werden zwischen der Genauigkeit des Verfahrens (dem praktischen Nutzen) und der Verständlichkeit (dem Nutzen für das Verständnis des Phänomens, das vorhergesagt werden soll).

8. Fazit und Ausblick

Mehr als 5000 neue wissenschaftliche Ergebnisse werden im Schnitt jeden Tag veröffentlicht, das sind mehr als zwei Millionen Studien pro Jahr. Von dieser unfassbaren Menge neuer Erkenntnisse schafft nur ein kleiner Teil über die Berichterstattung von Massenmedien den Weg in die Öffentlichkeit und trägt zum öffentlichen Bild der Wissenschaft bei. Diese Arbeit hat sich der Frage gewidmet, nach welchen Kriterien diese wenigen Studien von Wissenschaftsjournalisten ausgewählt werden. Damit wird ein etablierter und seit Jahrzehnten erforschter Kernbereich der Journalismusforschung adressiert, die Nachrichtenauswahl. Untersucht und erklärt werden sollte, wie Journalisten in einem kleinen und gut abgrenzbaren Teilbereich des Journalismus, der aktuellen Berichterstattung über neue wissenschaftliche Ergebnisse, Auswahlentscheidungen treffen. Dieser Bereich bietet sich besonders gut für die Untersuchung dieser Fragestellung an, da hier die idealen Voraussetzungen für eine umfangreiche Input-Output-Analyse vorliegen. Obwohl es bisher sehr selten eingesetzt wurde, ist dieses Untersuchungsdesign die beste Möglichkeit, um die konkrete Nachrichtenauswahl direkt zu untersuchen und zu erklären. Durch die digitale Veröffentlichung und Archivierung von Studien, die beispielsweise in multidisziplinären Literaturdatenbanken erfolgt, kann der Input in sehr großem Umfang erfasst werden. Auf der Outputseite kann die Berichterstattung über wissenschaftliche Studien durch altmetrics abgeschätzt werden. Verschiedene Anbieter

wie das Unternehmen Altmetric suchen automatisiert in verschiedenen Online-Quellen, darunter auch Medientitel, nach Erwähnungen von Forschungsprodukten. Diese Erwähnungen werden zu verschiedenen Scores zusammengefasst, von denen der Mainstream-Media-Score zur Abschätzung der Medienresonanz von Studien verwendet wurde.

Konkret untersucht wurde in der Input-Output-Analyse dann, welche der 1,7 Millionen Studien, die von August 2016 bis Juli 2017 in der Literaturdatenbank Scopus veröffentlicht wurden, in mindestens einem der über 2000 Medientitel erwähnt wurde, die Altmetric beobachtet. Als Zwischenergebnis konnte so die Struktur der Nachrichtenauswahl im Wissenschaftsjournalismus ermittelt werden. Um zu erklären, wie es zu dieser Struktur gekommen ist, welche Faktoren also für konkrete Auswahlentscheidungen verantwortlich sind, wurde ein innovatives methodisches Vorgehen verwendet. Dabei wurden die Bedeutung unterschiedlicher Merkmale der Studien in einem Prognoseverfahren geprüft. Mit diesem Ansatz können durch einen Vergleich der vorhergesagten mit den tatsächlich beobachteten Auswahlentscheidungen die Annahmen geprüft werden, die dem Vorhersagemodell zugrunde liegen. Eine genaue Vorhersage spricht dafür, dass die wirksamen Faktoren identifiziert und in das Modell integriert wurden, eine ungenaue Vorhersage dafür, dass wichtige Einflüsse übersehen und unwirksame Faktoren ins Modell integriert wurden.

Die theoretische Grundlage für das Prognoseverfahren ist ein Modell der Nachrichtenauswahl im Wissenschaftsjournalismus als Entscheidungssituation unter Komplexität. Dieses Modell soll verschiedene Schwächen der Ansätze beheben, die bisher für die Erklärung der Nachrichtenauswahl entwickelt wurden. In der Forschung zum allgemeinen Journalismus vorherrschend sind vor allem die Nachrichtenwerttheorie und der Gatekeeping-Ansatz. Insbesondere im Vergleich mit alternativen Ansätzen wurde offensichtlich, dass beide Ansätze die konkrete Entscheidung zwischen verschiedenen Handlungsalternativen vernachlässigen oder nicht berücksichtigen. Die Nachrichtenwertforschung erklärt zum Beispiel nicht, wie eine Auswahl in einer Situation getroffen werden soll, in der sehr viele Berichterstattungsanlässe mit ähnlichem Nachrichtenwert vorliegen. Die Gatekeeper-Forschung benennt dagegen eine Vielzahl von Einflussfaktoren auf Journalisten, die in Modelle mit verschiedenen hierarchisch angeordneten Ebenen eingeordnet wurden. Dafür wird selten betrachtet, wie diese Faktoren konkret zusammenspielen, um die Auswahl individueller Journalisten zu bestimmen.

Die Forschung zur Nachrichtenauswahl zum Wissenschaftsjournalismus ist bisher geprägt durch verschiedene verstreut vorliegende Einzelergebnisse, die nur in Einzelfällen in systematischere Modelle integriert wurden. Herausgestellt wurde neben einem Einfluss der

Themen und Inhalte von Studien vor allem die Bedeutung einer kleinen Gruppe renommierter Zeitschriften sowie der Pressearbeit wissenschaftlicher Institutionen. Zusammengenommen ist also ein theoretisches Modell erforderlich, das einerseits Journalisten als handelnde Akteure konzipiert, die in konkreten Situationen eine Entscheidung zwischen verschiedenen Alternativen treffen und andererseits die Erkenntnisse aus der Forschung zum Wissenschaftsjournalismus integrieren kann. Als Antwort auf diese Erfordernisse wurde ein Modell aus zwei Komponenten entwickelt. Die erste ist eine Anwendung der Theory of Planned Behavior auf den Wissenschaftsjournalismus. Dabei handelt es sich um eine Handlungstheorie, die das Handeln von Akteuren über deren Intentionen erklärt. Diese wiederum ergeben sich aus ihren Handlungseinstellungen, der wahrgenommenen sozialen Norm und der wahrgenommenen Kontrolle über ihr Handeln. Dadurch wird der Fokus auf die Wahrnehmung der Journalisten gelegt. Gleichzeitig wurden Erkenntnisse von Schimank zur Komplexität von Entscheidungssituationen auf den Wissenschaftsjournalismus übertragen. Demnach liegt in der Situation der Nachrichtenauswahl Komplexität in der sozialen, sachlichen und zeitlichen Dimension vor. Diese Komplexität macht eine rationale Entscheidung unmöglich und zwingt die Journalisten zu alternativen Handlungsstrategien. Zu nennen waren hier vor allem die „simple-minded“ search und Rationalitätsfiktionen, an denen das Handeln ausgerichtet wird. Beide Handlungsweisen führen dazu, dass nur wenige Quellen überhaupt für die Auswahl in Betracht gezogen werden. Zusammengenommen wird die Nachrichtenauswahl durch die Komplexität der Entscheidungssituation auf einen Bruchteil der Alternativen beschränkt, zwischen denen Journalisten dann basierend auf ihren Intentionen wählen. Eingefasst werden kann dieser Ansatz in ein Modell soziologischer Erklärung, in dem über die Logik der Aggregation auch das Entstehen sozialer Strukturen, in diesem Fall Mustern der Berichterstattung mit dem Handeln einzelner Akteure erklärt werden kann.

Die zentralen Faktoren in diesem Modell sind der Publikationsort, die Verfügbarkeit einer Pressemitteilung und der Inhalt der Studien. Alle drei Faktoren wurden daher für die Entwicklung eines Vorhersagemodells verwendet. Dieses basiert auf einem Machine Learning-Verfahren, sogenannten Random Forests. Entwickelt wurden sechs verschiedene Modelle, in denen die Menge verwendeter Variablen und der Grenzwert variiert wurden, ab dem von Medienaufmerksamkeit gesprochen wurde. Durch den Vergleich der Modelle kann einerseits der Einfluss der verschiedenen Faktoren für die Genauigkeit der Vorhersage geprüft werden und andererseits die Leistung des Verfahrens bei der Prognose seltener Fälle hoher Aufmerksamkeit. Insgesamt war die Vorhersage immer dann besonders genau, wenn keine Aufmerksamkeit vorhergesagt wurde. Das heißt, die Studien, die keine Aufmerksamkeit

erhielten, wurden in der Mehrzahl der Fälle richtig erkannt und die Vorhersage, dass eine Studie keine Aufmerksamkeit erhält, war meistens richtig. Ungenauer war das Verfahren dagegen, wenn prognostiziert wurde, dass eine Studie Aufmerksamkeit erhält. Das zeigte sich besonders bei höheren Grenzwerten, bei denen auf die positiven Vorhersagen kein Verlass mehr war.

Für das vorgestellte theoretische Modell bedeutet das zweierlei: die erste Komponente kann schon relativ gut belegt werden, die Menge und Spezialität der Studienergebnisse und die Komplexität, die sich daraus ergibt, führen dazu, dass ein Großteil neuer Studien nie für die Auswahl in Betracht gezogen wird. Die Vorauswahl orientiert sich an dem Publikationsort und der Pressearbeit von wissenschaftlichen Institutionen. Die zweite Komponente, die bis auf einen eher diffusen Einfluss des Themas auch noch nicht weiter spezifiziert wurde, kann noch präzisiert werden, um die Auswahl unter den verbliebenen Alternativen zu erklären. Dafür bieten sich Erweiterungen des Prognoseverfahrens an, die weitere Faktoren integrieren. Der große soziale Einfluss von Wissenschaftsverlagen und der Pressearbeit wissenschaftlicher Institutionen, der sich in den Ergebnissen zeigt, könnte vor aktuellen Veränderungen im Wissenschaftssystem zu einer Gefahr für die Leistungsfähigkeit des Journalismus werden. Wenn vor dem Hintergrund einer zunehmenden Konkurrenz zwischen Wissenschaftlern und Institutionen, der durch Entwicklungen wie das New Public Management verstärkt wird, Selbstbeschreibungen der Wissenschaft, denen Journalisten vertrauen, immer weniger verlässlich sind, kann das die Reputation des Journalismus beeinträchtigen.

Welche praktischen Konsequenzen man aus diesen Ergebnissen nun ziehen könnte, ist fraglich. Denn allein aus der Erkenntnis, dass der Wissenschaftsjournalismus sich an Indikatoren orientiert, die möglicherweise zurzeit ihren Wert verlieren, kann man noch nicht ableiten, wie es besser gehen sollte. Denn dass bestimmte Heuristiken notwendig sind, um zu einer irgendwie begründeten Auswahl der Millionen von Studien zu kommen, die jedes Jahr neu erscheinen, ist klar. Und die Publikationsmenge im wissenschaftlichen System wird sich wohl kaum verringern, eher sogar noch weiter vergrößern. Ob es andere, bessere Strategien oder Methoden gibt, mit denen relevante, interessante und verlässliche Ergebnisse identifiziert werden können, muss Gegenstand einer praxisnahen Forschung zum Wissenschaftsjournalismus sein. Diese Untersuchung hat eine Grundlage für solche Forschung geliefert. Einerseits durch die Verdeutlichung des Ausmaßes der Abhängigkeit von wissenschaftlichen Zeitschriften und institutioneller Pressearbeit und andererseits durch die Anwendung eines Verfahrens, das eine Lösung für das Auswahlproblem sein könnte. Denn Vorhersagemodelle, denen es in einer etwas anderen Zielsetzung gelingt, die Studien zu identifizieren, über die berichtet werden sollte (woran man das auch immer festmachen möchte), könnten ein Mittel sein, um dem

Wissenschaftsjournalismus zu größerer Unabhängigkeit von der Vorauswahl zu verhelfen, die die Wissenschaft für ihn trifft (vgl. Lehmkuhl 2019: 209–210).

Anhang

Tabelle 11: Inhaltsanalysen, in denen die Quellenauswahl von Wissenschaftsjournalisten untersucht wurde.

Studie	Van Trier et al 1994	Evans 1995	Semir 1996	Pahl 1998	Bubela, Caulfield 2004	Moriarty 2010	Suleski, Ibaraki 2010	Kiernan 2016
Methode	Inhaltsanalyse	Inhaltsanalyse	Inhaltsanalyse	Inhaltsanalyse	Inhaltsanalyse	Inhaltsanalyse	Inhaltsanalyse	Inhaltsanalyse
Wichtigste Quellen	NEJM Lancet	Science Nature	NEJM New Scientist	Nature Science	Science Nature	NEJM Journal of the National Cancer Institute	NEJM JAMA	Nature Science
	Nature BMJ	NEJM JAMA	Lancet BMJ Science Nature	NEJM Lancet New Scientist PNAS	Nature Genetics, Cell	JAMA Science Lancet Cancer	Science Nature Circulation Neurology	PNAS NEJM JAMA Archives of Internal Medicine Lancet
			JAMA	JAMA Nature Medicine BMJ		PNAS Journal of Clinical Oncology Annals of Internal Medicine CA: A Journal for Clinicians		
				Münchner Medizinische Wochenschrift		Nature		

Tabelle 13: Confusion Matrix des Vorhersagemodells ohne inhaltliche Informationen, Grenzwert bei MSM-Score ≥ 10 . Es wurden nur Studien verwendet, deren Score mindestens bei 1 liegt.

Gesamtzahl vorhergesagter Fälle: 7389	Vorhergesagte Klasse: keine Medienaufmerksamkeit	Vorhergesagte Klasse: Medienaufmerksamkeit
Echte Klasse: keine Medienaufmerksamkeit	RN: 4243	FP: 1143
Echte Klasse: Medienaufmerksamkeit	FN: 553	RP: 1450

Tabelle 14: Confusion Matrix des Vorhersagemodells ohne inhaltliche Informationen, Grenzwert bei MSM-Score ≥ 50 . Es wurden nur Studien verwendet, deren Score mindestens bei 1 liegt.

Gesamtzahl vorhergesagter Fälle: 30492	Vorhergesagte Klasse: keine Medienaufmerksamkeit	Vorhergesagte Klasse: Medienaufmerksamkeit
Echte Klasse: keine Medienaufmerksamkeit	RN: 4661	FP: 2302
Echte Klasse: Medienaufmerksamkeit	FN: 135	RP: 291

Tabelle 15: Confusion Matrix des Vorhersagemodells mit inhaltlichen Informationen, Grenzwert bei MSM-Score ≥ 10 .

Gesamtzahl vorhergesagter Fälle: 30492	Vorhergesagte Klasse: keine Medienaufmerksamkeit	Vorhergesagte Klasse: Medienaufmerksamkeit
Echte Klasse: keine Medienaufmerksamkeit	RN: 27278	FP: 1282
Echte Klasse: Medienaufmerksamkeit	FN: 485	RP: 1447

Tabelle 16: Confusion Matrix des Vorhersagemodells mit inhaltlichen Informationen, Grenzwert bei MSM-Score ≥ 50 .

Gesamtzahl vorhergesagter Fälle: 30492	Vorhergesagte Klasse: keine Medienaufmerksamkeit	Vorhergesagte Klasse: Medienaufmerksamkeit
Echte Klasse: keine Medienaufmerksamkeit	RN: 29542	FP: 522
Echte Klasse: Medienaufmerksamkeit	FN: 289	RP: 139

Tabelle 17: Confusion Matrix des Vorhersagemodells mit inhaltlichen Informationen, Grenzwert bei MSM-Score ≥ 10 . Es wurden nur Studien verwendet, deren Score mindestens bei 1 liegt.

Gesamtzahl vorhergesagter Fälle: 7389	Vorhergesagte Klasse: keine Medienaufmerksamkeit	Vorhergesagte Klasse: Medienaufmerksamkeit
Echte Klasse: keine Medienaufmerksamkeit	RN: 4595	FP: 791
Echte Klasse: Medienaufmerksamkeit	FN: 613	RP: 1390

Literatur

- Adie, Euan/Roe, William (2013): »Altmetric: enriching scholarly content with article-level discussion and metrics«, in: *Learned Publishing* 26 (1), S. 11-17.
- Adolf, Marian (2017): »Komplexität als Herausforderung der Kommunikationswissenschaft: von Landkarten, Strategien und Fallen«, in: *Medien & Kommunikationswissenschaft* 65 (3), S. 500-516.
- Ajzen, Icek (2012): »Martin Fishbein's legacy: The Reasoned Action Approach«, in: *The Annals of the American Academy of Political and Social Science* 640, S. 11-27.
- Allan, Stuart (2009): »Making science newsworthy: Exploring the conventions of science journalism«, in: Richard Holliman/Elizabeth Whitelegg/Eileen Scanlon et al. (Hg.), *Investigating science communication in the information age. Implications for public engagement and popular media*, Oxford, Milton Keynes: Oxford Univ. Press; Open Univ, S. 149-165.
- Althoff, Tim/Borth, Damian/Hees, Jörn/Dengel, Andreas (2013): »Analysis and forecasting of trending topics in online media streams«, in: Alejandro Jaimes (Hg.), *Proceedings of the 2013 ACM Multimedia Conference & co-located workshops*, New York, NY: ACM, S. 907-916.
- Amend, Elyse/Secko, David M. (2012): »In the face of critique: A metasynthesis of the experiences of journalists covering health and science«, in: *Science Communication* 34 (2), S. 241-282.
- Anderson, C. W. (2011): »Between creative and quantified audiences: Web metrics and changing patterns of newswork in local US newsrooms«, in: *Journalism* 12 (5), S. 550-566.
- Anderson, Melissa S./Martinson, Brian C./Vries, Raymond de (2007a): »Normative dissonance in science: results from a national survey of U.S. scientists«, in: *Journal of Empirical Research on Human Research Ethics* 2 (4), S. 3-14.
- Anderson, Melissa S./Ronning, Emily A./De Vries, Raymond/Martinson, Brian C. (2007b): »The perverse effects of competition on scientists' work and relationships«, in: *Science and Engineering Ethics* 13 (4), S. 437-461.

- (2010): »Extending the mertonian norms: Scientists' subscription to norms of research«, in: *Journal of Higher Education* 81 (3), S. 366-393.
- Arapakis, Ioannis/Cambazoglu, B. B./Lalmas, Mounia (2014): »On the feasibility of predicting news popularity at cold start«, in: Luca M. Aiello/Daniel McFarland (Hg.), *Social informatics. 6th international conference, SocInfo 2014*, Cham: Springer, S. 290-299.
- Arapakis, Ioannis/Cambazoglu, Berkant B./Lalmas, Mounia (2017): »On the feasibility of predicting popular news at cold start«, in: *Journal of the Association for Information Science and Technology* 68 (5), S. 1149-1164.
- Badenschier, Franziska/Wormer, Holger (2012): »Issue selection in science journalism: Towards a special theory of news values for science news?«, in: Simone Rödder/Martina Franzen/Peter Weingart (Hg.), *The sciences' media connection. Public communication and its repercussions*, Dordrecht: Springer, S. 59-85.
- Baerns, Barbara (1991): *Öffentlichkeitsarbeit oder Journalismus? Zum Einfluss im Mediensystem*, Köln: Wissenschaft und Politik.
- Bakkalbasi, Nisa/Bauer, Kathleen/Glover, Janis/Wang, Lei (2006): »Three options for citation tracking: Google Scholar, Scopus and Web of Science«, in: *Biomedical Digital Libraries* 3, S. 7.
- Bandari, Roja/Asur, Sitaram/Huberman, Bernardo A. (2012): »The pulse of news in social media: forecasting popularity«, in: *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*, Palo Alto, Calif.: AAAI Press.
- Bartlett, Christoph/Sterne, Jonathan/Egger, Matthias (2002): »What is newsworthy? Longitudinal study of the reporting of medical research in two British newspapers«, in: *British Medical Journal* 325 (7355), S. 81-84.
- Bengio, Yoshua/Ducharme, Réjean/Vincent, Pascal/Jauvin, Christian (2003): »A neural probabilistic language model«, in: *Journal of Machine Learning Research* 3 (6), S. 1137-1155.
- Berger, Jonah/Milkman, Katherine L. (2012): »What makes online content viral?«, in: *Journal of Marketing Research* 49 (2), S. 192-205.

- Berk, Richard A. (2006): »An introduction to ensemble methods for data analysis«, in: *Sociological Methods & Research* 34 (3), S. 263-295.
- Blei, David M./Ng, Andrew Y./Jordan, Michael I. (2003): »Latent Dirichlet allocation«, in: *Journal of Machine Learning Research* 3 (4-5), S. 993-1022.
- Blöbaum, Bernd (2017): »Wissenschaftsjournalismus«, in: Heinz Bonfadelli/Birte Fähnrich/Corinna Lüthje et al. (Hg.), *Forschungsfeld Wissenschaftskommunikation*, Wiesbaden: Springer VS, S. 221-238.
- Blöbaum, Bernd/Görke, Alexander/Wied, Kristina (2004): *Quellen der Wissenschaftsberichterstattung. Inhaltsanalyse und Befragung. Projektbericht*, Münster.
- Blöbaum, Bernd/Scheu, Andreas M./Summ, Annika/Volpers, Anna-Maria (2013): »Medien, Fächer und Politik. Wie Medien forschungspolitische Entscheidungen in verschiedenen wissenschaftlichen Disziplinen beeinflussen.«, in: Edgar Grande/Dorothea Jansen/Otfried Jarren et al. (Hg.), *Neue Governance der Wissenschaft. Reorganisation - externe Anforderungen - Medialisierung*, Bielefeld: transcript Verlag, S. 289-309.
- Böhle, Fritz/Wehrich, Margit (2009): »Ungewissheit, Uneindeutigkeit, Unsicherheit – Braucht die Theorie reflexiver Modernisierung eine neue Handlungstheorie?«, in: Margit Wehrich/Fritz Böhle (Hg.), *Handeln unter Unsicherheit*, Wiesbaden: VS Verlag für Sozialwissenschaften, S. 9-21.
- Bornmann, Lutz (2014): »Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics«, in: *Journal of Informetrics* 8 (4), S. 895-903.
- Bornmann, Lutz/Mutz, Rüdiger (2015): »Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references«, in: *Journal of the Association for Information Science and Technology* 66 (11), S. 2215-2222.
- Boumans, Jelle W./Trilling, Damian (2016): »Taking stock of the toolkit«, in: *Digital Journalism* 4 (1), S. 8-23.
- Brechman, Jean M./Lee, Chul-Joo/Cappella, Joseph N. (2009): »Lost in translation? A comparison of cancer-genetics reporting in the press release and its subsequent coverage in the press«, in: *Science Communication* 30 (4), S. 453-474.
- Breiman, Leo (2001): »Random forests«, in: *Machine Learning* 45 (1), S. 5-32.

- Brossard, Dominique (2013): »New media landscapes and the science information consumer«, in: Proceedings of the National Academy of Sciences of the United States of America 110 (Suppl 3), S. 14096-14101.
- Brühl, Rolf (2015): *Wie Wissenschaft Wissen schafft. Wissenschaftstheorie für Sozial- und Wirtschaftswissenschaften*, Konstanz: UVK.
- Bubela, Tania M./Caulfield, Timothy A. (2004): »Do the print media “hype” genetic research? A comparison of newspaper stories and peer-reviewed research papers«, in: CMAJ 170 (9), S. 1399-1407.
- Bucchi, Massimiano/Mazzolini, Renato G. (2003): »Big science, little news: Science coverage in the Italian daily press, 1946-1997«, in: Public Understanding of Science 12 (1), S. 7-24.
- Burggraaff, Christiaan/Trilling, Damian (2017): »Through a different gate: An automated content analysis of how online news and print news differ«, in: Journalism 40 (1), 146488491771669.
- Caple, Helen/Bednarek, Monika (2016): »Rethinking news values: What a discursive approach can tell us about the construction of news discourse and news photography«, in: Journalism 17 (4), S. 435-455.
- Castillo, Carlos/El-Haddad, Mohammed/Pfeffer, Jürgen/Stempeck, Matt (2014): »Characterizing the life cycle of online news stories using social media reactions«, in: Susan Fussell (Hg.), *Compilation publication of CSCW’14 proceedings & CSCW’14 companion*, New York, NY: ACM, S. 211-223.
- Chalmers, Alan F./Bergemann, Niels (2007): *Wege der Wissenschaft. Einführung in die Wissenschaftstheorie*, Berlin, Heidelberg: Springer.
- Chapa, Joaquin/Haq, Zeeshan/Cifu, Adam S. (2017): »Comparative analysis of the factors associated with citation and media coverage of clinical research«, in: Scientometrics 112 (3), S. 1271-1283.
- Chermak, Steven (1998): »Predicting crime story salience: The effects of crime, victim, and defendant characteristics«, in: Journal of Criminal Justice 26 (1), S. 61-70.
- Chermak, Steven/Chapman, Nicole M. (2007): »Predicting crime story salience: A replication«, in: Journal of Criminal Justice 35 (4), S. 351-363.

- Chiozza, Giacomo/Manzetti, Luigi (2015): »Perceptions of competence and the European economic crisis. A micro-level analysis«, in: *Political Research Quarterly* 68 (3), S. 457-473.
- Clark, Fiona/Illman, Deborah L. (2006): »A longitudinal study of the New York Times science times section«, in: *Science Communication* 27 (4), S. 496-513.
- Cobey, Kelly D./Lalu, Manoj M./Skidmore, Becky/Ahmadzai, Nadera/Grudniewicz, Agnes/Moher, David (2018): »What is a predatory journal? A scoping review«, in: *F1000Research* 7.
- Coleman, Renita/McCombs, Maxwell/Shaw, Donald/Weaver, David (2009): »Agenda Setting«, in: Karin Wahl-Jorgensen/Thomas Hanitzsch (Hg.), *The Handbook of Journalism Studies*, New York: Routledge, S. 147-160.
- Dernbach, Beatrice/Godulla, Alexander/Sehl, Annika (Hg.) (2019): *Komplexität im Journalismus*, Wiesbaden: Springer Fachmedien.
- Donsbach, Wolfgang (2004): »Psychology of news decisions«, in: *Journalism* 5 (2), S. 131-157.
- Dumas-Mallet, Estelle/Smith, Andy/Boraud, Thomas/Gonon, François (2017): »Poor replication validity of biomedical association studies reported by newspapers«, in: *PLoS One* 12 (2).
- Dumas-Mallet, Estelle/Tajika, Aran/Smith, Andy/Boraud, Thomas/Furukawa, Toshiaki A./Gonon, François (2019): »Do newspapers preferentially cover biomedical studies involving national scientists?«, in: *Public Understanding of Science* 28 (2), S. 191-200.
- Eilders, Christiane (1996): *Nachrichtenfaktoren und Rezeption. Eine empirische Analyse zur Auswahl und Verarbeitung politischer Information (= Studien zur Kommunikationswissenschaft, Band 20)*, Opladen: Westdeutscher Verlag.
- (2006): »News factors and news decisions. Theoretical and methodological advances in Germany«, in: *Communications* 31 (1), S. 5-24.
- (2016): »Journalismus und Nachrichtenwert«, in: Martin Löffelholz/Liane Rothenberger (Hg.), *Handbuch Journalismustheorien*, Wiesbaden: Springer VS, S. 431-442.
- Einsiedel, Edna F. (1992): »Framing science and technology in the Canadian press«, in: *Public Understanding of Science* 1 (1), S. 89-102.

- Elmer, Christina/Badenschier, Franziska/Wormer, Holger (2008): »Science for everybody? How the coverage of research issues in German newspapers has increased dramatically«, in: *Journalism & Mass Communication Quarterly* 85 (4), S. 878-893.
- Engelmann, Ines (2010): »Journalistische Instrumentalisierung von Nachrichtenfaktoren. Einflüsse journalistischer Einstellungen auf simulierte Issue-, Quellen- und Statement-Entscheidungen«, in: *Medien & Kommunikationswissenschaft* 58 (4), S. 525-543.
- (2011): »Journalistische Selektionskriterien und Selektionsentscheidungen. Methodische Implikationen für die Anwendung einer Wert-Erwartungstheorie«, in: Olaf Jandura/Thorsten Quandt/Jens Vogelgesang (Hg.), *Methoden der Journalismusforschung*, Wiesbaden: VS Verlag für Sozialwissenschaften, S. 141-154.
- (2012): »Nachrichtenfaktoren und die organisationsspezifische Nachrichtenselektion. Eine Erweiterung der Nachrichtenwerttheorie um die Meso-Ebene journalistischer Organisationen«, in: *Medien & Kommunikationswissenschaft* 60 (1), S. 41-63.
- (2016a): Gatekeeping (= Konzepte. Ansätze der Medien- und Kommunikationswissenschaft, Band 16), Baden-Baden: Nomos.
- (2016b): »Journalismus und alltagsrationale Nachrichtenauswahl«, in: Martin Löffelholz/Liane Rothenberger (Hg.), *Handbuch Journalismustheorien*, Wiesbaden: Springer VS, S. 457-471.
- Entwistle, Vikki (1995): »Reporting research in medical journals and newspapers«, in: *British Medical Journal* 310 (6984), S. 920-923.
- Esser, Hartmut (2007): »Der Handlungsbegriff in der modernen Soziologie«, in: Klaus-Dieter Altmeyden/Thomas Hanitzsch/Carsten Schlüter (Hg.), *Journalismustheorie: Next Generation. Soziologische Grundlegung und theoretische Innovation*, Wiesbaden: VS Verlag für Sozialwissenschaften, S. 27-46.
- Evans, William (1995): »The mundane and the arcane: Prestige media coverage of social and natural science«, in: *Journalism & Mass Communication Quarterly* 72 (1), S. 168-177.
- Fanelli, Daniele (2009): »How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data«, in: *PloS One* 4 (5), e5738.
- (2010): »Do pressures to publish increase scientists' bias? An empirical support from US States data«, in: *PloS One* 5 (4), e10271.

- Fengler, Susanne/Ruß-Mohl, Stephan (2008): »Journalists and the information-attention markets«, in: *Journalism* 9 (6), S. 667-690.
- Fenner, Martin (2013): »What can article-level metrics do for you?«, in: *PLoS Biology* 11 (10), e1001687.
- Fernandes, Kelwin/Vinagre, Pedro/Cortez, Paulo (2015): »A proactive intelligent decision support system for predicting the popularity of online news«, in: Francisco Pereira/Penousal Machado/Ernesto Costa et al. (Hg.), *Progress in artificial intelligence. 17th Portuguese Conference on Artificial Intelligence, EPIA 2015*, Cham: Springer, S. 535-546.
- Fishbein, Martin/Ajzen, Icek (2010): *Predicting and changing behavior. The reasoned action approach*, New York: Psychology Press.
- Franzen, Martina (2015): »Der Impact Faktor war gestern. Altmetrics und die Zukunft der Wissenschaft«, in: *Soziale Welt* 66 (2), S. 225-242.
- Frerichs, Stefan (2016): »Journalismus als konstruktives Chaos«, in: Martin Löffelholz/Liane Rothenberger (Hg.), *Handbuch Journalismustheorien*, Wiesbaden: Springer VS, S. 191-200.
- Fretwurst, Benjamin (2008): *Nachrichten im Interesse der Zuschauer. Eine konzeptionelle und empirische Neubestimmung der Nachrichtenwerttheorie*, Konstanz: UVK.
- Friedrichsmeier, Andres/Fürst, Silke (2012): »Neue Governance als Wettbewerb um Sichtbarkeit. Zur veränderten Dynamik der Öffentlichkeits- und Medienorientierung von Hochschulen«, in: *die hochschule* 21 (2), S. 46-64.
- Friedrichsmeier, Andres/Geils, Matthias/Kohring, Matthias/Laukötter, Esther/Marcinkowski, Frank (2013): *Organisation und Öffentlichkeit von Hochschulen. Forschungsreport 1/2013 des Arbeitsbereichs Kommunikation – Medien – Gesellschaft*, Münster.
- Friedrichsmeier, Andres/Laukötter, Esther/Marcinkowski, Frank (2015): »Hochschul-PR als Restgröße. Wie Hochschulen in die Medien kommen und was ihre Pressestellen dazu beitragen«, in: Mike S. Schäfer (Hg.), *Wissenschaftskommunikation im Wandel*, Köln: von Halem, S. 128-151.

- Fröhlich, Romy/Rüdiger, Burkhard (2006): »Framing political public relations: measuring success of political communication strategies in Germany«, in: *Public Relations Review* 32 (1), S. 18-25.
- Galtung, Johan/Ruge, Mari H. (1965): »The structure of foreign news: The presentation of the Congo, Cuba and Cyprus crises in four Norwegian newspapers«, in: *Journal of Peace Research* 2 (1), S. 64-90.
- Gonon, Francois/Bezard, Erwan/Boraud, Thomas (2011): »Misrepresentation of neuroscience data might give rise to misleading conclusions in the media: the case of attention deficit hyperactivity disorder«, in: *PloS One* 6 (1), e14618.
- Gonon, François/Konsman, Jan-Pieter/Cohen, David/Boraud, Thomas (2012): »Why most biomedical findings echoed by newspapers turn out to be false: The case of attention deficit hyperactivity disorder«, in: *PloS One* 7 (9), e44275.
- Granado, António (2011): »Slaves to journals, serfs to the web: The use of the internet in newsgathering among European science journalists«, in: *Journalism* 12 (7), S. 794-813.
- Grande, Edgar/Jansen, Dorothea/Jarren, Otfried/Schimank, Uwe/Weingart, Peter (2013): »Die neue Governance der Wissenschaft. Zur Einleitung«, in: Edgar Grande/Dorothea Jansen/Otfried Jarren et al. (Hg.), *Neue Governance der Wissenschaft. Reorganisation - externe Anforderungen - Medialisierung*, Bielefeld: transcript Verlag, S. 15-45.
- Grieneisen, Michael L./Zhang, Minghua (2012): »A comprehensive survey of retracted articles from the scholarly literature«, in: *PloS One* 7 (10), e44118.
- Grimes, David R./Bauch, Chris T./Ioannidis, John P. A. (2018): »Modelling science trustworthiness under publish or perish pressure«, in: *Royal Society Open Science* 5 (1), S. 171511.
- Grimmer, Justin/Stewart, Brandon M. (2013): »Text as data: The promise and pitfalls of automatic content analysis methods for political texts«, in: *Political Analysis* 21 (3), S. 267-297.
- Guenther, Lars/Froehlich, Klara/Ruhrmann, Georg (2015): »(Un)certainty in the news: Journalists' decisions on communicating the scientific evidence of nanotechnology«, in: *Journalism & Mass Communication Quarterly* 92 (1), S. 199-220.

- Guenther, Lars/Ruhrmann, Georg (2013): »Science journalists' selection criteria and depiction of nanotechnology in German media«, in: *Journal of Science Communication* 12 (03).
- (2016): »Scientific evidence and mass media: Investigating the journalistic intention to represent scientific uncertainty«, in: *Public Understanding of Science* 25 (8), S. 927-943.
- Guerini, Marco/Pepe, Alberto/Lepri, Bruno (2012): »Do linguistic style and readability of scientific abstracts affect their virality?«, in: *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*, Palo Alto, Calif.: AAAI Press.
- Günther, Elisabeth/Quandt, Thorsten (2016): »Word counts and topic models: Automated text analysis methods for digital journalism research«, in: *Digital Journalism* 4 (1), S. 75-88.
- Guyatt, Gordon H./Oxman, Andrew D./Vist, Gunn E./Kunz, Regina/Falck-Ytter, Yngve/Alonso-Coello, Pablo/Schünemann, Holger J. (2008): »GRADE: an emerging consensus on rating quality of evidence and strength of recommendations«, in: *British Medical Journal* 336 (7650), S. 924-926.
- Häder, Michael (2015): »Wissenschaftstheorie«, in: Michael Häder (Hg.), *Empirische Sozialforschung. Eine Einführung*, Wiesbaden: Springer Fachmedien Wiesbaden, S. 17-68.
- Hammarfelt, Björn (2014): »Using altmetrics for assessing research impact in the humanities«, in: *Scientometrics* 101 (2), S. 1419-1430.
- Hanitzsch, Thomas (2009): »Zur Wahrnehmung von Einflüssen im Journalismus. Komparative Befunde aus 17 Ländern«, in: *Medien & Kommunikationswissenschaft* 57 (2), S. 153-173.
- Hansen, Anders (1994): »Journalistic practices and science reporting in the British press«, in: *Public Understanding of Science* 3 (2), S. 111-134.
- Harcup, Tony/O'Neill, Deirdre (2001): »What is news? Galtung and Ruge revisited«, in: *Journalism Studies* 2 (2), S. 261-280.
- (2017): »What is news? News values revisited (again)«, in: *Journalism Studies* 18 (12), S. 1470-1488.

- Haustein, Stefanie/Peters, Isabella/Bar-Ilan, Judit/Priem, Jason/Shema, Hadas/Terliesner, Jens (2014): »Coverage and adoption of altmetrics sources in the bibliometric community«, in: *Scientometrics* 101 (2), S. 1145-1163.
- Heinrichs, Harald/Petersen, Imme (2009): »Mediatisierte Politikgestaltung? Medien, Expertise und politische Entscheidungsprozesse in wissenschaftsbasierten Themenfeldern«, in: Hans P. Peters (Hg.), *Medienorientierung biomedizinischer Forscher im internationalen Vergleich. Die Schnittstelle von Wissenschaft & Journalismus und ihre politische Relevanz*, Jülich: Forschungszentrum Jülich, S. 227-302.
- Helfer, Luzia/van Aelst, Peter (2016): »What makes party messages fit for reporting? An experimental study of journalistic news selection«, in: *Political Communication* 33 (1), S. 59-77.
- Hensinger, Elena/Flaounas, Ilias/Cristianini, Nello (2013): »Modelling and predicting news popularity«, in: *Pattern Analysis and Applications* 16 (4), S. 623-635.
- Hesselmann, Felicitas/Graf, Verena/Schmidt, Marion/Reinhart, Martin (2017): »The visibility of scientific misconduct: A review of the literature on retracted journal articles«, in: *Current Sociology Review* 65 (6), S. 814-845.
- Hijmans, Ellen/Pleijter, Alexander/Wester, Fred (2003): »Covering Scientific Research in Dutch Newspapers«, in: *Science communication* 25 (2), S. 153-176.
- Hodgetts, Darrin/Chamberlain, Kerry/Scammell, Margaret/Karapu, Rolinda/Waimarie Nikora, Linda (2008): »Constructing health news: Possibilities for a civic-oriented journalism«, in: *Health* 12 (1), S. 43-66.
- Hofer, Harald (2009): »Handlung und Legitimation im Zuge reflexiver Modernisierung«, in: Margit Wehrich/Fritz Böhle (Hg.), *Handeln unter Unsicherheit*, Wiesbaden: VS Verlag für Sozialwissenschaften, S. 139-147.
- Hothorn, Torsten/Hornik, Kurt/Zeileis, Achim (2006): »Unbiased recursive partitioning: A conditional inference framework«, in: *Journal of Computational and Graphical Statistics* 15 (3), S. 651-674.
- Ioannidis, John P. A. (2005): »Why most published research findings are false«, in: *PLoS Medicine* 2 (8), e124.

- Kamel, Samah O. M./Nour, Mohamed (2018): »Feature selection methods for predicting the popularity of online news: comparative study, and a proposed method.«, in: *Journal of Theoretical and Applied Information Technology* 96 (19), S. 6969-6980.
- Katzenbach, Christian/Pentzold, Christian (2017): »Theoriearbeit in der Kommunikationswissenschaft zwischen Komplexitätssteigerung und Komplexitätsreduzierung«, in: *Medien & Kommunikationswissenschaft* 65 (3), S. 483-499.
- Keneshloo, Yaser/Wang, Shuguang/Han, Eui-Hong/Ramakrishnan, Naren (2016a): »Predicting the popularity of news articles«, in: Sanjay C. Venkatasubramanian/Wagner Meira (Hg.), *Proceedings of the 2016 SIAM International Conference on Data Mining*, Philadelphia, PA: Society for Industrial and Applied Mathematics, S. 441-449.
- Keneshloo, Yaser/Wang, Shuguang/Han, Eui-Hong S./Ramakrishnan, Naren (2016b): »Predicting the shape and peak time of news article views«, in: James Joshi (Hg.), *2016 IEEE International Conference on Big Data*, Piscataway, NJ: IEEE, S. 2400-2409.
- Kepplinger, Hans M. (1998): »Der Nachrichtenwert der Nachrichtenfaktoren«, in: Christina Holtz-Bacha/Helmut Scherer/Norbert Waldmann (Hg.), *Wie die Medien die Welt erschaffen und wie die Menschen darin leben. Für Winfried Schulz*, Wiesbaden: VS Verlag für Sozialwissenschaften, S. 19-38.
- Kepplinger, Hans M./Bastian, Rouwen (2000): »Der prognostische Gehalt der Nachrichtenwert-Theorie«, in: *Publizistik* 45 (4), S. 462-475.
- Kepplinger, Hans M./Ehmig, Simone C. (2006): »Predicting news decisions. An empirical test of the two-component theory of news selection«, in: *Communications* 31 (1), S. 25-43.
- Kern, Christoph/Klausch, Thomas/Kreuter, Frauke (2019): »Tree-based machine learning methods for survey research«, in: *Survey Research Methods* 13 (1), S. 73-93.
- Khan, Aasim/Worah, Gautam/Kothari, Mehul/Jadhav, Yogesh H./Nimkar, Anant V. (2018): »News popularity prediction with ensemble methods of classification«, in: *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Piscataway, NJ: IEEE, S. 1-6.
- Kiernan, Vincent (2003): »Embargoes and science news«, in: *Journalism & Mass Communication Quarterly* 80 (4), S. 903-920.

- (2016): »Source diversity among journals cited in Science Times«, in: Public Understanding of Science 25 (2), S. 198-206.
- Kim, Joon H./Mantrach, Amin/Jaimes, Alejandro/Oh, Alice (2016): »How to compete online for news audience«, in: Balaji Krishnapuram/Mohak Shah/Alex Smola et al. (Hg.), KDD2016. 22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining, New York, NY: Association for Computing Machinery Inc. (ACM), S. 1645-1654.
- Kong, Junli/Wang, Baocang/Liu, Caiyun/Wu, Gaofei (2018): »An approach for predicting the popularity of online security news articles«, in: 2018 IEEE Conference on Communications and Network Security (CNS), Piscataway, NJ: IEEE, S. 1-6.
- Le, Quoc/Mikolov, Tomas (2014): »Distributed representations of sentences and documents«, in: 31st International Conference on Machine Learning, ICML 2014, S. 2931-2939.
- Lee, Angela M./Lewis, Seth C./Powers, Matthew (2014): »Audience clicks and news placement: A study of time-lagged influence in online journalism«, in: Communication Research 41 (4), S. 505-530.
- Lee, Jong G./Moon, Sue/Salamatian, Kave (2010): »An approach to model and predict the popularity of online contents with explanatory factors«, in: Xiangji J. Huang/Orland Hoerber (Hg.), IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010, Piscataway, NJ: IEEE, S. 623-630.
- Lee, Jong G./Moon, Sue/Salamatian, Kavé (2012): »Modeling and predicting the popularity of online contents with Cox proportional hazard regression model«, in: Neurocomputing 76 (1), S. 134-145.
- Lehmkuhl, Markus (2008): »Typologie des Wissenschaftsjournalismus«, in: Holger Hettwer/Markus Lehmkuhl/Holger Wormer et al. (Hg.), WissensWelten. Wissenschaftsjournalismus in Theorie und Praxis, Gütersloh: Verlag Bertelsmann Stiftung, S. 301-321.
- (2018): „Power-Law“ Verteilungen der Selektivität im Nachrichtenjournalismus über naturwissenschaftliche Studien., Jahrestagung der Deutschen Gesellschaft für Publizistik und Kommunikationswissenschaft, Offene Panels. Mannheim.
- (2019): »Komplexität der Wissenschaft als Herausforderung für den Wissenschaftsjournalismus«, in: Beatrice Dernbach/Alexander Godulla/Annika Sehl (Hg.), Komplexität im Journalismus, Wiesbaden: Springer Fachmedien, S. 203-212.

- Lehmkuhl, Markus/Promies, Nikolai (2019): „Power-Law“-Verteilungen der Selektivität im Nachrichtenjournalismus über naturwissenschaftliche Studien. Unveröffentlichtes Manuskript.
- Lerman, Kristina/Hogg, Tad (2010): »Using a model of social dynamics to predict popularity of news«, in: Michael Rappa (Hg.), Proceedings of the 19th International Conference on World Wide Web. WWW'10, New York, NY: ACM Digital Library; ACM, S. 621.
- Liu, Caiyun/Wang, Wenjie/Zhang, Yuqing/Dong, Ying/He, Fannv/Wu, Chensi (2017): »Predicting the popularity of online news based on multivariate analysis«, in: 2017 IEEE International Conference on Computer and Information Technology (CIT): IEEE, S. 9-15.
- Loh, Wei-Yin (2014): »Fifty years of classification and regression trees«, in: International Statistical Review 82 (3), S. 329-348.
- Long, Fei/Xu, Meixia/Li, Yulei/Wu, Zhihua/Ling, Qiang (2018): »XiaoA: a robot editor for popularity prediction of online news based on ensemble learning«, in: Zhongzhi Shi/Cyriel Pennartz/Tiejun Huang (Hg.), Intelligence Science II. Third IFIP TC 12 International Conference, ICIS 2018, Cham: Springer International Publishing, S. 340-350.
- Lundman, Richard J. (2003): »The newsworthiness and selection bias in news about murder: Comparative and relative effects of novelty and race and gender typifications on newspaper coverage of homicide«, in: Sociological Forum 18 (3), S. 357-386.
- MacLaughlin, Ansel/Wihbey, John/Smith, David A. (2018): »Predicting news coverage of scientific articles«, in: Proceedings of the Twelfth International Conference on Web and Social Media, ICWSM 2018: AAAI Press.
- Maier, Daniel/Waldherr, A./Miltner, P./Wiedemann, G./Niekler, A./Keinert, A./Pfetsch, B./Heyer, G./Reber, U./Häussler, T./Schmid-Petri, H./Adam, S. (2018a): »Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology«, in: Communication Methods and Measures 12 (2-3), S. 93-118.
- Maier, Michaela/Glogger, Isabella/Retzbach, Joachim/Stengel, Karin (2018b): Nachrichtenwerttheorie, Baden-Baden: Nomos.
- Manning, Christopher D./Raghavan, Prabhakar/Schütze, Hinrich (2009): Introduction to information retrieval, Cambridge: Cambridge Univ. Press.

- Meho, Lokman I./Yang, Kiduk (2007): »Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar«, in: *Journal of the American Society for Information Science and Technology* 58 (13), S. 2105-2125.
- Mellor, Felicity (2015): »Non-news values in science journalism«, in: Brian Rappert/Brian Balmer (Hg.), *Absence in Science, Security and Policy. From Research Agendas to Global Strategy*, London: Palgrave Macmillan, S. 93-113.
- Mikolov, Tomas/Chen, Kai/Corrado, Greg/Dean, Jeffrey (2013a): Efficient estimation of word representations in vector space, <http://arxiv.org/pdf/1301.3781v3>.
- Mikolov, Tomas/Sutskever, Ilya/Chen, Kai/Corrado, Greg/Dean, Jeffrey (2013b): »Distributed representations of words and phrases and their compositionality«, in: *Advances in Neural Information Processing Systems*.
- Mohammadi, Ehsan/Thelwall, Mike (2014): »Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows«, in: *Journal of the Association for Information Science and Technology* 65 (8), S. 1627-1638.
- Moher, David/Shamseer, Larissa/Cobey, Kelly D./Lalu, Manoj M./Galipeau, James/Avey, Marc T./Ahmadzai, Nadera/Alabousi, Mostafa/Barbeau, Pauline/Beck, Andrew/Daniel, Raymond/Frank, Robert/Ghannad, Mona/Hamel, Candyce/Hersi, Mona/Hutton, Brian/Isupov, Inga/McGrath, Trevor A./McInnes, Matthew D. F./Page, Matthew J./Pratt, Misty/Pussegoda, Kusala/Shea, Beverley/Srivastava, Anubhav/Stevens, Adrienne/Thavorn, Kednapa/van Katwyk, Sasha/Ward, Roxanne/Wolfe, Dianna/Yazdi, Fatemeh/Yu, Ashley M./Ziai, Hedyeh (2017): »Stop this waste of people, animals and money«, in: *Nature* 549 (7670), S. 23-25.
- Molek-Kozakowska, Katarzyna (2017): »Communicating environmental science beyond academia: Stylistic patterns of newsworthiness in popular science journalism«, in: *Discourse & Communication* 11 (1), S. 69-88.
- Moniz, Nuno/Torgo, Luis/Eirinaki, Magdalini (2016): »Time-based ensembles for prediction of rare events in news stream«, in: Francesco Bonchi/Josep Domingo-Ferrer/Ricardo Baeza-Yates et al. (Hg.), *16th IEEE International Conference on Data Mining Workshops*, Piscataway, NJ: IEEE, S. 1066-1073.
- Moniz, Nuno/Torgo, Luís/Rodrigues, Fátima (2014): »Resampling approaches to improve news importance prediction«, in: Hendrik Blockeel (Hg.), *Advances in intelligent data*

- analysis XIII. 13th International Symposium, IDA 2014, Leuven, Belgium, October 30-November 1, 2014, Cham: Springer, S. 215-226.
- Moriarty, Cortney M./Jensen, Jakob D./Stryker, Jo E. (2010): »Frequently cited sources in cancer news coverage: a content analysis examining the relationship between cancer news content and source citation«, in: *Cancer causes & control* 21 (1), S. 41-49.
- O'Neill, Deirdre/Harcup, Tony (2009): »News values and selectivity«, in: Karin Wahl-Jorgensen/Thomas Hanitzsch (Hg.), *The Handbook of Journalism Studies*, New York: Routledge, S. 161-174.
- Östgaard, Einar (1965): »Factors influencing the flow of news«, in: *Journal of Peace Research* 2 (1), S. 39-63.
- Ouyang, Shuxin/Li, Chenyu/Li, Xueming (2016): »A peek into the future: predicting the popularity of online videos«, in: *IEEE Access* 4, S. 3026-3033.
- Page, Scott E. (2015): »What sociologists should know about complexity«, in: *Annual Review of Sociology* 41, S. 21-41.
- Pahl, Carola (1998): »Die Bedeutung von Wissenschaftsjournalen für die Themenauswahl in den Wissenschaftsressorts deutscher Zeitungen am Beispiel medizinischer Themen«, in: *Rundfunk und Fernsehen* 46 (2-3), S. 243-253.
- Patterson, Thomas E./Donsbach, Wolfgang (1996): »News decisions: Journalists as partisan actors«, in: *Political Communication* 13 (4), S. 455-468.
- Post, Senja/Maier, Michaela (2016): »Stakeholders' rationales for representing uncertainties of biotechnological research«, in: *Public Understanding of Science* 25 (8), S. 944-960.
- Pürer, Heinz (2014): *Publizistik- und Kommunikationswissenschaft*, Konstanz: UVK.
- Raupp, Juliana (2008): »Der Einfluss von Wissenschafts-PR auf den Wissenschaftsjournalismus«, in: Holger Hettwer/Markus Lehmkuhl/Holger Wormer et al. (Hg.), *WissensWelten. Wissenschaftsjournalismus in Theorie und Praxis*, Gütersloh: Verlag Bertelsmann Stiftung, S. 379-392.
- Reinemann, Carsten (2007): »Subjektiv rationale Akteure: Das Potenzial handlungstheoretischer Erklärungen für die Journalismusforschung«, in: Klaus-Dieter Altmeppen/Thomas Hanitzsch/Carsten Schlüter (Hg.), *Journalismustheorie: Next*

- Generation. Soziologische Grundlegung und theoretische Innovation, Wiesbaden: VS Verlag für Sozialwissenschaften, S. 47-67.
- Reinemann, Carsten/Baugut, Philip (2016): »Journalismus als subjektiv rationales Handeln im sozialen Kontext«, in: Martin Löffelholz/Liane Rothenberger (Hg.), Handbuch Journalismustheorien, Wiesbaden: Springer VS, S. 309-321.
- Rezaeenour, Jalal/Eili, Mansoureh Y./Hadavandi, Esmaeil/Roozbahani, Mohammad H. (2018): »Developing a new hybrid intelligent approach for prediction online news popularity«, in: International Journal of Information Science and Management 16 (1), S. 71-87.
- Riesmeyer, Claudia (2007): Wie unabhängig ist Journalismus? Zur Konkretisierung der Determinationsthese, Konstanz: UVK.
- Rosen, Cecilia/Guenther, Lars/Froehlich, Klara (2016): »The question of newsworthiness: A cross-comparison among science journalists' selection criteria in Argentina, France, and Germany«, in: Science Communication 38 (3), S. 328-355.
- Ruhrmann, Georg/Woelke, Jens/Maier, Michaela et al. (Hg.) (2003): Der Wert von Nachrichten im deutschen Fernsehen, Wiesbaden: Springer Fachmedien.
- Russell Neuman, W./Guggenheim, Lauren/Mo Jang, S./Bae, Soo Y. (2014): »The dynamics of public attention: agenda-setting theory meets big data«, in: Journal of Communication 64 (2), S. 193-214.
- Sande, Øystein (1971): »The Perception of Foreign News«, in: Journal of Peace Research 8, S. 221-237.
- Schäfer, Mike S. (2011): »Sources, characteristics and effects of mass media communication on science: A review of the literature, current trends and areas for future research«, in: Sociology Compass 5 (6), S. 399-412.
- Schäfer, Mike S./Ivanova, Ana/Schmidt, Andreas (2014): »What drives media attention for climate change? Explaining issue attention in Australian, German and Indian print media from 1996 to 2010«, in: International Communication Gazette 76 (2), S. 152-176.
- Schimank, Uwe (2005): Die Entscheidungsgesellschaft. Komplexität und Rationalität der Moderne (= Hagener Studententexte zur Soziologie), Wiesbaden: VS Verlag für Sozialwissenschaften.

- (2009): »Die „reflexive Moderne“: eine wohlbekannte Entscheidungsgesellschaft«, in: Margit Wehrich/Fritz Böhle (Hg.), *Handeln unter Unsicherheit*, Wiesbaden: VS Verlag für Sozialwissenschaften, S. 77-93.
- (2010): »Reputation statt Wahrheit: Verdrängt der Nebencode den Code?«, in: *Soziale Systeme* 16 (2), S. 57.
- Scholl, Armin/Loosen, Wiebke (2019): »Die Komplexität der Komplexitätsforschung«, in: Beatrice Dernbach/Alexander Godulla/Annika Sehl (Hg.), *Komplexität im Journalismus*, Wiesbaden: Springer Fachmedien, S. 15-30.
- Schulz, Winfried (1976): *Die Konstruktion von Realität in den Nachrichtenmedien. Analyse der aktuellen Berichterstattung (= Alber-Broschur Kommunikation, Band 4)*, Freiburg: Alber.
- Schwartz, Lisa M./Woloshin, Steven/Andrews, Alice/Stukel, Therese A. (2012): »Influence of medical journal press releases on the quality of associated newspaper coverage: retrospective cohort study«, in: *British Medical Journal* 344, d8164.
- Schwarz, Andreas (2006): »The theory of newsworthiness applied to Mexico's press. How the news factors influence foreign news coverage in a transitional country«, in: *Communications* 31 (1), S. 135.
- Segal, Mark/Xiao, Yuanyuan (2011): »Multivariate random forests«, in: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1 (1), S. 80-87.
- Semir, Vladimir de (1996): »What is newsworthy?«, in: *The Lancet* 347 (9009), S. 1163-1166.
- Semir, Vladimir de/Ribas, Christina/Revuelta, Gemma (1998): »Press releases of science journal articles and subsequent newspaper stories on the same topic«, in: *JAMA* 280 (3), S. 294.
- Serong, Julia/Koppers, Lars/Luschmann, Edith/Molina Ramirez, Alejandro/Kersting, Kristian/Rahnenführer, Jörg/Wormer, Holger (2017): »Öffentlichkeitsorientierung von Wissenschaftsinstitutionen und Wissenschaftsdisziplinen«, in: *Publizistik* 62 (2), S. 153-178.
- Shamseer, Larissa/Moher, David/Maduekwe, Onyi/Turner, Lucy/Barbour, Virginia/Burch, Rebecca/Clark, Jocalyn/Galipeau, James/Roberts, Jason/Shea, Beverley J. (2017):

- »Potential predatory and legitimate biomedical journals: can you tell the difference? A cross-sectional comparison«, in: *BMC Medicine* 15 (1), S. 28.
- Shen, Cenyu/Björk, Bo-Christer (2015): »'Predatory' open access: a longitudinal study of article volumes and market characteristics«, in: *BMC Medicine* 13, S. 230.
- Shoemaker, Pamela J. (2006): »News and newsworthiness: A commentary«, in: *Communications* 31 (1), S. 105-111.
- Shoemaker, Pamela J./Cohen, Akiba A. (2006): *News around the world. Content, practitioners, and the public*, New York, NY: Routledge.
- Shoemaker, Pamela J./Reese, Stephen D. (1996): *Mediating the message. Theories of influences on mass media content*, White Plains N.Y.: Longman.
- Shoemaker, Pamela J./Vos, Tim P./Reese, Stephen D. (2009): »Journalists as gatekeepers«, in: Karin Wahl-Jorgensen/Thomas Hanitzsch (Hg.), *The Handbook of Journalism Studies*, New York: Routledge, S. 73-87.
- Shulman, Benjamin/Sharma, Amit/Cosley, Dan (2016): »Predictability of popularity: gaps between prediction and understanding«, in: *Proceedings of the Tenth International AAAI Conference on Web and Social Media*, Palo Alto, California: AAAI Press, S. 348-357.
- Siravuri, Harish V./Alhoori, Hamed (2017): »What makes a research article newsworthy?«, in: *Proceedings of the Association for Information Science and Technology* 54 (1), S. 802-803.
- Smith, Richard (2006): »Peer review: a flawed process at the heart of science and journals«, in: *Journal of the Royal Society of Medicine* 99 (4), S. 178-182.
- Staab, Joachim F. (1990a): *Nachrichtenwert-Theorie. Formale Struktur und empirischer Gehalt (= Alber-Broschur Kommunikation, Band 17)*, Freiburg: Alber.
- (1990b): »The role of news factors in news selection: a theoretical reconsideration«, in: *European Journal of Communication* 5 (4), S. 423-443.
- Steen, R. G./Casadevall, Arturo/Fang, Ferric C. (2013): »Why has the number of scientific retractions increased?«, in: *PloS One* 8 (7), e68397.
- Strobl, Carolin/Malley, James/Tutz, Gerhard (2009): »An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests«, in: *Psychological methods* 14 (4), S. 323-348.

- Strömbäck, Jesper/Karlsson, Michael/Hopmann, David N. (2012): »Determinants of news content: Comparing journalists' perceptions of the normative and actual impact of different event properties when deciding what's news«, in: *Journalism Studies* 13 (5-6), S. 718-728.
- Stryker, Jo E. (2002): »Reporting medical information: Effects of press releases and newsworthiness on medical journal articles' visibility in the news media«, in: *Preventive Medicine* 35 (5), S. 519-530.
- Sud, Pardeep/Thelwall, Mike (2014): »Evaluating altmetrics«, in: *Scientometrics* 98 (2), S. 1131-1143.
- Suleski, Julie/Ibaraki, Motomu (2010): »Scientists are talking, but mostly to each other: a quantitative analysis of research represented in mass media«, in: *Public Understanding of Science* 19 (1), S. 115-125.
- Summ, Annika/Volpers, Anna M. (2016): »What's science? Where's science? Science journalism in German print media«, in: *Public Understanding of Science* 25 (7), S. 775-790.
- Sumner, Petroc/Vivian-Griffiths, Solveiga/Boivin, Jacky/Williams, Andrew/Bott, Lewis/Adams, Rachel/Venetis, Christos A./Whelan, Leanne/Hughes, Bethan/Chambers, Christopher D. (2016): »Exaggerations and caveats in press releases and health-related science news«, in: *PloS One* 11 (12), e0168217.
- Sumner, Petroc/Vivian-Griffiths, Solveiga/Boivin, Jacky/Williams, Andy/Venetis, Christos A./Davies, Aimée/Ogden, Jack/Whelan, Leanne/Hughes, Bethan/Dalton, Bethan/Boy, Fred/Chambers, Christopher D. (2014): »The association between exaggeration in health related science news and academic press releases: retrospective observational study«, in: *British Medical Journal* 349, g7015.
- Szabo, Gabor/Huberman, Bernardo A. (2010): »Predicting the popularity of online content«, in: *Communications of the ACM* 53 (8), S. 80.
- Tatar, A./Antoniadis, P./Amorim, M. d. de/Fdida, S. (2012): »Ranking news articles based on popularity prediction«, in: *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2012, Piscataway, NJ: IEEE, S. 106-110.

- Tatar, Alexandru/Amorim, Marcelo D. de/Fdida, Serge/Antoniadis, Panayotis (2014a): »A survey on predicting the popularity of web content«, in: *Journal of Internet Services and Applications* 5 (1).
- Tatar, Alexandru/Antoniadis, Panayotis/Amorim, Marcelo D. de/Fdida, Serge (2014b): »From popularity prediction to ranking online news«, in: *Social Network Analysis and Mining* 4 (1), S. 249.
- Tatar, Alexandru/Leguay, Jérémie/Antoniadis, Panayotis/Limbourg, Arnaud/Amorim, Marcelo D. de/Fdida, Serge (2011): »Predicting the popularity of online articles based on user comments«, in: Rajendra Akerkar (Hg.), *Proceedings of the International Conference on Web Intelligence, Mining and Semantics, WIMS'11*, New York: ACM, S. 1.
- Thelwall, Mike/Haustein, Stefanie/Larivière, Vincent/Sugimoto, Cassidy R. (2013): »Do altmetrics work? Twitter and ten other social web services«, in: *PloS One* 8 (5), e64841.
- Trilling, Damian/Tolochko, Petro/Burscher, Björn (2017): »From newsworthiness to shareworthiness: How to predict news sharing based on article characteristics.«, in: *Journalism & Mass Communication Quarterly* 94 (1), S. 38-60.
- Tsagkias, Manos/Weerkamp, Wouter/Rijke, Maarten de (2009): »Predicting the volume of comments on online news stories«, in: David Cheung (Hg.), *Proceedings of the 18th ACM conference on Information and knowledge management*, New York, NY: ACM, S. 1765.
- Tsfati, Yariv/Cohen, Jonathan/Gunther, Albert C. (2011): »The influence of presumed media influence on news about science and scientists«, in: *Science Communication* 33 (2), S. 143-166.
- Uhlemann, Ingrid A. (2012): *Der Nachrichtenwert im situativen Kontext. Eine Studie zur Auswahlwahrscheinlichkeit von Nachrichten*, Wiesbaden: VS Verlag für Sozialwissenschaften.
- Urry, John (2005): »The complexity turn«, in: *Theory, Culture and Society* 22 (5), S. 1-14.
- van Atteveldt, Wouter/Peng, Tai-Quan (2018): »When communication meets computation: Opportunities, challenges, and pitfalls in computational communication science«, in: *Communication Methods and Measures* 12 (2-3), S. 81-92.

- van Canneyt, Steven/Leroux, Philip/Dhoedt, Bart/Demeester, Thomas (2018): »Modeling and predicting the popularity of online news based on temporal and content-related features«, in: *Multimedia Tools and Applications* 77 (1), S. 1409-1436.
- van Trigt, Anke M./Jong-Van Den Berg, Lolkje T.W. de/Haijjer-Ruskamp, Flora M./Willems, Jaap/Tromp, Theo F.J. (1994): »Journalists and their sources of ideas and information on medicines«, in: *Social Science & Medicine* 38 (4), S. 637-643.
- van Trigt, Anke M./Jong-Van Den Berg, Lolkje T.W. de/Voogt, Linda M./Willems, Jaap/T.Dirk Tromp, F. J./Haijjer-Ruskamp, Flora M. (1995): »Setting the agenda: does the medical literature set the agenda for articles about medicines in the newspapers?«, in: *Social Science & Medicine* 41 (6), S. 893-899.
- Vu, Hong T. (2014): »The online audience as gatekeeper: The influence of reader metrics on news editorial selection«, in: *Journalism* 15 (8), S. 1094-1110.
- Waldherr, Annie (2012): *Die Dynamik der Medienaufmerksamkeit. Ein Simulationsmodell (= Schriftenreihe politische Kommunikation und demokratische Öffentlichkeit, Band 4)*, Baden-Baden: Nomos.
- (2014): »Emergence of news waves: a social simulation approach«, in: *Journal of Communication* 64 (5), S. 852-873.
- Wallace, Byron C./Paul, Michael J./Elhadad, Noémie (2015): *What predicts media coverage of health science articles?*, AAAI Workshop on the World Wide Web and Public Health Intelligence, Austin, Texas.
- Wang, Michael T. M./Bolland, Mark J./Gamble, Greg/Grey, Andrew (2015): »Media coverage, journal press releases and editorials associated with randomized and observational studies in high-impact medical journals: a cohort study«, in: *PloS One* 10 (12), e0145294.
- Weitkamp, Emma/Eidsvaag, Torill (2014): »Agenda building in media coverage of food research. Superfoods coverage in UK national newspapers«, in: *Journalism Practice* 8 (6), S. 871-886.
- Welbers, Kasper/van Atteveldt, Wouter/Kleinnijenhuis, Jan/Ruigrok, Nel/Schaper, Joep (2016): »News selection criteria in the digital age: Professional norms versus online audience metrics«, in: *Journalism* 17 (8), S. 1037-1053.

- Wendelin, Manuel/Engelmann, Ines/Neubarth, Julia (2017): »User rankings and journalistic news selection«, in: *Journalism Studies* 18 (2), S. 135-153.
- White, David M. (1950): »The “Gate Keeper”: a case study in the selection of news«, in: *Journalism Quarterly* 27 (4), S. 383-390.
- Wilkie, Tom (1996): »Sources in science: who can we trust?«, in: *The Lancet* 347 (9011), S. 1308-1311.
- Woloshin, Steven/Schwartz, Lisa M. (2002): »Press releases. Translating research into news«, in: *JAMA* 287 (21), S. 2856.
- Woloshin, Steven/Schwartz, Lisa M./Casella, Samuel L./Kennedy, Abigail T./Larson, Robin J. (2009): »Press releases by academic medical centers: not so academic?«, in: *Annals of Internal Medicine* 150 (9), S. 613-618.
- Wormer, Holger (2008): »“Wie seriös ist Dr. Boisselier?” - Quellen und Recherchestrategien für Themen aus Wissenschaft und Medizin«, in: Holger Hettwer/Markus Lehmkuhl/Holger Wormer et al. (Hg.), *WissensWelten. Wissenschaftsjournalismus in Theorie und Praxis*, Gütersloh: Verlag Bertelsmann Stiftung, S. 322-340.
- Yavchitz, Amélie/Boutron, Isabelle/Bafeta, Aida/Marroun, Ibrahim/Charles, Pierre/Mantz, Jean/Ravaud, Philippe (2012): »Misrepresentation of randomized controlled trials in press releases and news coverage: a cohort study«, in: *PLoS Medicine* 9 (9), e1001308.
- Zahedi, Zohreh/Costas, Rodrigo/Wouters, Paul (2014): »How well developed are altmetrics? A cross-disciplinary analysis of the presence of ‘alternative metrics’ in scientific publications«, in: *Scientometrics* 101 (2), S. 1491-1513.
- Zhang, Ye/Willis, Erin/Paul, Michael J./Elhadad, Noémie/Wallace, Byron C. (2016): »Characterizing the (perceived) newsworthiness of health science articles: a data-driven approach«, in: *JMIR Medical Informatics* 4 (3), e27.
- Zhong, Bu/Newhagen, John E. (2009): »How journalists think while they write: a transcultural model of news decision making«, in: *Journal of Communication* 59 (3), S. 587-608.