

PHILIPP MÜLLER / STEFAN GEISS / CHRISTIAN SCHEMER /  
TERESA K. NAAB / CHRISTINA PETER (HRSG.)

METHODEN UND  
FORSCHUNGSLOGIK DER  
KOMMUNIKATIONSWISSENSCHAFT

**15** **Dynamische Prozesse  
der öffentlichen  
Kommunikation**

Methodische Herausforderungen

HW

**Bibliografische Information der Deutschen Nationalbibliothek**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

Philipp Müller / Stefan Geiß / Christian Schemer /  
Teresa K. Naab / Christina Peter (Hrsg.)  
*Dynamische Prozesse der öffentlichen Kommunikation.*  
*Methodische Herausforderungen*  
Köln: Halem, 2019

Die Reihe *Methoden und Forschungslogik der Kommunikationswissenschaft* wird herausgegeben von Prof. Werner Wirth.

Alle Rechte, insbesondere das Recht der Vervielfältigung und Verbreitung sowie der Übersetzung, vorbehalten. Kein Teil des Werkes darf in irgendeiner Form (durch Fotokopie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung des Verlages reproduziert oder unter Verwendung elektronischer Systeme (inkl. Online-Netzwerken) gespeichert, verarbeitet, vervielfältigt oder verbreitet werden.

© 2019 by Herbert von Halem Verlag, Köln

ISSN 1863-4966

ISBN (PRINT): 978-3-86962-404-4  
ISBN (PDF): 978-3-86962-405-1

Den Herbert von Halem Verlag erreichen Sie auch im Internet unter <http://www.halem-verlag.de>  
E-Mail: [info@halem-verlag.de](mailto:info@halem-verlag.de)

SATZ: Herbert von Halem Verlag  
LEKTORAT: Imke Hirschmann  
DRUCK: docupoint GmbH, Magdeburg  
GESTALTUNG: Claudia Ott Grafischer Entwurf, Düsseldorf  
Copyright Lexicon ©1992 by The Enschedé Font Foundry  
Lexicon® is a Registered Trademark of The Enschedé Font Foundry.

METHODEN UND FORSCHUNGSLOGIK  
DER KOMMUNIKATIONSWISSENSCHAFT

Philipp Müller / Stefan Geiß / Christian Schemer /

Teresa K. Naab / Christina Peter (Hrsg.)

# Dynamische Prozesse der öffentlichen Kommunikation

Methodische Herausforderungen

HERBERT VON HALEM VERLAG



## INHALTSVERZEICHNIS

PHILIPP MÜLLER / STEFAN GEISS / CHRISTIAN SCHEMER / TERESA K. NAAB / CHRISTINA PETER Die Dynamik öffentlicher Kommunikationsprozesse als Herausforderung für die empirische Forschung	9
MARCUS MAURER / JOHANNES DAXENBERGER / MATTHIAS ORLIKOWSKI / IRYNA GUREVYCH Argument Mining: A New Method for Automated Text Analysis and its Application in Communication Science	18
MARIO HAIM Capturing the Dynamics of Online News	38
MARKO BACHL An Evaluation of Retrospective Facebook Content Collection from Political Facebook Pages	57
PABLO JOST / FRANK MANGOLD / PASCAL JÜRGENS Herausforderungen und Potenziale der Nutzung digitaler Beobachtungsdaten auf Facebook – eine longitudinale Perspektive	73
SIMON KRUSCHINSKI / PASCAL JÜRGENS / BIRGIT STARK / MARCUS MAURER / CHRISTIAN SCHEMER In Search of the Known Unknowns. The Methodological Challenges in Developing a Heuristic Multi-Feature Framework for Detecting Social Bot Behavior on Facebook	103

DANIELA STOLTENBERG / DANIEL MAIER Descriptive Methods for Investigating Dynamics in Online Networks	147
PHILIPP K. MASUR Capturing Situational Dynamics and Processes: Strengths and Pitfalls of the Experience Sampling Method	177
ANNA SOPHIE KÜMPEL Dynamik im Blick: Die qualitative Beobachtung mit Post- Exposure-Walkthrough als Verfahren für die Rekonstruktion individueller Navigations- und Selektionshandlungen auf sozialen Netzwerkseiten (sns)	216
MARTIN WETTSTEIN / ANDREAS FAHR / ALEXANDER ORT Beyond Eyeballing: Automatische Mustererkennung in Prozessdaten	239
CHRISTIAN SCHEMER / STEFAN GEISS / PHILIPP MÜLLER Applying the Reinforcing Spirals Model to Dynamic Communication Phenomena: Conceptual and Statistical Pitfalls	262
MICHAEL SCHARKOW / MARKO BACHL Stable Attitudes and Behaviors as Boundary Conditions for Testing Reinforcing Spirals Models	286
LUKAS P. OTTO / FABIAN THOMAS Experience Sampling und multivariate Wachstumsmodelle zur Erfassung kurzfristiger selbstverstärkender Dynamiken in der Kommunikationswissenschaft	301
Autorinnen und Autoren	325







PHILIPP MÜLLER / STEFAN GEISS / CHRISTIAN SCHEMER /  
TERESA K. NAAB / CHRISTINA PETER

## Die Dynamik öffentlicher Kommunikationsprozesse als Herausforderung für die empirische Forschung

Nahezu alle Untersuchungen aus dem Bereich der Kommunikationsforschung befassen sich implizit oder explizit mit Prozessen. Sie fragen danach, wie kommunikative Phänomene entstehen, wie sie sich im Zeitverlauf entwickeln, welche gesellschaftlichen und individuellen Effekte sie auslösen und welche Ursachen ihrer Entstehung und ihrem Wandel zugrunde liegen. Nicht zuletzt entstehen auch die Kommunikationsakte selbst innerhalb eines Netzwerks dynamischer Interaktionsprozesse zwischen KommunikatorInnen und RezipientInnen, Botschaften und technischen Kanälen. Botschaften in der öffentlichen Kommunikation sind eben keine isolierbaren Einheiten, sondern stets eingebettet in eine Folge sozialer Handlungen und vorausgegangener sowie antizipierter Kommunikationsakte.

Während die Kommunikationswissenschaft konzeptuell schon seit langer Zeit der Prozesshaftigkeit öffentlicher Kommunikation mit Ansätzen wie dem Multi-Step-Flow (WEIMANN 1982), dem dynamisch-transaktionalen Ansatz (FRÜH/SCHÖNBACH 2005) oder dem Reinforcing-Spirals-Modell (SLATER 2015) Rechnung trägt, scheint die empirische Forschung hinterherzuhinken. Die Umsetzung längsschnittlicher Forschungsdesigns ist nicht nur zeit- und kostenintensiv, sie wirft auch eine ganze Reihe neuer Fragen bzw. zu lösender Probleme auf, z. B. die nach den richtigen Messzeitfenstern und dem passenden Zeitversatz, nach der Ziehung einer möglichst geeigneten Stichprobe, die über den gesamten Untersuchungszeitraum zur Verfügung steht, nach adäquaten Formen der (im Zeitverlauf möglichst

äquivalenten) Messung und einer angemessenen Auswertungsstrategie, die aus Effizienzgründen oft nicht ohne Aggregations Schritte bzw. Datenreduktion auskommt. Insofern ist die empirisch arbeitende Kommunikationswissenschaft mit mannigfaltigen methodischen Herausforderungen konfrontiert, wenn sie die theoretisch unterstellten Prozesse möglichst valide und reliabel abbilden möchte.

Zusätzlich verschärft hat sich diese Problematik durch den universellen ›Game Changer‹ unserer Zeit, das Internet. Internetbasierte Kommunikation zeichnet sich durch kontinuierlichen Informationsfluss und vielfältige Interaktionsmöglichkeiten ohne technologisch bedingten Zeitversatz aus. Sie nivelliert zudem weitgehend die geografisch-räumlichen Einschränkungen des Kommunikationsflusses. Es lässt sich also attestieren, dass öffentliche Kommunikationsprozesse durch die viel beschworene Digitalisierung noch einmal dynamischer und komplexer geworden sind. *Dynamischer* ist die Online-Kommunikation in dem Sinne, dass sie durch eine noch schnellere Abfolge einer immer größeren Menge potenziell aufeinander bezogener bzw. miteinander interagierender Kommunikationsakte besteht, die noch schwerer empirisch fassbar ist. *Komplexer* ist sie insofern, als im Rahmen dieser zeitlich und räumlich entgrenzten Masse an Kommunikationsakten immer unklarer ist, in welchen Zeiträumen, mit welchen Frequenzen, zwischen welchen Kommunizierenden und an welchen Stellen des scheinbar unbeschränkten digitalen Kommunikationsraumes sich die für eine spezifische sozialwissenschaftliche Fragestellung relevanten Kommunikationsprozesse vollziehen, geschweige denn wie sie vernetzt sind und wie sie sich empirisch fassen lassen. Folglich ist es wichtiger und gleichzeitig schwieriger geworden, diese Prozesse abzubilden. Die Kommunikationsforschung steht dadurch vor einer ganzen Reihe neuer Herausforderungen, die sich zu den noch nicht zufriedenstellend gelösten Problemen der Erfassung der grundsätzlichen Prozesshaftigkeit öffentlicher Kommunikation gesellen, die uns in der Offline-Welt beschäftigen. Sie reagiert mit einer methodischen Ausdifferenzierung, die unter anderem in der Entwicklung neuer digitaler Datenerhebungs- und Auswertungsverfahren unter dem Schlagwort ›Computational Methods‹ besteht.

Zusammenfassend lässt sich attestieren, dass Kommunikationsprozesse von wachsender Dynamik und Komplexität geprägt sind und Theorien verstärkt Dynamik in den Fokus nehmen. Gleichzeitig sind konventionelle empirische Methoden bei der Abbildung längsschnittlicher Abfolgen von Kommunikationsprozessen oft überfordert, während die Möglichkeiten

der Computational Methods gerade erst ausgelotet werden. Dies hat uns veranlasst, die Jahrestagung der DGPK-Fachgruppe Methoden der Publizistik- und Kommunikationswissenschaft, die im September 2017 in Mainz stattgefunden hat, unter das Motto ›Dynamische Prozesse in der Publizistik- und Kommunikationswissenschaft: Methodische Herausforderungen‹ zu stellen.

Schon im Vorfeld der Tagung wurden wir darauf hingewiesen, dass die Begriffsschöpfung ›dynamische Prozesse‹ doch einen Pleonasmus darstelle, da Prozesse per definitionem dynamisch seien. Unsere Wortwahl beruht darauf, dass wir Prozesse als (im Prinzip) voneinander isolierbare Vorgänge verstehen, die sich dabei gegenseitig beeinflussen, überlagern oder unabhängig voneinander entwickeln können und die gemeinsam eine bestimmte beobachtbare Entwicklung über die Zeit erzeugen – letztere bezeichnen wir als ›Dynamik‹. Es erscheint uns darüber hinaus notwendig, auf die Dynamik von Kommunikationsprozessen gesondert hinzuweisen, da wir in ihr eine der größten Herausforderungen der empirischen Kommunikationsforschung im Internetzeitalter sehen und der empirischen Kommunikationsforschung und ihrer Methodenentwicklung hier auch ein Defizit attestieren würden.

Zwar nimmt – unserem Eindruck nach – der Anteil längsschnittlicher Studien in unserem Forschungsfeld zu und damit auch der Anteil solcher Projekte, deren empirisches Design prinzipiell dazu geeignet ist, der Prozesshaftigkeit öffentlicher Kommunikation Rechnung zu tragen. Jedoch erscheint es uns fraglich, ob damit bereits eine Stufe erreicht ist, auf der wirklich davon gesprochen werden kann, dass kommunikative Dynamiken ausreichend abgebildet werden. Es genügt hierfür nicht, die Prozesshaftigkeit unseres Untersuchungsgegenstands zur Kenntnis zu nehmen und darauf aufbauend längsschnittliche Forschungsdesigns zu entwickeln, die sich mit drei bis fünf Messzeitpunkten begnügen und deren Startpunkt und Zeitversatz in Ermangelung einer fundierteren Informationsgrundlage oft auf der Basis von Vermutungen gewählt werden. Es genügt auch nicht, zeitverzögerte Effekte von Ursache- und Wirkungsvariablen als dynamische Prozesse zu titulieren, weil damit nur die Beziehung zweier Zustände zu unterschiedlichen Zeitpunkten untersucht wird, aber nicht die Dynamik des Prozesses selbst, der sich im betreffenden Zeitraum entfaltet hat. Es erscheint vielmehr unerlässlich, die zeitlichen Dynamiken ganz unterschiedlicher Kommunikationsphänomene wirklich zu durchdringen, um ein adäquates empirisches Abbild von ihnen zu schaffen.

Die Herstellung eines Gleichgewichts zwischen forschungsökonomisch vertretbarem Erhebungsaufwand und einer ausreichend kleinteiligen Datenlage, die nicht Gefahr läuft, wichtige Ausschnitte, Veränderungsstufen oder Kippmomente dynamischer Kommunikationsprozesse zu verpassen, stellt dabei eine der größten Herausforderungen dar.

Der vorliegende Band widmet sich diesem Themenkomplex. Aus ganz unterschiedlichen Forschungsperspektiven beleuchten die hier versammelten Beiträge den Stand der Methodenentwicklung bezüglich der Dynamik von Kommunikationsprozessen, skizzieren Problemfelder und machen Lösungsangebote. Die Beiträge sind auf verschiedenen Stufen des Forschungsprozesses angesiedelt und befassen sich mit einer Vielzahl methodischer Ansätze wie der automatisierten Inhaltsanalyse, der Netzwerkanalyse oder der qualitativen Beobachtung. Sie eint die Suche nach innovativen Lösungen für ein gemeinsames Problem, nämlich die (zunehmende) Dynamik öffentlicher Kommunikationsprozesse adäquat abzubilden. Der Band dokumentiert damit einerseits die Mainzer Jahrestagung, enthält aber auch einige Weiterentwicklungen von Tagungsbeiträgen sowie punktuelle Ergänzungen des Tagungsprogramms.

Neben Überblicken und Systematisierungen von Problemlagen verschiedener Forschungsfelder oder Methodenbereiche finden sich im Band auch empirische Lösungsvorschläge für konkrete methodische Probleme aus verschiedenen Forschungsfeldern der öffentlichen Kommunikation. Viele Beiträge vereinen Überlegungen zur Datengewinnung mit solchen zur Datenauswertung und bilden daher bereits in sich verschiedene Stufen des Forschungsprozesses ab. Die Reihung der Beiträge im Band folgt daher einer eher losen inhaltlichen Sortierung anhand der Forschungsgegenstände.

Den Aufschlag liefern MARCUS MAURER, JOHANNES DAXENBERGER, MATTHIAS ORLIKOWSKI und IRYNA GUREVICH, die sich in einem Überblicksaufsatz mit der Argument-Mining-Methode beschäftigen. Dies ist ein in der Informatik entwickeltes, auf Machine-Learning-Algorithmen beruhendes Verfahren zur überwachten automatisierten Extraktion von Argumentation aus Textkorpora. Insbesondere für längsschnittliche Fragestellungen über einen längeren Zeitraum stößt die manuelle Inhaltsanalyse aufgrund der großen Menge zu codierenden Materials an ihre Grenzen. Um die langfristigen Dynamiken medial vermittelter Diskurse im Zeitverlauf nachzuzeichnen und dabei nicht bloß Berichterstattungsgegenstände oder Valenzen, sondern inhaltliche Argumentationsmuster zu erfassen, kann Argument Mining daher ein vielversprechender Ansatz sein.

Der Aufsatz stellt die Entwicklung dieses Ansatzes, seiner Potenziale und Limitationen zusammenfassend vor und diskutiert seinen Mehrwert für die Kommunikationswissenschaft und ihre spezifischen Fragestellungen.

Die Dynamik von Medieninhalten stellt auch neue Herausforderungen an die Stichprobenziehung für Inhaltsanalysen. MARIO HAIM befasst sich in seinem Beitrag über das Erfassen der Dynamik von Online-Nachrichten mit den Aktualisierungsintervallen von Online-Nachrichtenseiten. Anhand einer Analyse von fünf deutschen Nachrichtenportalen untersucht er, in welcher Frequenz neue Texte publiziert werden und Aktualisierungen bereits erschienener Texte erfolgen. Hieraus leitet er Empfehlungen für die Forschungspraxis im Rahmen von Online-Inhaltsanalysen ab.

Mit einer ähnlichen Zielrichtung beschäftigt sich MARKO BACHL mit der retrospektiven Erfassung von Social-Media-Posts. Er vergleicht die Ergebnisse einer nachträglichen Erhebung von politischen Facebook-Posts mit variierendem zeitlichen Abstand mit den Ergebnissen einer zeitnahen Datenerfassung. Seine Ergebnisse zeigen, dass eine nicht unerhebliche Menge von Posts bereits nach wenigen Monaten wieder gelöscht worden sind und entsprechend bei einer retrospektiven Datenerhebung nicht mehr erfasst worden wären. Hieraus leitet er Implikationen für konkrete Forschungsvorhaben ab, spricht sich aber auch für eine institutionalisierte, kontinuierliche Datenerhebung relevanter Inhalte der öffentlichen Kommunikation durch die Sozialforschung aus.

Auch PABLO JOST, FRANK MANGOLD und PASCAL JÜRGENS suchen, mit einem etwas anderen Fokus, nach dem richtigen Zeitpunkt für das Sampling von Facebook-Posts. Ihnen geht es im Besonderen um die dynamische Entwicklung von Interaktionen mit geposteten Inhalten, also um Shares, Likes und andere Reaktionen. Im Vordergrund steht die Frage, innerhalb welcher Zeiträume eine Sättigung der zu erwartenden Reaktionen erreicht ist und mit welchem Zeitversatz bzw. mit wie vielen Messzeitpunkten diese Popularitätsindikatoren folglich erhoben werden sollten, wenn man deren Entwicklung im Zeitverlauf nachzeichnen will. Die Ergebnisse sensibilisieren zudem dafür, dass Studien, die Facebook-Posts lediglich zu einem Zeitpunkt erfassen, und dies womöglich bereits kurze Zeit nach ihrer Veröffentlichung, mit der Auswertung von Popularitätsindikatoren zurückhaltend sein sollten, da diese über einen längeren Zeitraum dynamisch anwachsen.

Um Facebook-Inhalte geht es auch SIMON KRUSCHINSKI, PASCAL JÜRGENS, BIRGIT STARK, MARCUS MAURER und CHRISTIAN SCHEMER. Sie suchen im Rahmen eines größer angelegten Forschungsprojekts nach Möglichkei-

ten zur automatisierten Identifikation von Social-Bot-Accounts. In ihrem Beitrag stellen sie die hierfür bisher in der Forschungsliteratur eingesetzten Verfahren vor und unterziehen sie einer kritischen Bewertung. Im Anschluss präsentieren sie eine eigene Lösung zur Bot-Identifikation. Sie plädieren für die Kombination verschiedener Indikatoren. Zentral für ihre Detektionsstrategie ist die automatisierte Analyse der abgesetzten Posts hinsichtlich inhaltlicher Dubletten und zeitlicher Verteilungsmuster. Diesen Ansatz wollen sie in Zukunft zudem um thematische Indikatoren sowie Merkmale der Follower-Netzwerke der abgesetzten Accounts erweitern.

Für die Netzwerkanalyse, die sich mit kommunikativen Querverweisen und Beziehungen befasst, sollten zeitliche Dynamiken in Zukunft ganz grundsätzlich eine stärkere Rolle spielen. Zu diesem Schluss kommen DANIELA STOLTENBERG und DANIEL MAIER in ihrem Überblicksaufsatz zur Erfassung von Dynamiken in der Online-Netzwerkanalyse. Die meisten kommunikationswissenschaftlichen Netzwerkanalysen begnügen sich demnach bisher mit querschnittlichen Momentaufnahmen oder aggregierten longitudinalen Daten in einer Weise, die ihre zugrunde liegenden temporalen Strukturen ignorieren. Die Autoren erkennen hierin ein klares Manko und sprechen sich für die Implementierung zeitlicher Dynamiken in die Analysen aus, um etwa Fragestellungen aus dem Bereich der Diffusion von Inhalten adäquater adressieren zu können. Sie stellen verschiedene Ansätze vor und bewerten ihre Verwendbarkeit für unterschiedliche Datenstrukturen und Fragestellungen.

Eine Einführung in ein noch relativ junges methodisches Anwendungsfeld zur Erfassung von Dynamiken der Mediennutzung liefert auch PHILIPP MASUR. Er setzt sich mit der Experience-Sampling-Methode auseinander, einem Ansatz, der es erlaubt, Mediennutzung und damit in Verbindung stehende Konstrukte in situ anstatt retrospektiv zu messen. Im ersten Teil seines Beitrags stellt er die Methode, ihre Voraussetzungen und potenziellen Anwendungsgebiete vor. Im zweiten Teil wertet er einen Beispieldatensatz unter Verwendung einer Mehr-Ebenen-Analyse aus und illustriert daran die Konsequenzen verschiedener Auswertungsentscheidungen. Zudem untersucht er mögliche Designeffekte auf die Antwortmuster der Befragten.

Der Beitrag von ANNA SOPHIE KÜMPEL stellt ebenfalls einen Ansatz zur situationsbezogenen Erhebung von Mediennutzung dar. Sie greift jedoch auf ein Laborverfahren zurück und kombiniert qualitative Beobachtungen von Rezeptionsepisoden auf Social Networking Sites (SNS) mit der retrospektiven Rekonstruktion der zugrunde liegenden Selektionsentschei-

dungen mit den Probanden. Die Autorin argumentiert, dass mithilfe eines solchen qualitativen Mehr-Methoden-Ansatzes ein besseres Bild der intra-individuellen Prozesse der Rezeption dynamischer Medieninhalte (wie den Newsfeeds von Social-Media-Plattformen) entstehen kann als durch jede Form der standardisierten Ex-post-Befragung oder durch Logfile-Analysen.

Die begleitende Beobachtung von Rezeptionsprozessen steht auch im Mittelpunkt des Beitrags von MARTIN WETTSTEIN, ANDRAS FAHR und ALEXANDER ORT. Sie befassen sich mit der Frage, wie sich kontinuierliche quantitative Daten über eine Vielzahl von Rezeptionsepisoden, z.B. das dokumentierte Nutzungsverhalten auf einzelnen Websites oder physiologische Reaktionen während der Rezeption eines Filmbeitrags, sinnvoll auswerten lassen. Denn solche kontinuierlichen Datenströme werden in der Forschungspraxis bisher allzu oft nach relativ einfachen Regeln, bzw. per Augenschein, zu größeren Einheiten aggregiert. Die Autoren schlagen eine automatisierte Lösung zur Mustererkennung in Verlaufsdaten vor, die Forschende dazu ermächtigt, datengetrieben anstatt per bloßem Augenschein Verlaufsmuster und Sequenzen innerhalb von Rezeptionsepisoden zu identifizieren.

Ein abschließender Block von Beiträgen befasst sich mit den dynamischen Zusammenhängen zwischen Medienrezeption und Medienwirkung. In einem Überblicksaufsatz diskutieren CHRISTIAN SCHEMER, STEFAN GEISS und PHILIPP MÜLLER Michael Slaters Reinforcing-Spirals-Ansatz, der Mediennutzung einerseits und Einstellungen bzw. Verhaltensweisen von Individuen andererseits in einem wechselseitigen Abhängigkeitsverhältnis aus Medienwirkung und Medienselektion modelliert. Entgegen Slaters eigener Intention wird das Modell oft so verstanden, als prognostiziere es eine, von Drittvariablen weitestgehend unabhängige, dynamische wechselseitige Eskalation der beteiligten Variablen im Zeitverlauf. Allerdings scheitern empirische Studien regelmäßig daran, derartige Effekte zu beobachten. Dies lässt sich teilweise durch konzeptuelle Missverständnisse erklären, resultiert jedoch auch aus einer oft nicht adäquaten Auswertungslogik. Der Beitrag fasst die konzeptuellen und statistischen Grundlagen dieser Missverständnisse zusammen und entwickelt anschließend Vorschläge für Konzeptualisierungen und Auswertungslogiken, die den theoretischen Grundannahmen des Modells gerecht werden.

Dazu passend stellen MICHAEL SCHARKOW und MARKO BACHL empirische Befunde zur zeitlichen Stabilität von Variablen aus den Bereichen ›Mediennutzung‹ und ›politische Einstellungen‹ vor, die die Erwartbarkeit



von eskalatorischen Reinforcing-Spirals-Zusammenhängen mit bedeutsamen Effektstärken infrage stellen. Auch diese Befunde lassen also eher ein sich gegenseitig verstärkendes Equilibrium aus Mediennutzung und beeinflussten Variablen erwarten als eine immer weiter voranschreitende wechselseitige Beeinflussung von Einstellungen und Mediennutzungsmustern. Der Beitrag schließt mit einer Reihe von Implikationen für die empirische Anwendung des Ansatzes.

Eine solche Anwendung präsentieren im abschließenden Kapitel des Bandes LUKAS OTTO und FABIAN THOMAS. Sie nutzen die Experience-Sampling-Methode, um kurzfristige selbstverstärkende Dynamiken zwischen Medienrezeption und von ihr beeinflussten Variablen abzubilden. Das angewandte Design erlaubt eine genauere Analyse der Dynamiken zwischen den beteiligten Variablen als das klassische Längsschnittdesign mittels Panel-Befragung, welches mit wesentlich größeren Messintervallen arbeitet. Für die Auswertung ihrer Daten verwenden die Autoren ein multivariates Wachstumskurvenmodell, dessen Anwendungsgebiete und Voraussetzungen im Beitrag ebenfalls ausführlich diskutiert werden. Allerdings finden sich auch in dieser Studie keine eskalatorischen Reinforcing-Spirals-Zusammenhänge. Auf der Basis dieser Befundlage diskutieren die Autoren abschließend die Übertragbarkeit des Reinforcing-Spirals-Ansatzes auf wechselseitige Verstärkungsprozesse mit kurzem Zeitversatz.

Wie zu erwarten war, bieten dieser Band und die Beiträge, die er versammelt, keine endgültigen Antworten – solche für sich zu reklamieren könnte nur eine Mogelpackung sein. Vielmehr bieten die Beiträge spannende Ansätze und inspirierende Ideen rund um die Herausforderungen, die Dynamiken und Prozesshaftigkeit an die Methodik und Methodenentwicklung der Kommunikationswissenschaft herantragen. Die damit verbundenen Problemstellungen werden das Fach und seine Methodendebatten sicherlich noch lange Zeit beschäftigen.

## Literatur

- FRÜH, W.; K. SCHÖNBACH: Der dynamisch-transaktionale Ansatz III: Eine Zwischenbilanz. In: *Publizistik*, 50, 2005, S. 4-20. doi: 10.1007/s11616-005-0115-7
- SLATER, M. D.: Reinforcing spirals model: Conceptualizing the relationship between media content exposure and the development and

maintenance of attitudes. In: *Media Psychology*, 18, 2015, S. 370-395.  
doi: 10.1080/15213269.2014.897236

WEIMANN, G.: On the importance of marginality: One more step into the  
two-step flow of communication. In: *American Sociological Review*, 47,  
1982, S. 764-773. doi: 10.2307/2095212

MARCUS MAURER / JOHANNES DAXENBERGER /  
MATTHIAS ORLIKOWSKI / IRYNA GUREVYCH

## Argument Mining: A New Method for Automated Text Analysis and its Application in Communication Science

*Abstract:* In the age of online communication, automated methods for analyzing media content become more and more important in the social sciences. However, methods for automated content analysis used in communication science thus far are still limited to a small set of highly standardized categories like the central topic or the tone of a message and, therefore, fail to reach the complexity of manual codebooks. In this paper, we introduce argument mining, a new method for automated text analysis recently developed in computer science. Based on machine learning algorithms, Argument Mining detects full arguments in media messages. Moreover, the quality of these arguments can be measured. We give a brief introduction in the method and its underlying assumptions, discuss recent research in computer science based on Argument Mining, give an overview about various possible future applications of Argument Mining in communication science, and mention some limitations concerning the transfer of Argument Mining techniques to communication science issues.

*Keywords:* Automated Content Analysis, Argument Mining, Machine Learning, Argument Quality

## 1. Automated text analysis in communication science

Since the seminal works of Bernard Berelson (1952), content analysis has developed into a well-established method to measure the content of media messages in a more or less objective way. Until recently, in communication science and other social sciences, content analysis was primarily done manually with the help of trained human coders. This has changed for at least two reasons. First, in the age of online communication the number of messages relevant for communication scientists has dramatically increased. In traditional content analyses, findings based on samples consisting of a few hundreds or thousands of articles from especially far-reaching news media outlets were perceived as sufficient to make generalized claims about overall media coverage. Nowadays, the amount of news media outlets has become much more diverse and variable across media. Consequently, generalizing findings from a small sample of well-established news media may be misleading. Moreover, weblogs, (political) websites, and especially social media have become an important source of (political) information and are, therefore, perceived as relevant enough to include them in content analytic studies. As the population of online messages is unknown, creating a representative sample of e.g. websites or social media posts is hardly possible. Manually coding the full population of millions of online messages per day is unrealistic. Second, in computer science various techniques for automated content analysis that can also be applied to media messages have recently been developed. This resulted in a variety of open source tools for automated content analysis easily available from the internet. Consequently, communication scientists started to use these tools for their own research.

Compared to manual coding, automated content analysis has several advantages and disadvantages. On the one hand, millions of media messages can be analyzed quickly and – given that no human coders are needed – at low costs. Moreover, when media content is analyzed by computer programs, reliability issues do not seem to occur.<sup>1</sup> While human coders even after an intensive coder training still show some variations in their coding behavior, especially when latent variables like the tone of a message are concerned, no or very little variations exist when media mes-

1 However, as some of the computerized methods are probabilistic, repeated iterations will not show exactly the same results, which can also be seen as a reliability problem.

sages are analyzed by computer programs. On the other hand, analyzing media messages with the help of computer programs may cause validity problems. As computers are not to the same extent as human coders able to identify the connotative or latent meaning of a message, computerized content analyses may not be as valid as analyses done by human coders. This holds especially true when media messages use irony, which even human coders often can hardly uncover (e.g. LEWIS/ZAMITH/HERMIDA 2013). Moreover, while communication scientists for a long time have been used to develop more and more complex categories and codebooks for manual content analysis, especially open source tools for automated analysis are limited to relatively simple and highly standardized tasks like identifying the central topic or the general tone of a message. Consequently, communication scientists are to some degree tempted to reduce the complexity of their research questions to make them fit to the needs of tools for automated content analysis.

Automated content analysis methods used in communication science can roughly be distinguished in dictionary-based and machine learning methods (for a more elaborate classification see GRIMMER/STEWART 2013). Using dictionary-based methods, researchers start their analysis by defining a set of words (dictionary) which is applied to the media messages under examination. The occurrence of single words or co-occurrence of two or more words from the dictionary is used to automatically identify the topic, the central frame, or the tone (>sentiment<) of a message (e.g. GRO-SHEK/AL-RAWI 2013; GREUSSING/BOOMGARDEN 2017). In a study like that, for example, words like >unemployment<, >jobs<, >economic growth<, or >company< would be indicating that the central topic of a media message is the economy. Moreover, words like >winning<, >great<, or >up< would suggest a positive tone of the message. As these analyses usually ignore the context of the words under examination, especially dictionary-based sentiment analyses are quite shaky (e.g. >economic growth goes up< and >the unemployment rate goes up< might both be classified as a positive sentiment). A similar but much more elaborated method is Semantic Network Analysis (e.g. VAN ATTEVELDT/KLEINNIJENHUIS/RUIGROK 2008). Based on dictionaries including relevant actors and verbs, grammar parsers are used to identify the grammatical relations of evaluative statements (who evaluates whom with which tone? e.g., >Angela Merkel criticizes the Green Party<). However, especially in this case developing a dictionary including all relevant terms is a very time-consuming process.

Using machine learning methods, researchers often start with findings from manual content analysis or crowdsourcing procedures that are used to train a machine learning algorithm (supervised machine learning). Put simply, by comparing documents with their >correct< assessments by human coders, the algorithm step by step learns to classify new documents correctly. The validity of the computer-generated findings is finally assessed by comparing them to the findings produced by manual content analysis, which are treated as correct classifications. The most frequently employed algorithm in communication science is the latent Dirichlet allocation (LDA) which is used to identify the latent thematic structure of media messages (topic modeling; for an overview see MAIER et al. 2018). LDA topic modeling uses unsupervised machine learning. No pre-defined dictionary or manual coding is needed. The reason is that the algorithm calculates the probability with which a given topic appears in a media message based on the frequency of relevant terms occurring in this message (relating them to the frequency in which they occur in other media messages). Comparing the findings of dictionary-based text analysis and LDA topic modeling to findings from human coding, Guo et al. (2016) found that the LDA-based analysis performed much better than the dictionary-based approach.

Moreover, a few communication scientists have recently started to use machine learning algorithms in other contexts. On the one hand, there are several established open or closed source analytics platforms that are increasingly used by communication scientists. One of them is Crimson Hexagon, a platform that has been developed to identify the optimistic or pessimistic sentiment of media messages (HOPKINS/KING 2010; SU et al. 2016). Another is the spaCy module in Python, which has e.g. been used to analyze the linguistic complexity of different news media (TOLOCHKO/BOOMGARDEN 2017). On the other hand, communication scientists started to develop their own machine learning algorithms especially suitable for research questions from that field. For example, algorithms have been developed to identify frames (BURSCHER et al. 2014) or news factors (TRILLING/TOLOCHKO/BURSCHER 2018) in media messages.

Taken together, we can conclude that communication science increasingly uses computerized methods for content analysis and increasingly uses machine learning methods instead of dictionary-based methods. However, it is also obvious that the methods of automated content analysis used in communication science thus far are still limited to a small set of highly standardized categories and, therefore, fail to reach the complexity of

manual codebooks. In this paper, we will introduce methods for automated content analysis recently developed in computer science enabling communication scientists to conduct more complex analyses. Argument Mining uses machine learning algorithms to identify argumentative structures in digital texts. One of the applications of argument mining is to automatically assess the quality of arguments, e.g. whether an argument contains a substantial claim or a justification. We will explain the method, discuss some examples of analyses using Argument Mining recently conducted in computer science, and give an overview of possible future applications of Argument Mining in communication science.

## 2. Text Classification and Argument Mining: An Introduction

Before introducing the inner working of Argument Mining, we will give a brief introduction to statistical machine learning for text. As the term suggests, statistical machine learning learns heuristics over properties of texts, which help them to perform ›soft‹ decisions (as opposed to the ›hard‹ decisions taken using a set of hand-coded rules as in dictionary-based text analysis). Although it has been around since several decades, text classification (SEBASTIANI 2002) and text mining, i.e. the automatic assessment of certain properties of text, have gained significant attention in several disciplines lately. The recent popularity of automatic approaches to text understanding (or machine reading) can be attributed largely to the (resurrected) success of neural network methods in general and so-called ›deep learning‹ (i.e. deep neural networks with many layers) in particular. The success of these methods in text classification has followed their successful application to image processing, where deep learning methods yielded massive breakthroughs (LECUN/BENGIO/HINTON 2015).

### *Levels and types of classification*

Text mining and variants of this field of research (text analytics, text or document classification) is most frequently used to categorize entire documents (e-mail classification, sentiment analysis of tweets, etc.) or to structure the content of documents or document collections. The latter also may involve information retrieval (the detection of relevant information

from text collections, given a user-defined query), a research field of its own which we will not cover here. Textual documents can be (and are) defined on many levels: from short texts (e.g. comments on social media platforms) or single sentences up to very long documents such as entire books. Applications adding structure to textual documents on the level of (one or several) words or characters are often referred to as sequence tagging, as they seek to identify (<tag> individual portions (<sequences>) of the text with labels (e.g. part-of-speech, named entities, etc.). Although slightly simplifying, in the following we will focus on methods for short text classification (sentence or short document level) and sequence tagging. These are used frequently in current state-of-the-art Natural Language Processing (NLP, i.e. computer science research for automatically processing language resources<sup>2</sup>) applications in general and in Argument Mining (PELDSZUS/STEDE 2013), a subfield of NLP, in particular.

### *Supervised and unsupervised methods*

While the above-mentioned LDA method for topic modeling is an example for a popular unsupervised machine learning application, in the following we will focus on supervised methods. Unsupervised machine learning <learns> from the structure or content of (typically large volumes of) documents alone, while supervised machine learning requires manual coding of the category of interest (e.g. the sentiment of a document). One of the major disadvantages of unsupervised methods is that their validity and reliability are much harder to assess. The topics retrieved by an LDA analysis of a large collection of documents cannot be evaluated against a ground-truth but only by measuring their impact in downstream applications, e.g. a human expert searching for relevant content in the collection. To the opposite, in supervised machine learning, the algorithm is intended to learn to reproduce the manual coding of a human expert and can thus be evaluated using that same manual coding.

In practice, human experts or crowd workers code a subset of the entire document collection (typically a few thousand examples depending on the size of the collection and the level of classification, see above). The resulting ground-truth datasets is then divided into training and test data. The

2 E.g., machine translation, sentiment analysis, and other forms of content analysis.



machine learning algorithm is then trained on the former; the resulting model is applied to the test data (coded labels from human experts removed) resulting in automatic codes which can finally be compared to the human codes. Resulting validity of the model is often expressed as accuracy (share of correctly coded examples overall), precision (share of relevant examples<sup>3</sup>), recall (share of relevant examples that have been detected by the model<sup>4</sup>), or F1 scores (weighted average of precision and recall).

### *Algorithms for supervised learning*

The core of any program for automatic coding of textual documents is obviously the algorithm that does the inference from textual input to the desired output categories (labels or tags). Simple algorithms (close to the above-mentioned dictionary-based methods) learn a relationship between isolated words or sequences of words (so called *n*-grams). More complex methods also consider information above the level of *n*-grams (e.g. syntactic properties such as part-of-speech), or meta-information about a document (e.g. the timestamp of a tweet), or any other information that can be automatically extracted (e.g. the number of emojis in a tweet). Algorithms such as decision trees or support vector machines (SVM) learn to combine such portions of information (often referred to as *features*) in a smart way. Many of these types of approaches either learn to regress over features as variables (e.g. Logistic regression), or convert them into a high-dimensional vector space in which individual documents can be separated based on their categories (e.g. SVM). Another group of methods are so-called *artificial neural networks* which model complex relationships between in- and outputs using one or several layers of a network of neurons (inspired by the human brain). The above-mentioned deep neural networks abstract over specific input samples by transforming them in several (*deep*) layers. The output layer determines the result of the automatic coding. At training time, deep neural networks (DNNs) need to learn a large set of parameters (activation functions, batch sizes, etc.). DNNs take input in the form of vectors, thus, in text classification, documents need to be transformed into a meaningful vector representation. To this

3 True positives divided by the sum of true and false positives.

4 True positives divided by the sum of true positives and false negatives.

end, so-called ›word embeddings‹, have been applied with success lately. Word embeddings basically resemble the meaning of a word by projecting it into an optimized continuous vector space which is learned from the context in which a word appears. Popular methods to generate word embeddings are `word2vec` (MIKOLOV et al. 2013) and `fasttext` (BOJANOWSKI et al. 2016). Similar methods can learn embeddings for sentences or short documents, e.g. `infsent` (CONNEAU et al. 2017). Many architectures for DNNs are available, including ones for sequence tagging or sentence/short document classification (see above). Popular methods nowadays include Convolutional Neural Nets (CNNs) and Long Short-Term Memory networks (LSTM), or combinations of both (MA/HOY 2016).

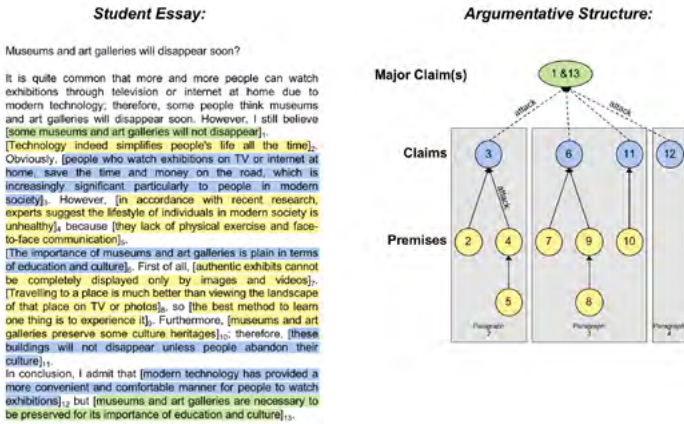
### *Argument Mining as text classification*

Argument Mining is the automatic analysis of argumentative structures in natural language discourse (VAN EEMEREN et al. 2014). Theories from philosophy and logic use various definitions for the concept of ›argument‹, but most theories agree that an argument consists of at least a claim (or conclusion) and one or more premises (reasons, evidence) connected to the claim. As such, Argument Mining includes, for example, the automatic identification of claims, but also the identification of more complex structures in which assertions must be linked with one another as well as with suitable premises. Such algorithms can detect, e.g. which arguments are mentioned for or against nuclear energy – regardless of how the arguments were put forth. In addition, Argument Mining can also be used to detect the quality of arguments, e.g. whether they are supported by evidence or refer to previous arguments.

Several Argument Mining methods operate on the level of entire sentences. This is based on the simplifying assumption that single sentences can be coded as individual (non-)arguments. To avoid making this assumption, arguments can also be coded as sequences of words, i.e. any span of text. The latter requires sequence tagging methods (see above). For an arbitrary text (in the example below, a student essay) the method finds different types of arguments (e.g. claims) and links them in a meaningful way (e.g. attacking each other). An example student essay and the resulting argumentative structure are displayed in Fig. 1.

The problem of detecting argumentative structures in text can be broken down into several sub-steps (STAB/GUREVYCH 2017a). First, argumentative

FIGURE 1  
An example essay and the resulting argumentative structure according to the claim-premise model



Source: Taken from Gurevych et al. 2016

text parts have to be recognized (i.e. sentences, clauses, or spans of words). Second, the relevant parts have to be classified (for example >claims< or >premises<). Finally, the relationships among the argumentative components are analyzed and classified (e.g. as >supporting< or >attacking<). We discuss an exemplary approach to this below.

### *Detecting argument quality*

Argument Mining can also be used to detect the quality of arguments. The argumentative structure of a text can be an indicator for certain aspects of argument quality, e.g. by counting the number of claims made or measuring the number of premises to support or attack claims. However, how to measure quality of arguments is a highly debated topic. Since the early works of Aristotle on logic, arguments are typically assessed along the dimensions of Ethos, Pathos, and Logos. Parts of these dimensions have been empirically assessed, e.g. the convincingness of arguments in online debates (HABERNAL/GUREVYCH 2016). Wachsmuth et al. (2017) present an empirical study on the reliability of several dimensions of argument quality and the approaches to automatically learn them. Despite these works,

to date, there is rather little agreement on the suitability of approaches to detect argument quality in written text.

Several empirical approaches are dealing with the absence of quality in arguments, by analyzing flaws in the argumentation. Recent studies include recognizing insufficiently supported arguments (STAB/GUREVYCH 2017b) and detecting fallacies (HABERNAL et al. 2018). Since these approaches measure specific weaknesses of the argumentation, their manifestation in texts can be coded more reliably. We report on one example in the next section.

### 3. Argument Mining in computer science: Recent applications

This paragraph gives an overview about recent research from computer science, which showcase practical approaches to detecting argumentative structures and argument quality.

One of the most successful models (EGER/DAXENBERGER/GUREVYCH 2017) for detecting argumentative structures solves this problem by means of a DNN. On a record of 402 student essays, this model correctly recognizes about 70 percent of the argumentative text spans. The corpus consists of student essays written in response to controversial topics such as competition or cooperation – which is better? (EGER/DAXENBERGER/GUREVYCH 2017). The input to the algorithm is a sequence of words coded using so called BIO-tags (begin-in-out). Each argumentative component is tagged by labeling the position of a word with respect to the component (B, I, or O), the respective component type (one of Major Claim, Claim, or Premise), its relation to previous or subsequent components (support, attack, pro, con), and the distance  $d$  to that related component (absolute or relative). Using this tagging scheme, any sequence tagging algorithm can be used to learn the automatic coding. The authors found that a DNN sequence tagging system such as a BiLSTM (an LSTM which reads the input from both beginning to end and reverse) with additional information on label dependencies from a Conditional Random Field (CRF et al. 2001) classifier and a CNN which captures information on the character level (embeddings on character-level are helpful for cases in which words are not found in the word embedding dictionary) performs best in this context.

While this model is useful to detect and analyze argumentative structures in student essays, its application to more heterogeneous text doc-

uments remains doubtful. Various approaches have used the coding of Stab and Gurevych (2017a) to label text types other than student essays, e.g. online participation projects (LIEBECK/ESAU/CONRAD 2016). However, models trained on these datasets are usually not applicable to other domains or text types (DAXENBERGER et al. 2017). One potential solution to this obstacle is to use a more general coding scheme, an approach discussed in the following paragraph.

As opposed to the model presented in the previous section, Stab, Miller, and Gurevych (2018) propose a coding scheme that strictly works on sentence level. Their system is intended to detect arguments (a text »expressing evidence or reasoning that can be used to either support or oppose a given topic«) in any kind of discourse. Thus, the unit of classification in this model is a sentence and a topic – the latter being a string of arbitrary length (one word or a full sentence). The coding scheme proposed in this work covers three basic types of sentences: no-argument, argument-pro, and argument-con. Consider the following examples for the topic »nuclear energy« from Stab et al. (2018):

1. »Nuclear fission is the process that is used in nuclear reactors to produce energy using element called uranium.«
2. »It has been determined that the amount of greenhouse gases have decreased by almost half because of the prevalence in the utilization of nuclear power.«

While sentence 1. is stating a factual claim about the topic that is not expressing any (positive or negative) stance towards it (thus, no-argument), 2. expresses (implicit) positive stance, stating a factual claim as evidence (thus, argument-pro according to STAB et al. 2018).

The authors used crowdsourcing to code random sentences extracted from documents retrieved from Google search results (the topic used as query term). Overall, the dataset they created consists of about 28,000 sentences for eight topics. Using this scheme, the agreement both among a group of experts, as well as among crowdworker and experts was Cohen's  $\kappa$  0.72, showing that this task can be solved with high reliability.

Stab et al. (2018) also trained a DNN to detect sentences as no/pro/con-arguments in web documents. They used a BiLSTM architecture (see above), encoding the sentence in question in the following way: the embedding of each word  $w_i$  of the sentence is concatenated with the cosine similarity between  $w_i$  and the word embeddings (average) of the words contained in the topic. To understand the validity of the automatic coding approach, the

authors carried out experiments both on documents from a single topic, as well as on documents from different topics (training on documents from seven topics and testing on documents from the remaining topic). Their best model achieved an F1 score (macro, i.e. averaged over the three codes) of 0.74 for in-topic experiments (average across topics), and 0.66 across topics (again, average).

Habernal et al. (2018) analyze a special type of fallacy in argumentative contexts, so-called ad hominem arguments, which target the opponent personally instead of refuting his or her counter-arguments. The authors carry out an empirical study using web-based argumentation data from Change My View on Reddit.<sup>5</sup> Change My View allows users to start a discussion thread by detailing an opinion that they hold, e.g. ›We need to restructure our global economy to save the planet from climate change‹. Other users then try to convince the original opinion holder to change his or her view. If a post successfully changes any discussion participant's view, the argument is awarded so-called ›delta points‹. The discussion is moderated strictly, so that Habernal et al. (2018) could harness the moderators' decisions to generate datasets with examples of ad hominem for their experiments. First, they study ad hominem arguments in isolation and estimate how accurately humans can identify them based only on a post's text using a comparison of the moderators' and crowd-sourced annotations. They find that a CNN is able to decide whether a post is ad hominem or not with a level of accuracy comparable to the human annotators (81.0% vs. 87.8% correctly classified instances). Secondly, the authors analyze the influence of the topic's characteristics on the occurrence of ad hominem arguments targeted directly at the original post. In particular, they crowdsourced ratings for the level of controversy and reasonableness of the original post and find that more controversial topics are associated with ad hominem arguments, but that there is no strong statistical pattern between reasonableness and ad hominem. They also build systems to automatically predict the reasonableness and controversy of an original post based on its text. In both cases a regression model combining a DNN and a topic model (CNN-LDA) performs best but has considerably lower scores when predicting reasonableness.

5 <https://www.reddit.com/r/changemyview/>

In our own yet unpublished project about claim type detection in political debates, we have created a corpus based on the publicly available transcripts of speeches held in the German Bundestag. A sample of speeches held between 2010 and 2015 was manually coded with claim type annotations (SCHIAPPA/NORDIN 2013) on sentence level. Here, a claim can be descriptive (fact claims, true or false), evaluative (value claims, good or bad) or prescriptive (policy claims, referring to a specific course of action). The annotation process was carried out by students as part of a content analysis class project. The final corpus consists of 38 individual speeches and 1,914 sentences. Of all coded sentences, 14 percent are policy claims, 34 percent value claims, 28 percent fact claims, and 24 percent do not contain a claim.

Based on this corpus, we have developed a baseline system to perform claim type classification, i.e. given a sentence, decide if it contains a claim and if so, determine its type, otherwise label it as non-claim. The system uses a CNN architecture for sentence classification (KIM 2014). Due to the relatively small dataset, we evaluate the system in a 5-fold cross-validation setting, i.e. we split the corpus into five sections of approximately equal length and alternately use one of these as test data and the respective other four as training data. The results show that this comparatively simple and hardly task-specific system is able to correctly classify a reasonable number of sentences. The averaged results for all claim types resulted in F1 scores of 75.24 percent (micro) and 73.87 percent (macro).<sup>6</sup>

#### 4. Conclusion: Applying Argument Mining to communication research

While in communication science various methods of automated content analysis are used, most of them are still limited to the analysis of a small set of standard categories like the topics or the tone/sentiment of a message. Argument Mining, a new method for automated text analysis recently developed in computer science, allows for much more detailed analyses. Instead of simply measuring, e.g. the tone of coverage, Argument Mining measures the occurrence of full arguments used in a variety of (media)

6 The experiments used to produce these results are based on the Tensorflow implementation given here: <https://github.com/dennybritz/cnn-text-classification-tf> (in Python).

messages. Moreover, the quality of these arguments can be measured, e.g. based on the claims used or the presence of a justification. Identifying arguments as well as identifying argument quality concerns various theories and research questions that are in the center of communication research.

The use of arguments in media messages, for example, can serve as an elaborated indicator of media bias (see SCHÖNBACH 1977, for an early approach). In most present research, communication scientists use relatively simple indicators of media bias like the visibility of political candidates or parties and the tone of media coverage (HOPMANN/VAN AELST/LEGNANTE 2011). However, due to the demand that news should be presented in a balanced way, mass media may not simply cover a certain candidate constantly with a positive tone but rather prefer the candidate's arguments in a political conflict over the arguments of other candidates. By constantly repeating and confirming them, they may establish a certain view in a political conflict that later e.g. influences voting behavior in a much more subtle way (REINEMANN et al. 2013). In the same way, when analyzing social media messages, Argument Mining can be used to uncover public opinion expressed in social media by measuring the frequency of the occurrence of pro- and con-arguments in a given political conflict. Again, measuring full arguments may be a much more precise indicator of public opinion than simply measuring the tone of a message. Moreover, exactly knowing the arguments frequently used in political conflicts may also help (political) strategists to develop effective counterarguments.

Identifying argument quality is in the center of various normative theories of democracy. They consider (political) debates as an opportunity for deliberative discussions in which the quality of arguments is crucial for (political) decision-making (e.g., HABERMAS 1991). For communication scientists, analyzing the quality of political discourses is relevant in various contexts. First, similar to the study on claim detection above, this concerns the debate strategies of politicians, e.g. the analysis of argument quality in political talk shows (e.g., SCHULTZ 2006) or the use of rhetorical strategies in televised debates (e.g., REINEMANN/MAURER 2005). Second, communication scientists have also recently started to deal with the deliberativeness of news media coverage. This concerns, e.g. the question which media systems and media formats support a rational political discourse (e.g., WESSLER/RINKE 2014). Moreover, by combining the analysis of political debates and media coverage about these debates, one could analyze how effective different kinds of arguments are. This concerns, e.g., the question which



argumentative strategies used in political debates increase the likelihood of being cited in the news media afterwards (e.g., JOST/SÜLFLOW/MAURER 2015). Third, political debates in social media or user comments to online news articles can be analyzed. Using Argument Mining, the frequently examined question how elaborated online comments are (e.g., KSIĄZEK 2018), can be answered on a large basis including thousands of different platforms and using elaborated indicators of argument quality.

Taken together, Argument Mining promises to allow deeper insights into media content compared to other methods of automated text analysis that are already used in communication science as it analyzes full arguments including argument quality instead of relatively simple dictionary-based features. Moreover, the initial effort for doing a study with Argument Mining is relatively low as it is based on machine learning algorithms and no dictionary needs to be developed in advance. However, the transfer of Argument Mining techniques to communication science issues is by no means trivial, because mass media content is often less clearly structured than essays or scientific publications. The main problem here is that arguments in different types of discourse can take very different forms, and that the evaluation of argument quality crucially depends on contextual knowledge. A prerequisite for improving the automatic procedures are suitable training data. The creation of such resources is also the subject of research, since there are hardly any coding schemes for argumentative structures or argument quality that are designed for heterogeneous discourse types (DAXENBERGER et al. 2017). Nevertheless, recent research in NLP has recognized this problem and addresses it in several ways: generically applicable coding schemes as the one proposed by Stab et al. (2018) are suitable for heterogeneous types of discourse (at the price of a slightly simplifying concept of argument). Another promising direction is to leverage multi-task learning to combine (training) data from multiple datasets (SCHULZ et al. 2018). This both helps to address the problem of domain dependency as the datasets might come from multiple types of discourse, as well as to increase the number of training examples.

## References

- BERELSON, B.: *Content Analysis in Communication Research*. Glencoe [The Free Press] 1952
- BOJANOWSKI, P.; E. GRAVE; A. JOULIN; T. MIKOLOV: *Enriching Word Vectors with Subword Information*, 2016. arXiv Preprint arXiv:1607.04606
- BURSCHER, B.; D. ODIJK; R. Vliegenthart; M. DE RIJKE; C. DE VREESE: Teaching the computer to code frames in news: Comparing two supervised machine learning approaches to Frame Analysis. In: *Communication Methods & Measures*, 8, 2014, pp. 190-206
- CONNEAU, A.; D. KIELA; H. SCHWENK; L. BARRAULT; A. BORDES: Supervised learning of universal sentence representations from natural language inference data. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 681-691
- DAXENBERGER, J.; S. EGER; I. HABERNAL; C. STAB; I. GUREVYCH: What is the essence of a claim? Cross-domain claim identification. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 2055-2066
- EGER, S.; J. DAXENBERGER; I. GUREVYCH: Neural end-to-end learning for computational argumentation mining. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 2017, pp. 11-22
- GREUSSING, E.; H. BOOMGARDEN: Shifting the refugee narrative? An automated frame analysis of Europe's 2015 refugee crisis. In: *Journal of Ethnic and Migration Studies*, 43, 2017, pp. 1749-1774
- GRIMMER, J.; B. M. STEWART: Text as data: The promise and pitfalls of automatic content analysis methods for political texts. In: *Political Analysis*, 21, 2013, pp. 267-297
- GROSHEK, J.; A. AL-RAWI: Public sentiment and critical framing in social media content during the 2012 U.S. presidential campaign. In: *Social Science Computer Review*, 31, 2013, pp. 563-576
- GUO, L.; C. J. VARGO; Z. PAN; W. DING; P. ISHWAR: Big social data analytics in journalism and mass communication: Comparing dictionary-based text analysis and unsupervised topic modeling. In: *Journalism & Mass Communication Quarterly*, 93, 2016, pp. 332-359
- GUREVYCH, I.; C. REED; N. SLONIM; B. STEIN: *NLP approaches to computational argumentation*. Tutorial held at the 2016 annual meeting of the Association for Computational Linguistics, 2016

- HABERMAS, J.: *Erläuterungen zur Diskursethik*. Frankfurt/M. [Suhrkamp] 1991
- HABERNAL, I.; J. ECKLE-KOHLER; I. GUREVYCH: Argumentation mining on the web from information seeking perspective. In: *Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and Natural Language Processing*, 2014, pp. 26-39
- HABERNAL, I.; I. GUREVYCH: What makes a convincing argument? Empirical analysis and detecting attributes of convincingness in web argumentation. In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2016, pp. 1214-1223
- HABERNAL, I.; H. WACHSMUTH; I. GUREVYCH; B. STEIN: Before name-calling: Dynamics and triggers of ad hominem fallacies in web argumentation. In: *Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2018
- HAHN, U.; M. OAKSFORD: A Bayesian approach to informal argument fallacies. In: *Synthese*, 152, 2006, pp. 207-236
- HOPKINS, D. J.; G. KING: A method for automated nonparametric content analysis for social science. In: *American Journal of Political Science*, 54, 2010, pp. 229-247
- HOPMANN, D. N.; P. VAN AELST; G. LEGNANTE: Political balance in the news: A review of concepts, operationalizations and key findings. In: *Journalism*, 13, 2011, pp. 240-257
- JOST, P.; M. SÜLFLOW; M. MAURER: »Das können Sie morgen in der Zeitung lesen« – Wie die Massenmedien über die Generaldebatten im Bundestag berichten und wie die Politik darauf reagiert. In: *Publizistik*, 61, 2015, pp. 5-20
- KIM, Y.: Convolutional neural networks for sentence classification. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, 2014, pp. 1746-1751
- KIRSCHNER, C.; J. ECKLE-KOHLER; I. GUREVYCH: Linking the thoughts: Analysis of argumentation structures in scientific publications. In: *Proceedings of the 2nd Workshop on Argumentation Mining held in Conjunction with the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2015, pp. 1-11
- KSIAZEK, T. B.: Commenting on the news. Explaining the degree and quality of user comments on news websites. In: *Journalism Studies*, 19, 2018, pp. 650-673

- LAFFERTY, J. D.; A. MCCALLUM; F. C. N. PEREIRA: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: *Proceedings of the Eighteenth International Conference on Machine Learning (ICML '01)*, 2001, pp. 282-289
- LECUN, Y.; Y. BENGIO; G. E. HINTON: Deep learning. In: *Nature*, 521(7553), 2015, pp. 436-444
- LEWIS, S. C.; R. ZAMITH; A. HERMIDA: Content analysis in an era of big data: A hybrid approach to computational and manual methods. In: *Journal of Broadcasting & Electronic Media*, 57, 2013, pp. 34-52
- LIEBECK, M.; K. ESAU; S. CONRAD: What to do with an airport? Mining arguments in the German online participation project Tempelhofer Feld. In: *Proceedings of the 3rd Workshop on Argument Mining*, 2016, pp. 144-153
- MA, X.; E. HOVY: End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, Volume 1: Long Papers, 2016, pp. 1064-1074
- MAIER, D.; A. WALDHERR; P. MILTNER; G. WIEDEMANN; A. NIEKLER; A. KLEINERT; B. PFETSCH; G. HEYER; U. REBER; T. HÄUSSLER; H. SCHMID-PETRI; S. ADAM: Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. In: *Communication Methods & Measures*, 2018
- MIKOLOV, T.; K. CHEN; G. CORRADO; J. DEAN: Efficient estimation of word representations in vector space. In: *CoRR*, 1301.3781, 2013. <http://arxiv.org/abs/1301.3781>
- MOCHALES-PALAU, R.; M.-F. MOENS: Argumentation mining: The detection, classification and structure of arguments in text. In: *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, 2009, pp. 98-107
- PELDSZUS, A.; M. STEDE: From argument diagrams to argumentation mining in texts: A survey. In: *International Journal of Cognitive Informatics and Natural Intelligence*, 7(1), 2013, pp. 1-31
- REINEMANN, C.; M. MAURER: Unifying or polarizing? Short-term effects and post-debate consequences of different rhetorical strategies in televised political debates. In: *Journal of Communication*, 55, 2005, pp. 775-794
- REINEMANN, C.; M. MAURER; T. ZERBACK; O. JANDURA: *Die Spätentscheider. Medieneinflüsse auf kurzfristige Wahlentscheidungen*. Wiesbaden [Springer VS] 2013

- SCHIAPPA, E.; J. P. NORDIN: *Argumentation: Keeping faith with reason*. Boston/MA [Pearson] 2013
- SCHÖNBACH, K.: *Trennung von Nachricht und Meinung. Empirische Untersuchung eines journalistischen Qualitätskriteriums*. Freiburg [Karl Alber Verlag] 1977
- SCHULTZ, T.: *Geschwätz oder Diskurs? Die Rationalität politischer Talkshows im Fernsehen*. Köln [Herbert von Halem] 2006
- SCHULZ, C.; S. EGER; J. DAXENBERGER; T. KAHSE; I. GUREVYCH: Multi-task learning for argumentation mining in low-resource settings. In: *Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2018
- SEBASTIANI, F.: Machine learning in automated text categorization. In: *ACM Computing Surveys*, 34, 2002, pp. 1-47
- STAB, C.; I. GUREVYCH: Parsing argumentation structures in persuasive essays. In: *Computational Linguistics*, 43, 2017a, pp. 619-659
- STAB, C.; I. GUREVYCH: Recognizing insufficiently supported arguments in argumentative essays. In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, 2017b, pp. 980-990
- STAB, C.; T. MILLER; I. GUREVYCH: Cross-topic argument mining from heterogeneous sources using attention-based neural networks. In: *Arxiv Preprint, arXiv:1802.05758*, 2018
- SU, L. Y.-F.; M. A. CACCIATORE; X. LIANG; D. BROSSARD; D. A. SCHEUFELE; M. A. XENOS: Analyzing public sentiments online: Combining human- and computer-based content analysis. In: *Information, Communication & Society*, 20, 2017, pp. 406-427
- TOLOCHKO, P.; H. BOOMGARDEN: Analysis of linguistic complexity in professional and citizen media. In: *Journalism Studies*, 19, 2018, pp. 1786-1803
- TRILLING, D.; P. TOLOCHKO; B. BURSCHER: From newsworthiness to shareworthiness: How to predict news sharing based on article characteristics. In: *Journalism & Mass Communication Quarterly*, 94, 2017, pp. 38-60
- VAN ATTEVELDT, W.; J. KLEINNIJENHUIS; N. RUIGROK: Parsing, semantic networks, and political authority. Using syntactic analysis to extract semantic relations from Dutch newspaper articles. In: *Political Analysis*, 16, 2008, pp. 428-446

- VAN EEMEREN, F. H.; B. GARSSSEN; E. C. W. KRABBE; A. F. SNOECK  
HENKEMANS; B. VERHEIJ; J. H. M. WAGEMANS: *Handbook of  
Argumentation Theory*. Wiesbaden [Springer vs] 2014
- WACHSMUTH, H.; N. NADERI; Y. HOU; Y. BILU; V. PRABHAKARAN;  
T. ALBERDINGK THIJM; G. HIRST; B. STEIN: Computational  
argumentation quality assessment in natural language. In: *Proceedings  
EACL*, 2017, pp. 176-187
- WESSLER, H.; E. M. RINKE: Deliberative performance of television news in  
three types of democracy: Insights from the United States, Germany,  
and Russia. In: *Journal of Communication*, 64, 2014, pp. 827-851

MARIO HAIM

## Capturing the Dynamics of Online News

*Abstract:* A changing media environment and decreasing revenue requires online journalism to adapt to the dynamics of an online *Marketplace of Attention* (WEBSTER 2014) by publishing fast and often, appealing to users, and through a multitude of distribution channels. Yet, this adaptation entails challenges to journalism research since such dynamics in online news are difficult to capture. Thus, this methodological study sets out to provide a more thorough understanding of news outlets' publishing and updating patterns to derive adequate recommendations for data capturing within the dynamics of online news. Data originates from an automated online observation of five exemplary news outlets over the course of six months, resulting in 53,356 captured articles and 260,963 captured homepages. The results highlight the underlying assumption that online news follow different dynamics than legacy news, illustrating online news' constant and short-lived but at the same time almost static character since updates on single articles are rare. For future studies, the findings suggest capturing online news continuously and on rather short notice, homepages and individual articles deserve separate and differentiated consideration.

*Keywords:* online news, data capture, sampling, web scraping, observation, content analysis

Online journalism faces several challenges these days. Most media organizations' revenue is decreasing due to a changing media environment and changing media use patterns. News are increasingly consumed online, on the go, and preponderantly after stumbling upon it on social net-

working sites (NEWMAN et al. 2017; NIELSEN/SCHRÖDER 2014; RÖSSLER/HAUTZER/LÜNICH 2014). The willingness to pay for online news is stagnantly low, and advertising revenues alone are insufficient (CHYI/LEE 2013; FLETCHER/NIELSEN 2017; PICARD 2014). One strategy to tackle these challenges (KÜNZLER 2016; LOBIGS 2018) is that online journalism is required to adapt to the dynamics of the online *Marketplace of Attention* (WEBSTER 2014) by publishing fast and often, appealing to users, and through a multitude of distribution channels (ANDERSON 2011a, 2011b; HELMOND 2015; NIELSEN/GANTER 2018; ZAMITH 2018). This adaptation to online dynamics also entails various challenges to journalism research, both theoretically and methodologically.

Theoretically, traditional understandings of news as periodic and public goods seem partly passé as – contrary to printed newspapers – an online news outlets' ›edition‹ from today does not necessarily differ completely from yesterday's version (NEUBERGER/NUERNBERGK 2015: 203). Moreover, news not only need to address human users but also need to aim at automated crawlers to be picked up by in order to gain (additional) publicity within intermediaries' platforms such as social networking sites (DICK 2011; NIELSEN/GANTER 2018). Hence, prior studies indicate that news outlets' homepages – online counterparts to printed front pages – are maintained at short intervals and individual articles are published at strategically opportune times of the day through well-chosen distribution channels. Once published, they are updated and extended over time (CARLSON 2007; HANUSCH 2015; HARDER/SEVENANS/VAN AELST 2017; MITCHELSTEIN/BOCZKOWSKI 2009; SALTZIS 2012).

Methodologically, such dynamics in online news are difficult to capture. While data capturing of news is traditionally ›bound to continual affordances of analogue communication forms‹ (KARLSSON/SJØVAAG 2016: 178), online news have become ›liquid‹ (KARLSSON/SJØVAAG 2016: 181f.), rendering it impossible to determine – and thus capture – a single homepage or an individual article. That is, recent studies have shown that individual articles might very well be repositioned or repackaged due to audience feedback and organizational routines (TANDOC/FERRUCCI 2017; TANDOC/JENNER 2016). Yet, despite these apprehensions, little is known about the actual degree of online news' ›liquidity‹. For example, while scholars agree on the necessity to capture news outlets' homepages frequently, currently employed capturing intervals range from fifteen minutes (e.g., BRIGHT/NICHOLLS 2014) to three hours (e.g., LEE/LEWIS/POWERS 2014), sometimes



including (e.g., ZAMITH 2016), sometimes excluding (e.g., BOCZKOWSKI/MITCHELSTEIN/WALTER 2011) weekends.

This study seeks to provide adequate recommendations for data capturing within the dynamics of online news. In the remainder of this chapter, current endpoints and endeavors for capturing the dynamics of online news are examined. Then, results on publishing patterns and update frequencies from a six-month period of automatically observing various online news outlets are presented. Ultimately, these findings are discussed in order to derive suggestions on adequate online news capturing.

## 1. Capturing online dynamics

Capturing and archiving online content is troublesome. Online documents, such as single websites, images, or stylesheets, are stored decentralized and can change rapidly. Since there is no gold standard for archiving online documents or, let alone, online documents' single versions at certain points in time, it is almost impossible to draw adequate samples of online news post-hoc. In order to establish archives of their own, some resource-rich news outlets have set out to archive their own online publications. While some news outlets even provide public access to parts of their archives (e.g., *The Guardian*, *The New York Times*, *Die Zeit*), these outlet-specific archives typically do neither contain homepages nor multiple versions of online documents. Instead, they provide the final version of individual articles' raw texts (i.e., without styling information). In a more universal approach, community endeavors such as archive.org seek to store every publicly available online document including some single versions at certain points in time. However, due to this endeavor's massive scale, online documents can neither be crawled nor stored in every single version. Thus, for full access to single versions of online news, adequate archives need to be built manually.

Three common channels serve this purpose in research projects. First, application programming interfaces (APIs; BURGESS/BRUNS 2012) provide easy-to-use access to structured information. Yet, APIs are only provided by some resource-rich news outlets to directly access their archives. Relatedly, APIs are common among intermediaries (e.g., Facebook, Twitter), which provide indirect access to news outlets' shared (article) links. However, this indirect approach lacks article completeness (KEYLING/HAIM 2014).

In addition, APIs are dependent upon a provider's will and may change unexpectedly. Second, a majority of studies captures and archives news items through RSS feeds as this channel comprises the least technological resistance (BAKKER 2012). While RSS feeds need to be provided by the news outlet itself, almost all outlets offer such feeds. Yet, RSS feeds lack context, such as styling information and an article's position on the website, as they only provide individual articles' raw texts. Third, web scraping provides an almost natural way of capturing news (MARRES/WELTEVREDE 2013) as it »accesses web pages, finds specified data elements on the page, extracts them, transforms them if necessary, and finally saves these data as a structured data set« (BOEING/WADDELL 2017: 3). Web scraping thus acts almost like a human user and thereby overcomes the drawback of missing context. However, in contrast to APIs and RSS feeds, scraping provides reactive access and – due to the necessity to extract information from unstandardized HTML code (>boilerplate removals<) – comprises the technologically most challenging approach. Moreover, some websites legally refuse permission to automatically and systematically capture their online content. Despite these shortcomings, web scraping has emerged as the most extensive (KEYLING/HAIM 2014) and direct channel to manually archive online news.

In the process of scraping news outlets' websites including their homepage as well as individual articles, timestamps are of the essence and thus need to be stored. However, between an article's actual publication time or a homepage's modification time and the scraping process delays might occur introducing minor inaccuracies to the archive. Two alternatives are available. First, technological standards require publication and modification timestamps to be included inside every transmission's HTTP header (RFC 7232, 2014). However, such timestamps are primarily used to indicate a document's caching lifetime. In addition, these timestamps are reported by the document's original web server and due to their customizability are not applicable as reliable sources of a news item's publication or modification time. Second, journalistic norms ask news outlets to disclose such times. Yet, while most outlets promote a publication time, formats and willingness to also disclose modification times vary. Taken together, neither of the two alternatives provides a reliable source of timestamps. In turn, date and time of the scraping process itself provide a reliable source of such timestamps and are thus preferable, although they can only serve as approximation.

## 2. Research questions

This study seeks to provide adequate recommendations for data capturing within the dynamics of online news. Therefore, a thorough understanding of news outlets' publishing and updating dynamics is necessary. As such, this study first looks at individual articles' publishing patterns and hence asks:

RQ1: How often are new articles being published on news outlets' homepages?

RQ2: When are new articles being published on news outlets' homepages?

Second, following the aforementioned assumptions on journalism's necessity to update articles regularly, follow-up capturing might be necessary. However, article updates may be of various kind and due to a wide variety of reasons. That is, article updates may affect its main content or aim at an article's more prominent characteristics, such as its headline or main image. Moreover, article updates may be due to, among other reasons, corrections (SALTZIS 2012), recent developments (WIDHOLM 2016), or assumptions about how different headlines or images might affect users' attention toward the article (TANDOC 2014; VU 2014). I hence ask:

RQ3: How often are articles' (a) headlines and (b) main images being updated?

RQ4: When are articles' (a) headlines and (b) main images being updated?

Third, this study looks at update patterns of homepages. Prior evidence indicates that an article's position on a news outlet's homepage may change over time, not only because of newly published articles (i.e., the selection of articles), but also because of changing levels of ascribed journalistic importance (i.e., the arrangement of articles; BOCZKOWSKI et al. 2011). Such behavior implies the necessity to also regularly capture news outlets' homepages. Thus, the current study ultimately asks:

RQ5: How often are news outlets' homepages' (a) selections and (b) arrangements being updated?

RQ6: When are news outlets' homepages' (a) selections and (b) arrangements being updated?

## 3. Method

Data originates from an automated online observation of five exemplary news outlets over the course of six months. The objects under observation

are both news outlets' homepages and individual articles. Data collection took place from December 1, 2016, until May 31, 2017, in the outlets' respective time zones.

All outlets were scraped automatically every five minutes. For this project, two NodeJS-based scrapers were developed, which allow for a fast and server-sided web content evaluation using JavaScript. That is, these scrapers act almost like ordinary web browsers in that they fully load and interpret a website in order to inject JavaScript code which in turn can be used to identify and extract single entities on the website using CSS selectors. All tools are publicly available from <https://github.com/MarHai/dissertation>.

The first scraper captured all five news outlets' homepages every five minutes and extracted the main editorial articles. This includes all articles that are selected and arranged editorially. Put differently, sponsored content as well as algorithmically selected content (e.g., the list of most-clicked articles) were ignored. Moreover, sole video content and image galleries were ignored as well.

The second scraper captured all articles that had been identified by the homepage scraper and extracted their headline and image. Importantly, both headline and image were taken from the articles' individual article pages rather than from the news outlets' homepages. For the outlets under observation this procedure ensured that for every article containing an image, the image was actually captured, since not all articles are presented with an image on the news outlets' homepages whereas images are available in almost every case on the articles' individual article pages. Every article was checked for updates for seven days subsequent to its initial publication.

### 3.1 News outlets

The study is part of a larger research project on the dynamics of online news in the realm of ubiquitous quantified audience attention. As such, the five news outlets under investigation have been chosen (a) to represent highly regarded outlets that serve as role models to others (HARDER/SEVENANS/VAN AELST 2017); (b) to map exemplars within a most different systems design (HANITZSCH 2009); and (c) to depict a variety of organizational circumstances, such as political relevance, reach, and economic conditions (JANDURA/KÖSTERS 2017). The sample thus ultimately consists

of four commercial outlets from either Germany or the u.s. with an either regional or national reach: *Berliner Morgenpost* (German regional), *Zeit online* (German national), *Washington Times* (u.s. regional), and *Washington Post* (u.s. national). In addition, *Tagesschau.de* as a public broadcasting service (PBS) outlet from Germany has also been included, constituting also a legally special case as *Tagesschau.de* is prohibited from spending too much online-specific endeavors.

### 3.2 Variables

All data was stored in a relational MySQL database. As such, separate data sets were used for homepages and individual articles.

#### 3.2.1 Homepages

Every news outlets' homepage was scraped regularly, resulting in a multitude of individual homepage captures, each of which consists of a selection of articles in a certain arrangement. While the selection was based on the articles' URLs, the arrangement of articles was operationalized through ordinal numbering of articles on the homepage, counting from top to bottom and left to right.

In order to identify homepage updates, captured articles were assigned a numeric unique identification number. To compare the selection of articles, all depicted articles' identification numbers were ordered ascendingly and concatenated. If the resulting chain of numbers did not match the previously captured homepage's selection of articles, an update on the selection of articles was coded. To compare the sole arrangement of articles, the identification numbers were kept in their respective order for concatenation prior to the comparison.

For instance, if articles 4, 9, 33, 41 and 55 were scraped at time 1 and articles 9, 33, 41, 4 and 55 were scraped at time 2, they were chained without sorting to 4-9-33-41-55 and 9-33-41-4-55; as both strings of numbers are not identical, there has been a change in arrangement. However, after sorting, the chaining results in 4-9-33-41-55 and 4-9-33-41-55, respectively; hence, the articles were only rearranged but the selection of articles remained the same.

### 3.2.2 Articles

Articles were identified through their normalized URLs. As such, the URL's protocol (HTTP or HTTPS) as well as any optional URL parameters (suffixed key-value pairs) were removed. Shortened URLs were resolved into their usual format. While identifying articles through their URL has been applied before (e.g., ZAMITH 2016), it also carries the danger of capturing individual articles multiple times if a news outlet rewrites its articles' URLs in order to become more attractive to search engines. However, prior analyses on systematic URL rewriting did not reveal any such patterns.

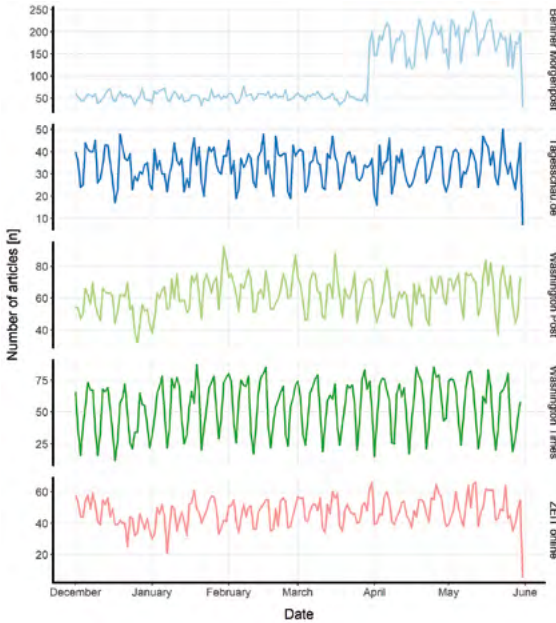
Captured headlines were compared case-sensitively to their prior version in order to identify headline updates. If a captured headline did not match its prior version for the article under investigation, this was coded as an article update. Images were reduced to their MD5 sums before comparing them to their prior version to identify updates. MD5 sums represent a »128-bit »fingerprint« or »message digest« of the input« (RIVEST 1992) and allow for actual file rather than filename comparisons.

## 4. Results

### 4.1 Article publication (RQ<sub>1</sub>, RQ<sub>2</sub>)

Overall, the data collection over the course of six months yielded 53,356 articles. To investigate RQ<sub>1</sub>, the number of article publications per day was analyzed. While absolute numbers of articles per news outlet vary, publishing patterns are very much alike (Fig. 1). News outlets published less on weekends and during prominent holidays (i.e., Christmas break). While these patterns do suggest a differentiated view on weekdays and weekends, a total of 22 percent of all articles was published on either Saturday or Sunday. This clearly calls for an inclusion of weekends. On average, news outlets published  $M = 58.8$  ( $SD = 37.7$ ;  $Mdn = 51.0$ ) articles per day. Importantly, *Berliner Morgenpost* started using live blogs in late March, which regularly updates their URL, most updates of which had been counted as a new article. Also, the *Washington Post* increased their number of published articles per day from the inauguration of Donald Trump (January) onwards.

FIGURE 1  
Share of published articles per day

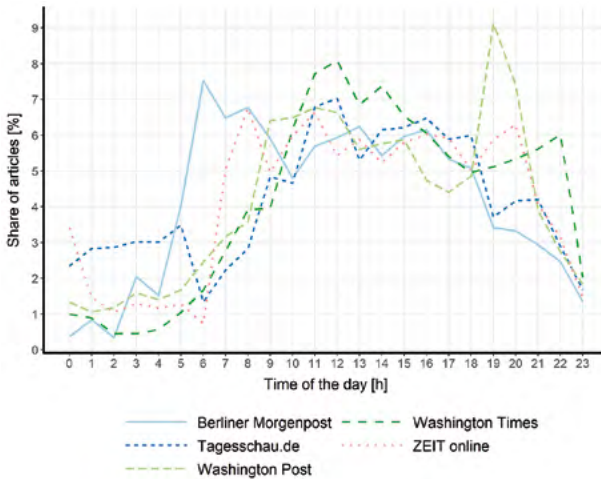


Note: *Berliner Morgenpost* started using live blogs in late March which regularly update their URL, each update of which has been counted as a new article.

For answering RQ<sub>2</sub>, it was analyzed how the publishing times of articles are distributed over the day. Reducing the perspective on published articles over the course of a day, again, similarities become apparent (Fig. 2). While most articles were published between eight o'clock in the morning and ten o'clock in the evening, peaks mostly indicating editorial conferences. Notable exceptions occur for the *Washington Post* (late peaks) and *Berliner Morgenpost* (early peak). Personal conversation with the author suggests that both peaks are based on audience analytics – the *Washington Post* has seen a rise in late-night media use whereas *Berliner Morgenpost* serves plenty of commuters (HAIM 2019). Methodologically, the constantly high level of publications along with an outlet-specific variety of peaks also calls for constant data capture since no single points in time for valid and reliable capture could be identified.

FIGURE 2

## Share of published articles over the course of a day



Note: The time of the day is depicted according to each news outlet's headquarter's time zone.

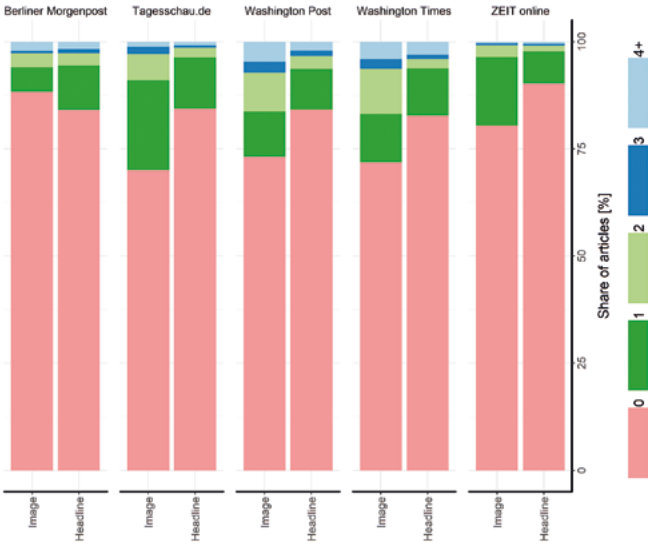
#### 4.2 Article updates (RQ<sub>3</sub>, RQ<sub>4</sub>)

In order to answer RQ<sub>3</sub>, I looked at the share of articles in which the headline and the image(s) were updated. The majority of articles' headlines (82.8-90.2%) and main images (70.1-88.3%) were not updated at all. On a low overall level updating, images were updated more often than headlines, with the exception of *Berliner Morgenpost*. However, most updated articles received only one update (Fig. 3).

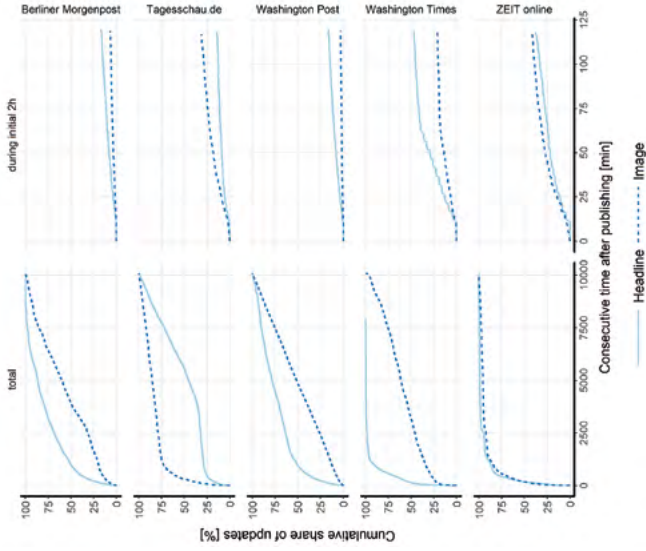
RQ<sub>4</sub> looks at the timing of updating headlines and images. Article updates were applied rather briefly after an article's publication (Fig. 4). That is, two hours after publication, 21.7 percent of all headline updates were already applied whereas only 10.0 percent of all image updates were applied. While the quantitative analysis does not allow for an interpretation of the types of updates, personal conversation as well as individual qualitative in-depth analyses suggest that image updates oftentimes happen due to the availability of better material (e.g., from press agencies); in contrast, headline updates typically depict the result of A/B tests or are a journalist's reaction to audience analytics (HAIM 2019).



**FIGURE 3**  
**Number of updates that individual articles receive**



**FIGURE 4**  
**Intervals between article updates**

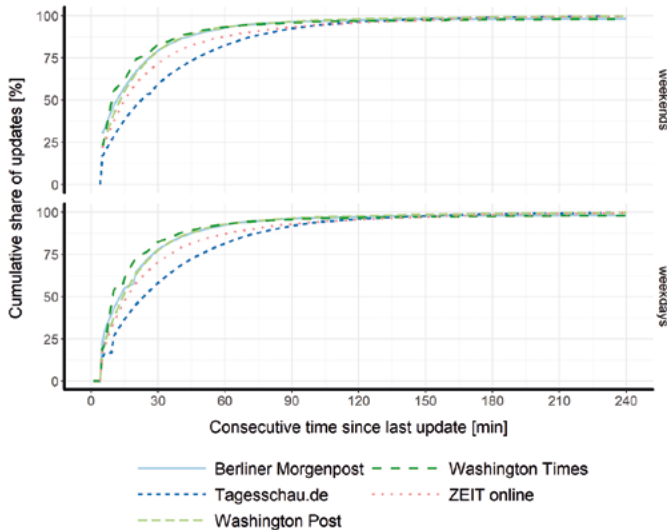


Note: The left column depicts updating intervals of up to one week (10,080 minutes), the right column the updating dynamics during the initial two hours (120 minutes).

### 4.3 Homepage updates (RQ5, RQ6)

Scraping all five news outlets' homepages every five minutes over six months yielded 260,963 homepage captures, each of which consists of a selection of articles arranged in a certain order. Importantly, every news outlets' homepage was updated at least every two hours, be it on weekends, throughout nights, or during prominent holidays (Fig. 5). On average, homepages are updated every 6.1 minutes ( $SD = 33.1$ ;  $Mdn = 5.0$ ). Individual outlets do not differ much.

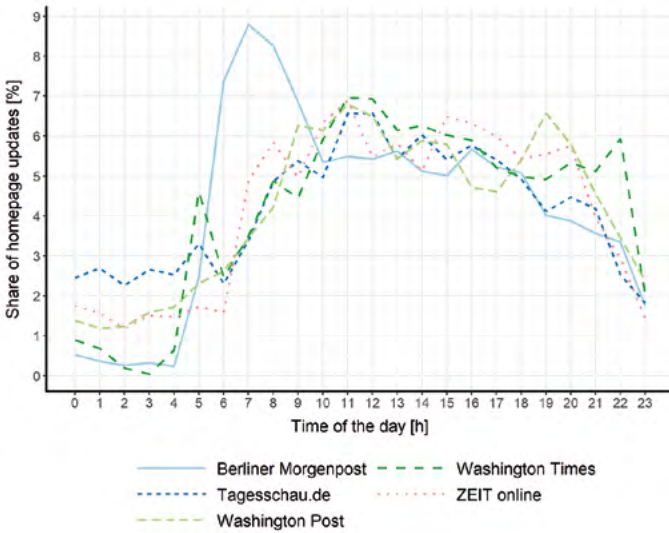
FIGURE 5  
Intervals between homepage updates



Driven by article publications, the news outlets' homepages received updates over the whole course of the day without significant peaks across all outlets (Fig. 6). Exceptionally, as explained in a personal conversation with the author, *Berliner Morgenpost* serves many commuters, which in turn leads to an increase of published articles and homepage updates during morning hours (HAIM 2019). Exclusive arrangement updates played a minor role, with only 6.2 percent of all homepage updates affecting solely the arrangement but not the selection of articles.

In turn, individual published articles stayed on news outlet’s homepages for an average of 9.7 hours ( $SD=17.2$ ;  $Mdn=6.3$ ). This finding differs quite a bit depending on the news outlet. That is, while average articles at the relatively short homepage of *Zeit Online* only survive for about 5.2 hours ( $SD=7.3$ ;  $Mdn=3.4$ ), the *Washington Times* provides much more space and thus allows articles to remain there for 12.6 hours ( $SD=29.9$ ;  $Mdn=7.1$ ). *Tagesschau.de* ( $M=9.2$ ;  $SD=6.7$ ;  $Mdn=8.3$ ), *Berliner Morgenpost* ( $M=9.6$ ;  $SD=8.8$ ;  $Mdn=7.4$ ), and the *Washington Post* ( $M=11.4$ ;  $SD=17.3$ ;  $Mdn=8.1$ ) range in between.

FIGURE 6  
Share of homepage updates over the course of a day



Note: The time of the day is depicted according to each news outlet’s headquarter’s time zone.

## 5. Discussion

The results of this study highlight the underlying assumption that online news follow different dynamics than legacy news. Although this study does not provide a direct comparison of online and legacy news, the current descriptive insights clearly depict constant and short-lived but at

the same time almost static news, which seem to follow audience expectations. Three main methodological conclusions can be drawn for future online-news capturing.

First, online news flow is unceasing. Online news' publishing times are distributed more or less equally over the news day. While data collections for empirical journalism studies are oftentimes guided by anchoring points, such as editorial conferences or prime time, the current findings imply that no such time of the day of journalistic makeover can be identified. That said, less news are published on weekends and during the night. This also gives rise to the admittedly provocative assumption that news are obviously more likely to be published when audiences will most probably read them, rather than when newsworthy events occur.

Second, online news are short-lived news, at least regarding their placement on the front page. With an article's average lifetime on a news outlet's homepage of roughly ten hours and homepage update cycles of two hours tops, data capturing needs to be quick. The very brief intervals between homepage updates even indicate that for an entire capture of homepages the five-minute scraping intervals were just short enough.

Third, online news items are almost static in content. While individual articles are being repositioned on a news outlet's homepage, the headlines and main images of most articles either do not change at all or only once within the first ten minutes after publication. Hence, for headlines and main images, regular scraping of individual articles currently seems less essential than capturing the URLs of the articles on the home page.

Limitations of the current study are threefold. First, the outlets under observation were specifically chosen due to research-economic constraints. Yet, although the present five outlets do not account for any representativeness, all five outlets are considered digitally innovative with varying reach and background. The rather uniform results support this argumentative line of selection. The current results are a workable point of reference. It is commendable, however, to run checks of the outlet-specific patterns in any study seeking to capture online news content. Second, scraping followed a five-minute interval. As such, the current study's methodological approach already limits the findings of this methodological study. This gets especially clear when looking at article and homepage updates, both of which take place within rather short intervals. Again, this limitation is mainly due to research-economic constraints as the research servers' Internet connection required the pretested five-minute interval to load

all observed articles and homepages. Future studies with the aim of fully capturing such dynamics on even shorter notice should thus rely on cloud infrastructure providers with backbone Internet access. Third, web scraping provides reactive access to websites and thus itself affects news outlets under observation. That is, scraping – like any human user – is subject to on-site mechanisms, such as A/B testing or personalization. Thereby, the equal use of all articles can be expected to level any upcoming biases. In addition, while personalization features were not in place on the outlets under observation, A/B testing assigns recurring users (or scrapers) to one version of a headline or image. Hence, during an (invisible) A/B testing phase, scraping is only presented with one of the tested versions. If, after some time, the A/B testing results suggest to change something, an article update was captured. Yet, as this update only originates from A/B testing, it can be considered an update *sui generis*, just as any other update that has not been suggested by A/B testing results.

Taken together, the current study's findings suggest considering online journalism a dynamic field of research that does not fully comply with its legacy alter ego's assumptions. Online news should be captured continuously and on rather short notice, homepages and individual articles deserve separate and differentiated consideration. Homepages are extremely dynamic and should be captured often, whereas articles seem to be relatively static and can usually be captured at a lower rate.

## References

- ANDERSON, C. W.: Between creative and quantified audiences: Web metrics and changing patterns of newswork in local us newsrooms. In: *Journalism*, 12(5), 2011a, pp. 550-566. <https://doi.org/10.1177/1464884911402451>
- ANDERSON, C. W.: Deliberative, agonistic, and algorithmic audiences: Journalism's vision of its public in an age of audience transparency. In: *International Journal of Communication*, 5, 2011b, pp. 529-547
- BAKKER, P.: Aggregation, content farms and huffinization: The rise of low-pay and no-pay journalism. In: *Journalism Practice*, 6(5-6), 2012, pp. 627-637. <https://doi.org/10.1080/17512786.2012.667266>
- BOCZKOWSKI, P. J.; E. MITCHELSTEIN; M. WALTER: Convergence across divergence: Understanding the gap in the online news choices of

- journalists and consumers in Western Europe and Latin America. In: *Communication Research*, 38(3), 2011, pp. 376-396. <https://doi.org/10.1177/0093650210384989>
- BOEING, G.; P. WADDELL: New insights into rental housing markets across the United States: Web scraping and analyzing Craigslist rental listings. In: *Journal of Planning Education and Research*, 37(4), 2017, pp. 457-476. <https://doi.org/10.1177/0739456X16664789>
- BRIGHT, J.; T. NICHOLLS: The life and death of political news measuring the impact of the audience agenda using online data. In: *Social Science Computer Review*, 32(2), 2014, pp. 170-181. <https://doi.org/10.1177/0894439313506845>
- BURGESS, J.; A. BRUNS: Twitter archives and the challenges of »big social data« for media and communication research. In: *M/C Journal*, 15(5), 2012. <http://journal.media-culture.org.au/index.php/mcjournal/article/view/561>
- CARLSON, M.: Order versus access: News search engines and the challenge to traditional journalistic roles. In: *Media, Culture & Society*, 29(6), 2007, pp. 1014-1030. <https://doi.org/10.1177/0163443707084346>
- CHYI, H. I.; A. M. LEE: Online news consumption. In: *Digital Journalism*, 1(2), 2013, pp. 194-211. <https://doi.org/10.1080/21670811.2012.753299>
- DICK, M.: Search engine optimisation in UK news production. In: *Journalism Practice*, 5(4), 2011, pp. 462-477. <https://doi.org/10.1080/17512786.2010.551020>
- FIELDING, R.; J. RESCHKE: RFC 7232, 2014. <https://tools.ietf.org/html/rfc7232>
- FLETCHER, R.; R. K. NIELSEN: Paying for online news. In: *Digital Journalism*, 5(9), 2017, pp. 1173-1191. <https://doi.org/10.1080/21670811.2016.1246373>
- HAIM, M.: *Die Orientierung von Online-Journalismus an seinen Publika: Anforderung, Antizipation, Anspruch*. Wiesbaden [Springer vs] 2019. <https://doi.org/10.1007/978-3-658-25546-6>
- HANITZSCH, T.: Zur Wahrnehmung von Einflüssen im Journalismus. Komparative Befunde aus 17 Ländern. In: *Medien & Kommunikationswissenschaft*, 57(2), 2009, pp. 153-173. <https://doi.org/10.5771/1615-634X-2009-2-153>
- HANUSCH, F.: *Audience impacts on journalistic culture. The role of web analytics in Australian journalism*. Presented at the Future of Journalism, Cardiff, 2015

- HARDER, R. A.; J. SEVENANS; P. VAN AELST: Intermedia agenda setting in the social media age: How traditional players dominate the news agenda in election times. In: *The International Journal of Press/Politics*, 22(3), 2017, pp. 275-293. <https://doi.org/10.1177/1940161217704969>
- HELMOND, A.: The platformization of the web: Making web data platform ready. In: *Social Media + Society*, 1(2), 2015. <https://doi.org/10.1177/2056305115603080>
- JANDURA, O.; R. KÖSTERS: Neue Medienumgebungen, andere Auswahlkriterien? In: *Publizistik*, 62(1), 2017, pp. 25-41. <https://doi.org/10.1007/s11616-017-0320-1>
- KARLSSON, M.; H. SJØVAAG: Content analysis and online news. Epistemologies of analysing the ephemeral Web. In: *Digital Journalism*, 4(1), 2016, pp. 177-192. <https://doi.org/10.1080/21670811.2015.1096619>
- KEYLING, T.; M. HAIM: *Zur Stichprobenqualität von Online-Nachrichtenmedien*. Presented at the 16. Tagung der Fachgruppe ›Methoden der Publizistik und Kommunikationswissenschaft‹ der DGPK, München, October 2014
- KÜNZLER, M.: Befreien Digitalisierung und Automatisierung den Journalismus aus der Zwangsjacke der Kulturökonomie? In: JARREN, O.; C. STEININGER (Eds.): *Journalismus jenseits von Markt und Staat. Institutionentheoretische Ansätze und Konzepte in der Publizistik- und Kommunikationswissenschaft. Festschrift für Marie Luise Kiefer*. Baden-Baden [Nomos] 2016, pp. 219-226
- LEE, A. M.; S. C. LEWIS; M. POWERS: Audience clicks and news placement: A study of time-lagged influence in online journalism. In: *Communication Research*, 41(4), 2014, pp. 505-530. <https://doi.org/10.1177/0093650212467031>
- LOBIGS, F.: Wirtschaftliche Probleme des Journalismus im Internet: Verdrängungsängste und fehlende Erlösquellen. In: NUERNBERGK, C.; C. NEUBERGER (Eds.): *Journalismus im Internet. Profession – Partizipation – Technisierung*. 2nd ed. Wiesbaden [Springer vs] 2018, pp. 295-334
- MARRES, N.; E. WELTEVREDE: Scraping the social? Issues in live social research. In: *Journal of Cultural Economy*, 6(3), 2013, pp. 313-335. <https://doi.org/10.1080/17530350.2013.772070>
- MITCHELSTEIN, E.; P. J. BOCZKOWSKI: Between tradition and change. In: *Journalism*, 10(5), 2009, pp. 562-586. <https://doi.org/10.1177/1464884909106533>

- NEUBERGER, C.; C. NUERNBERGK: Verdatete Selbstbeschreibung der Gesellschaft – Über den Umgang des Journalismus mit Big Data und Algorithmen. In: SÜSSENGUTH, F. (Ed.): *Die Gesellschaft der Daten – Über die digitale Transformation der sozialen Ordnung*. Bielefeld [Transcript] 2015, pp. 199-224
- NEWMAN, N.; R. FLETCHER; A. KALOGEROPOULOS; D. A. L. LEVY; R. K. NIELSEN: *Digital news report 2017*. Oxford, 2017. [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web\\_o.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web_o.pdf)
- NIELSEN, R. K.; S. A. GANTER: Dealing with digital intermediaries: A case study of the relations between publishers and platforms. In: *New Media & Society*, 20(4), 2018, pp. 1600-1617. <https://doi.org/10.1177/1461444817701318>
- NIELSEN, R. K.; K. C. SCHRØDER: The relative importance of social media for accessing, finding, and engaging with news. In: *Digital Journalism*, 2(4), 2014, pp. 472-489. <https://doi.org/10.1080/21670811.2013.872420>
- PICARD, R. G.: Twilight or new dawn of journalism? In: *Journalism Studies*, 15(5), 2014, pp. 500-510. <https://doi.org/10.1080/1461670X.2014.895530>
- RIVEST, R. L.: RFC 1321, 1992. <https://tools.ietf.org/html/rfc1321>
- RÖSSLER, P.; L. HAUTZER; M. LÜNICH: Mediennutzung im Zeitalter von Social Navigation. In: LOOSEN, W.; M. DOHLE (Eds.): *Journalismus und (sein) Publikum*. Wiesbaden [Springer vs] 2014, pp. 91-112
- SALTZIS, K.: Breaking news online. How news stories are updated and maintained around-the-clock. In: *Journalism Practice*, 6(5-6), 2012, pp. 702-710. <https://doi.org/10.1080/17512786.2012.667274>
- TANDOC, E. C. J.: Journalism is twerking? How web analytics is changing the process of gatekeeping. In: *New Media & Society*, 16(4), 2014, pp. 559-575. <https://doi.org/10.1177/1461444814530541>
- TANDOC, E. C. J.; P. R. FERRUCCI: Giving in or giving up: What makes journalists use audience feedback in their news work? In: *Computers in Human Behavior*, 68, 2017, pp. 149-156. <https://doi.org/10.1016/j.chb.2016.11.027>
- TANDOC, E. C. J.; M. JENNER: Analysing analytics: How journalists' role conceptions influence how they use audience metrics. In: *Journal of Applied Journalism & Media Studies*, 5(3), 2016, pp. 423-439. [https://doi.org/10.1386/ajms.5.3.423\\_1](https://doi.org/10.1386/ajms.5.3.423_1)



- VU, H. T.: The online audience as gatekeeper: The influence of reader metrics on news editorial selection. In: *Journalism*, 15(8), 2014, pp. 1094-1110. <https://doi.org/10.1177/1464884913504259>
- WEBSTER, J. G.: *The marketplace of attention: How audiences take shape in a digital age*. Cambridge/MA [MIT] 2014
- WIDHOLM, A.: Tracing online news in motion. Time and duration in the study of liquid journalism. In: *Digital Journalism*, 4(1), 2016, pp. 24-40. <https://doi.org/10.1080/21670811.2015.1096611>
- ZAMITH, R.: On metrics-driven homepages. In: *Journalism Studies*, 19(8), 2016, pp. 1116-1137. <https://doi.org/10.1080/1461670X.2016.1262215>
- ZAMITH, R.: Quantified audiences in news production. A synthesis and research agenda. In: *Digital Journalism*, 6(4), 2018, pp. 418-435. <https://doi.org/10.1080/21670811.2018.1444999>

MARKO BACHL

## An Evaluation of Retrospective Facebook Content Collection from Political Facebook Pages

*Abstract:* The content on Facebook pages is obviously subject to changes. However, there is often a considerable time lag between the sampling period and the collection date in studies on public Facebook communication. The present evaluation study aims to assess whether – and if so, how – problematic this common practice is in terms of unavailable content. I investigated whether a sample of content items, which were posted to 408 German political Facebook pages in March 2017, could be retrospectively collected in June 2017, September 2017, and March 2018. 27 percent of the 132,068 test items were no longer accessible after 12 months. The deletion of complete pages was rare, but some prominent pages were not available retrospectively. Posts by the pages themselves were far more likely to be available compared to content, which was posted by private accounts to the pages. There were only small differences between the political parties, but there was substantial variation between the distinct pages. I conclude with a discussion of implications for single research projects (applying timely data collection whenever possible) and for communication research and related disciplines (considering institutionalized data collection efforts).

*Keywords:* Dynamic content, Data collection, Content analysis, Archive, Facebook

## 1. Introduction

Today content from the Internet is an established research object in communication research (SKALSKI/NEUENDORF/CAJIGAS 2017). Studies on user-generated online content also have become increasingly popular (NAAB/SEHL 2017). One important challenge in online content analysis is the initial content collection, because Internet content is inherently dynamic: It can change over time or even disappear completely. McMillan (2000) aptly titled her early review of online content analyses *The microscope and the moving target* (ibid.: 80). 17 years later, Naab and Sehl (2017) still concluded that »several of the challenges posed by the continuously changing nature of the content are not fully met« (ibid.: 1256). In this report, I look at the retrospective collectability of political content on one relevant online platform. Facebook is the online social network site with the greatest reach in most countries and it is considered an important platform for political communication. Scholars investigated, for example, how political parties communicate on Facebook (ARZHEIMER 2015; MAGIN et al. 2016; STIER et al. 2018; STIER et al. 2017), how users interact with posts from media organizations' pages (SCHMIDT et al. 2017), or which (media) sources were referred to by parties and their followers on political pages (BACHL 2018).

Lacking a systematic review of the divergences between the sampling period and time of data collection in the literature, I will provide a comparison in some selected studies. This serves to illustrate the practical relevance of the problem: How much time passes between the original publication of the Facebook content and its collection? In an own study, in December 2016 (BACHL 2018), I collected posts, comments, and replies, which were posted on AfD-related Facebook pages during the year 2016. Arzheimer (2015) was interested in all posts by the party and the users on the main AfD page since its inception in March 2013. The posts were retrieved in July 2014. The sampling period of Stier et al. (2017) was December 2014 to August 2016, all content was collected at the end of the sampling period. Schmidt et al. (2017) proceeded similarly for their sampling period of six years (January 2010 to December 2015). Magin et al. (2016) investigated content from political party pages in the four weeks before the 2013 German and Austrian national elections. The content collection took place in November 2015. The first »Social Media Monitoring« project of GESIS relied on a purely retrospective collection in November 2013 for the sample period from January 2009 to October 2013 (KACZMIREK et al. 2014). The sec-

ond edition of the project for the Bundestag election 2017 again initially planned to use data from a retrospective collection. Luckily, they conducted an additional collection in which the data for each post was downloaded eight days after its creation for a shorter sampling period (August 15th to September 30th, 2017). A comparison of both data sets revealed that the retrospectively collected data set missed a non-negligible amount of content compared to the rolling collection. The authors consequently decided to publish the latter data set (STIER et al. 2018).

Retrospectively collected content may be incomplete compared to the population of content, which was present during the sampling period. Both the administrators of the public pages and the individual users can decide to modify or delete the content at any time and for many reasons. The complete deletion of prominent pages is the clearest example. Hannelore Kraft, the former minister-president of North Rhine-Westphalia, and Frank-Walter Steinmeier, the current Bundespräsident of Germany, recently deleted their public profiles.<sup>1</sup> Their decisions made it impossible to retrospectively collect all relevant content for a comprehensive study on the use of Facebook during the 2017 election campaign in North Rhine-Westphalia or for a study about changes in the Bundespräsident's online communication over the course of his career. Beyond such obvious problems, the deletion of many single content items across pages and users can similarly interfere with the validity of retrospectively collected data. The consequences may be particularly problematic if the missingness occurred not at random. The assumption, that specific content (e.g., hate speech, arguments against a party on its own page, communication mishaps) is more likely to be deleted, seems highly plausible. The results would then not only be less precise, but also biased. With this in mind, I attempt to answer two straightforward research questions:

1. Which proportion of the content items on selected political Facebook pages can be retrospectively collected after three to four, six to seven, and eleven to twelve months?
2. How do the odds to retrospectively collect a content item differ between content types, parties, and pages?

The first question simply wants to know which shares of the original content would still have been available if the collection was conducted later

1 Steinmeier now uses a new profile: <https://www.facebook.com/Bundespraesident.Steinmeier/>

in time. The exact dates of the three test collections are not especially meaningful, as they were chosen for pragmatic reasons (conference submission, conference presentation, writing of this manuscript). They are, however, sufficiently typical to provide useful points of reference for comparison with other studies. The second question differentiates the messages based on their content types and based on the pages and the parties, to which the pages belong. With regard to content types, I distinguish posts by the pages themselves from posts by individual accounts. Comments and replies were treated as separate content types. This distinction provides points of reference for studies which focus on the messages of political communicators (posts from pages) and for studies which are interested in user-generated content (user posts, comments, replies) on political Facebook pages. In addition, the hierarchical structure of the content items on Facebook must be considered when thinking about patterns of content deletion. Each post is published on one page, each comment refers to one post, and each reply refers to one comment. If, for example, a post is deleted, all comments to this post and their replies also vanish. It can therefore be expected that content items, which are lower in the hierarchy, are more likely to be missing because of the deletion of higher-order items. The comparison of parties and pages is indicative to gauge to which extent studies on certain political parties or studies which focus on specific pages are affected by retrospectively non-available content items.

It should be noted that the present study does not try to explain *why* some content items cannot be retrieved after a given time period. The reasons why any given item is missing is practically impossible to infer from publicly available data. An unavailable content item could have been deleted by the creator, by the page's administrators, or on behalf of Facebook. Similarly, a user with all of his or her content items could have been banned from one page or from the whole platform. He or she could also have decided to change the privacy settings, or delete or deactivate his or her whole account. The simple goal of the present study is to provide estimates of the magnitudes of losses which have to be expected when collecting content from German political Facebook pages retrospectively. Any interpretations beyond this descriptive aim are only informed speculation. A survey of page administrators or Facebook users would be necessary for a thorough investigation of the underlying causes.

## 2. Methods

The present analysis is based on a larger data collection effort in the run-up to the 2017 Bundestag election. All posts, comments, and replies on 422 political Facebook pages were collected every 6 hours since late February 2017 via the Facebook Graph API.<sup>2</sup> The pages were sampled to represent the most important pages of the major political parties in Germany and, additionally, of the PEGIDA movement.<sup>3</sup> Pages of the AfD were oversampled because of a research interest in populist-right online communication. For this evaluation study, I selected a stratified subsample of all content items which were published in March 2017. A maximum of 25 posts a) from each page itself and b) from users to each page were randomly selected. Then a maximum of 50 comments to each post and, in turn, a maximum of 50 replies to each comment were randomly selected. The procedure resulted in a sample of 408 pages (the other pages did not post or receive any content in March 2017), 8,032 posts from pages (i.e., by an administrator of the page in the name of the page), 1,013 posts from users, 69,182 comments, and 53,841 replies (overall 132,068 content items). Each content item on Facebook has a unique alphanumeric identifier, which was used to check whether it could still be retrieved in three test runs on June 12th, 2017, September 13th, 2017, and March 14th, 2018.

Two shortcomings of the procedure should be noted right away: First, the original data collection already missed content which was posted *and* deleted in the 6-hour windows between two collection runs. This is relevant for research questions on content which is presumably deleted quickly by the pages' moderators, such as hate speech and trolling. It may also preclude items which were published by error or immediately received unfavorable responses and were therefore quickly deleted. Research interests which inherently include the risk of content deletion will most likely require an even closer tracking of new content. Second, I did not systematically evaluate technical inconsistencies of the data retrieval procedure. In each evaluation period, all content items were checked several times and the results differed very slightly in each run. About 0.5 percent to 1.5 percent of all items had divergent results in one run compared to the next one about 10 minutes

2 <https://developers.facebook.com/docs/graph-api>

3 For the sake of simplicity, I refer to the organizations behind the Facebook pages as ›parties‹. This also includes the PEGIDA movement and its pages, although PEGIDA is of course not a political party.

later. A cursory review of the inconsistencies did not suggest any systematic losses, but the technical reliability of the procedure surely remains a topic that deserves further attention. I decided to use the results from the first run in each test period for the subsequent analyses. About 1 percent of the items in the study showed highly implausible results which are most likely due to (technical) errors in data retrieval.<sup>4</sup> 186 content items were found only in the third test but neither in the first nor in the second. 163 items were found only in the second test. 636 items were not found in the first but in both later tests. 545 items were missing only in the second test.

A public repository on the Open Science Framework documents some accompanying materials to this article: <https://osf.io/an9u5/>. It contains a minimal data set, which is sufficient to reproduce the statistical analyses presented herein. However, I cannot make the full data set publicly available because of privacy and data protection concerns and the Facebook terms and services. The variable `id` in the public data set can be used to retrieve the content items which were in the initial sample for this evaluation study. A continuation of the study over a longer period is made possible in principle with this information. The OSF page also contains an *R* (R CORE TEAM 2017) script, which documents data collection and preparation. *RFacebook* (BARBERA et al. 2017) was used for data collection. Data management and preparation were mainly conducted with *dplyr* (WICKHAM/FRANCOIS 2016) and *tidyr* (WICKHAM 2017). The original manuscript was typeset with *papaja* (AUST/BARTH 2017) and contains all code necessary to reproduce the statistical analyses. They were conducted with *lme4* (BATES et al. 2015) and *arm* (GELMAN/SU 2016) and the plots were created with *ggplot2* (WICKHAM 2009), *ggstance* (HENRY/WICKHAM/CHANG 2016), and *egg* (AUGUIE 2017).

### 3. Results

Before turning to the research questions, we take a look at the pages whose complete March 2017 content was no longer collectable after one year (Tab. 1). This concerned 11 of the 408 pages. The deletion of the content on the four

4 Another explanation would be that the items come from users who deactivated and then reactivated their Facebook accounts. While a small amount of divergences could be caused by such user actions, I doubt that they occur that frequently.

pages of individual politicians was presumably motivated by major changes in their careers. Hannelore Kraft resigned from all public and party offices after the 2017 NRW election. Frank-Walter Steinmeier was elected Bundespräsident in March 2017. Michael Fuchs did not stand for re-election and left the Bundestag after the 2017 election. All three pages were no longer publicly available. The page of Alexander Dobrindt seems to have been reactivated after March 2017. No own content was published during the sampling period of the evaluation study, only some comments by other users were found. They were all removed during the September 2017 and the March 2018 evaluation runs, maybe in an attempt to >start fresh< with new communication activities after the transition from federal minister to head of the CSU parliamentary group. All four examples suggest (albeit with a tiny sample size of politicians) that the deactivation or a major overhaul of the public Facebook pages of politicians shortly after important career changes can pose a threat to the completeness of retrospectively collected data. The deletion of whole pages did not matter much quantitatively, but there is a real chance that any crucial page's content will no longer be available.

TABLE 1

### Pages without collectable content items after March 2017

Page	Identified in
Frank-Walter Steinmeier	June 2017
Hannelore Kraft	June 2017
BadenWürttemberg freiheitlich-patriotisch-traditionsbewusst*	September 2017
Junge Alternative Lüneburg	September 2017
Alexander Dobrindt, MdB **	March 2018
Alternative für Deutschland – Ortsgruppe Rostock	March 2018
Alternative für Deutschland – Unterstützer Kreisverband Ennepe-Ruhr	March 2018
Alternative für Deutschland AfD – Ostholstein	March 2018
Die Linke. SDS Berlin	March 2018
Michael Fuchs	March 2018
Pegida NRW	March 2018

Note. \* All content items were deleted, page still exists; \*\* Only userposts to page in March 2017. All content items were deleted, page still exists.

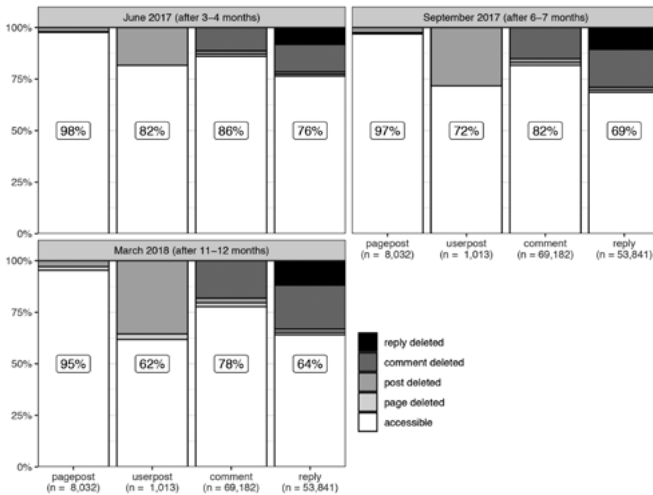
The disappearance of most of the pages from political organizations concerned smaller, local chapters. They may be less important for many typical



research questions. However, the inability to retrospectively collect the content on the PEGIDA-related ›BadenWürttemberg freiheitlich-patriotisch-traditionsbewusst‹ page uncovered a – at least to me – surprising pattern of deletions. The page *regularly* removes all older posts every few months. The prevalence of this phenomenon beyond the present sample and the underlying motivation have to remain unclear for now. However, it is obvious how such a systematic revision of a page’s publication history may invalidate the data from a retrospective collection even if the page itself remains available.

Research question 1 asked about the shares of content items which could be collected retrospectively. Overall, 17 percent of the 132,068 test items were no longer accessible after three to four months, 23 percent after six to seven months, and 27 percent after eleven to twelve months. A more detailed picture of the temporal volatility of Facebook content has to take the path dependencies due to the hierarchical organization of the content items into account.

FIGURE 1  
Shares of accessible and deleted content items

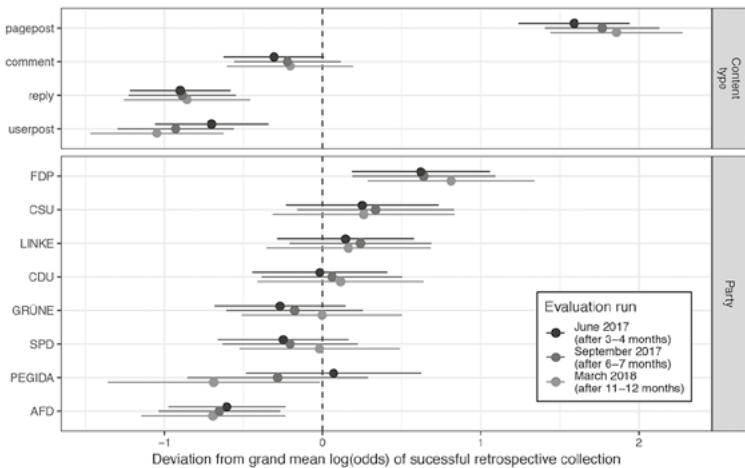


Notes: The percentages report the shares of items in a category, which could be accessed in the evaluation run.

Figure 1 presents these dependencies organized by content type. The white areas show the shares of items in a category, which could be accessed in the evaluation run. The areas in different shades of gray indicate the shares

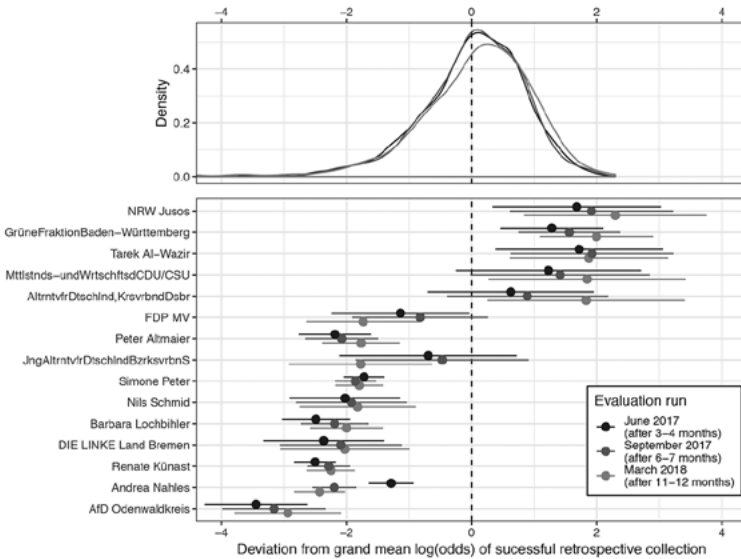
of items categorized by the hierarchically highest item which was also missing. The first column in each panel clearly indicates that posts by the pages themselves were most stable, with only 5 percent of inaccessible posts even after about one year. The political organizations in this sample mostly seem to have refrained from removing their own content. User-generated content was far more volatile. This especially holds for posts by users to the topmost level of a page. Almost one fifth of them could no longer be collected after three to four months. About one year after publication, over one third was unavailable. Comments and replies were also quite volatile, with one fifth or one third deletions after one year, respectively. The inaccessibility of higher-order content elements, mainly of comments, was an important reason for the disappearance of replies. About 20 percent of all replies could no longer be accessed after one year because the respective comment had been removed.

FIGURE 2  
Differences in retrospective accessibility between content types and parties



Notes: The figure shows point estimates and intervals of  $\pm 2$  standard errors. Zero is the grand mean log(odds) of successful retrospective collection in each evaluation run (June 2017: 2.76 or, converted to a probability, .94; September 2017: 2.31 or .91; March 2018: 1.88 or .87; note that the grand means are not identical with the simple shares from above). Positive values indicate above-average odds, negative values below-average odds. The model includes content type, parties, and pages. The results for pages are shown in Figure 3.  $n_{\text{observations}} = 121,864$ ;  $n_{\text{pages}} = 408$ ;  $n_{\text{parties}} = 8$ ;  $n_{\text{content types}} = 4$ .

FIGURE 3  
Differences in retrospective accessibility between pages



Notes: The upper panel shows the density of the distributions of the point estimates of all pages. The lower panel shows the point estimates and intervals of  $\pm 2$  standard errors for the 15 most extreme pages (with the exception of the pages in Table 1) in March 2018. The names of some pages were abbreviated: Mttlstnds-undWirtschftsdCDU/CSU: Mittelstands- und Wirtschaftsvereinigung der CDU / CSU; AltrntvfrDtschlnD,KrsvrbdDabr: Alternative für Deutschland, Kreisverband Duisburg; JngAltrntvfrDtschlnDBezkrsvrbns: Junge Alternative für Deutschland Bezirksverband Südbaden; Zero is the grand mean log(odds) of successful retrospective collection in each evaluation run (June 2017: 2.76 or, converted to a probability, .94; September 2017: 2.31 or .91; March 2018: 1.88 or .87; note that the grand means are not identical with the simple shares from above). Positive values indicate above-average odds, negative values below-average odds. The model includes content type, parties, and pages. The results for content types and parties are shown in Figure 2.  $n_{\text{observations}} = 121,864$ ;  $n_{\text{pages}} = 408$ ;  $n_{\text{parties}} = 8$ ;  $n_{\text{content types}} = 4$ .

Research question 2 asked about differences between content types, parties, and pages. Figure 2 summarizes the results for content types and parties. The first panel confirms the descriptive results from Figure 1. Posts by the pages themselves were far more stable than user-generated content. Userposts and replies had the lowest odds of being collected retrospectively. The differences between the parties and PEGIDA were rather small compared to the differences between content types or between pages. Content items on FDP-related pages could be retrieved retrospectively somewhat

more often, the content on AfD-related pages was somewhat more likely to vanish. The pattern of differences did not change substantially over the course of the three evaluation periods. Only the deletion rates on PEGIDA pages have grown slightly stronger compared to the overall trend.

Figure 3 presents the differences between the pages. The upper panel shows how the odds to retrospectively collect a content item vary between all pages. The considerable variation is comparable in magnitude to the differences between the content types. The distribution is skewed somewhat to the left. Overall, the rates of successful retrospective collection were high on most pages, but some pages lost substantial parts of their content. The lower panel presents the 15 pages with the strongest absolute deviations from the grand mean of successful retrospective collection in the last evaluation run after one year (with exception of the pages in Table 1 which had no retrospectively collectable content). The highly diverse set of pages mirrors the limited differences between the parties (Fig. 2, lower panel). Both the positive and negative tails of the distribution include pages from different parties as well as pages from prominent politicians and small, local party chapters. There seems to be no obvious pattern to distinguish between categories of pages with highly stable or more volatile content. The patterns of differences were quite similar for all three evaluation runs. Among the notable pages in the bottom panel, only the accessibility of content items on Andrea Nahles' page showed a substantial change between the June and the September 2017 evaluation runs compared to the overall trend.

#### 4. Discussion

In sum, the evaluation study detected substantial volatility in the content which was posted on a large set of German political Facebook pages in March 2017. The deletion of complete pages did not matter much quantitatively. However, the absence of some pages (Hannelore Kraft, Frank-Walter Steinmeier) from the data set would have had devastating consequences for common research projects in political communication. Posts, which were published by the pages themselves, were far more likely to be available retrospectively than user-generated content (userposts, comments, replies). There were only small differences between the political parties, but there was substantial variation between the distinct pages. None of the findings are overly surprising. It is, after all, well-known that content on the Inter-

net and especially user-generated content are subject to changes. However, the common knowledge seems to have only limited consequences for the empirical practice of Facebook studies. The empirical evidence presented in this report will hopefully inspire critical reflection and methodological improvements toward less ad hoc data collection practices.

The results have implications for the interpretation of studies which employed retrospective content collection, for researchers who plan to conduct studies on Facebook content, and for communication research and related fields which are interested in social media content. Past Facebook studies, like the ones listed in the introduction, obviously cannot change their data collection procedure. Similarly, researchers, who wish to conduct a study on public Facebook communication during a past sampling period, have no viable alternative to collecting the content retrospectively. In these instances, the implications of the present evaluation study depend on the research interest. If a study is primarily interested in the content, which was published by the pages themselves (i.e., their posts), retrospective content collection seems to be an acceptable approach. However, the prominent examples of completely unavailable pages reminded us that not every research question on parties' and politicians' Facebook communication activities can be answered with retrospectively collected data. Plausibility checks of the data sets are certainly in order. Checking for deleted data in the absence of a full data set for comparison is, of course, mostly educated guesswork. For political pages, contrasting the collected data with a timeline of politically relevant events and with a list of known relevant politicians is one possible starting point. If the Facebook page of a major politician is missing from the sample or if party or politician pages do not contain posts during important election campaigns, then skepticism is certainly in order. The untested assumption of complete data is not prudent, even for posts by the pages.

Studies on party-user interactions or user-generated content, which applied retrospective content collection, should certainly be taken with a grain of salt. The reported results strongly suggest that substantial shares of userposts, comments, and replies on political Facebook pages vanish from the platform. The present study did not try to identify the (likely) causes for the deletion of user-generated content. It can only be assumed that a combination of active moderation by the pages' administrators and actions taken by the users were responsible for the lower odds of retrospective availability. Whatever the reasons might be – the results clearly show

that the user-generated content which could be retrieved after one year is an incomplete and most likely biased representation of the original content.

The most straightforward consequence for future research projects is to use more timely and repeated data collection procedures. It is my impression that researchers routinely apply the data collection approaches from traditional content analysis to the »moving target« (MCMILLAN 2000: 80) of online and social media communication. Relatively easy access via APIs and user-friendly tools like *Facepager* (JÜNGER/KEYLING 2018) convey the look and feel of traditional media archives such as Nexis or WISO.<sup>5</sup> This has lured some researchers into collecting data from Facebook like from curated archives, where retrospective data collection is the inherent purpose of the service. If one wants to pick an analogy from established research methods, it makes much more sense to think about the data collection plan of a Facebook study like the one of a longitudinal survey. Obviously, nobody would believe that asking about respondents' attitudes from last year is a sensible research design decision. If one »asks« a political Facebook page via the API about its content from a year ago, the answer is likely to be similarly unreliable. A timely data collection procedure does not pose a major challenge for many typical research projects in political communication. The most important pages are often known in advance, as are the sampling periods of interest, which are usually defined by election dates or other long-term expectable events. Timely data collection should be considered the standard approach if these conditions hold.<sup>6</sup>

Finally, the uncertainty about the retrospective accessibility of public Facebook communication highlights the need to think about institutional initiatives to reliably archive online and social media content of public interest and to make it available for non-commercial research purposes. One obvious solution would be a closer cooperation with the platforms themselves, in this case, Facebook. The recent revelations surrounding the data collection practices of Cambridge Analytica, which were at least in part

5 <https://www.nexis.com> and <https://www.wiso-net.de>

6 Another practical reason to favor timely over retrospective data collection procedures in Facebook studies is a bug in the Facebook Graph API, which to the best of my knowledge is still unresolved at the time of this writing: <https://developers.facebook.com/bugs/1838195226492053/>. In short, the collection of past posts from a page will sometimes be incomplete. The bug is mostly relevant for pages with high activity. Although this issue is not related to the empirical results presented in this report, it should be mentioned here, because it makes large-scale retrospective collections at least infeasible, if not impossible.

made possible by a researcher who passed on data collected for scientific purposes,<sup>7</sup> have led to mixed developments in this regard. On the one hand, recent updates of the Graph API have already seriously limited the ability of independent researchers to collect publicly accessible data from Facebook.<sup>8</sup> On the other hand, Facebook agreed to a research partnership, in which previously unattainable data sets from within Facebook can be analyzed by independent scholars.<sup>9</sup> It remains to be seen whether the data sets from this initiative will be sufficiently robust to changes in the original content. The description of the first data set mentions that the data from deleted users cannot not be included – even if the data set is only available in aggregated form and does not contain information about individual users. Another possibility would be a coordinated data collection effort – if the scope of public data access remains sufficient for such an endeavor. The ›Social Media Monitoring‹ initiative by GESIS (KACZMIREK et al. 2014; STIER et al. 2018) is an interesting step in this direction. Yet the project is also confronted with many legal and ethical challenges beyond the technical implementation, ranging from data and privacy protection to copyright and the platforms' terms and conditions. Some aggregated results can be retrieved via <http://mediamonitoring.gesis.org/>, but the distribution of the raw or at least more detailed data is not possible at the moment.

It is beyond the scope of the present evaluation study to outline an institutional solution to the challenges of changing online and social media content. Yet the empirical results make it clear that such initiatives would certainly be appreciated by many researchers. At present, the best practice is an own timely data collection effort and I strongly encourage its application whenever possible. But not every research group has the resources to implement and operate an adequate infrastructure. Because of this, and also because we do not always know in advance which content we would like to collect, some Facebook (and more generally, social media) studies will continue to rely on retrospective content collection. The present evaluation study has shown that this approach may be acceptable in some instances (especially if the focus is on professional communicators' publi-

7 See, for example, the extensive coverage by *The Guardian*: <https://www.theguardian.com/news/series/cambridge-analytica-files>

8 <https://developers.facebook.com/blog/post/2018/04/04/facebook-api-platform-product-changes/>

9 <https://socialscience.one/>

cation activities) but will probably lead to substantially and systematically incomplete data sets. At the very least, we should take this into account when evaluating the results of our studies.

## References

- ARZHEIMER, K.: The AfD: Finally a successful right-wing populist eurosceptic party for Germany? In: *West European Politics*, 38(3), 2015, pp. 535-556. doi:10.1080/01402382.2015.1004230
- AUGUIE, B.: *Egg: Extensions for >ggplot2<, to align plots, and set panel sizes*. 2017. <https://CRAN.R-project.org/package=egg>
- AUST, F.; M. BARTH: *papaja: Create APA manuscripts with R Markdown*. 2017. <https://github.com/crsh/papaja>
- BACHL, M.: (Alternative) media sources in AfD-centered Facebook discussions. In: *Studies in Communication and Media*, 7(2), 2018, pp. 128-142. doi:10.5771/2192-4007-2018-2-128
- BARBERA, P.; M. PICCIRILLI; A. GEISLER; W. VAN ATTEVELDT: *Rfacebook: Access to facebook api via R*. 2017. <https://CRAN.R-project.org/package=Rfacebook>
- BATES, D.; M. MÄCHLER; B. BOLKER; S. WALKER: Fitting linear mixed-effects models using lme4. In: *Journal of Statistical Software*, 67(1), 2015, pp. 1-48. doi:10.18637/jss.v067.i01
- GELMAN, A.; Y.-S. SU: *Arm: Data analysis using regression and multi-level/hierarchical models*. 2016. <https://CRAN.R-project.org/package=arm>
- HENRY, L.; H. WICKHAM; W. CHANG: *Ggstance: Horizontal >ggplot2< components*. 2016. <https://CRAN.R-project.org/package=ggstance>
- JÜNGER, J.; T. KEYLING: *Facepager. An application for generic data retrieval through APIs*. 2018. <https://github.com/strohne/Facepager/>
- KACZMIREK, L.; P. MAYR; R. VATRAPU; A. BLEIER; M. S. BLUMENBERG; T. GUMMER; A. HUSSEIN; K. KINDER-KURLANDA; K MANSHAEI; M. THAMM; K. WELLER; A. WENZ; C. WOLF: *Social media monitoring of the campaigns for the 2013 German Bundestag elections on Facebook and Twitter*. GESIS-Working Papers. 2014. <http://nbn-resolving.de/urn:nbn:de:0168-ss0ar-381955>
- MAGIN, M.; N. PODSCHUWEIT; J. HASSLER; U. RUSSMANN: Campaigning in the fourth age of political communication. A multi-method study on the use of Facebook by German and Austrian parties in the 2013



- national election campaigns. In: *Information, Communication & Society*, 20(11), 2016, pp. 1698-1719. doi:10.1080/1369118X.2016.1254269
- MCMILLAN, S. J.: The microscope and the moving target: The challenge of applying content analysis to the World Wide Web. In: *Journalism & Mass Communication Quarterly*, 77(1), 2000, pp. 80-98. doi:10.1177/107769900007700107
- NAAB, T. K.; A. SEHL: Studies of user-generated content: A systematic review. In: *Journalism*, 18(10), 2017, pp. 1256-1273. doi:10.1177/1464884916673557
- R CORE TEAM: *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. 2017. <https://www.R-project.org/>
- SCHMIDT, A. L.; F. ZOLLO; M. DEL VICARIO; A. BESSI; A. SCALA; G. CALDARELLI; H. E. STANLEY; W. QUATTROCIOCCHI: Anatomy of news consumption on Facebook. In: *Proceedings of the National Academy of Sciences*, 114(12), 2017, pp. 3035-3039. doi:10.1073/pnas.1617052114
- SKALSKI, P. D.; K. A. NEUENDORF; J. A. CAJIGAS: Content analysis in the interactive media age. In: K. A. NEUENDORF (Ed.): *The content analysis guidebook* (2nd ed.). Los Angeles [Sage] 2017, pp. 201-242
- STIER, S.; A. BLEIER; M. BONART; F. MÖRSHEIM; M. BOHLOULI; M. NIZHEGORODOV; L. POSCH; J. MAIER; T. ROTHMUND; S. STAAB: *Systematically monitoring social media: The case of the German federal election 2017*. GESIS-Working Papers. 2018. <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-56149-4>
- STIER, S.; L. POSCH; A. BLEIER; M. STROHMAIER: When populists become popular: comparing Facebook use by the right-wing movement Pegida and German political parties. In: *Information, Communication & Society*, 20(9), 2017, pp. 1365-1388. doi:10.1080/1369118X.2017.1328519
- WICKHAM, H.: *Ggplot2: Elegant graphics for data analysis*. New York [Springer] 2009. <http://ggplot2.org>
- WICKHAM, H.: *Tidyr: Easily tidy data with >spread()< and >gather()< functions*. 2017. <https://CRAN.R-project.org/package=tidyr>
- WICKHAM, H.; R. FRANCOIS: *Dplyr: A grammar of data manipulation*. 2016. <https://CRAN.R-project.org/package=dplyr>

PABLO JOST / FRANK MANGOLD / PASCAL JÜRGENS

## Herausforderungen und Potenziale der Nutzung digitaler Beobachtungsdaten auf Facebook – eine longitudinale Perspektive

*Abstract:* Bei der Interaktion mit Inhalten auf sozialen Netzwerken wie Facebook hinterlassen Nutzer digitale Spurendaten (bspw. Likes). Diese können in Ihrer aggregierten Form Aufschluss über Hinwendung oder gar Bewertung von Inhalten geben und damit die in der Medienwirkungsforschung etablierten (experimentellen) Befragungen um nicht-reaktive Online-Analysen ergänzen. Bisherige Studien beschränken sich allerdings auf die Analyse vorwiegend positiv konnotierter Popularity-Cues und erfassen diese erst retrospektiv im Aggregat. Unklar bleibt, über welchen Zeitraum sich die Interaktionen erstrecken und welche Dynamiken – angetrieben durch algorithmische Rahmenbedingungen sowie soziale Prozesse – deren Zustandekommen beeinflussen. Der Beitrag sucht diese Lücke mittels hochfrequenter Erhebung von Reactions zu Facebook-Posts deutscher Nachrichtenmedien zu schließen. Die Ergebnisse zeigen, dass Nutzer insbesondere in den ersten Stunden nach Veröffentlichung mit einem Beitrag interagieren und die Sättigung des Wachstums der Reactions durchschnittlich nach rund zwölf Stunden erreicht ist. Wird ein Post zu einer Tageszeit veröffentlicht, in der wenige Nutzer online sind, verlängert sich der Wachstumszyklus. Zudem lässt sich die Verteilung der verschiedenen Reaction-Typen bereits eine Stunde nach Veröffentlichung recht präzise voraussagen, ein Meinungsumschwung zeigt sich nicht mehr. Daraus ergibt sich, dass die wesentlichen Determinanten des letztlichen Stimmungsbilds in stabilen Faktoren zu suchen sind, die bereits bei der Publikation feststehen. Der Beitrag ordnet die Ergebnisse vor dem Hin-

tergrund sozialer und technischer Rahmenbedingungen ein und erläutert darüber hinaus methodische Implikationen für zukünftige Studien.

*Keywords:* digitale Spurendaten, Facebook, hochfrequente Datenerhebung, Algorithmus

## 1. Perspektiven der non-reaktiven Messung mittels digitaler Spurendaten

Ein zentrales Anliegen der Medienwirkungsforschung ist, den Einfluss von Medieninhalten auf die Einstellungen von Rezipienten zu messen. Für diesen Zweck werden meist Befragungsstudien – häufig in experimentellen Designs – durchgeführt. Diese sind mit einer Reihe bekannter Nachteile (bspw. Reaktivität der Befragung, forcierte Rezeptionssituation, künstliches Stimulusmaterial) verbunden. Die zunehmende Nutzung digitaler Medien bietet die Möglichkeit, (experimentelle) Befragungen durch nicht-reaktive Online-Analysen zu ersetzen. Dabei werden Daten verwendet, welche die Rezipienten durch die Nutzung von Webinhalten generieren (SCHOBER et al. 2016). Grundsätzlich kann man zwischen Daten unterscheiden, die durch das schiere Aufrufen von Inhalten anfallen (z. B. auf Wikipedia), und solchen, die durch die intentionale Interaktion mit Webinhalten (z. B. durch Bewertung oder Kommentierung von Inhalten) entstehen. Zudem unterscheiden sich die verschiedenen Datenquellen hinsichtlich ihrer (öffentlichen) Sichtbarkeit sowie hinsichtlich der Möglichkeit, sie bestimmten Nutzern zuzuordnen.

Auf Facebook fallen bei der Interaktion mit Online-Inhalten beispielsweise Likes oder Shares an. Diese aggregierten Zustimmungsbekundungen – sogenannte ›Popularity-Cues‹ – sind öffentlich sichtbar und können als Indikatoren für aggregierte Zustimmung (AV) zu Medieninhalten erhoben werden. Mittels Inhaltsanalysen werden Botschaftsmerkmale (UV) erfasst, deren Wirkung dann als relativer Einfluss auf die aggregierte Zahl der Popularity-Cues gemessen werden kann (zum Überblick siehe PORTEN-CHEE et al. 2018). Im Gegensatz zu klassischen Methoden sind hierbei die Untersuchungsgegenstände manifest, das Erhebungsverfahren ist nicht-reaktiv und die Anzahl der untersuchten Inhaltsmerkmale wird lediglich durch deren natürliches Vorkommen begrenzt. Beispiele für die Anwendung finden sich in der Forschung zur Wirkung von politischen Kommunikati-

onsstrategien: So lässt sich die Zahl der Likes, Shares und Kommentare auf Statusmeldungen von Parteien und Politikern mitunter durch die Verwendung von positiven Argumenten und Humor (HEISS/SCHMUCK/MATTHES 2018) oder Negativität erklären (BENE 2017). Auch zur Überprüfung der Nachrichtenwerttheorie bietet sich die Nutzung von digitalen Spurendaten an. Eine plattformübergreifende Untersuchung zeigt beispielsweise einen positiven Einfluss von verschiedenen Eigenschaften von Artikeln (darunter Nachrichtenfaktoren) auf die Zahl der Shares auf Facebook und Retweets auf Twitter (TRILLING/TROLOCHKO/BURSCHER 2016).

Das Vorgehen der genannten Studien beruht allerdings auf der voraussetzungs-vollen Annahme, dass die erhobenen Popularity-Cues ausschließlich die (aggregierten) Einstellungen der Nutzer zum Inhalt der Botschaften oder zu deren Sender widerspiegeln. Durch diese Annahme ergeben sich zwei grundlegende Limitationen der bisherigen Forschung: 1) Es wurde bislang primär das Ausmaß der positiven Reaktionen auf Botschaften erfasst, was der Konnotation der erfassten Popularity-Cues geschuldet ist. In vielen Fällen ist darüber hinaus die positive Bedeutung lediglich zugeschrieben, wenn wie etwa im Fall von Retweets lediglich von einem Indikator von Aufmerksamkeit ausgegangen werden kann (JUNGHERR et al. 2017). 2) Allen genannten Studien ist gemein, dass die Popularity-Cues erst retrospektiv im Aggregat erfasst wurden. Damit bilden sie lediglich das Ergebnis eines Prozesses ab, in welchem die Interaktionen zustande gekommen sind.

Dabei scheint gerade die zeitliche Dimension von besonderer Relevanz: So ist bisher größtenteils unklar, über welchen Zeitraum sich die Interaktionen erstrecken bzw. wann eine Sättigung der Interaktionen erreicht ist, also kaum noch neue oder keine Interaktionen mehr hinzukommen. Dies ist mitunter aus forschungspraktischen Gesichtspunkten wichtig, da beispielsweise bei zu geringem zeitlichen Abstand zwischen Veröffentlichung eines Inhalts und der Datenerhebung unter Umständen spätere Interaktionen unberücksichtigt bleiben. Andererseits kann es bei einer größeren Zeitspanne zu Datenverlusten durch Löschung kommen (BACHL 2018). Zudem versperrt eine retrospektive Erhebung den Einblick in die dynamische Entwicklung der Interaktionen. Diese wird sowohl durch die technischen Rahmenbedingungen sozialer Online-Netzwerke (Algorithmen und Benutzeroberfläche) als auch durch soziale Prozesse beeinflusst. So begünstigt die nutzerseitige Interaktion mit Nachrichten deren potenzielle Sichtbarkeit im Newsfeed anderer Nutzer, was wiederum die Wahrscheinlichkeit von Likes erhöht (BUCHER 2012; DEVITO 2016). Weiter

ist die Tageszeit ein wesentlicher Kontextfaktor für die Mediennutzung der Rezipienten (z. B. KLINGLER/TURECEK 2016; SCHNAUBER 2017) und in Konsequenz auch für die Interaktion mit Medieninhalten im Speziellen. Damit könnte der Zeitpunkt der Veröffentlichung einer Botschaft unabhängig von deren inhaltlichen Merkmalen zu Unterschieden in der Zahl der Interaktionen führen. Zudem wirken die Interaktionen selbst als Hinweisreize, welche nachfolgende Interaktionen (bspw. Reactions) beeinflussen bzw. auslösen können (z. B. SALGANIK/DODDS/WATTS 2006).

Die vorliegende Studie begegnet diesem Defizit. Sie illustriert, wie sich mittels einer hochfrequenten Erhebung der Reactions klären lässt, welche Rolle zeitlichen Aspekten zukommt. Dazu erläutern wir zunächst die zentrale Bedeutung von Algorithmen sowie von sozialen Kontextfaktoren für die Entstehung und das Wachstum der Reactions. Anschließend gehen wir auf die technischen Anforderungen der hochfrequenten Datenerhebung ein. Schließlich ermitteln wir mittels Fixed-Effekt-Modellen die Dauer und Sättigung des Wachstums der Reactions sowie den Einfluss des Publikationszeitpunkts auf jene Prozesse. Abschließend diskutieren wir die Implikationen unserer Ergebnisse für zukünftige Kommunikationsforschung mittels digitaler Beobachtungsdaten und geben Hinweise für anschließende Forschung.

## 2. Facebook als Untersuchungsgegenstand

Facebook ist das soziale Online-Netzwerk mit der größten Verbreitung, sowohl weltweit als auch in Deutschland. Beinahe die Hälfte der deutschen Internetnutzer sind dort mit einem Profil präsent (ALLFACEBOOK.DE 2017). Zwar sind die Facebook-Mitglieder verglichen mit der Gesamtheit der Internetnutzer etwas jünger, unterscheiden sich jedoch in ihren politischen Einstellungen nicht signifikant (STARK/MAGIN/JÜRGENS 2017), was für die Erforschung der nutzerseitigen Reaktionen auf öffentliche Kommunikation mit politischem oder gesellschaftlichem Bezug von besonderer Relevanz ist. Neben zahlreichen Unternehmen und Unterhaltungsangeboten sind nahezu alle überregionalen und ein überwiegender Teil der regionalen Nachrichtenanbieter mit einem Profil auf Facebook vertreten und verbreiten dort ihre Inhalte. Auch unter politischen Akteuren ist dieses Netzwerk das meistgenutzte: 96 Prozent der Bundestagsabgeordneten sind auf Facebook präsent, während Twitter nur rund zwei Drittel der Abgeordneten

nutzen (SCHMIDT 2017). Damit bietet Facebook nicht nur den Zugang zu den beobachtbaren Interaktionen einer breiten Bevölkerungsschicht, sondern zudem ein vielfältiges Angebot verschiedener Botschaften, die potenziell zur inhaltsanalytischen Erfassung herangezogen werden können.

Facebook bietet seinen Nutzern die Möglichkeit, Inhalte weiterzubreiten (Shares) oder ihre Zustimmung und ihr Wohlgefallen durch Likes zum Ausdruck zu bringen (FACEBOOK 2018). Nutzerbefragungen verweisen darauf, dass sowohl Likes als auch Shares überwiegend als Ergebnis einer positiven Evaluation von Botschaften verstanden werden (z. B. CHOI 2016; HAYES/CARR/WOHN 2016; MEIER/ELSWELER/WILSON 2014). Demnach vermochten Studien, welche von Likes und Shares auf mögliche Wirkungen von Botschaftsinhalten schlossen, Reaktionen von überwiegend positiver Valenz abzubilden. Dieses Defizit wird jedoch durch die Weiterentwicklung der Interaktionsmöglichkeiten auf Facebook zumindest partiell beseitigt. Mit der Einführung der »Reactions« im Jahr 2016 erweiterte das größte Online-Netzwerk das Spektrum der möglichen Reaktionen und gab damit Nutzern die Möglichkeit auf »glückliche, traurige, witzige oder zum Nachdenken anregende Inhalte« passend reagieren zu können; neben dem bekannten Like können Nutzer nun auch Reaktionen wie »Love, Haha, Wow, Traurig und Wütend« auswählen (FACEBOOK 2016).<sup>1</sup> Damit decken die Reactions ein breiteres Spektrum an Gefühls- bzw. Meinungsäußerungen ab und erweitern das Potenzial der non-reaktiven Medienwirkungsforschung zu Popularity-Cues auf Facebook entscheidend.

### 3. Funktionsweise des Facebook-Algorithmus

Auf Facebook können private Nutzer, Nachrichtenanbieter, politische und gesellschaftliche Akteure unbegrenzt Inhalte veröffentlichen. Einige dieser Inhalte sind dann im Newsfeed verbundener Nutzer sichtbar. Bereits im Jahr 2013 wurden auf Facebook täglich 4,75 Mrd. Inhalte – Statusmeldungen, Fotos, Videos und Kommentare – veröffentlicht. Damals hatte das Netzwerk weltweit rund 1,1 Mrd. Nutzer (ALFERT 2015). Seitdem hat sich die Zahl der Nutzer nahezu verdoppelt: Im letzten Quartal von

1 Diese Funktionen werden von den Nutzern zunehmend eingesetzt – bis März 2017 wurden die Reactions rund 300 Mrd. Mal verwendet (FUTUREBIZ 2017).

2017 waren 2,1 Mrd. Nutzer auf Facebook angemeldet. Davon sind rund 1,4 Mrd. Nutzer täglich aktiv. Es scheint plausibel, dass die Zahl der veröffentlichten Inhalte demnach zugenommen hat (ALLFACEBOOK.DE 2017). Die schiere Zahl dieser Inhalte führt dazu, dass ein durchschnittlicher Nutzer potenziell mit Hunderten Inhalten am Tag in seinem Newsfeed konfrontiert würde.<sup>2</sup> Die Zahl der Beiträge ist vor allem davon abhängig, wie viele Freunde ein Nutzer hat und wie vielen Seiten er »folgt«. So wird angenommen, dass die Zahl der Inhalte für Nutzer mit einem größeren Netzwerk (mehr Freunde und/oder abonnierte Seiten) schnell in den dreistelligen Bereich rücken kann (ALLFACEBOOK.DE 2017).

Um eine Überflutung mit irrelevanten Informationen zu vermeiden, werden Webinhalte häufig an die individuellen Präferenzen der Nutzer angepasst. Dieser Prozess wird unter den Begriff der Personalisierung gefasst (CHURCHILL 2013). Basis personalisierter Empfehlungssysteme (»Recommender Systems«) können explizite (bspw. angegebene Vorlieben) oder implizite (bspw. aufgezeichnetes Verhalten) individuelle Präferenzen der Nutzer sein. Diese Informationen werden zunächst gesammelt und anschließend zu individuellen Nutzerprofilen verdichtet (BOZDAG 2013). Die Verfahren zur Personalisierung fußen dabei im Wesentlichen auf zwei Elementen: Relevant sind zum einen die zur Personalisierung herangezogenen Merkmale der Inhalte und zum anderen das mathematische Modell – der sogenannte »Algorithmus« –, der das Gewicht der herangezogenen Merkmale bestimmt (JÜRGENS/STARK/MAGIN 2014). Der Algorithmus gleicht dann vorherige Entscheidungen des Nutzers (bspw. Kauf eines Produktes) anhand der Inhaltsmerkmale ab und präsentiert neue Inhalte anhand von Übereinstimmungen (»Content-Based Filtering«). Informationen zum beobachteten Verhalten anderer Nutzer können als zusätzliche Informationsquelle herangezogen werden (»Collaborative Filtering«). Dabei wird davon ausgegangen, dass sich Nutzer, deren Verhalten in der Vergangenheit Übereinstimmungen aufweist, auch zukünftig ähnlich entscheiden werden (BOZDAG 2013). Je mehr Informationen über einen Nutzer zur Verfügung stehen und je mehr Inhaltsmerkmale im Algorithmus berücksichtigt werden, desto treffender sind die daraus abgeleiteten Empfehlungen.

2 Andere Schätzungen gehen von rund 1.500 Meldungen aus, die potenziell im Newsfeed eines durchschnittlichen Nutzers landen würden (MCGEE 2013).

Auch Facebook nutzt einen Algorithmus, um seinen Nutzern individuell passende Inhalte zu liefern. Dieser wurde von Facebook unter dem Namen ›EdgeRank‹ erstmals im Jahr 2010 vorgestellt.<sup>3</sup> Der EdgeRank basiert auf drei Grundparametern, die Einfluss darauf haben, welche Posts im Newsfeed eines Nutzers angezeigt werden (BUCHER 2012; KINCAID 2010): (1) Die Verbundenheit (›affinity‹) referiert auf die Beziehung, die zwischen dem Urheber eines Inhalts und potenziellen Empfängern besteht. Diese Beziehung definiert sich über die vergangene Interaktion zwischen den Nutzern und/oder deren Inhalten. Likt, teilt oder kommentiert beispielsweise ein Nutzer regelmäßig Inhalte eines anderen Nutzers, so werden dessen Beiträge auch in Zukunft mit höherer Wahrscheinlichkeit im eignen Newsfeed landen. (2) Die Bedeutsamkeit (›weight‹) bezieht sich auf die Interaktionen, die ein bestimmter Inhalt bereits erzielt hat. Wird mit einem Beitrag nach der Veröffentlichung interagiert, so misst der Newsfeed-Algorithmus diesem mehr Relevanz bei. (3) Das Alter (›decay‹) eines Inhalts determiniert ebenfalls die Relevanz eines Inhalts. Je älter ein Inhalt ist, desto niedriger wird dessen Relevanz eingestuft. Insbesondere die beiden letztgenannten Referenzen für die Relevanzzuschreibung des Facebook-Algorithmus verweisen auf die Relevanz der zeitlichen Dimension.

#### 4. Zeitliche Einflüsse auf die Anzahl der Interaktionen

Die Grundgesamtheit der Nutzer, die mit Inhalten auf Facebook interagieren, wird einerseits durch den Algorithmus bestimmt. Es gilt: Je relevanter ein Post ist, desto mehr Nutzern wird er im Newsfeed angezeigt, desto mehr Nutzer werden potenziell mit dem Post interagieren. Das

3 Allerdings ist der Facebook-Algorithmus keinesfalls statisch, sondern wird kontinuierlich weiterentwickelt. Nach Angaben eines Mitarbeiters sollen beim Nachfolger des EdgeRanks rund 100.000 Merkmale zur Bestimmung der individuellen Relevanz von Inhalten herangezogen werden (MCGEE 2013). Zusammensetzung der Faktoren sowie deren Gewichtung bleiben freilich ein gut gehütetes Firmengeheimnis (GILLESPIE 2014; kritisch hierzu TUFEKCI 2015). Die Funktionsweise des Algorithmus entzieht sich damit einer direkten wissenschaftlichen Analyse. Dem begegnet DeVito (2016) mit einer Kombination aus quantitativer und qualitativer Inhaltsanalyse von insgesamt 186 von Facebook zwischen 2006 und 2014 veröffentlichten Dokumenten (u. a. Pressemitteilungen und Patentanmeldungen), die sich mit der Entwicklung des Newsfeed befassen. Die Analyse bestätigt die Relevanz der bereits im EdgeRank dominanten Merkmale für die Gewichtung von Inhalten.



Alter eines Posts hängt negativ mit der vom Algorithmus zugeschriebenen Relevanz zusammen. Das bedeutet, dass mit zunehmendem Alter eines Posts die Zahl der Nutzer, in deren Newsfeed der Post angezeigt wird, über die Zeit hinweg abnimmt. Dem folgend, kann angenommen werden, dass auch die Zahl der neuen Interaktionen (Reactions) mit zunehmendem Alter abnimmt. Dieser Effekt wurde bereits für andere Online-Plattformen nachgewiesen. So findet bei der Plattform Digg.com ein Großteil der Interaktionen bereits in den ersten 24 Stunden statt, während bei YouTube auch nach 30 Tagen keine Sättigung eintritt (SZABO/HUBERMAN 2008). Andere Untersuchungen finden nach einigen Tagen ein verlangsamtes Wachstum der Views auf YouTube (z. B. CHENG/DALE/LIU 2008; für einen Überblick siehe KEYLING 2017). Die Unterschiede sind mitunter durch die Plattformarchitektur zu erklären. So gab es auf Digg.com eine Startseite, auf der neue Inhalte präsentiert werden; verlässt eine Meldung diese Startseite, so verringert sich die Zahl neuer Interaktionen drastisch (SZABO/HUBERMAN 2008). Auch auf YouTube führt die Präsenz von Videos in sogenannten ›Top-Listen‹ zu einem rapiden Anstieg der Viewzahlen (FIGUEIREDO et al. 2014).<sup>4</sup> Die Newsfeed-Architektur von Facebook spricht dafür, dass die Posts den Nutzern durch den Algorithmus nur für eine begrenzte Dauer angezeigt werden und die Zeitspanne zwischen Veröffentlichung und einem deutlichen Abflauen der Interaktionen eher kurz sind. Allerdings ist unklar, wann diese Sättigung auf Facebook tatsächlich erreicht ist (FF1).

Andererseits kann Facebook nur an jene Nutzer Inhalte ausspielen, die das Netzwerk nutzen also zu einem bestimmten Zeitpunkt online sind. Insbesondere vor dem Hintergrund tageszeitabhängiger Mediennutzungsmuster scheint der Veröffentlichungszeitpunkt entscheidend dafür zu sein, ob Nutzer mit einem Post interagieren. So zeigen Szabo und Huberman (2008), dass Nutzer auf Digg.com insbesondere in den Morgen- und Abendstunden mit Inhalten interagieren. Inhalte, die zu diesen Zeiten veröffentlicht werden, erhalten zudem insgesamt mehr Interaktionen. Inhalte, die zu Zeiten veröffentlicht werden, an denen Nutzer wenig interagieren, können diesen Rückstand offenbar nicht aufholen. Auch das Interakti-

4 Auch inhaltliche Merkmale bestimmen die Länge der Aufmerksamkeitszyklen. So verlieren Musikvideos weniger an Aktualität (und haben entsprechend einen eher linearen Zuwachs an Interaktionen) als nachrichtliche Inhalte (KEYLING 2017).

onsverhalten auf anderen Plattformen wie beispielsweise reddit.com lässt vergleichbare Schlüsse zu (BOZDAG 2013). In Deutschland ist die Internetnutzung im Tagesverlauf vergleichsweise stabil (SCHNAUBER 2017). Erst gegen 21 Uhr nimmt die Nutzung merklich ab und steigt dann wieder ab 6 Uhr langsam wieder an (KLINGLER/TURECEK 2016; VERBAND PRIVATER RUNDSTUND UND TELEMEIDIEN 2018). Daher ist anzunehmen, dass auf Facebook in den Abendstunden sowie in der Nacht weniger deutschsprachige Nutzer auf Facebook aktiv sind. Beiträge deutscher Nachrichtenanbieter, die zu dieser Zeit veröffentlicht werden, haben demnach eine geringere Chance auf Interaktionen in der ersten Phase nach der Veröffentlichung. Dies könnte dazu führen, dass jene Beiträge vom Newsfeed-Algorithmus als weniger relevant eingestuft werden. Daher scheint es plausibel, dass Beiträge, die in den Abendstunden veröffentlicht werden, eine geringere Zahl an Reactions aufweisen. Alternativ könnte man annehmen, dass der Newsfeed-Algorithmus bei der Einschätzung der die Zahl der initialen Interaktionen ins Verhältnis zur Größe des potenziellen Empfängerkreises (also zu der Zahl zum Veröffentlichungszeitpunkt aktiven Nutzer) setzt. In diesem Falle sollte das Wachstum der Reactions lediglich über die Nacht abgeschwächt werden und in den Morgenstunden wieder ansteigen. So sollte sich nur die Zeitspanne bis zur Sättigung des Wachstums nicht aber die Gesamtzahl der Reactions zwischen den Posts in Abhängigkeit zum Veröffentlichungszeitpunkt unterscheiden. Wir möchten beide Möglichkeiten in Betracht ziehen (FF2).

Neben der Zahl der Reactions (als Merkmal der Relevanzzuschreibung) ist auch das Verhältnis der verschiedenen Reactions zueinander (als Hinweis auf Meinungsverteilung) von Interesse. Neben der schieren Menge kann das Verhältnis verschiedener Reaction-Typen weitere Interaktion beeinflussen. So können aggregierte Zustimmungsbekundungen zur Abschätzung des Meinungsklimas herangezogen werden. Entsprechend den Annahmen der Theorie der Schweigespirale (NOELLE-NEUMANN 1974), kann ein zur eigenen Meinung als dissonant wahrgenommenes Meinungsklima dazu führen, dass die Redebereitschaft abnimmt (MATTHES/KNOLL/VON SIKORSKI 2018). So sind Nutzer weniger bereit, sich online in einer Diskussion zum Mobbing von Homosexuellen zu positionieren, deren Kommentare mehrheitlich nicht ihrer Meinung entsprechen (GEARHART/ZHANG 2014). Findet ein Nutzer beispielsweise einen Beitrag problematisch, würde also mit einem ›Angry‹ reagieren, während die prävalenten Reactions überwiegend positiv sind, so könnte dies Nutzer von der Interaktion abhalten.

Umgekehrt sollte eine Übereinstimmung von vorhandenen Reactions und eigener Bewertung die Bereitschaft erhöhen. Die empirische Evidenz für Schweigespiralprozesse in Online-Netzwerken durch Popularity Cues ist jedoch allenfalls schwach. Dies liegt einerseits daran, dass Nutzer das Meinungsklima kaum aus aggregierten Zustimmungsbekundungen extrapolieren (LEE/JANG 2010; PORTEN-CHEÉ/EILDERS 2015). Dies könnte der Tatsache geschuldet sein, dass in experimentellen Untersuchungen bis dato lediglich Likes und andere positive Reaktionen manipuliert wurden, wohingegen entsprechende negative Bewertungen fehlen (NEUBAUM/KRÄMER 2017; WINTER/BRÜCKNER/KRÄMER 2015). Zum anderen unterscheiden sich Kommentieren und Liken. Die sogenannte ›Click-Speech‹ hinterlässt – im Vergleich zum Teilen oder Kommentieren – weniger öffentliche Spuren. Die Gefahr der sozialen Isolation ist somit geringer und der Einfluss eines gegenläufigen Meinungsklimas auf die ›Klickbereitschaft‹ schwächer (MATTHES/KNOLL/SIKORSKI 2018; PORTEN-CHEÉ et al. 2018). Studien zeigen, dass ein zur eigenen Einstellung abweichendes Meinungsklima Nutzer nicht davon abhält (PANG et al. 2016) oder gar motiviert, Inhalte zu liken und ihre Meinung ohne die Drohung direkter Sanktionen mit vergleichsweise geringem Aufwand kundzutun (PORTEN-CHEÉ/EILDERS 2015). Daher scheinen Reactions eine niedrighschwellige Möglichkeit, seinen eigenen Standpunkt auch gegen die vermeintliche Mehrheitsmeinung zu äußern. Wir möchten dennoch prüfen, ob das Verhältnis der verschiedenen Reactions zueinander im Zeitverlauf variiert (FF3).

## 5. Datenerhebung und Analyse

Mit dem Ziel, die zeitliche Entwicklung von Reactions abzubilden, und dies auch bereits in der Anfangsphase kurz nach dem Erscheinen eines Posts zu tun, rückt die Präzision der Erhebung mitsamt der Konsequenzen für das Studiendesign in den Fokus. Grundsätzlich haben technisch automatisierte Erfassungen von digitalen Spurendaten mit zahlreichen potenziellen Fehlerquellen zu kämpfen (vgl. JÜRGENS/JUNGHERR 2016). Im konkreten Anwendungsfall für Zeitreihenanalysen können wir neben den gängigen Gütekriterien empirischer Forschung vier besonders wichtige Aspekte identifizieren. (1) *Zugänglichkeit*: An erster Stelle müssen die gesuchten Daten überhaupt zugänglich sein, was im Fall der Reactions durch die Reactions-Endpunkte der Facebook Graph-API prinzipiell ge-

geben ist.<sup>5</sup> (2) *Minimierung und Dokumentation von technischen Fehlern*: Bei der systematischen Erhebung ist es zweitens wichtig, technische Fehler (wie Störungen der API, der Netzwerkverbindung oder des verwendeten Programms) nicht nur zu minimieren, sondern auch sichtbar zu machen, um statistischen Fehlinterpretationen (etwa durch die fälschliche Annahme von null Aktivität bei Störungen) vorzubeugen. (3) *Synchronität*: Drittens – und das ist für longitudinale Analysen der wichtigste Aspekt – muss versucht werden, zeitliche Verzerrungen weitestgehend auszuschließen. Das bedeutet primär, dass bei der Erhebung differenzielle Latenzen minimiert werden – dass also Reactions von zwei unterschiedlichen Posts, die gleichzeitig gespeichert werden, auch tatsächlich Daten vom selben Zeitpunkt widerspiegeln. Gelingt diese Kontrolle nicht, dann wird die Zeit sozusagen mit zweierlei Maß gemessen; verdeckte Fehler, nicht-lineare Verzerrungen und Über- sowie Unterschätzungen von Effekten können die Folge sein.

In der Praxis ist der Anspruch der simultanen Erfassung nicht perfekt, aber immerhin annähernd einlösbar. Da die angestrebte Genauigkeit unmittelbar mit dem Aufwand des Designs zusammenhängt, lohnt sich eine grobe Schätzung der erwarteten Größenordnungen. Am unteren Ende wird die Messgenauigkeit von Facebook-Daten von der nicht perfekten Synchronisation der API begrenzt. Die Datenbanken der meisten Plattformen sind nicht zeitgleich auf demselben Stand, sondern verteilen Änderungen (wie z. B. das Hinzufügen oder Abziehen von Reactions, Kommentaren usw.) sukzessiv. In der Datenbankterminologie sind sie daher nicht »consistent«, sondern »eventually consistent« (LU et al. 2015), und Anfragen von Informationen können um mehrere Sekunden veraltete Daten liefern (BRONSON et al. 2013). Unter Einbezug der restlichen Messlatenz (Bearbeitung durch Facebook-Server, Netzwerkübertragung) scheint es daher nicht zielführend, die Zahl an Reactions häufiger als einmal pro Minute zu erfassen.

Am oberen Ende ist die angestrebte Messgenauigkeit von der erwarteten zeitlichen Dynamik des Phänomens bestimmt. Je schneller ein zu untersuchender Prozess abläuft, desto hochfrequenter muss er gemessen werden. Studien zu Nachrichtenaggregatoren (WU/HUBERMAN 2007) zeigen, dass sich Aufmerksamkeitsdynamiken insbesondere in den ersten 100 Minuten abspielen (vgl. Abb. 3, Seite 90). Im Fall von Facebook ist zusätzlich

5 Dies galt zur Zeit der Entstehung der vorliegenden Untersuchung. Wir werden auf die seit 2018 bestehenden Restriktionen am Ende des Beitrags eingehen.

in Betracht zu ziehen, dass Echtzeitbenachrichtigungen (Notifications) von Webseite und Apps die Geschwindigkeit von Nutzerinteraktionen unmittelbar nach der Veröffentlichung erhöhen dürften. Zusammenfassend erscheint plausibel, dass eine Erfassung der zeitlichen Entwicklung von Facebook-Reactions einerseits im Minutentakt erfolgen sollte, um hypothetische Mikrodynamiken abbilden zu können, und andererseits mindestens einen vollständigen Tag abdecken muss, um spezifische Effekte von Tageszeiten einbeziehen zu können.

(4) *Kontextkontrolle*: Letztlich ist bei der Wahl der Stichprobe ebenfalls wichtig, der großen Streubreite an verschiedenen Kontexten auf Facebook Rechnung zu tragen. Sowohl in der inhaltlichen Gestaltung als auch in der Nutzerpopulation und Aktivität von Seiten liegen selbst innerhalb eines Mediensamples große Unterschiede. Hinzu kommen die bekannten Schwankungen durch die Nachrichtenlage (etwa bei Großereignissen wie Terroranschlägen, die Nachrichten mit hoher Reichweite, Geschwindigkeit und Konsonanz auslösen). Fassen wir alle vier Kriterien für die Güte der Erhebung zusammen (Zugänglichkeit, Minimierung und Dokumentation von Fehlern, Synchronität und Kontextkontrolle), so folgt als ideales Design eine zeitgleiche, synchronisierte Erfassung von verschiedenen Seiten über mehrere Tage. Da keine fertig nutzbare Software existiert, die diese Parameter erfüllt, wurde für die Studie ein eigener Facebook-Crawler programmiert.

Grundsätzlich sind für die zuverlässige Abbildung von zeitlichen Entwicklungen zwei Einschränkungen entscheidend: Die API bildete jeweils nur den Status Quo zum Zeitpunkt der Abfrage ab; vorausgegangene Änderungen waren nicht verfügbar. Das heißt, gelöschte und verborgene Inhalte lassen sich nachträglich nicht erfassen – die API bestätigte noch nicht einmal deren frühere Existenz. Ebenso lagen Inhalte, die verändert wurden, nur in der jeweils aktuellen Fassung vor. Die Reactions waren darüber hinaus von der besonderen Einschränkung betroffen, dass sie von Facebook nicht einzeln mit Datums- und Zeitangaben versehen werden. Zu jedem Zeitpunkt konnte die Gesamtzahl sowie eine Liste der einzelnen Reactions mit dazugehörigen Nutzern angefragt werden; ob diese jedoch alle zur selben Zeit oder im Verlauf entstanden sind, blieb offen<sup>6</sup>.

6 <https://developers.facebook.com/docs/graph-api/reference/v3.2/object/reactions>

Die aus dieser Zeitagnostik resultierende empirische Unschärfe wurde umgangen, indem Forscher in vorgegebenen Zeitabständen jeweils Versionen derselben Daten anfordern, um diese dann zu Zeitreihen zu kombinieren. Bei Inhalten, die kein offizielles Datum tragen, dient der Zeitpunkt der Abspeicherung als Näherungswert. Dieses Verfahren geht immer mit einer inhärenten Abwägung zwischen zwei wünschenswerten Zielen einher: Einerseits ist eine möglichst hochfrequente Abfrage sinnvoll, um ›unsichtbare‹ Änderungen zwischen zwei Datenpunkten zu minimieren. Werden beispielsweise jede Stunde Kommentare zu einem Post erfasst, dann fehlen solche, die innerhalb dieser Stunde sowohl publiziert als auch wieder entfernt wurden. Andererseits lässt sich in dem von Facebook vorgegebenen Anfragevolumen nur eine bestimmte Zahl an Posts bzw. Seiten beobachten. Je häufiger diese aktualisiert werden sollen, desto weniger Seiten/Posts können in die Stichprobe eingehen.

Der Crawler erfasste schließlich minütlich die Reactions auf alle Posts, die zwischen dem 3.6.2017 und dem 5.6.2017 auf den Seiten von insgesamt 17 deutschen Medienangeboten<sup>7</sup> veröffentlicht wurden. In jeder Minute wurden also zwei Schritte durchgeführt: (1) Der Abruf von neuen Posts auf den 17 Seiten und (2) der Abruf von neuen Reactions auf alle bis zu diesem Zeitpunkt bekannten Posts. Die Anzahl der im zweiten Schritt auftretenden Anfragen an die API wuchs im Laufe der Erhebung kontinuierlich: Aus der gleichbleibenden Taktung über drei Tage hinweg ergaben sich letztlich 556 Posts über 17 Seiten und somit am Ende 556 Anfragen pro Minute bzw.  $556 \cdot 60 = 33.360$  Anfragen pro Stunde. Diese Obergrenze an Datendurchsatz war letztlich durch die von Facebook zugelassene Kontingentierung bedingt. Bei Einsatz von zusätzlichen API-Accounts oder einer rotierenden Stichprobe an Posts (mithilfe eines ›Fensters‹, bei dem also Posts nach X Tagen aus dem Sample genommen werden) könnten auch längere und/oder breitere Erhebungen realisiert werden. Als Grundlage der vorliegenden Studie dienen schließlich insgesamt 258.736 Reactions auf die 556 Facebook-Posts.

Eine weitere methodische Herausforderung bei der Analyse der Reactions auf Facebook-Posts ist die statistische Modellierung hierarchischer

7 *Focus Online, Faz.Net – Frankfurter Allgemeine Zeitung, Tagesschau, Die Welt, Bild, Zeit Online, Süddeutsche Zeitung, SWR Aktuell, Spiegel Online, Stern, RTL Aktuell, taz, die Tageszeitung, MDR Aktuell, RTL 2 News, ZDF heute, SAT.1 Nachrichten, Frankfurter Rundschau.*

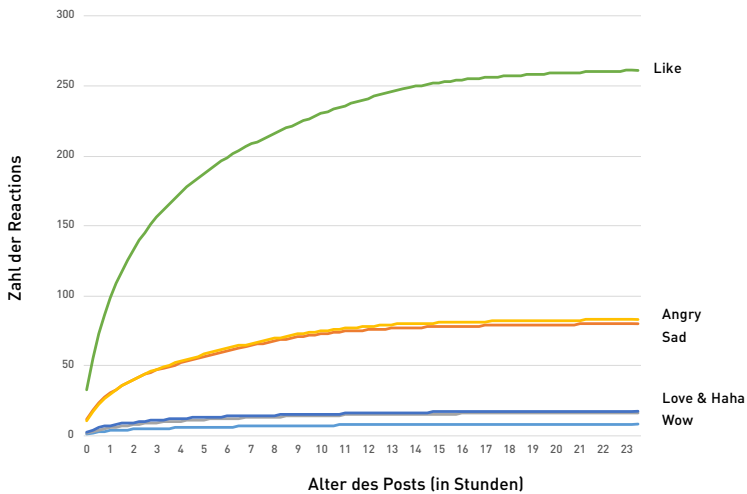
Datenstrukturen. So sind im vorliegenden Fall die 556 Posts (Level 1-Einheiten) nach den Facebook-Seiten der 17 betrachteten deutschen Medienangebote (Level 2-Kontexte) geschachtelt (LEGEWIE 2012). Würden wir diese Datenhierarchien unangemessen modellieren oder sogar gänzlich vernachlässigen, könnte es zu fehlgeleiteten Inferenzschlüssen und Verzerrungen der Effektschätzungen kommen. Für die Modellierung der Schachtelungen der Posts in den Seiten stehen mit Mehrebenenverfahren einerseits sowie Fixed-Effect-Regressionsmodellen (FE-Modelle) andererseits zwei Alternativen zur Verfügung. Mehrebenenverfahren stellen die Datenhierarchie durch die Schätzung verschiedener (ebenspezifischer) Residualvarianzkomponenten in Rechnung. Die Fixed-Effect-Estimation berücksichtigt die Facebook-Seiten demgegenüber, indem deren Unterschiede durch dichotome (seitenspezifische) Prädiktoren kontrolliert werden. Dieser Ansatz ist konzeptionell äquivalent zu einem Mehrebenenansatz mit Kontextmittelwertzentrierung. Der zentrale Vorteil der Mehrebenenmodellierung ist, dass spezifische Eigenschaften der Facebook-Seiten mit den Eigenschaften der einzelnen Posts in Bezug gesetzt werden können. Allerdings stellen Mehrebenenverfahren höhere Anforderungen an die Datengrundlage als FE-Modelle, denn ihre zuverlässige Schätzung erfordert die Berücksichtigung einer größeren Zahl von Facebook-Seiten. Ebenfalls ermöglichen FE-Modelle eine effektivere Drittvariablenkontrolle: »It is hard to overstate the gain in identifying power«, denn sie ermöglichen »safeguarding causal inferences against bias arising from the presence of unmeasured confounders« (GANGL 2010: 34). Mit anderen Worten können wir die Effekte der Posts-Eigenschaften (z. B. Zeitpunkt der Veröffentlichung) auf sämtliche Unterschiede zwischen den Facebook-Seiten (z. B. höhere Interaktionsraten bei manchen Seiten) kontrollieren, auch wenn wir – was gemeinhin nicht der Fall ist – nicht alle zentralen Unterschiedsvariablen kennen. Nachfolgend nutzen wir FE-Modelle, da die Zahl der unsererseits berücksichtigten Facebook-Seiten recht gering ist und der inhaltliche Akzent des vorliegenden Beitrags eher auf den Facebook-Posts denn auf den Facebook-Seiten liegt.

## 6. Ergebnisse

Wir gehen davon aus, dass Nutzer vor allem dann mit Posts interagieren, wenn diese in ihrem Newsfeed sichtbar sind. Diese Sichtbarkeit wird von

der Relevanzzuschreibung des Newsfeed-Algorithmus bestimmt. Die Relevanzzuschreibung sinkt mit zunehmendem Alter des Posts. Daher vermuten wir, dass die Zahl neuer Reactions im Zeitverlauf abnimmt. Die von uns erhobenen Daten bestätigen diese Annahme (s. Abb. 1). Die Zahl der Reactions auf die Posts nimmt insbesondere in den ersten Stunden nach der Veröffentlichung zu und flacht dann kontinuierlich ab. Dieses Muster findet sich bei allen Typen der Reactions in ähnlicher Form (s. Abb. 2).

ABBILDUNG 1  
Reactions im Zeitverlauf nach Medien



N = 556 Posts

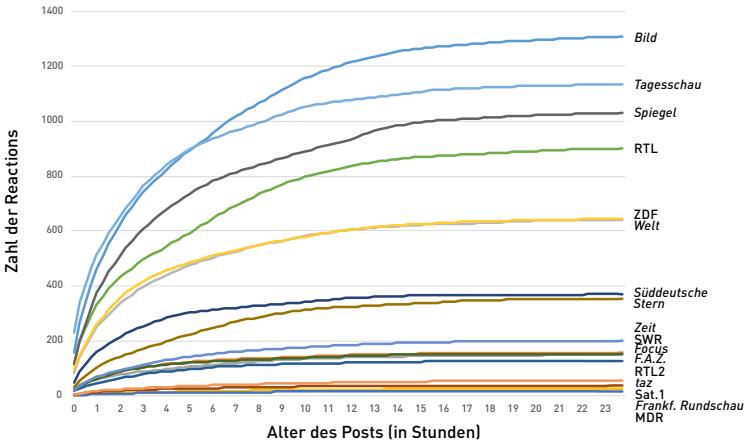
Weiter ist von Interesse, wann eine Sättigung des Wachstums der Reactions erreicht ist (FF1). Wir sprechen von einem gesättigten Wachstum, wenn ein Post 95 Prozent der Gesamtanzahl seiner Reactions erreicht hat.<sup>8</sup> Im Durchschnitt beträgt die Zeitspanne zwischen Veröffentlichung und Sättigung rund 12 Stunden ( $M = 711$  min;  $SD = 312$  min). Allerdings variiert der Sättigungsprozess zwischen den einzelnen Beiträgen: So gibt es Posts, die bereits direkt nach ihrer Veröffentlichung keine neuen Reactions mehr

8 Dieser Schwellenwert scheint sinnvoll, da kaum noch nennenswerte Veränderungen in der relativen Zusammensetzung der Reactions erwartbar sind.



ABBILDUNG 2

## Reactions im Zeitverlauf nach Reaction-Typ



N=556 Posts

erhalten, und solche, die erst nach über zwei vollen Tagen 95 Prozent aller Reactions erreichen (Range=1-3076 min). Bei diesen Ausreißern mit früher Sättigung handelt es sich um Posts, die sich durch eine besonders geringe Zahl an Reactions auszeichnen.

Zudem nehmen wir an, dass der Veröffentlichungszeitpunkt Auswirkungen darauf hat, wie intensiv die Nutzer mit einem Post direkt nach seiner Veröffentlichung interagieren (FF2). Vor dem Hintergrund einer tageszeitabhängigen Intensität der Mediennutzung scheint es plausibel, dass in den Abend- und Nachtstunden veröffentlichte Posts innerhalb der ersten Stunde weniger Reactions erhalten. Im Schnitt erhält ein Post unseres Samples in der ersten Stunde nach Veröffentlichung 154 Reactions ( $SD=312$ , Range=1-3076). Tabelle 1 zeigt die durchschnittliche Zahl aller Reactions auf einen Post sowie die durchschnittliche Zahl der unmittelbaren Reactions auf einen Post in den ersten 60 Minuten nach seiner Veröffentlichung in Abhängigkeit der Uhrzeit der Veröffentlichung des Posts. Die Zahl der Reactions auf einen Post während der ersten 60 Minuten nach seiner Veröffentlichung variiert in Abhängigkeit vom Veröffentlichungszeitpunkt signifikant ( $F=3,262$ ,  $df=7$ ,  $p<0,001$ ). Posts, die vor 6 Uhr abgesetzt werden, erfahren innerhalb der ersten 60 Minuten weniger Interaktion als solche, die am Tag und in den Abendstunden veröffentlicht werden. Eine

Ausnahme bilden Veröffentlichungen um die Mittagszeit; zu dieser Zeit findet die geringste Interaktion mit den von uns untersuchten Inhalten statt. Entgegen unserer Annahme erreichen am Abend veröffentlichte Posts das höchste Maß unmittelbarer Reactions. Demgegenüber hängt die Gesamtzahl der Reactions ( $M = 466$ ,  $SD = 977$ ,  $\text{Range} = 1\text{--}9885$ ) auf einen Post nicht signifikant mit dem Zeitpunkt der Veröffentlichung zusammen ( $F = 0,837$ ,  $df = 7$ ,  $p = 0,554$ ). Signifikant hängt hingegen die Zeitspanne nach der Veröffentlichung eines Posts, innerhalb der 95 Prozent aller Reactions auf den Post erfolgen, mit dem Veröffentlichungszeitpunkt zusammen ( $F = 6,173$ ,  $df = 7$ ,  $p < 0,001$ ).

TABELLE 1

### Reactionsverläufe in Abhängigkeit vom Veröffentlichungszeitpunkt

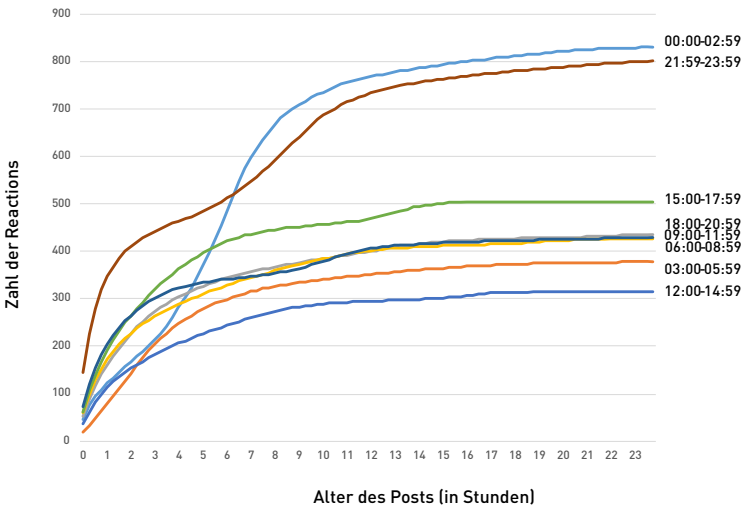
Zeitpunkt	Anzahl Posts	Gesamtzahl der Reactions nach 60 Minuten $M = 154$ [%]	Gesamtzahl der Reactions $M = 466$	Alter des Posts in Minuten bei 95% aller Reactions $M = 711$
00:00 – 02:59	19	110 [13]	834	749
03:00 – 05:59	47	65 [17]	380	502
06:00 – 08:59	103	140 [32]	433	646
09:00 – 11:59	85	153 [36]	426	798
12:00 – 14:59	82	99 [32]	312	702
15:00 – 17:59	84	166 [33]	498	719
18:00 – 20:59	88	183 [42]	432	718
21:00 – 23:59	48	320 [40]	807	879
<i>Signifikanz</i>				
<i>F</i>		3,262	0,837	6,173
<i>df</i>		7	7	7
<i>P</i>		< 0,001	0,554	< 0,001

$N = 556$  Posts, Signifikanztests basieren auf FE-Modellen unter Berücksichtigung der Schachtelung von Posts in Seiten. Mittwerte wurden zur intuitiveren Interpretierbarkeit nicht um die Seitenunterscheide adjustiert.

Die detaillierte Aufschlüsselung impliziert, dass Posts am späten Abend zwar in den ersten 60 Minuten nach der Veröffentlichung in vergleichsweise hohem Maße unmittelbare Reactions erhalten; das Wachstum der Zahl der Reactions flacht dann durch die tages- bzw. nachtzeitbedingte (Nicht-)Nutzung von Facebook temporär ab. Erst in den frühen Morgenstunden steigt die Zahl der Reactions mit der Aktivität der Nutzer weiter an. Dies gilt

ebenfalls für Posts, die während der frühen Nachtstunden veröffentlicht werden, und jene, die vor den interaktionsärmeren Mittagsstunden veröffentlicht werden und deren Sättigung sich auf einen späteren Zeitpunkt verschiebt. Die Ergebnisse deuten demnach eher auf eine Unterbrechung des Wachstums als auf eine generelle Abschwächung hin (FF2). Diese Interpretation legt auch der Verlauf der Wachstumskurven (s. Abb. 3) sowie die recht zuverlässige Vorhersage der Gesamtzahl der Reactions bereits eine Stunde nach Veröffentlichung (s. Tab. 2) nahe.

ABBILDUNG 3  
**Reactions im Zeitverlauf nach Veröffentlichungszeitpunkt**



N = 556 Posts

Abschließend prüfen wir, ob sich die Verhältnisse der verschiedenen Typen der Reactions im Zeitverlauf verändern. Tabelle 2 legt dar, wie gut sich der finale Anteil eines Reaction-Typs an der Gesamtzahl der Reactions anhand seines Anteils innerhalb der ersten 5, 15, 30 und 60 Minuten nach der Veröffentlichung des Posts prognostizieren lässt. Wir stellen die Ergebnisse der betreffenden Fixed-Effekt-Regressionen in Form von standardisierten Regressionskoeffizienten dar, die allesamt für  $p < .001$  signifikant sind. Es zeigt sich, dass sich der finale Anteil eines Reaction-Typs an der Gesamtzahl der Reactions bereits relativ zuverlässig anhand

der in den ersten fünf Minuten nach Veröffentlichung des Posts erfolgten Reactions vorhersagen lässt. Die Vorhersagekraft nimmt wiederum mit der zunehmenden Lebensdauer des Posts noch weiter zu, sodass sich der finale Anteil eines Reaction-Typs spätestens 60 Minuten nach Veröffentlichung des Posts nahezu perfekt vorhersagen lässt. Es kommen also auch später noch weitere Reactions hinzu, aber diese führen nicht mehr zu einer Verschiebung des Meinungsbildes, das sich in der ersten Stunde nach Veröffentlichung des Posts herauskristallisiert.

TABELLE 2

**Prognosen der Gesamtzahl der Reactions und  
Gesamtanteile der Reaction-Typen**

Zeit nach Veröffent- lichung	Anteil an allen Reactions						
	Gesamt	LIKE	WOW	SAD	HAHA	ANGRY	LOVE
05 min	.639	.769	.498	.775	.637	.733	.503
15 min	.767	.861	.719	.889	.795	.868	.818
30 min	.845	.927	.813	.932	.871	.928	.813
60 min	.906	.969	.903	.975	.917	.963	.881

N=556 Posts; standardisierte FE-Regressionskoeffizienten unter Berücksichtigung der Schachtelung von Posts in Seiten; alle  $p < .001$ .

7. Zusammenfassung und Diskussion

Die Verwendung von Popularity Cues erlaubt es, Einstellungen zu Bot-schaftsinhalten nicht-reaktiv zu erfassen. Insbesondere die Reactions von Facebook bieten hinsichtlich ihrer Differenzierung die Möglichkeit, ein breiteres Spektrum an positiven und negativen Reaktionen auf Medien-inhalte zu beobachten. Um die Aussagekraft der Reactions richtig ein-schätzen zu können, sollten jedoch zunächst die Eigendynamiken un-tersucht werden, die sich im Verlauf ihrer Entstehung ergeben. Diese können durch die Architektur der Plattform und der Funktionsweise des Newsfeed-Algorithmus sowie sozialer Prozesse bedingt werden. Die Un-tersuchung berücksichtigt dies sowohl bei der Erhebung als auch bei der statistischen Modellierung.

Die Auswertung unserer Daten basiert auf Fixed-Effekt-Regressionsmodellen, wobei die untersuchten Posts (Level 1) in den Facebook-Seiten der verschiedenen Medien geschachtelt sind. Durch Mittelwertzentrierung der Variablen auf Postebene halten wir die Unterschiede zwischen den verschiedenen Medienseiten konstant. Wie bereits im Methodenkapitel angeklungen, ist ein solches Vorgehen grundsätzlich anzuraten, da Seiteneigenschaften direkten Einfluss auf die Reactions haben können. So determiniert beispielsweise die Zahl der Fans zu einem großen Maß die potenzielle Reichweite und damit auch die Wahrscheinlichkeit von Interaktionen mit den Posts.

Die Ergebnisse unserer Untersuchung lassen darauf schließen, dass das Wachstum der Reactions einem degressiven Verlauf folgt. Damit legt die Annahme nahe, dass die Newsfeed-Architektur von Facebook sowie der korrespondierende Algorithmus dafür sorgen, dass veröffentlichte Inhalte mit zunehmendem Alter aus den Newsfeeds der Nutzer verschwinden und die Nutzer entsprechend seltener mit den Inhalten interagieren. Die Sättigung des Wachstums ist in durchschnittlich 12 Stunden erreicht. Diese Sättigung setzt bei Posts, die vor oder in interaktionsärmeren Zeiträumen (Mittagszeit, Abend- und Nachtstunden) veröffentlicht werden, tendenziell später ein. In unseren Daten war die Sättigung jedoch nach maximal 51 Stunden erreicht. Aus diesen Beobachtungen lassen sich Empfehlungen für die Datenerhebung ableiten, welche bei zukünftigen Studien berücksichtigt werden sollten.

So scheint eine kontinuierliche hochfrequente Erhebung der Reactions die ideale Lösung. Damit lassen sich zum einen mögliche Dynamiken abbilden; zum anderen können so auch gelöschte Inhalte analysiert bzw. deren Löschung dokumentiert werden. Alternativ ist auch eine Querschnitterhebung aller Daten zu einem Zeitpunkt möglich. Hier ist es jedoch geboten, den Abstand zwischen Veröffentlichung des letzten Posts des jeweiligen Samples und der Datenerhebung zu bestimmen. Hierbei gilt es zu bedenken, dass mit zunehmender Zeitspanne zwischen Veröffentlichung und Erhebung die Gefahr einer Löschung von Inhalten steigt. Vor diesem Hintergrund sowie den aus unseren Daten ermittelten Sättigungsverläufen scheint eine Erfassung der Reactions mit einer Latenz von beispielsweise 24 Stunden angemessen.

Zudem zeigen unsere Daten, dass die Verteilung der verschiedenen Reaction-Typen im Zeitverlauf stabil bleibt: Bereits eine Stunde nach Veröffentlichung lässt sich der Anteil des jeweiligen Reaction-Typs recht zuverlässig vorhersagen. Ein Meinungsumschwung findet offenbar nicht

mehr statt. Daraus ergibt sich, dass die wesentlichen Determinanten des letzten Stimmungsbilds in stabilen Faktoren zu suchen sind, die bereits bei der Publikation feststehen.

Anzumerken bleibt, dass die vorliegende Untersuchung in einem spezifischen Kontext stattfand. So bezogen sich einige der veröffentlichten Beiträge auf den Terroranschlag in London, welcher in den Abendstunden des 3.6.2017 stattfand. Dies hat sowohl Konsequenzen für die Interpretation unserer Befunde hinsichtlich des Einflusses des Veröffentlichungszeitpunktes und der Stabilität des Verhältnisses der verschiedenen Reaction-Typen: Zum Terroranschlag wurden in den Abend- und Nachtstunden zahlreiche Beiträge veröffentlicht. Der außerordentliche Nachrichtenwert dieses Ereignisses hat mutmaßlich zu einem überdurchschnittlichen Interesse an den Beiträgen und damit zu einer potenziell höheren Zahl an Reactions geführt. So bleibt offen, ob die Beiträge in den Abend- und Nachtstunden ohne Konfundierung von Inhaltsmerkmalen und Veröffentlichungszeitpunkt eine so hohe Zahl an Reactions erhalten hätten.

Weiterhin sind die Ereignisse in London von stark negativer Valenz. Im Unterschied zu anderen Ereignissen und Themen, deren Interpretationen durchaus kontroverser ausfallen, sind hier eher homogen negative Reactions zu erwarten. Die bereits eine Stunde nach Veröffentlichung beobachtete Stabilität des Verhältnisses der Reaction-Typen könnte somit zumindest partiell der Themenlage geschuldet sein. Zukünftige Forschung sollte den Einfluss verschiedener Themen testen oder zumindest deren Einfluss kontrollieren. So könnten bei ›neueren‹ oder strittigeren Themen, gegenüber denen viele Nutzer noch keine gefestigte Einstellung haben, die Entwicklungen der Reactions eine geringere Konstanz aufweisen.

Fernab der Kontextbedingungen der Untersuchung bleibt festzuhalten, dass wir lediglich Reactions in unserer Analyse berücksichtigt haben. Dabei blieben Kommentare und Shares unberücksichtigt. Der Newsfeed-Algorithmus von Facebook wertet diese Interaktionen mutmaßlich als gewichtiger bei der Zuschreibung von Relevanz und damit bei der potenziellen Weiterverbreitung von Inhalten. Zudem schätzen Nutzer das vorhandene Meinungsklima insbesondere anhand des Inhalts von Kommentaren unter Medieninhalten ab (LEE/JANG 2010), weswegen Effekte auf folgende Interaktionen wahrscheinlicher sind, als dies bei vorhandenen Reactions der Fall ist.

Neben den hier skizzierten Spezifika der vorliegenden Erhebung ergeben sich weitere Ansatzpunkte zur Forschung, um das Potenzial von

Popularity-Cues für die Kommunikationsforschung zu vergrößern. Wir haben sowohl die technischen als auch die sozialen Ursachen möglicher Dynamiken der Entstehung von Reactions theoretisch skizziert. Es scheint plausibel, dass sich diese Prozesse im Sinne einer Reinforcing Spiral (SLATER 2007) gegenseitig befeuern. In unserem Untersuchungsdesign war eine unabhängige Untersuchung dieser beiden Faktoren nicht möglich. Um deren Einfluss unabhängig untersuchen zu können, sollten zukünftige Untersuchungen von in der Kommunikationswissenschaft bislang eher unüblichen Methoden Gebrauch machen. Drei Ansätze sind hier hervorzuheben, weil sie geeignet sind, Einflüsse von intra-, interindividuellen und technischen Faktoren zu differenzieren. Erstens bietet sich bei technischen Plattformen an, die Funktionsweise durch systematisches ›Ausprobieren‹ von geeigneten Stimuli zu rekonstruieren. Beispielsweise ließe sich der Einfluss vorangegangener Interaktionen auf die Präsenz von Statusmeldungen bestimmter Seiten testen, indem automatisierte Nutzeraccounts systematisch mit den Inhalten bestimmter Seiten interagieren und anschließend deren Newsfeeds ausgewertet werden. In der Praxis ist zwar meist die Anzahl der prüfbar Inputs begrenzt; dennoch lassen sich für viele Probleme Gegensatzkonstellationen theoretisch herleiten. In Anlehnung an die Logik des Hypothesentests können beispielsweise eine Kontroll- und eine Experimentalbedingung nebeneinander untersucht werden. Ein solches Verfahren wurde beispielsweise angewandt, um den Einsatz von Personalisierungstechnologie bei Suchmaschinen zu untersuchen (HANNÁK et al. 2017; JÜRGENS/STARK/MAGIN 2015). Wenn eine Manipulation nicht möglich oder nicht zielführend ist, kann die Experimentallogik auch mit reinen Beobachtungsdaten angewandt werden – unter der Bedingung, dass ein bekannter externer Stimulus existiert, dessen Wirkung im Datensatz abgebildet werden kann. Ein solches natürliches Experiment muss dann versuchen, Störfaktoren weitgehend zu kontrollieren (da eine Randomisierung ja fehlt). Besonders nützlich hierfür sind wiederum Daten einer unbeeinflussten ›Kontrollgruppe‹. Diese Strategie verwenden Jürgens und Stark (2017), um in einem Datensatz von Reddit-Kommentaren den Einfluss der geänderten Zusammensetzung der Startseite auf die Aufmerksamkeit der Nutzer zu prüfen. Als Stimulus dient hier die extern induzierte, zeitlich bekannte Änderung; Kontrollgruppe sind die Teile der Startseite, die nicht geändert wurden, wohingegen die Änderungen selbst die Experimentalgruppe bilden. Schließlich bieten sich in nahezu allen Fällen Simulationsstudien an. In diesen werden (meist

simplifizierte) theoretische Annahmen auf einige wenige freie Parameter reduziert, deren Interaktionen durch ›blindes‹ Ausprobieren ermittelt werden. Üblich sind hier im statistischen Bereich Monte-Carlo-Analysen (SCHARKOW/BACHL 2017) und in der analytischen Soziologie sowie der Komplexitätsforschung Agentenbasierte Modelle (ABM, vgl. WALDHERR 2014). Um die beiden Methoden zielführend und vor allem systematisch einzusetzen, dürfen die Probleme weder zu einfach sein (dann sind schließlich analytische Lösungen möglich), noch zu schwer (etwa wenn zu viele Parameter variiert werden sollen) – denn dann ist die Simulation nicht mehr zeit- und kosteneffizient realisierbar.

## 8. Ausblick: Neue methodische Herausforderungen

Seit der Durchführung der vorliegenden Datenerhebung ist der Zugriff auf digitale Spurendaten durch APIs sowohl für die Wissenschaft als auch für sonstige Akteure stark eingeschränkt worden (vgl. BRUNS 2018). Die Erhebung von digitalen Beobachtungsdaten (›digital traces‹) erfolgte auf Facebook bisher durch die Graph API. Diese bot einen kontingentierte, automatisierbaren Zugriff auf Inhalte der Plattform (mit einer begrenzten Zahl an Anfragen pro Zeiteinheit, Facebook 2018: <https://developers.facebook.com/docs/graph-api/>). Soweit diese öffentlich (also für alle Facebook-Nutzer einsehbar) sind, konnten so vormals die auf Seiten veröffentlichten Posts, dazugehörige Kommentare und Metadaten wie Likes und die restlichen Reactions erfasst werden.

Im Zuge der Reaktionen auf den sogenannten ›Cambridge Analytica‹-Skandal hat Facebook die API von Instagram vorzeitig weitgehend geschlossen;<sup>9</sup> gleichzeitig wurde der Zugriff auf die Graph-API von Facebook deutlich eingeschränkt<sup>10</sup>. Ähnliche restriktive Entwicklungen zeichnen sich bei anderen Plattformen ab<sup>11</sup>. Zwei Änderungen sind in Bezug auf Facebook besonders einschneidend: Zum einen stehen nicht wie zuvor alle

9 <https://www.instagram.com/developer/changelog/>

10 <https://developers.facebook.com/blog/post/2018/04/04/facebook-api-platform-product-changes/>

11 <https://www.theverge.com/2018/4/6/17206524/twitter-tweetbot-twiterrific-apps-features-api-changes>



historischen Posts auf Seiten zur Verfügung, sondern nur noch die jeweils letzten 600<sup>12</sup>. Zum anderen erfordert der Zugriff – selbst auf ›öffentliche‹ Daten wie Kommentare auf Medienseiten – nun ein Review und Genehmigung durch Facebook<sup>13</sup>, die zumindest nach dem Kenntnisstand der Autoren für Wissenschaftler nicht erteilt wird. Vor dem Hintergrund der konträren Motivation von Facebook und der Wissenschaft (kommerzieller Erfolg, Abwendung von finanziellem und rechtlichem Schaden vs. generalisierbare, transparente und replizierbare Erkenntnis) erscheint eine bilaterale Lösung, die die Interessen beider Parteien angemessen berücksichtigt auf absehbare Zeit unrealistisch.

Einen Kompromissweg schlugen King und Persily (2018) in Form eines unabhängigen Gremiums vor, das über eingehende Anträge auf Datennutzung entscheidet. Den Rahmen der verfügbaren Daten gibt dabei Facebook vor; zudem werden die Rohdaten vor der wissenschaftlichen Analyse (mittels Differential Privacy) anonymisiert. Diese Konstruktion ist mittlerweile unter dem Namen ›Social Science One‹ ins Leben gerufen worden<sup>14</sup>; ihre Arbeit hat die in der Wissenschaft sehr kontrovers diskutierte<sup>15</sup> Institution bislang noch nicht aufgenommen.

Klar ist, dass die Diskussion um die gesellschaftlichen Auswirkungen von Intermediären die Relevanz von entsprechender Forschung verdeutlicht und diese gleichzeitig, quasi als Kollateralschaden, drastisch einschränkt. Alternative Methoden zur Datenerhebung in sozialen Netzwerkseiten existieren zwar (vgl. FREELON 2018), sind aber umständlich und rechtlich sowie ethisch nicht unproblematisch. Somit bleibt eine offizielle Lösung für den breiten wissenschaftlichen Zugang wünschenswert.

12 <https://developers.facebook.com/docs/graph-api/reference/v3.2/page/feed> (›The API will return a maximum of 600 ranked, published posts per year.«).

13 <https://developers.facebook.com/docs/apps/review>

14 <https://socialscience.one>

15 Vgl. u. a. die Konferenz ›Locked Out of Social Platforms‹: <https://blogit.itu.dk/ics2018/>

## Literatur

- ALFERT, N.: *Facebook in der Sozialen Arbeit*. Wiesbaden [Springer] 2015.  
<https://doi.org/10.1007/978-3-658-07100-4>
- ALLFACEBOOK.DE: *Der Facebook Newsfeed Algorithmus: Die Faktoren für die organische Reichweite im Überblick*. 4.8.2017. <https://allfacebook.de/pages/facebook-newsfeed-algorithmus-faktoren> [5.4.2018]
- ALLFACEBOOK.DE: *Nutzerzahlen: Facebook, Instagram und WhatsApp, Highlights, Umsätze, uvm.* (Stand November 2017). 2.11.2017. <https://allfacebook.de/toll/state-of-facebook> [4.11.2017]
- BACHL, M.: An Evaluation of Retrospective Facebook Content Collection. In: *Open Science Framework*. 2018. <https://doi.org/10.17605/OSF.IO/6TXGE>
- BADER, C.; K. FUCHS; C. GÜTHLEIN; O. KAMM; T. SCHUNK; M. MAURER: Die Wahl in 140 Zeichen Twitter als Kommunikationsplattform für Politik, Medien und Bürger im Bundestagswahlkampf 2013. In: *Politische Psychologie*, 4, 2015, S. 5-22
- BENE, M.: Go viral on the Facebook! Interactions between candidates and followers on Facebook during the Hungarian general election campaign of 2014. In: *Information, Communication & Society*, 20(4), 2017, S. 513-529. <https://doi.org/10.1080/1369118X.2016.1198411>
- BOZDAG, E.: Bias in algorithmic filtering and personalization. In: *Ethics and Information Technology*, 15(3), 2013, S. 209-227. <https://doi.org/10.1007/s10676-013-9321-6>
- BUCHER, T.: Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. In: *New Media & Society*, 14(7), 2012, S. 1164-1180. <https://doi.org/10.1177/1461444812440159>
- BRONSON, N.; Z. AMSDEN; G. CABRERA; P. CHAKKA; P. DIMOV; H. DING et al.: [TAO]: Facebook's Distributed Data Store for the Social Graph. Presented at the 2013 Annual Technical Conference. 2013, S. 49-60. <https://www.usenix.org/conference/atc13/technical-sessions/presentation/bronson>
- BRUNS, A.: After the >APICalypse<: Social media platforms and their fight against critical scholarly research. In: *Information, Communication & Society*, 2019. <https://doi.org/10.1080/1369118X.2019.1637447>
- CHENG, X.; C. DALE; J. LIU: Statistics and Social Network of YouTube Videos. In: *2008 16th International Workshop on Quality of Service*, 2008, S. 229-238. <https://doi.org/10.1109/IWQOS.2008.32>
- CHOI, J.: Why do people use news differently on SNSs? An investigation of the role of motivations, media repertoires, and technology cluster

- on citizens' news-related activities. In: *Computers in Human Behavior*, 54, 2016, S. 249-256. <https://doi.org/10.1016/j.chb.2015.08.006>
- CHURCHILL, E. F.: Putting the Person Back into Personalization. In: *Interactions*, 20(5), 2013, S. 12-15. <https://doi.org/10.1145/2504847>
- DEVITO, M. A.: From Editors to Algorithms: A values-based approach to understanding story selection in the Facebook news feed. In: *Digital Journalism*, 5(6), 2016, S. 753-773. <https://doi.org/10.1080/21670811.2016.1178592>
- FACEBOOK: *Facebook Reactions weltweit verfügbar*. 24.2.2016. <https://de.newsroom.fb.com/news/2016/02/facebook-reactions-weltweit-verfuegbar/> [12.6.2017]
- FACEBOOK: *What does it mean to »Like« something?* | *Facebook Help Center*. 2018. <https://www.facebook.com/help/110920455663362> [30.8.2018]
- FIGUEIREDO, F.; J. M. ALMEIDA; M. A. GONÇALVES; F. BENEVENUTO: *On the Dynamics of Social Media Popularity: A YouTube Case Study*. arXiv:1402.1777 [physics]. 2014. <http://arxiv.org/abs/1402.1777>
- FUTUREBIZ: *300 Milliarden Facebook Reactions in 12 Monaten*. 24.2.2017. <http://www.futurebiz.de/artikel/300-milliarden-facebook-reactions/> [13.6.2017]
- FREELON, D.: Computational Research in the Post-API Age. In: *Political Communication*, 35(4), 2018, S. 665-668. <https://doi.org/10.1080/10584609.2018.1477506>
- GANGL, M.: Causal Inference in Sociological Research. In: *Annual Review of Sociology*, 36(1), 2010, S. 21-47. <https://doi.org/10.1146/annurev.soc.012809.102702>
- GEARHART, S.; W. ZHANG: *Gay Bullying and Online Opinion Expression: Testing Spiral of Silence in the Social Media Environment*. In: *Social Science Computer Review*, 32(1), 2014, S. 18-36. <https://doi.org/10.1177/0894439313504261>
- GILLESPIE, T.: The Relevance of Algorithms. In: GILLESPIE, T.; P. J. BOCZKOWSKI; K. A. FOOT (Hrsg.): *Media Technologies*. Cambridge/MA [The MIT Press] 2014, S. 167-194. <https://doi.org/10.7551/mitpress/9780262525374.003.0009>
- HANNÁK, A.; P. SAPIEŻYŃSKI; A. M. KHAKI; D. LAZER; A. MISLOVE; C. WILSON: *Measuring Personalization of Web Search*. arXiv:1706.05011 [cs]. 2017. <http://arxiv.org/abs/1706.05011>
- HAYES, R. A.; C. T. CARR; D. Y. WOHN: *One Click, Many Meanings: Interpreting Paralinguistic Digital Affordances in Social Media*. In:

- Journal of Broadcasting & Electronic Media*, 60(1), 2016, S. 171-187. <https://doi.org/10.1080/08838151.2015.1127248>
- HEISS, R.; D. SCHMUCK; J. MATTHES: What drives interaction in political actors' Facebook posts? Profile and content predictors of user engagement and political actors' reactions. In: *Information, Communication & Society*, 2018. <https://doi.org/10.1080/1369118X.2018.1445273>
- HOLBACH, T.; M. MAURER: Wissenswerte Nachrichten. In: *Publizistik*, 59(1), 2014, S. 65-81. <https://doi.org/10.1007/s11616-013-0191-z>
- JUNGHERR, A.; H. SCHOEN; O. POSEGGA; P. JÜRGENS: Digital Trace Data in the Study of Public Opinion: An Indicator of Attention Toward Politics Rather Than Political Support. In: *Social Science Computer Review*, 35(3), 2017, S. 336-356. <https://doi.org/10.1177/0894439316631043>
- JÜRGENS, P.; A. JUNGHERR: *A Tutorial for Using Twitter Data in the Social Sciences: Data Collection, Preparation, and Analysis (SSRN Scholarly Paper No. ID 2710146)*. Rochester/NY [Social Science Research Network] 2016 <https://papers.ssrn.com/abstract=2710146>
- JÜRGENS, P.; B. STARK: The Power of Default on Reddit: A General Model to Measure the Influence of Information Intermediaries. In: *Policy & Internet*, 9(4), 2017, S. 395-419. <https://doi.org/10.1002/poi3.166>
- JÜRGENS, P.; B. STARK; M. MAGIN: Gefangen in der Filter Bubble? Search Engine Bias und Personalisierungsprozesse bei Suchmaschinen. In: STARK, B.; D. DÖRR; S. AUFENANGER (Hrsg.): *Die Googleisierung der Informationssuche. Suchmaschinen zwischen Nutzung und Regulierung*. Berlin [de Gruyter] 2014, S. 98-135
- JÜRGENS, P.; B. STARK; M. MAGIN: Vom Öffnen der Black Box. Messung von Personalisierung in computervermittelter Kommunikation. In: MAIREDER, A.; J. AUSSERHOFER; C. SCHUMANN; M. TADDICKEN (Hrsg.): *Digitale Methoden in der Kommunikationswissenschaft*. Berlin [Digital Communication Research] 2015, S. 251-270. <https://doi.org/10.17174/dcr.v2.11>
- KEYLING, T.: *Kollektives Gatekeeping*. Wiesbaden [Springer VS für Sozialwissenschaften] 2017. <https://doi.org/10.1007/978-3-658-16686-1>
- KINCAID, J.: *EdgeRank: The Secret Sauce That Makes Facebook's News Feed Tick*. 2010. <https://techcrunch.com/2010/04/22/facebook-edgerank/> [5.4.2018]
- KING, G.; N. PERSILY: *A new model for industry-academic partnerships*. 2018. <https://gking.harvard.edu/files/gking/files/partnerships.pdf>

- KLINGLER, W.; I. TURECEK: *Medienzeitbudgets und Tagesablaufverhalten*. Ergebnisse auf Basis der ARD/ZDF-Studie Massenkommunikation 2015. In: *Media Perspektiven*, (2), 2016, S. 98-107
- LEE, E.-J.; Y. J. JANG: What Do Others' Reactions to News on Internet Portal Sites Tell Us? Effects of Presentation Format and Readers' Need for Cognition on Reality Perception. In: *Communication Research*, 37(6), 2010, S. 825-846. <https://doi.org/10.1177/0093650210376189>
- LEGEWIE, J.: Die Schätzung von kausalen Effekten: Überlegungen zu Methoden der Kausalanalyse anhand von Kontexteffekten in der Schule. In: *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 64(1), 2012, S. 123-153. doi: 10.1007/s11577-012-0158-5
- LU, H.; K. VEERARAGHAVAN; P. AJOUX; J. HUNT; Y. J. SONG; W. TOBAGUS; S. KUMAR; W. LLOYD: Existential consistency: measuring and understanding consistency at Facebook. In: *Proceedings of the 25th Symposium on Operating Systems Principles – SOSP '15*. Monterey/CA [ACM Press] 2015, S. 295-310. <https://doi.org/10.1145/2815400.2815426>
- MATTHES, J.; J. KNOLL; C. VON SIKORSKI: The »Spiral of Silence« Revisited: A Meta-Analysis on the Relationship Between Perceptions of Opinion Support and Political Opinion Expression. In: *Communication Research*, 45(1), 2018, S. 3-33. <https://doi.org/10.1177/0093650217745429>
- MC GEE, M.: *EdgeRank Is Dead: Facebook's News Feed Algorithm Now Has Close To 100K Weight Factors*. 2013. <https://marketingland.com/edgerank-is-dead-facebooks-news-feed-algorithm-now-has-close-to-100k-weight-factors-55908> [9.4.2018]
- MEIER, F.; D. C. ELSWEILER; M. L. WILSON: More than Liking and Bookmarking? Towards Understanding Twitter Favouriting Behaviour. In: *Eighth International AAAI Conference on Weblogs and Social Media*. 2014. <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8094>
- NEUBAUM, G.; N. C. KRÄMER: Monitoring the Opinion of the Crowd: Psychological Mechanisms Underlying Public Opinion Perceptions on Social Media. In: *Media Psychology*, 20(3), 2017, S. 502-531. <https://doi.org/10.1080/15213269.2016.1211539>
- NOELLE-NEUMANN, E.: The Spiral of Silence A Theory of Public Opinion. In: *Journal of Communication*, 24(2), 1974, S. 43-51. <https://doi.org/10.1111/j.1460-2466.1974.tb00367.x>

- PANG, N.; S. S. HO; A. M. R. ZHANG; J. S. W. KO; W. X. LO; K. S. Y. TAN: Can spiral of silence and civility predict click speech on Facebook? In: *Computers in Human Behavior*, 64, 2016, S. 898-905. <https://doi.org/10.1016/j.chb.2016.07.066>
- PORTEN-CHEÉ, P.; C. EILDERS: Spiral of silence online: How online communication affects opinion climate perception and opinion expression regarding the climate change debate. In: *Studies in Communication Sciences*, 15(1), 2015, S. 143-150. <https://doi.org/10.1016/j.scoms.2015.03.002>
- PORTEN-CHEÉ, P.; J. HASSLER; P. JOST; C. EILDERS; M. MAURER: Popularity Cues in Online Media: Theoretical and Methodological Perspectives in Political Communication Research. In: *Studies in Communication Media*, 2018
- SALGANIK, M. J.; P. S. DODDS; D. J. WATTS: Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. In: *Science*, 311(5762), 2006, S. 854-856. <https://doi.org/10.1126/science.1121066>
- SCHARKOW, M.; M. BACHL: How measurement error in content analysis and self-reported media use leads to minimal media effect findings in linkage analyses: A simulation study. In: *Political Communication*, 34(3), 2017, S. 323-343. doi:10.1080/10584609.2016.1235640
- SCHARKOW, M.; J. VOGELGESANG: Measuring the Public Agenda using Search Engine Queries. In: *International Journal of Public Opinion Research*, 23(1), 2011, S. 104-113. <https://doi.org/10.1093/ijpor/edq048>
- SCHMIDT, V. J.-H.: Twitter-Nutzung von Kandidierenden der Bundestagswahl 2017. In: *Media Perspektiven*, 12, 2017, S. 616-629
- SCHNAUBER, A.: Medienselektion im Alltag: Die Rolle von Gewohnheiten im Selektionsprozess. Wiesbaden [vs Verlag für Sozialwissenschaften] 2017
- SCHOBER, M. F.; J. PASEK; L. GUGGENHEIM; C. LAMPE; F. G. CONRAD: Social Media Analyses for Social Measurement. In: *Public Opinion Quarterly*, 80(1), 2016, S. 180-211. <https://doi.org/10.1093/poq/nfv048>
- SLATER, M. D.: Reinforcing Spirals: The Mutual Influence of Media Selectivity and Media Effects and Their Impact on Individual Behavior and Social Identity. In: *Communication Theory*, 17(3), 2007, S. 281-303. <https://doi.org/10.1111/j.1468-2885.2007.00296.x>
- STARK, B.; M. MAGIN; P. JÜRGENS: *Ganz meine Meinung? Informationsintermediäre und Meinungsbildung – Eine Mehrmethodenstudie*

- am Beispiel von Facebook*. LfM-Dokumentation Band 55. Düsseldorf [Landesanstalt für Medien Nordrhein-Westfalen (LfM)] 2017
- SZABO, G.; B. A. HUBERMAN: *Predicting the Popularity of Online Content* (SSRN Scholarly Paper No. ID 1295610). Rochester/NY [Social Science Research Network] 2008. <https://papers.ssrn.com/abstract=1295610>
- TRILLING, D.; P. TOLOCHKO; B. BURSCHER: From Newsworthiness to Shareworthiness: How to Predict News Sharing Based on Article Characteristics. In: *Journalism & Mass Communication Quarterly*. 2016. <https://doi.org/10.1177/1077699016654682>
- TUFEKCI, Z.: Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency. In: *Colorado Technology Law Journal*, 13, 2015, S. 203-217
- VERBAND PRIVATER RUNDfunk UND TELEMedIEN: *Mediennutzung in Deutschland 2017*. Berlin, 2018
- WALDHERR, A.: Emergence of News Waves: A Social Simulation Approach. In: *Journal of Communication*, 64(5), 2014, S. 852-873. <https://doi.org/10.1111/jcom.12117>
- WINTER, S.; C. BRÜCKNER; N. C. KRÄMER: They Came, They Liked, They Commented: Social Influence on Facebook News Channels. In: *Cyberpsychology, Behavior and Social Networking*, 18(8), 2015, S. 431-436. <https://doi.org/10.1089/cyber.2015.0005>
- WU, F.; B. A. HUBERMAN: Novelty and collective attention. In: *Proceedings of the National Academy of Sciences*, 104(45), 2007, S. 17599-17601. <https://doi.org/10.1073/pnas.0704916104>

SIMON KRUSCHINSKI / PASCAL JÜRGENS /  
BIRGIT STARK / MARCUS MAURER /  
CHRISTIAN SCHEMER

## In Search of the Known Unknowns. The Methodological Challenges in Developing a Heuristic Multi-Feature Framework for Detecting Social Bot Behavior on Facebook

*Abstract:* Recent political events such as the 2016 U.S. election and the UK Brexit campaign have shown that automated activities by social bot accounts are no longer a marginal phenomenon. Researchers and companies try to detect the opaque and dynamic computational propaganda efforts by programming or using detection methods which draw on arbitrary social bot features often in combination with machine learning algorithms. Furthermore, studies about social bots have concentrated almost exclusively on Twitter.

On the one hand, this reveals the existence of a research gap regarding social bot activities on Facebook which is the more relevant information intermediary for public, media, and political actors alike. On the other hand, this poses several methodological challenges for social bot detection since the used methods are dependent on the platform's contexts and the availability and maintenance of a high-quality gold standard of human-annotated social bots.

In this article, we want to shed light on the challenges of social bot detection on Twitter and derive methodological implications for researching automated behavior on Facebook. By drawing on a large-scale study of Facebook bots in the German federal election 2017, we propose an alter-



native, more theory-driven detection approach that focuses on text duplication as the primary social bot feature to investigate digital astroturfing campaigns which spread political content in high volume and frequency. In conjunction with large-scale data collection, stronger data sharing, and replication attempts, such analytical strategies could serve to establish long-term criteria for the identification of problematic use of platforms – whether by social bots or humans.

*Keywords:* social bots, computational methods, digital astroturfing, social network sites, German federal election, locality sensitive hashing.

## 1. Introduction

Social Network Sites (SNS) like Twitter or Facebook have changed the way how people interact, disseminate, and consume political information, thereby enabling the formation of a networked public sphere (PAPACHARISSI 2015). Ideally, people on SNS are exposed to diverse opinions, news, and ideas which should enable all users to participate in the political discourse or articulate their political interests. But the same features, algorithms, and infrastructures which promised to foster freedom of information and expression, can also be abused to accomplish goals that are dangerous or harmful to liberal democracy.

Especially the use of software automation and programmed scripts by political actors or political interest groups open up the potential of abusing the social media ecosystem for political propaganda or manipulation. One form of this computational propaganda are social bots, which are SNS accounts equipped with scripts to automate communication tasks while trying to mimic human users. They can be controlled completely or partly by computer algorithms which are programmed to fake organic, grassroots activity, so-called digital astroturfing (KOVIC et al. 2018). This includes inflating SNS users' follower counts, skewing online posts' popularity cues (such as likes, shares or comments) or spreading various political information in high volume and frequency to disrupt discussions or spread disinformation. Although the exact number of social bots on major SNS is dynamic and unknown, recent political events such as the 2016 U.S. election and the UK Brexit campaign have shown that automated activities through bot accounts are no longer a marginal phenomenon. Studies

estimate that around 9 to 15 percent of Twitter's 319 million monthly active users in 2017 (VAROL et al. 2017) and up to 9.5 percent of Instagram's one billion users in 2018 (ALBERGOTTI/KURANDA 2018) were bot accounts. Facebook (2018) revealed that 3 to 4 percent of its 2.2 billion monthly active users in 2018 were >fake accounts<.<sup>1</sup>

In the center of the growing body of research about social bots are the questions about their definition (see, e.g., GORWA/GUILBEAULT 2018; STIEGLITZ et al. 2017), algorithmic/technological features (see, e.g., BASTOS/MERCEA 2017; HEGELICH/JANETZKO 2016; SHAO et al. 2017), influence on public discourse on SNS (see, e.g., BESSI/FERRARA 2016; BRONIATOWSKI et al. 2018; ROSS et al. 2019), and harm for political participation processes like elections or referenda (see, e.g., STELLA/FERRARA/DE DOMENICO 2018; COSSU/LABATUT/DUGUÉ 2016; WOOLLEY/HOWARD 2018). All of these debates are grounded on the greatest challenge for researchers and social media companies alike: the detection of social bots. Three key issues make this a complicated endeavor: First of all, the malicious use of social bots by political actors or campaigns on SNS is covert and can be operated by a wide range of political actors. These hidden activities make the study of social bots extremely challenging especially if using established research methods. However, the researchers' access to and the amount of SNS data which is needed for more complex – so called computational – analyses is severely limited, especially for Facebook. Furthermore, social bots' increasingly dynamic and sophisticated behavior blurs the boundaries between bot- and human-like activity, exacerbating the methodological challenge of a valid separation. All in all, researchers and companies deal with these problems by programming and using tools or algorithms with questionable metrics for bot detection which often result in biased evidence for social bots' distribution or influence and misleading implications for politics and society.

In this article, we want to shed light on the contextual, platform and methodological challenges and their implications for research on social bots. We draw from a large-scale study of bot behavior on Facebook in the German federal election 2017. In a first step, we will define social bots as a way of computational propaganda and show how they are strategically

1 Facebook uses the label >fake accounts< for all accounts that violate Facebook's Community Standards in terms of >pretending to be something or someone that doesn't exist< (FACEBOOK 2018). This includes hybrid, social bot, sockpuppet, hacked, or impersonation accounts.

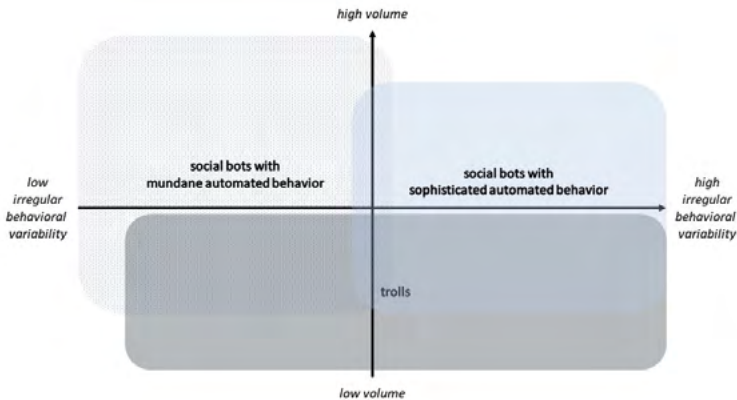
employed by political actors in election campaigns on SNS to mimic bottom-up activity by autonomous individuals. Then, we will explore existing methodological challenges of bot detection and introduce a heuristic multi-feature framework for the analysis of bot activities on Facebook. Our detection framework draws on the theoretical assumption that social bots on Facebook are programmed with certain features and patterns that reflect their programmers intentions and give preliminary insights into its development and tests. In a third step, an architecture to crawl Facebook data in a high amount and on a high frequency will be described. We will conclude with an outlook for the currently ongoing modelling of the analysis framework's remaining features and look ahead beyond our immediate findings.

## 2. Social bots as (semi-)automated software agents on SNS

The task of designing machines to imitate human behavior in such a way that humans can no longer tell the difference between a human and a machine has been a central endeavor in computer science from its outset (TURING 1950). Looking at the growing body of bot-related research varying definitions for the term social bot exist. While some authors highlight their technical structure or algorithms, others focus on intent, functions or uses (for a detailed overview, see GORWA/GUILBEAULT 2018 and STIEGLITZ et al. 2017). Basically, the word ›bot‹ is a reduction of ›robot‹ which refers to software or physical artificial agents equipped with computer instructions such as codes, scripts, or algorithms to automate tasks for saving time and energy of a human author by parsing and organizing »information at great speeds, saving human actors from doing the work« (HOWARD/WOOLLEY/CALO 2018). In this regard, bots (sometimes also called sybils) can be defined as automated software agents in online environments that are coded with algorithms to »run continuously, formulate decisions, act upon those decisions without human intervention and are able to adapt to the context they operate in« (GORWA/GUILBEAULT 2018). In the course of the changing online environment, developments in data processing and artificial intelligence (AI), the algorithms, functions, uses, and labeling of bots became more diverse (LEONARD 1997). This ranges from bots which execute unnoticed information managing tasks like gathering, indexing,

or archiving information on websites (scrapers, crawlers, spiders) to reactive social interaction tasks like responding to human requests about specific topics in narrow and primarily private online environments (chatbots).

FIGURE 1  
Categorization of automated communication behavior on SNS



With the emergence of SNS, programmers created accounts and equipped them with similar algorithms to crawl or scrape information and automate interactions with other user accounts or content on these platforms while trying to simulate human (communication) behavior (FERRARA et al. 2016; HOWARD/WOOLLEY/CALO 2018).<sup>2</sup> Those social bots can be used on behalf of human actors to observe and act upon a given SNS environment in order to achieve certain goals with different functions – either completely autonomous, semi-autonomous if automated activity is assisted by human

2 Technically speaking, coders write or access pre-made social bot scripts in any development environment and run this software on a server. In order to deploy the social bot on the SNS, the programmer connects an SNS account to the software by using the SNS's so called >Application Programming Interface< (API). This allows the social bot software to observe and communicate with the SNS environment in real time and make the connected account react to it in line with the programmed script.

intervention (›hybrids‹), or as networks (›botnets‹) in coordination with other bot accounts which are controlled remotely by connected entities.

Depending on the used scripts, we propose to categorize social bots along the two axes volume and irregular behavioral variability (see Fig. 1). Social bots with mundane automated communication behavior may consist of simple scripts which allow them to ›listen‹ to the activity on SNS and react to it with predetermined interactions or messages. They can exert basic functions, like following users, posting, liking, reporting, or sharing content at given times or as triggered responses to specific keywords, accounts, or hashtags. Often, algorithms are used to do these tasks with high volume and frequency or in coordination with other social bots which is known as spamming. To simulate human (communication) behavior, ›mundane‹ social bots are programmed with basic algorithms that make them active only at predetermined times or on a randomized schedule.<sup>3</sup> Therefore, we expect social bots with mundane automated behavior to have a rather low irregular behavioral variability and exert tasks with moderate to very high volume (see Fig. 1). With recent developments in data processing and AI more advanced social bots can be employed and execute sophisticated automated communication behavior (see Fig. 1). These functions draw on machine learning algorithms which can (half-)automatically learn to ›understand‹ and derive ›meaning‹ from human communication by analyzing websites, SNS activities, or datasets trying to adequately simulate human internet users (FREITAS et al. 2015). Especially, natural language processing algorithms could make social bots search the internet for information and enter discussions by generating thematically fitting responses or off-topic interest around popular hashtags or trending topics to make them look credible (LOWE 2019; RADFORD et al. 2019). Furthermore, they can be programmed to emulate »the human temporal signature of content production and consumption – including circadian patterns of daily activity and temporal spikes of information generation« (FERRARA et al. 2016). Thus, we assume social bots with sophisticated automated behavior to have a rather high variability of irregular behavior which is executed in moderate to medium high volume (see Fig. 1). The programming effort and skills for building these kinds of social bots are much higher than for program-

3 See Bot Repository (2019) or Oentaryo, Murdopo and Prasetyo (2016) for repositories of labelled ›mundane‹ social bots on Twitter.

ming ›mundane‹ ones, so that social bot programming with ›methods in context of machine- or deep-learning currently only play a minor role‹ (ASSENMACHER et al. 2019: 1).<sup>4</sup> But chances are quite high that this might change in future (NEUDERT 2018).

Although bots can be deployed in different online environments like community forums, chat platforms, or messaging apps, recent media coverage and research emphasize social bot activities and are therefore focused on SNSs' environments – especially on Twitter (see, e.g., KIND et al. 2017; WOOLLEY/HOWARD 2018). Three major reasons might account for that: (1) *Data access and automation through Application Programming Interfaces*: First, SNS enabled code-to-code connections and automation through third party apps or software via so-called Application Programming Interfaces (APIs). APIs act as a ›doorway‹ which allow third parties with the right ›key‹ to access, receive, and send data from or to the SNSs' databases or perform common tasks on the SNSs' front-end (e.g., following, sharing, commenting, or liking) in an automated way. While APIs were established to attract developers, who can make third party applications or software more interactive (SCHONFELD 2010), programmers started to use these ›entry points‹ to equip registered SNSs' accounts with ›social bot scripts‹ and control them in large scale with software like TweetDeck. The APIs vary in their functionality or permissiveness and have terms of services (TOS) which describe the rights and responsibilities when accessing them. Therefore, APIs lay the technical foundation for social bots and decide over their proliferation in dependence of their functionality and how SNS enforce their TOS. Thus, SNS like Twitter or Instagram are expected to have significantly more social bots than Facebook while at the same time offer better data access for researchers – due to their relatively easy-to-use API and open policies on automation (for a detailed discussion see the next section). (2) *SNS as a relevant source of information*: In the last decade, there is a relentless growth in the use of SNS worldwide so that human interaction – whether for entertainment, cultural or political purposes – increasingly takes place on these platforms. Furthermore, SNS have become an important source of news for hundreds of millions of individuals worldwide (NEWMAN et al. 2018). This presents ›real incentives – including economic and political ones – to

4 In comparison to social bots, automated communication behavior by human trolls – who sow discord on the internet by posting inflammatory or irrelevant messages to provoke an emotional response – can be expected to range from a low to high variability of irregular behavior which is executed in rather low volume (see Fig. 1).

design algorithms that exhibit human-like behavior« (FERRARA et al. 2016) and implement them on SNS for one's own benign or manipulative purposes. (3) *The algorithmic structure of SNS*: As information intermediaries, SNS filter, sort, and personalize information with the help of algorithms (JÜRGENS/STARK 2017). Based on the evaluation of hundreds of data points (e.g., user information, interaction histories, reactions of other users, features of content, etc.) these algorithms decide whether or how prominent information are ranked in every individual's news feeds, ultimately leading to a post's virality (O'BRIEN 2018). Especially, user interactions like clicks, likes, shares, and comments influence these decision-making algorithms while at the same time serve as users' indicators »of the perceived relevance and the [...] evaluation of news items and political messages« (PORTEN-CHEÉ et al. 2018: 224). Therefore, social bots' ability to game these algorithms by interacting with content in high volume and frequency can give »the false impression that some piece of information, regardless of its accuracy, is highly popular and endorsed by many« (FERRARA et al. 2016: 98) and may ultimately manipulate individual information orientation and political opinion formation.

These three reasons make SNS a fertile ground for automated behavior – especially in the realm of political campaign communication. Because whether it is printed media, radio, television, or the internet – communication channels and technological tools have ever been used for political propaganda purposes (JOWETT/O'DONNELL 2015) and scholars have been concerned with its study and effects upon public opinion – at least since the work of Lazarsfeld (1941) or Lasswell (1948). This will be highlighted in the next section, where we will discuss main strategies, functions, and possible effects of social bots for disguised propaganda purposes in election campaigns.

### 3. Social bots as disguised computational propaganda in election campaigns

With the above mentioned reasons for the increasing relevance of SNS in terms of access by third parties and as platforms for the public to interact, disseminate, and consume political information, individuals started to program social bots to undertake political tasks to the point that these automated scripts became part of contemporary political communication strategies and election campaigns worldwide (BRADSHAW/HOWARD 2018; HOWARD/WOOLLEY/CALO 2018; STELLA/FERRARA/DOMENICO 2018). Hence, social bots can be

used by political actors who represent all kinds of political viewpoints, ideologies, or even monetary interests including governments, political parties or their members and supporters, campaign strategists or campaigners, political interest or terrorist groups, media outlets or journalists, or individual citizens (WOOLLEY 2016). Politicized social bots can be used in a benign or even useful way by monitoring public information, reporting political and media malpractice, or shining light on police misconduct (DIAKOPOULOS 2018; STIEGLITZ et al. 2017). However, a growing body of research analyzes potentially harmful uses of social bots in political communication. Of particular concern, there are disguised computational propaganda activities,<sup>5</sup> which use algorithms, automation, and human curation to purposefully shape perceptions, manipulate cognitions, and direct behavior over SNS to achieve a response that promotes the desired intent of the propagandist (FARKAS/NEUMAYER 2018; JOWETT/O'DONNELL 2015; WOOLLEY/HOWARD 2017). Especially during times of electoral campaigning, when political discourse intensifies (LAZARSFELD/BERELSON/GAUDET 1944; NOELLE-NEUMANN/KEPPLINGER/DONSBACH 1999), parties or candidates enhance their range of political communication to get the widest possible attention for their agendas, policies, and goals (KEPPLINGER/MAURER 2005; MCCOMBS 2004) and political advertising become highly relevant for persuading or mobilizing the electorate (BLUMLER/KAVANAGH 1999; SCHMITT-BECK 2000). In these periods, political actors take advantage of SNSs' mode of large-scale proliferation of user-generated content, third party access, anonymity, and algorithmic logics to use social bots for disguised propaganda strategies (WOOLLEY/HOWARD 2018). One of these strategies is digital astroturfing, which is »a form of manufactured, deceptive and strategic top-down activity on the Internet initiated by political actors that mimics bottom-up activity by autonomous individuals« (KOVIC et al. 2018: 71). Within this strategy, social bots can act in the following three ways to fulfil digital astroturfing's general goals of either supporting or opposing policies or political actors:

(1) Social bots can *manipulate follower or like counts of political actors' SNS accounts*. This rather common strategy in social media marketing is based on social bots which are programmed to inflate the number of followers or

5 According to FARKAS and NEUMAYER (2018) disguised propaganda on SNS can be divided into different subcategories: While identifiable (white), obfuscated (gray), and impersonated (black) propaganda relates to the source's deception, vertical (hierarchical) and horizontal (nonhierarchical) propaganda describes the mode of organization.



likes of a target account and create the assumption that these accounts are popular or credible (CRESCI et al. 2015). As aggregated popularity indicator, the follower and like count helps informing human perception and forming personal opinion based on limited information (PORTEN-CHEÉ et al. 2018). Its manipulation can invoke or reinforce cognitive biases, where social bots could trigger bandwagon effects according to which individuals emulate the consensus among a critical mass of peers as their focal choice (SIMON 1954; SUNDAR 2008). Previous studies show, that SNSs' users perceive accounts with large numbers of followers to be more credible and trustworthy than accounts with low numbers (DE VEIRMAN/CAUBERGHE/HUDDERS 2017; LEE/SUNDAR 2013; LIN/SPENCE/LACHLAN 2016). Furthermore, this manipulation could increase political actor's perceived social network capital which affects how SNSs' users evaluate and react to their accounts' information during elections (MURTHY et al. 2016; SANDIM et al. 2018). This strategy can be executed in high volume by social bots or botnets with mundane automated behavior. To appear human-like, the involved social bots often follow each other or form a network that lends credibility to each member (YANG et al. 2019). Moreover, they might retweet, share, like, or favor tweets or posts of the target or other accounts. This aims at the second digital astroturfing effort by social bots:

(2) The *manipulation of posts' or tweets' aggregated popularity cues*. For this strategy, social bots are programmed to apply large volumes of popularity cues to political online messages and boost their total count of likes, shares, comments, or reactions in order to label them as (un)popular or to simulate public support or opposition towards them (FERRARA et al. 2016; SHAO et al. 2017). These aggregated numbers represent metric information about previous users' reactions or their evaluation of items on SNS and can be easily grasped by other users while scrolling through their news feeds (HAIM/KÜMPEL/BROSIOUS 2018). Therefore, they are important for individual information orientation or in assessing how other individuals on SNS think. The manipulation of SNS items' aggregated popularity cues by social bots may mislead attention for, selection of and reaction to affected posts or tweets (MESSING/WESTWOOD 2014; SUNDAR/KNOBLOCH-WESTERWICK/HASTALL 2007) and impact the forming or expression of personal opinions (ROSS et al. 2019; NEUBAUM/KRÄMER 2017; PORTEN-CHEÉ/EILDERS 2019). Additionally, the automated interaction with SNSs' items in high volume and frequency skew their prominence or ranking in the news feeds and thus create media effects in their own right due to the algorithmic logics of SNS (JÜRGENS/STARK 2017). This digital astroturfing strategy has also special implications for journalists

who use SNS to gather information, check other news organizations' stories, and use the quantification of likes and shares or in the form of >trending topics< to assess audience preferences and public interests for their reporting (HAIM 2019; TANDOC/VOS 2015). As with the first strategy, this technique can be deployed by rather mundane social bots who disguise in botnets or with human-like activity.

(3) Social bots can *generate and spread various political content in high volume and frequency* in order to manufacture false consensus and give the illusion that large numbers of people approve or disapprove a policy or political actor. This strategy is based on social bots' ability to create posts or tweets >on their own< and retweet or share information at given times or as triggered response to specific keywords, phrases, hashtags, or accounts (SHAO et al. 2017; WOOLLEY/HOWARD 2018). In that matter, political content of all kinds can be spread. It may be language- or picture-based, convey support or rejection in various degrees, or consist of made-up or factual information. The massive spread of these information-rich individual user reactions may open up new ways to influence selection, evaluation or opinion-forming processes (ROSS et al. 2019; THIELTGES et al. 2018). This is due to SNSs users drawing conclusions about the perceived opinion climate on a topic by assessing other users' messages or shares (LEE/JAN 2010) and, as a consequence, silence other users because of the fear of isolation (ROSS et al. 2019; SOFFER/GORDONI 2018). In addition to these effects on the perceived opinion climate, social bots' manufactured spread of certain content may influence the understanding of topics and the attitudes of recipients: Studies showed that rude and uncivil comments lead to aggressive cognitions, stereotypical or polarized attitudes (ANDERSON et al. 2014), or change the willingness to participate in a discussion (ZIEGELE et al. 2018). While mundane social bots may spread these messages in large quantities as duplicates or share them as botnets, more sophisticated social bots are not that obvious and include variations or mixed up text fragments to cover up. Last, social bots of both types can execute a combination of all three described digital astroturfing strategies.

The success of these strategies is based on the premise that human users perceive, or at least process, social bots' activity as human communication behavior. Then, the automated digital astroturfing efforts are able to exploit the vulnerabilities that stem from human's information processing, social and cognitive biases. On the one hand, there is evidence that the majority of users cannot reliably distinguish between automated generated communication and human communication (EVERETT/NURSE/EROLA 2016;

JUNG et al. 2017). On the other hand, effects are likely to underlie specific boundary conditions and as we outlined earlier, most social bots are currently programmed to execute mundane automated behavior with a rather low irregular behavioral variability and a very high volume. Therefore, we assume that this limits their potential persuasive power or as Kovic et al. (2018: 81) conclude: »Not only could there be no effect, but there could actually be a negative effect (from the point of view of the initiating actor) [...] [if] people become aware of [these] persuasion attempts and thus develop persuasion knowledge, [...] and react negatively.«

All in all, it remains difficult to evaluate behavioral, cognitive and attitudinal consequences of social bots' digital astroturfing efforts empirically. One of the most important underlying reasons for that, the methodological challenges of bot detection, will be discussed in the coming sections followed by a first insight into how we tackled some problems connected to this and developed a multi-feature framework for detecting social bot behavior on Facebook.

#### 4. Methodological challenges of social bot detection

Many authors agree that social bot detection is very difficult – especially if scientific standards of validity are to be satisfied (FERRARA et al. 2016; GORWA/GUILBEAULT 2018; TUCKER et al. 2018). This is due to the following two reasons:

(1) *The availability of SNSs' data is severely limited in access and amount:* Facebook and other platforms could offer rich opportunities for social scientific research since they have the ability to offer APIs for convenient access to behavioral data from a space with homogenous rules (such as interface design, methods of self-expression, etc.). As we've outlined earlier, limitations in access to and amount of retrievable data through the publicly available APIs yield important constraints and can, if not handled carefully, incur biased research results. On the one hand, limitations include technical challenges of public APIs which restrict the retrieval of specific data types (technical, reach, user, or search result data), in specific time periods and by the number of queries. On the other hand, legal and ethical limitations arise when scholars extract large-scale content automatically. This is because it may invade individual privacy, exerts a substantial bandwidth toll on target sites and is forbidden by almost every SNSs' TOS (FREELON 2018). These limitations cause

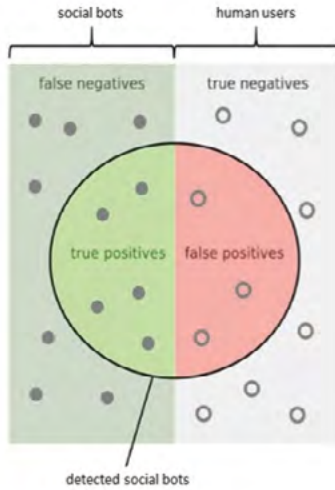
problems because a key prerequisite for social media analysis and social bot detection alike is a well-defined dataset – or at least one whose limitations are well-known. Unmanaged bias in the data collection process – whether introduced by the platform or the data collection process – has the potential to severely skew descriptive and inferential insights or hinder researchers to fully map the appearance, amount, scale, and activity of social bots. In this regard, researching social bots on Facebook poses a special problem because of its particularly restricted API access.<sup>6</sup> For example, Facebook limits the amount of data that can be collected as a function of the number of permitted requests per hour (rate limit), the number of accounts used to submit requests, the number of computers employed, the storage capacity and so on. A lack of true random sampling access places strong restrictions on generalizable inference. And finally, the high rate and frequency of social bots' automated communication behavior leads to potential temporal skew, since not every piece of information can be retrieved simultaneously. As a result »virtually all studies of bot activity have taken place on Twitter« (GORWA/GUILBEAULT 2018: 15) which is an artifact of its more accessible API.

(2) *The ›ground truth‹ problem or why any bot detection method lacks validity:* There are numerous cues, or so-called ›features‹, that can be used to differentiate between social bot and human accounts. These features can contain several descriptors of the SNS accounts' meta-data, friend network, content and language, sentiment, or temporal activity. Some of these features lend themselves more easily to analysis, while others do not; some do have meaningful theoretical underpinnings while others appear to work ›magically‹, without any obvious explanation. Many of the core ambiguities in social bot detection stem from the fact that there is no objective, agreed-upon, operational definition of a social bot and its characterizing features. This results in what can be termed the ›ground truth‹-problem of social bot detection. Because of the lack of generally valid social bot features and the dynamic of automated communication behavior researchers can never be certain that an account is truly a social bot (DAVIS et al. 2016). But the selection of accurately classified social bots is a critical step for developing and training valid detection methods or algorithms (YANG et al. 2019).

6 Recent incidents like the misuse of private data on Facebook in the so called ›Cambridge Analytica scandal‹ also refer to the access through APIs and the SNS' weak API policies (ALBRIGHT 2018). As a result, Facebook changed their policies and continuously restricted its APIs' functionalities limiting the freedom of researchers (see BRUNS 2018; FREELON 2018).

Hence, they hinge on the successful identification of social bot accounts by human coders, which are furthermore not particularly good at identifying automated accounts and behavior (EDWARDS et al. 2014; JUNG et al. 2017). For one thing, this poses a challenge for machine learning models that use human-labeled training data (DAVIS et al. 2016). For the other, it seriously limits validity of any social bot detection method, namely precision (true positives divided by detected social bots) and recall (true positives divided by all social bots) (see Fig. 2). This means that researchers have to rely on face validity in the context of social bot detection methods.

FIGURE 2  
Precision and recall in social bot detection



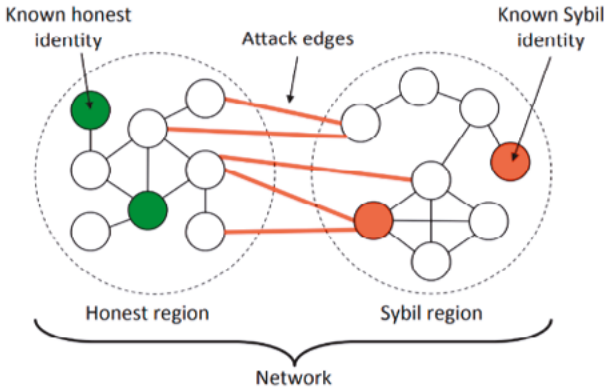
5. Social bot detection methods and their shortcomings

In light of these problems, social bot research draws on a great variety of detection techniques. Based on their underlying methods, we will explain (1) *structure-based*, (2) *crowdsourcing-based*, and (3) *feature-based* social bot detection techniques and their shortcomings in the following.

(1) *Structure-based social bot detection*: Structure-based detection methods of social bots (also graph-based detection) use (semi-)automated techniques to analyze the network structure of accounts on SNS for separating human users from social bots (BOSHMAF et al. 2013; GAO et al. 2015; GONG/FRANK/MITTAL 2014). In that regard, SNS are modelled as social graphs, with nodes representing user accounts and edges representing social relationships (see Fig. 3). This detection approach assumes that human users build big communities with strong-trust relationships on SNS and avoid interacting with unknown accounts, e.g. social bots respectively sybils. Such communities preserve a strong level of homophily tendency, meaning that two linked nodes are likely to have similar attributes. As we have outlined before, social bots connect more frequently to other bot accounts and build their own communities as they are in need of social ties to appear trustworthy and increase the possibility to interact with humans (FERRARA et al. 2016). Therefore, sybil identities often exhibit unique topological features which make them distinguishable from honest identities. Drawing on the intuition of node centrality, nodes that have proportionally smaller degree centrality indices when compared to others, represent isolated identities that are likely to be sybils (BOSHMAF/BEZNOSOV/RIPEANU 2013). This makes them visible to structure-based detection techniques. Their goal is to identify network clusters of user profiles with disproportionately high closeness and betweenness around legitimate users. For that, graph theory provides mathematical structures to model pairwise relationships between different entities.

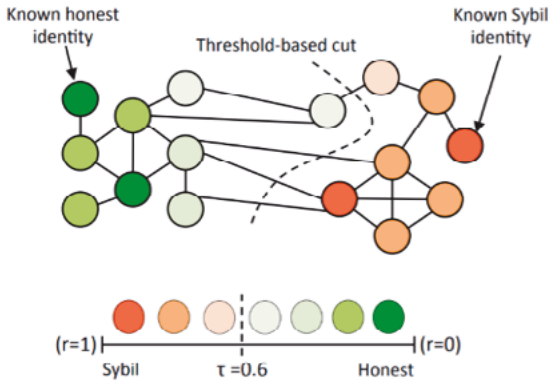
Structure-based detection techniques focus especially on so-called sybil regions (GONG/FRANK/MITTAL 2014), which are multiple social bot accounts controlled by an adversary to impersonate different identities and launch an orchestrated attack on honest users (see Fig. 3). The edges that connect the honest region and the sybil region are called attack edges (see Fig. 3). According to graph-theoretic assumptions, it is hard for attackers to set up links to honest users in a social graph where edges represent strong trust relationships between users (GAO et al. 2015). Therefore, social network-based detection techniques exploit the limited number of attack edges or relationships between sybil and honest identities (see Fig. 3). Building on graph inference, sybil identities are detected by using different mathematical approaches – for example Random Walks, Bayesian Networks, Markov Random Fields, or Loopy Belief Propagation. Figure 4 shows a structure-based detection model built by Boshmaf et al. (2013), where dark grey color indicates an identity that is less likely to be a sybil. A threshold

FIGURE 3  
The network system model with honest (left) and sybil regions (right)



Source: Boshmaf/Beznosov/Ripeanu 2013: 469

FIGURE 4  
Structure-based detection model



Source: Boshmaf/Beznosov/Ripeanu 2013: 470

partitions the network graph into two non-empty, node-disjoint subgraphs or regions: the honest and sybil region.

Although these techniques have pioneered the use of social network structure for social bot detection, there are several limitations to be aware of. First, the underlying assumption of big communities with strong trust relationships oversimplifies the social network structures in real world SNSs. For Facebook, Leskovec, Lang, Dasgupta, and Mahoney (2009) show that instead of one big community or cluster, users build multiple and hardly overlapping communities for different purposes (education, career, or geography) which might include networks with weak trust relationships. This enables an adversary to provoke attacks. Experiments on Facebook prove that over 20 percent of legitimate users accept friendship requests indiscriminately, and over 60 percent accept requests from accounts with at least one contact in common (BOSHMAF et al. 2013). For other SNSs such as Twitter, Tumblr or Instagram the connection and interaction with strangers is one of their main features. Unlike on Facebook, users on Twitter often use pseudonyms and try to increase their social capital by following back anyone who follows them (GHOSH et al. 2012). In these circumstances, large numbers of attack edges exist causing network-based detection techniques to yield high false-negative rates (missing real bots). Second, there are limits in the applicability and performance of detecting social bots with sophisticated automated behavior solely from network structure information. Alvisi et al. (2013) show that SNS contains many groups of legitimate users who fell prey to naive social bots. If more sophisticated social bots can mimic community structures' features of the honest region it is almost impossible to detect them with social network-based techniques only (FERRARA et al. 2016). Therefore, some authors (BOSHMAF et al. 2016; DAVIS et al. 2016) propose detection techniques which combine social network information with other features (see segment 3).

(2) *Crowdsourcing-based social bot detection*: Crowdsourcing-based social bot detection uses human annotators or so-called turkers<sup>7</sup> to distinguish between human and social bot accounts. Supporters of this approach assume that social bot detection is a feasible task for trained coders because they can outperform data-driven detection techniques in evaluating a wide

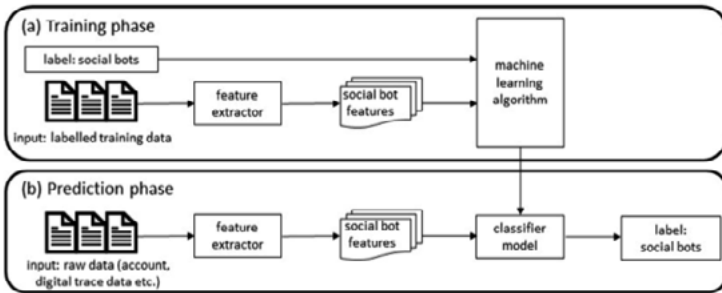
7 Turkers are human users who perform tasks on crowdsourcing platforms (e.g. Amazon Mechanical Turk) in exchange for money.



variety and dynamic change of profile information, interaction patterns, communication styles, and corresponding anomalies (see, e.g., CRUMP/MCDONNELL/GURECKIS 2013; SOUVATZIS 2018). Wang et al. (2012) state that »careful users can apply intuition to detect even small inconsistencies or discrepancies in the details of a user profile«. In the process of detection, multiple users are shown the same SNSs profiles which are labeled as real or fake accounts. Although there are parallels to content analysis, the methodologic rigor and systematic procedure is not that sophisticated. For Facebook and RenRen, Wang et al. (2012) use a large corpus of ground-truth social bots to test both the expert annotators' and turkers' efficacy at detecting bot accounts only from the profile information. The authors found that well-motivated annotators produced exceptionally good detection rates with near zero false-positives. Turkers missed more social bot accounts, but still produced near zero false-positives.

Despite its proven efficacy crowdsourcing-based social bot detection is not employed very often. This is due to the following downsides: Everett, Nurse, and Erola (2016) showed that social bot detection is a hard task for most users. Therefore, annotators require significant training to accurately detect them. Especially, the growing amount and sophistication of social bots would make crowdsourcing-based social bot detection techniques very expensive if not impractical for platforms with billions of users worldwide.

FIGURE 5  
Supervised machine learning process for social bot detection



(3) *Feature-based social bot detection*: The main idea behind feature-based social bot detection is that human users and social bots can be distinguished by key characteristics or behavioral patterns, e.g. humans produce unique genuine content or post at irregular intervals in a day-night-cycle or have an active follower network (STIEGLITZ et al. 2017). These characteristics can be encoded in features. While some companies (>botswatch<) and researchers (HOWARD/KOLLANYI 2016; KOLLANYI/HOWARD/WOOLLEY 2016; NEUDERT 2017) only use one feature like temporal tweet frequency (more than 50 tweets a day) for detecting social bots on Twitter, others use thousands of such characteristics (DAVIS et al. 2016). In the computer science literature, which arguably has shown the largest output and progress on the task of bot identification, there is a clear dominant strategy regarding the use of such features. Supervised machine learning algorithms (see Fig. 5) make use of a manually labelled training set of known social bots (the >ground truth< or >gold standard<) in order to >learn<, that is, approximate for the given data set, the relevance of all available features. This approach, which results in specific weights for each of the features, has the large benefit of making the best of all available information – which in most cases increases its performance over analyses of single types of features. Supervised methods build upon a plethora of algorithms including generalized linear models, decision trees, artificial neural networks or support vector machines (see, e.g., EFTHIMION/PAYNE/PROFERES 2018; OENTARYO/MURDOPO/PRASETYO 2016; STUKAL et al. 2017). Some of these supervised feature-based detection systems are also able to do real-time social bot detection (NASIM et al. 2018). One example is Botometer (DAVIS et al. 2016) which extracts over 1,000 characteristics (see Tab. 1) and uses state-of-the-art methods to assign scores to Twitter accounts that indicate the likelihood of a given account to be a social bot. These features are used in a >random forest< machine learning algorithm which was >trained< from a training data set by the research group's own team of human coders. Using an optimal threshold, Varol et al. (2017) measured false positive and false negative rates with Botometer at 0.15 and 0.11 respectively.

Supervised machine learning approaches learn to pick from a pre-defined set of features, but they cannot easily generalize to theory-level abstractions. This means that most of the mentioned systems have been purposefully designed and applied to one or a few datasets yielding relatively high precision and recall rates. If bot behavior changes, or if the training dataset was too small, too homogenous or unrepresentative, per-

TABLE 1  
List of features extracted by the Botometer framework

Screen name length	(***) Happiness scores of aggregated tweets		
Number of digits in screen name	(***) Valence scores of aggregated tweets		
User name length	(***) Arousal scores of aggregated tweets		
Time offset (sec.)	(***) Dominance scores of single tweets		
Default profile (binary)	(*) Happiness score of single tweets		
Default picture (binary)	(*) Valence score of single tweets		
Account age (days)	(*) Arousal score of single tweets		
User meta-data	Sentiment		
		(*) Dominance score of single tweets	
		(*) Polarization score of single tweets	
		(*) Entropy of polarization scores of single tweets	
		(*) Positive emoticons entropy of single tweets	
		(*) Negative emoticons entropy of single tweets	
		(*) Emoticons entropy of single tweets	
		(*) Positive and negative score ratio of single tweets	
		(*) Number of positive emoticons in single tweets	
		(*) Number of negative emoticons in single tweets	
		(*) Total number of emoticons in single tweets	
		Ratio of tweets that contain emoticons	
		Number of unique profile descriptions	
		(*) Profile description lengths	
		(*) Number of friends distribution	
(*) Number of followers distribution			
(*) Number of favorites distribution			
Number of friends (signal-noise ratio and rel. change)			
Number of favorites (signal-noise ratio and rel. change)			
Number of tweets (per hour and total)			
Number of retweets (per hour and total)			
Number of mentions (per hour and total)			
Number of replies (per hour and total)			
Number of retweeted (per hour and total)			
Friends (f)	Network (n)		
		Number of distinct languages	Number of nodes
		Entropy of language use	Number of edges (also for reciprocal)
		(*) Account age distribution	(*) Strength distribution
		(*) Time offset distribution	(*) In-strength distribution
		(*) Number of friends distribution	(*) Out-strength distribution
		(*) Number of followers distribution	Network density (also for reciprocal)
		(*) Number of tweets distribution	(*) Clustering coeff. (also for reciprocal)
		(*) Description length distribution	
		Fraction of users with default profile and default picture	
Content	Timing		
		(***) Frequency of POS tags in a tweet	(*) Time between two consecutive tweets
		(***) Proportion of POS tags in a tweet	(*) Time between two consecutive retweets
		(*) Number of words in a tweet	(*) Time between two consecutive mentions
		(*) Entropy of words in a tweet	

<sup>1</sup> We consider four types of connected users: retweeting, mentioning, retweeted, and mentioned.

<sup>2</sup> We consider three types of network: retweet, mention, and hashtag co-occurrence networks.

<sup>3</sup> Distribution types. For each distribution, the following eight statistics are computed and used as individual features: min, max, median, mean, std. deviation, skewness, kurtosis, and entropy.

<sup>4</sup> Part-Of-Speech (POS) tag. There are nine POS tags: verbs, nouns, adjectives, modal auxiliaries, pre-determiners, interjections, adverbs, wh-, and pronouns.

<sup>5</sup> For each feature, we compute mean and std. deviation of the weighted average across words in the lexicon.

Source: Davis et al. 2016

formance suffers. Since the programmers adapt social bots' behavior to the newest detection techniques and make them less conspicuous (HOWARD et al. 2018; NEUDERT 2018), the feature-based social bot detection systems may then have degraded performance or become unsuitable for the analysis of new datasets. Grimme et al. (2017) show that multiple features of a state-of-the-art detection tool like Botometer can be bypassed by programming social bots that exhibit human-like behavior. This is underlined by a preliminary analysis of Stukal et al. (2017) who demonstrate that their classifier loses about 20 percent of its precision if a training set from one year is applied to data from the following year, even in the same country. Moreover, Subrahmanian et al. (2016) showed that human coders hinge on the successful identification of social bot detection,

posing a further challenge for supervised machine learning models that use human-labeled training data.

In sum, we see some reason to be skeptical about the long-term benefit of arbitrary features in combination with machine learning algorithms for social bot detection. Among the inherent limitations, the dependency on the availability and maintenance of a high-quality gold standard seems most important, because these goals are especially hard to reach in this field. Not only are the mechanisms of social bots opaque and actively camouflaged; they also change frequently. Existing findings will of course remain valid, but the volatility of this ›arms race‹ interferes with an incremental progress of scientific understanding of social bots: Neither insights about the behavior nor about the prevalence of social bots are truly comparable. The context-specificity of existing machine learning models furthermore impedes cross-platform tracking of coordinated communication campaigns. Our conclusion is that machine learning might seem like the best option within the setting of a single case (i.e. a data set, a given time frame, user population, platform, content sample, and so on). In the larger context of a changing communication ecosystem, however, it seems more desirable to have less strict but more general common criteria.

Following this reasoning, we propose and advocate for an alternative approach: Drawing on theoretical reasoning about the ultimate manipulative effects of bot-transmitted communication, we derive a set of behavioral patterns which in turn lead to a framework of testable cues. This approach will be described in the following section.

## 6. Developing a heuristic multi-feature framework for detecting social bot behavior on Facebook

Given the challenges and shortcomings of social bot detection presented in the previous chapter, we give a preliminary insight into the development of our heuristic multi-feature bot detection framework for Facebook. We decided to study social bot behavior on Facebook instead of Twitter because of two main reasons: First, Facebook is the more relevant information intermediary for the public, media, and political actors alike to consume and disseminate political information – especially in Germany (NEWMAN et al. 2018). Second, since almost all social bot studies concentrate on Twitter we want to close a research gap. Therefore, we present a large-scale study of

Facebook bots in the German federal election 2017 in the following chapters. In a first step, we will briefly explain how we derived specific features and patterns from the comprehensive Twitter research literature and test for their various sensitivity to evaluate the extent to which a Facebook account exhibits similarity to the assumed social bot behavior. In a second step, we discuss the architecture of our high-frequency, high-volume crawler, which plays an integral part in the analytical framework.

## 7. Separating social bots and humans on Facebook

Our framework draws on the assumption that social bots on Facebook are programmed with certain features and patterns that reflect their mode of behavior and intentions, and that these programmed patterns lead to manifest patterns in the data which may be identified using statistical inference. The basic idea is to identify discernible cues which can be directly linked to illegitimate manipulation of communication channels, then analyze these features in combination and test for their various sensitivity to evaluate the extent to which a Facebook account exhibits similarity to the assumed social bot behavior.

The motivation for eschewing machine learning in favor of simpler metrics is perhaps best illustrated with a thought experiment. First, consider two hypothetical Facebook accounts, which both post seemingly authentic political comments but in very low volume, averaging maybe one comment per day. For the purpose of this example, we choose one of them to be a social bot while the other is a politically interested human. Given a sparsity of information (in the absence of any data outside the comments), it will be nearly impossible to discern which of them is the computer program, as long as there are no trivial linguistic or other cues. To put this observation in more generalized terms: The ability to differentiate between human and automation is determined by a trade-off between effort to detect, effort to masquerade and the volume of data about an account. Within this trade-off, methods may opt to err on any side of caution when flagging suspicious accounts: Following the basic logic of classification, a method biased towards finding the bot will in all likelihood also flag the human – a false positive. A strategy which, on the other hand, is biased towards correctly identifying the human will logically tend to miss the bot, creating a false negative.

The issue here is that in any case reported so far, humans by far outnumber bots,<sup>8</sup> and most of them have a low posting volume. False positives are therefore much more likely and more harmful than false negatives. To put it bluntly, if we use statistical tools to find low-volume social bots, then they are often not bots (false positives).

Going back to the thought experiment, we can also consider a case that is the very opposite of our first example: Take two accounts with a high posting volume of, say, 200 comments per day. Again, it is theoretically possible to program a social bot to produce this output, even with a certain face validity of the content. However, that task has certainly become much harder, because in addition to the authenticity of a single comment's text, the bot now needs to exhibit a credible variation across the comments in terms of content, spread across Facebook pages and over time. The same task is much easier for humans, who will need more time and effort than a bot (once it is programmed), but who can nevertheless reach the same output volume. In this case, the ratio of false negatives and false positives is much more favorable, since few human users are such active commenters: Automated high-volume accounts are thus more likely to be classified correctly. Then there is the additional argument regarding their effect: Even if we mislabel a human as bot, that account can still have a notable impact on readers due to its activity.

The specific power of social bots lies not only in quickly using one account, of course, but rather in an orchestration of many. Still, the observations from the second thought experiment remain valid: (1) The larger the group of accounts involved, the harder it becomes to feign authentic behavior (including interactions and timing). And (2) whatever the criterion for identification of bots, we can imagine a constellation of human users that behaves identically (for example in an online firestorm).

Together, these thought experiments illustrate that there is no empirical indicator which universally and reliably differentiates social bots from human users. They do also show that it is most fruitful, both from an empirical and from a theoretical perspective, to search for high volumes of homogenous behavior (see Fig. 1). We want to provide an additional argument for this particular focus on volume; one that does not rely on the features produced

8 Once that relation is reversed (once there are more bots than humans), these and other assumptions no longer hold. Such a case would warrant a thorough re-assessment, since, for example, bots might be predominantly interacting with bots, leading to self-reinforcing processes.

by automation but rather on the penultimate effects. Taking into account what social bots seek to achieve in digital astroturfing efforts, it becomes clear that the normative evaluation does not hinge on the technical mechanism but instead, one can argue, on the economic one. Just as it is possible to perform ›bot-like‹ communication campaigns by use of human force, we can imagine such campaigns being employed in legitimate and illegitimate ways. The Facebook pages of large newspapers, for example, frequently employ human moderators (and in some cases automated programs) which reply to user comments. The criterion that separates good from bad use is therefore not the mere use, but rather the use to bypass rules and mechanisms which safeguard healthy discussion. An actor that creates a botnet with the intention to feign an artificially inflated volume of attention (or support) for its cause (or criticism of a competing cause) violates several rules: It eschews (for whatever reason) using advertisements, the normal and legitimate form of intended mass influence. It misleads individuals by not identifying accounts as social bots. It violates the platforms' tos and (depending on the field) may also violate more specific laws such as financial, political, and commercial transparency and compliance rules. Note that especially with the increasing pressure exerted by platforms, the use of social bots is not necessarily a cheap strategy. Rather than being simply a fixed cost offset that can be scaled ›for free‹, larger bot campaigns do incur running costs. Maintenance, the re-creation of accounts that were disabled, the preparation of material, monitoring, and continuous adaptation require significant resources.<sup>9</sup> Following this reasoning, we are concerned with the superset of any kind of illegitimate digital astroturfing campaigns that follows these patterns rather than with social bots as a subtype of this.

Drawing on the existing literature and Facebook's specific architecture, we identified content duplication as the primary candidate feature to investigate. We will further elaborate on this in the following section. Beyond content duplication, we intend to extend the framework to three additional areas: A temporal analysis of posting behavior, network analysis, and topic modeling. However, because existing methods for the latter two rely on Twitter-specific data that is not available for Facebook, we will discuss them only briefly and defer a proper treatment to future work.

9 See Assenmacher et al. (2019) for a first discussion. We hope that further investigative work will uncover more of the economic details of this ecosystem of manipulation.

## 8. Content duplication as an indicator for social bots

Our first path of analysis aims at identifying manipulative actions through the content that was assumedly produced by social bots. Given that the data available from Facebook is mainly text and lacks the rich behavioral detail of Twitter data, this indirect approach seems prudent. We consider large-scale duplication of messages to be a highly valuable indicator for this, because it directly captures the intended effect (reaching many recipients, skewing the distribution of perceived opinion climate etc.; see Fig. 1) rather than proxy variables. Duplication also has the benefit of a high analytic precision, because there is no plausible cause for large-scale repetition other than conscious astroturfing.<sup>10</sup> Given the size of the dataset – approximately nine million comments during the month leading up to the election – clustering duplicates is not trivial, but still a manageable task. In principle, identifying identical comments could be as easy as computing a short, deterministic ID (known as a ›hash‹ in computer science) for the content of each one, and then comparing which comments have the same ID. However, such a naive approach would only cluster two comments if they are exactly identical, including case (upper- and lowercase), spaces, and punctuation. While some trivial bots might actually use the simple strategy of posting duplicate content, it is likely that social bots with more sophisticated automated behavior will modify their comments slightly in order to masquerade as humans. Moreover, humans mass-producing messages may re-use keywords and phrases, but will in all likelihood not copy and paste the exact same text. This makes our task somewhat harder: Instead of clustering comments through equal hashes, we need to measure their similarity. In computer science, there are several methods that are frequently used to compare text similarity. One of the simpler variants is ›edit distance‹ (LESKOVEC/RAJARAMAN/ULLMAN 2014: 95; NAVARRO 2001 shows the complexity of the task), which computes the number of charac-

<sup>10</sup> This is true for comments of sufficient length where the chance of random duplicates is low. Short comments, such as simply writing one emoji symbol or using common phrases in the vein of ›I agree‹, are an exception. We account for this by limiting our analysis to comments with a minimum length of 50 characters. Duplication furthermore captures quotations, as long as there is a negligible amount of added text. We posit that cases with a large number of similar quotations still qualify as potential coordinated activity.



ters that need to be changed in order to turn comment A into comment B. Another is the »cosine similarity« (ibid.), which represents documents as vectors in a space where each word is one dimension and compares them by measuring the angle between them. Even though these methods work well, there is a crucial issue that prevents their application to large data sets: They rely on pairwise comparisons. Picking every possible combination, even ignoring order, yields an unfeasible large number (9,000,000 over 2).

Reducing the amount of computation is possible, if one is willing to substitute probabilistic comparisons in place of exact comparisons. Instead of determining the exact similarity between two comments, it is possible to cluster them not based on their full text, but rather based on a shorter representation (again, a hash). This step reduces complexity but comes at a cost: There is a (low) probability that two items will be clustered even though they are not, in fact, similar. Likewise, there is the opposite (again, small) probability that two items will not be clustered, even though they should be. In order for the clustering to work, the hashes need to be similar when the original text is also similar. This is achieved by a method called »locality sensitive hashing« (LSH). Our analysis was performed using the »datasketch« library for python,<sup>11</sup> which implements LSH and picks optimal parameters to minimize the false positive and false negative rate.

Being able to identify different but similar comments on this large scale is a great benefit, with one caveat: The additional precision requires researchers to set a threshold at which two similar items are deemed duplicates. Since it does not present a dichotomous (identical/different) but rather a metric variable (similarity: 0-1), the cluster structure of comments depends on the similarity level at which we consider comments to be »similar enough«. We systematically tested values from 0.5 through 0.9 and manually inspected the results, settling for a similarity threshold of 0.8 which maximized variations in phrasing (capturing truly similar messages), while minimizing variations in meaning (flagging few messages that were substantially different with chance overlap).

An intuition for the impact of the threshold can be gained from a simple permutation test. For this illustration, we took a paragraph from a New York Times article. Using this text snippet (with 200 words in total) as the baseline, we iteratively replaced random words with different ones, until

11 See <https://github.com/ekzhu/datasketch>

the text was completely transformed (and unintelligible). On each step, we computed the similarity of the mutated paragraph against the original one. Once the differences were so large that the original and the mutated paragraph were placed into different clusters, we stopped the procedure and noted the number of iterations.

Because LSH compares overlapping chunks of text, it is not equally sensitive to all differences. When some modifications are made next to each other – for example by adding a short comment at the beginning or the end – then two messages might be considered identical. If, on the other hand, changes are spread rather evenly throughout the text, then they might be considered distinct. This means that the results from our permutation test will vary, which is why we ran the test 100 times for each value of the threshold.

The results show that a threshold of 0.9 is quite strict – on average, one needs to replace two words out of 200 to render two comments dissimilar. A threshold of 0.7 was instead considered too lenient: It requires an interchange of an average of four words. Taking into consideration that there is some variation, 0.8 seems an appropriate value for the type of text under consideration.

TABLE 2  
Permutation test for LSH threshold

Threshold	Iterations to divergence (average over 100 runs)	Range of iterations to divergence (min - max)
0.5	13.11	4 - 29
0.6	9.70	3 - 25
0.7	5.75	2 - 14
0.8	4.07	1 - 10
0.9	2.08	1 - 5

Here Table 2 shows an illustration from the permutation test. The original text chunk was:

»Businesses and government agencies in the United States have been targeted in aggressive attacks by Iranian and Chinese hackers who security experts believe have been energized by President Trump’s wi«

This was subsequently turned into:

»Businesses and government agencies in the United States have been targeted in aggressive attacks by Iranian and Chinese hackers who security experts believe have been energized *corporations* President Trump’s wi«

»Businesses and government agencies in the United *were* have been targeted in aggressive attacks by Iranian and Chinese hackers who security experts believe have been energized *corporations* President Trump's wi«

And a third and last change rendered the phrase so dissimilar that it was placed into separate clusters:

»Businesses and government agencies in the United *were* have been targeted in aggressive attacks *of* Iranian and Chinese hackers who security experts believe have been energized *corporations* President Trump's wi«

Preliminary results from our election dataset show that broad duplication is a very rare phenomenon. Out of the nine million comments, less than 1 percent are longer than 50 characters and similar enough to be clustered together at all. After merging clusters with more than 90 percent identical comments, we were left with 150 clusters comprising 5,088 comments. Most of these were very small, consisting of only the low-end cut-off of 20 occurrences. We consider these to be chance duplicates; they probably occur when users refer to an external source, such as when quoting media content or using common phrases. However, there are some outliers that stand out. One large group of similar comments comprises over 200 variations of a statement made by the right-wing politician Alice Weidel (from the AfD party). In the run-up to the elections, Weidel was accused to have employed an immigrant housekeeper without paying taxes,<sup>12</sup> which she denied vehemently. The duplicated statement was accompanied by calls to be spread and hence published by 130 users. Some of the comments had merely added statements at the end, while others embedded it in a block of text. Regardless of whether or not this cluster was produced by bots, there are signs of a campaign which produced a high number of focused messages across 40 Facebook pages. Other prominent clusters simply picked up the largest party's (CDU) campaign slogan ( $n=154$  occurrences, although some are clearly ironic), a polemic statement accusing Germany of disregarding historic injustice against Poland ( $n=103$ ), a link to a Wikipedia table listing German politicians who used to be members of the NSDAP ( $n=93$ ), and a sentence linking the established parties to a future ›demise‹ of the nation ( $n=85$ ).

Given these conservative but productive results, locality sensitive hashing seems to be a useful tool (out of a family of tools) that offers a robust,

12 <https://www.zeit.de/politik/deutschland/2017-09/afd-alice-weidel-asylbewerberin-schwarzarbeit>

theoretically plausible way of capturing broad-scale astroturfing campaigns, which spread political content in high volume and frequency. On the downside, LSH on its own is not able to detect social bots which manipulate follower and like counts of political actors' SNS accounts or SNS items aggregated popularity cues (see chapter 3). Therefore, we intend to extend our framework to three additional areas which we will discuss in chapter 10.<sup>13</sup>

## 9. Facebook data collection

Given our study's focus, we considered the specific characteristics of duplicate detection in order to make informed trade-offs while implementing the retrieval system. The result is a hierarchy of quality criteria that are optimized for the specific goal of identifying bots: It aims at maximizing the potential for clustering. Following is a list of the top three goals in order of descending importance, along with the trade-offs incurred.

(1) *Breadth*: The most important goal in the data collection was the breadth of collected material. Qua definition, bots aiming to seem human cannot be identified via queries of manifest features. There is no >#bot< hashtag, no publicly available flag on the account or single keyword that would yield only bot-related content. Therefore, bot accounts need to be (and are, in virtually every existing study) detected through a large-scale comparison of behavioral data from both bots and non-bots. To give an example: If there were a thousand bots that each published the same, identical slogan once, then none of them would be immediately obvious if scrutinized in isolation. Only in a clustering of the comments would the coordinated campaign become apparent. Hence, the first and foremost goal was to gather as much publicly available data as possible, from a space that was thematically as homogenous as possible. This meant crawling around 1,500 mostly German political and media pages, since these would be the most useful target audience for the German general election. Trade-off: The obvious drawback of this broad approach is its resource intensity. In contrast to smaller, more focused data collections of, say, one or two pages

13 For an overview of the field of similarity analyses, see LESKOVEC, RAJARAMAN, and ULLMAN (2014). Another interesting approach is the computer-linguistic tracing of individual phrases as LESKOVEC, BACKSTROM, and KLEINBERG (2009) have shown.

from media outlets, crawling 1,500 pages takes considerable time. Prioritizing breadth means limiting (B) frequency and (C) depth.

(2) *High Frequency*: A second limiting factor for detecting bot activity is the competing influence of different removal processes. There are several instances that may look out for, identify, and remove bot content: Facebook through automated processes and moderators, site owners via their moderation authority, users via the flagging mechanisms, and finally bot operators themselves by deleting comments or the user accounts. Each of those actors can lead to a removal of content that was previously accessible, hindering our ability to get a comprehensive collection. Therefore, it is desirable to poll Facebook's API frequently in order to detect new content as soon as possible. Due to the way user content is structured on Facebook, there are two steps that need to be performed: (1) For each page in the sample, fetch the list of newest posts and (2) for each of the N newest posts, fetch the list of newest comments until there are no new ones left. Our scheduling program was designed to run both steps independently. New posts were collected roughly every hour, which meant around 1,500 requests per hour. The posts for which we might scan for new comments are of course a much larger set, which is why the crawling was staggered: During the first day after publication of a post, new comments were fetched once every hour. During the first week after that, they were polled once every day. After that, we searched for new comments roughly every month. This, however, comes with a trade-off: Even though we are confident that the three-tiered schedule outlined above represents a good compromise in terms of efficiency and ability to catch new comments, it carries a number of specific limitations: (1) Any comment posted and deleted in between crawling intervals will not be picked up. (2) Because crawling tasks are not executed synchronously but rather spread out over time without explicit ordering, the timing of crawls will differ between posts. For example, one post's comments might have been captured for the first time two minutes after publication and then again 1.1 hours after publication, while another's schedule might rather have been 1 minute and 1.5 hours after publication. This implementation was optimized for throughput, not ordering: Rather than wait for some minutes and then perform many tasks at once, requests to the API were queued up and then executed as fast as possible, with priority given to pages first and comments second.

(3) *Depth*: Finally, within the limits allowed by maximizing the first two goals, it makes sense to gather additional information that might be useful in identifying bots. In contrast to other platforms such as Twitter,

where user profiles and friend lists are publicly accessible, Facebook has imposed strong barriers to accessing that kind of information, especially in scale. Personal information is only available to developers if they produce an app where users agree to share that data. This is not only problematic in terms of research ethics and unfeasible for the size of our dataset. It is also extremely unlikely that bot operators would agree to share the data from their fake profiles. Therefore, we limited our collection to the likes and shares of posts. Third in position in the hierarchy of goals, these were collected once approximately 15 minutes after a post's publication – if there was enough capacity available.

In summary, our data collection was designed to focus on generating a shallow but broad dataset with high frequency crawling. This is a concession to the limited set of features available through Facebook's API, which has since been reduced even further (BRUNS 2018). A more thorough description of the data collection strategy (as well as newer limitations) may be found in the chapter by Jost et al. in this volume, which is based on the same crawling methodology.

## 10. Outlook

There are three additional types of features that lend themselves to the same effect-oriented analysis as duplicate content: Temporal distributions, network features and topics. We will outline their central proposition very briefly.

*Temporal features.* Recent studies suggest that the temporal signature of bots' posting behavior can differ significantly regarding their level of sophistication (GURAJALA et al. 2016; CHU et al. 2012). Again, the basic assumption is that large-scale social bots (whether single highly active accounts or a large number of synchronized low activity accounts) will have a hard time hiding systematic patterns in their temporal behavior. A first simple criterion is the circadian rhythm formed by our day and night schedule. Simply because most people are asleep at night, the usage of social networks within the respective world regions drops during those hours. A social bot with mundane automated behavior would not pay attention to that kind of human behavior. Given that such temporal patterns are rather obvious, bot authors will try and hide by either mirroring human volume patterns or following some random model (such as a Poisson distribution).

A second type of temporal feature lies in synchronicity. Under the assumption that automated posting behavior is similar within a group of bots but different from other users, we can search for groups with low internal and high external variance. Long-term similar posting times point towards orchestration, as might sudden, synchronous spikes. Such analyses again aim at clustering users and can be performed statistically through techniques such as sequence methods (CRESCI et al. 2015) and Dynamic Time Warping (CHAVOSHI/HAMOONI/MUEEN 2016).

*Network features.* Network research showed that the structure of friend/follow relationships and interactions help significantly to detect political astroturfing (RATKIEWICZ et al. 2011). Because the Facebook API does not allow for crawling the friends of users we can only test the latter through co-posting, answering, and mentioning. In a network defined by these indicators we can identify strongly connected communities that have few connections to the rest of the user base. This is usually done by computing a ratio between the internal and external link density and optimizing the network partition so that the ratio is as high as possible. Algorithms such as modularity optimization or clique percolation are frequently used, with other approaches gaining prominence (for example spectral clustering). Thorough discussions of these methods can be found in Newman (2003), Zweig (2016), and Jürgens (2012).

*Topic features.* Following the assumption that digital astroturfing efforts intend to either support or oppose policies or political actors by faking public support, social bots may be very active only on specific sites, use specific hashtags, and post partisan content. Furthermore, social bots on Twitter use trending topics to jump on the bandwagon and maximize the visibility of their tweets because many people are tracking these specific topics at given times. Therefore, it seems prudent to employ topic modeling techniques (such as Latent Dirichlet Allocation, BLEI/NG/JORDAN 2003) to identify specific interconnected words or phrases and cluster documents related to them in different topics automatically in a huge corpus of unstructured text. Latent Dirichlet Allocation seems to be a promising technique for the analysis of topical social bot communication because of its highly interpretable topics. Furthermore, it is suitable for analyzing large corpora, like our Facebook dataset, due to dimensionality reduction as it transforms sparse document-term matrices into fixed low dimensional document-topic matrices.

## 11. Conclusion

As outlined in the first part of this paper, the literature on the characterization, use, and detection of social bots is replete with numerous very different approaches. This is not only because of the creativity of researchers, but also due to the arms race of bot creators adapting their programs to evade detection. Diverging from a clear trend, we argued for a theory-driven approach that does not primarily seek to identify mechanistic behavior for its own sake, but rather aims at isolating and looking for the effects of broad computational propaganda campaigns, which spread political content in high volume and frequency. Of special interest are so-called digital astroturfing efforts which strategically employ social bots' automated communication behavior on SNS to mimic bottom-up activity by autonomous individuals in order to either support or oppose policies or political actors and deceive the public. We discussed two main components for detecting social bots in digital astroturfing efforts which spread political content in high volume and frequency on Facebook: (1) A broad capture of a large portion of the public discussion through a custom Facebook crawling program, and (2) a computational strategy for clustering similar comments on a large scale. Our preliminary analysis of this technique found a very low amount of duplicated comments. Although, there were some statements which were adapted and spread several hundred times across the dataset, most clusters contained similar, short phrases which seem to stem from individual users' posting rather than a coordinated campaign.

Looking ahead beyond our immediate findings, we outlined several fundamental challenges that will shape the future of research on social bots. The most important insight is that researchers in this field are operating in the context of a continuing competition between bot operators (which might be state actors) and other actors trying to suppress them, most notably the platform companies Facebook and Twitter. These antagonists are locked in an arms race that fuels continuous innovation, leading to technological and operative innovations, mutation of strategies and active attempts to mask automated behavior. Those conditions not only make it difficult to accurately detect social bots at any given point: Worse, they impede incremental research progress as well as the replication and transfer of insights. Complicating this matter is the strong (excessive, we argue) reliance on machine learning methods, whose reliability hinges on the difficult task of hand-annotation. Rather than the panacea they appear



to be, these methods currently exacerbate the overly case-specific focus of most bot studies.

In presenting an alternative, more theory-driven approach that focuses on text duplication, we hope to show a more constructive way forward. In conjunction with large-scale data collection, stronger data sharing and replication attempts, such analytical strategies could serve to establish long-term criteria for the identification of problematic use of platforms – whether by social bots or humans. In this vein, we hope for other researchers to take up and extend the framework beyond the scope of the German elections, and beyond the platform of Facebook.

## References

- ALBERGOTTI, R.; S. KURANDA: *Instagram's Growing Bot Problem*. 2018. <https://www.theinformation.com/articles/instagrams-growing-bot-problem>
- ALBRIGHT, J.: *The Graph API: Key Points in the Facebook and Cambridge Analytica Debacle*. 2018. <https://medium.com/tow-center/the-graph-api-key-points-in-the-facebook-and-cambridge-analytica-debacle-b69fe692d747>
- ALLEM, J. P.; E. FERRARA: Could social bots pose a threat to public health? In: *American Journal of Public Health*, 108(8), 2018, pp. 1005-1006
- ALVISI, L.; A. CLEMENT; A. EPASTO; S. LATTANZI; A. PANCONESI: SoK: The Evolution of Sybil Defense via Social Networks. In: *2013 IEEE Symposium on Security and Privacy (SP)*: 19-22 May 2013, San Francisco, California, USA. Piscataway/NJ [IEEE] 2013, pp. 382-396. <https://doi.org/10.1109/SP.2013.33>
- ANDERSON, A. A.; D. BROSSARD; D. A. SCHEUFELE; M. A. XENOS; P. LADWIG: The »nasty effect«: Online incivility and risk perceptions of emerging technologies. In: *Journal of Computer-Mediated Communication*, 19(3), 2014, pp. 373-387. <https://doi.org/10.1111/jcc4.12009>
- ASSENMACHER, D.; L. ADAM; L. FRISCHLICH; H. TRAUTMANN; C. GRIMME: *Openbots. An empirical study on automated programs in social media*. 2019. <https://arxiv.org/pdf/1902.06691.pdf>
- BASTOS, M. T.; D. MERCEA: The Brexit botnet and user-generated hyperpartisan news. In: *Social Science Computer Review*, 0894439317734157, 2017

- BESSI, A.; E. FERRARA: Social Bots Distort the 2016 US Presidential Election Online Discussion. In: *First Monday*, 21(11), 2016. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2982233](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2982233)
- BLEI, D. M.; A. Y. NG; M. I. JORDAN: Latent dirichlet allocation. In: *Journal of Machine Learning Research*, 3, 2003, pp. 993-1022. <http://dx.doi.org/10.1162/jmlr.2003.3.4-5.993>
- BLUMLER, J. G.; D. KAVANAGH: The third age of political communication: Influences and features. In: *Political communication*, 16(3), 1999, pp. 209-230
- BOSHMAF, Y.; D. LOGOTHETIS; G. SIGANOS; J. LERÍA; J. LORENZO; M. RIPEANU; K. BEZNOSOV; H. HALAWA: Íntegro: Leveraging victim prediction for robust fake account detection in large scale OSNs. In: *Computers & Security*, 61, 2016, pp. 142-168. <https://doi.org/10.1016/j.cose.2016.05.005>
- BOSHMAF, Y.; K. BEZNOSOV; M. RIPEANU: Graph-based sybil detection in social and information systems. In: *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2013, pp. 466-473
- BOSHMAF, Y.; I. MUSLUKHOV; K. BEZNOSOV; M. RIPEANU: Design and analysis of a social botnet. In: *Computer Networks: The International Journal of Computer and Telecommunications Networking*, 57(2), 2013, pp. 556-578. <https://doi.org/10.1016/j.comnet.2012.06.006>
- BOT REPOSITORY: *Bot Repository Datasets*. 2019. <https://botometer.iuni.iu.edu/bot-repository/datasets.html>
- BRADSHAW, S.; P. N. HOWARD: *Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation*. 2018. <http://comprop.oii.ox.ac.uk/research/cybertroops2018/>
- BRONIATOWSKI, D. A.; A. M. JAMISON; S. QI; L. ALKULAIB; T. CHEN; A. BENTON; S. C. QUINN; M. DREDZE: Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. In: *American journal of public health*, 108(10), 2018, pp. 1378-1384
- BRUNS, A.: *Facebook Shuts the Gate after the Horse Has Bolted, and Hurts Real Research in the Process*. 2018. <https://medium.com/@Snurb/facebook-research-data-18662cf2cacb>
- CHAVOSHI, N.; H. HAMOONI; A. MUEEN: DeBot: Twitter Bot Detection via Warped Correlation. In: *Proceedings of 16th International Conference on Data Mining*, 2016. pp. 817-822. <https://doi.org/10.1109/ICDM.2016.0096>

- CHU, Z.; S. GIANVECCHIO; H. WANG; S. JAJODIA: Detecting automation of Twitter accounts: Are you a human, bot, or cyborg? In: *IEEE Transactions on Dependable and Secure Computing*, 2012, pp. 811-824. <https://doi.org/10.1109/TDSC.2012.75>
- CONFESSORE, N.; G. DANCE; R. HARRIS; M. HANSEN: *The Follower Factory*. 2018. <https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html>
- COSSU, J.-V.; V. LABATUT; N. DUGUÉ: A review of features for the discrimination of twitter users: Application to the prediction of offline influence. In: *Social Network Analysis and Mining*, 6(1), 2016, p. 25
- CRESCI, S.; R. DI PIETRO; M. PETROCCHI; A. SPOGNARDI; M. TESCONI: Fame for sale: Efficient detection of fake Twitter followers. In: *Decision Support Systems*, 80, 2015, pp. 56-71
- CRUMP, M. J. C.; J. V. MCDONNELL; T. M. GURECKIS: Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. In: *PLoS One*, 8(3), e57410, 2013. <https://doi.org/10.1371/journal.pone.0057410>
- DAVIS, C. A.; O. VAROL; E. FERRARA; A. FLAMMINI; F. MENCZER: BotOrNot. In: BOURDEAU, J.; J. A. HENDLER; R. N. NKAMBOU; I. HORROCKS; B. Y. ZHAO (Eds.): *WWW '16 companion: Proceedings of the 25th International Conference on World Wide Web: May 11-15, 2016, Montreal, Canada*. Republic and Canton of Geneva, New York/NY: International World Wide Web Conferences Steering Committee; ACM. 2016, pp. 273-274. <https://doi.org/10.1145/2872518.2889302>
- DE VEIRMAN, M.; V. CAUBERGHE; L. HUDDERS: Marketing through Instagram influencers: the impact of number of followers and product divergence on brand attitude. In: *International Journal of Advertising*, 36(5), 2017, pp. 798-828
- DIAKOPOULOS, N.: *Bots and the future of automated accountability*. 2018. [https://www.cjr.org/tow\\_center/prepare-to-welcome-our-accountability-bot-overlords.php](https://www.cjr.org/tow_center/prepare-to-welcome-our-accountability-bot-overlords.php)
- EDWARDS, C.; A. EDWARDS; P. R. SPENCE; A. K. SHELTON: Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on Twitter. In: *Computers in Human Behavior*, 33, 2014, pp. 372-376. <https://doi.org/10.1016/j.chb.2013.08.013>
- EFTHIMION, P. G.; S. PAYNE; N. PROFERES: Supervised Machine Learning Bot Detection Techniques to Identify Social Twitter Bots.

- In: *SMU Data Science Review*, 1(2), 2018. <https://scholar.smu.edu/cgi/viewcontent.cgi?article=1019&context=datasciencereview>
- EVERETT, R. M.; J. R. C. NURSE; A. EROLA: *The anatomy of online deception: what makes automated text convincing?* ACM. 2016. [http://dl.acm.org/ft\\_gateway.cfm?id=2851813&type=pdf](http://dl.acm.org/ft_gateway.cfm?id=2851813&type=pdf)
- FACEBOOK: *Facebook Publishes Enforcement Numbers for the First Time*. 2018. <https://newsroom.fb.com/news/2018/05/enforcement-numbers/>
- FARKAS, J.; C. NEUMAYER: Disguised Propaganda from Digital to Social Media. In: HUNSINGER, J.; L. KLAstrup; M. M. ALLEN (Eds.): *Second International Handbook of Internet Research*. Dordrecht [Springer Netherlands] 2018, pp. 1-17. [https://doi.org/10.1007/978-94-024-1202-4\\_33-1](https://doi.org/10.1007/978-94-024-1202-4_33-1)
- FERRARA, E.; O. VAROL; C. DAVIS; F. MENCZER; A. FLAMMINI: The rise of social bots. In: *Communications of the ACM*, 59(7), 2016, pp. 96-104. <https://doi.org/10.1145/2818717>
- FRELON, D.: Computational research in the post-API age. In: *Political Communication*, 35(4), 2018, pp. 665-668
- FREITAS, C.; F. BENEVENUTO; S. GHOSH; A. VELOSO: *Reverse engineering socialbot infiltration strategies in Twitter*. Paper presented at the Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Paris, France, 2015, pp. 25-32
- GAO, P.; N. Z. GONG; S. KULKARNI; K. THOMAS; P. MITTAL: *SybilFrame: A Defense-in-Depth Framework for Structure-Based Sybil Detection*. 2015. <http://arxiv.org/pdf/1503.02985v2>
- GHOSH, S.; B. VISWANATH; F. KOOTI; N. K. SHARMA; G. KORLAM; F. BENEVENUTO; N. GANGULY; K. P. GUMMADI: Understanding and Combating Link Farming in the Twitter Social Network. In: *ACM*, 2012, pp. 61-70. <https://doi.org/10.1145/2187836.2187846>
- GONG, N. Z.; M. FRANK; P. MITTAL: SybilBelief: A Semi-Supervised Learning Approach for Structure-Based Sybil Detection. In: *IEEE Transactions on Information Forensics and Security*, 9(6), 2014, pp. 976-987. <https://doi.org/10.1109/TIFS.2014.2316975>
- GORWA, R.; D. GUILBEAULT: Unpacking the Social Media Bot: A Typology to Guide Research and Policy. In: *Policy & Internet*, 40(3), 2018, p. 420. <https://doi.org/10.1002/poi3.184>
- GRIMME, C.; M. PREUSS; L. ADAM; H. TRAUTMANN: Social Bots: Human-Like by Means of Human Control? In: *Big Data*, 5(4), 2017, pp. 279-293. <http://arxiv.org/pdf/1706.07624v1>

- GURAJALA, S.; J. S. WHITE; B. HUDSON; B. R. VOTER; J. N. MATTHEWS:  
Profile characteristics of fake Twitter accounts. In: *Big Data & Society*,  
3(2), 2016, pp. 1-13. <https://doi.org/10.1177/2053951716674236>
- HAIM, M.; A. S. KÜMPEL; H. B. BROSIUS: Popularity cues in online  
media: A review of conceptualizations, operationalizations, and  
general effects. In: *SCM Studies in Communication and Media*, 7(2), 2018,  
pp. 186-207
- HAIM, M.: *Die Orientierung von Online-Journalismus an seinen Publika.  
Anforderung, Antizipation, Anspruch*. Wiesbaden [Springer vs] 2019
- HEGELICH, S.; D. JANETZKO: Are social bots on Twitter political actors?  
Empirical evidence from a Ukrainian social botnet. In: *Proceedings  
of the Tenth International Conference on Weblogs and Social Media  
(ICWSM-2016)*, Cologne, Germany. Palo Alto/CA [The AAAI Press] 2016,  
pp. 579-582
- HOWARD, P. N.; B. KOLLANYI: *Bots, #StrongerIn, and #Brexit: Computational  
Propaganda during the UK-EU Referendum*. 2016. [http://arxiv.org/  
pdf/1606.06356v1](http://arxiv.org/pdf/1606.06356v1)
- HOWARD, P. N.; S. WOOLLEY; R. CALO: Algorithms, bots, and political  
communication in the US 2016 election: The challenge of automated  
political communication for election law and administration.  
In: *Journal of Information Technology & Politics*, 15(2), 2018, pp. 81-93.  
<https://doi.org/10.1080/19331681.2018.1448735>
- HWANG, T.; I. PEARCE; M. NANIS: Social bots: Voices from the fronts.  
In: *Interactions*, 19(2), 2012, pp. 38-45
- JOWETT, G.; V. O'DONNELL: *Propaganda & persuasion* (6. Edition). Los  
Angeles, London, New Delhi, Singapore, Washington DC [SAGE] 2015
- JUNG, J.; H. SONG; Y. KIM; H. IM; S. OH: Intrusion of software robots into  
journalism: The public's and journalists' perceptions of news written  
by algorithms and human journalists. In: *Computers in human behavior*,  
71, 2017, pp. 291-298
- JÜRGENS, P.; B. STARK: The Power of Default on Reddit: A General Model  
to Measure the Influence of Information Intermediaries. In: *Policy &  
Internet*, 9(4), 2017, pp. 395-419. <https://doi.org/10.1002/poi3.166>
- JÜRGENS, P.: Communities of Communication: Making Sense of the  
>Social< in Social Media. In: *Journal of Technology in Human Services*, 30(3-4),  
2012, pp. 186-203. <https://doi.org/10.1080/15228835.2012.746079>
- KEPPLINGER, H. M.; M. MAURER: *Abschied vom rationalen Wähler. Warum  
Wahlen im Fernsehen entschieden werden*. Freiburg i.B. [Alber] 2005

- KIND, S.; M. BOVENSCHULTE; S. EHRENBURG-SILIES; T. JETZKE; S. WEIDE: *Social Bots: TA-Vorstudie. TAB-Horizon-Scanning*; Nr. 3. Berlin: Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB), 2017
- KOLLANYI, B.; P. N. HOWARD; S. C. WOOLLEY: *Bots and Automation over Twitter during the U.S. Election* (COMPROP DATA MEMO No. 2016.4). 2016. <http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2016/11/Data-Memo-US-Election.pdf>
- KOVIC, M.; A. RAUCHFLEISCH; M. SELE; C. CASPAR: Digital astroturfing in politics: Definition, typology, and countermeasures. In: *Studies in Communication Sciences*, 18(1), 2018, pp. 69-85. <https://doi.org/10.24434/j.scoms.2018.01.005>
- LAZARSELD, P. F.; B. BERELSON; H. GAUDET: *The people's choice: how voter makes up his mind in a presidential campaign*. New York [Columbia University Press] 1944
- LEONARD, A.: *Bots: The origin of new species. Computers/science & technology*. London [Penguin] 1997
- LEE, E. J.; Y. J. JANG: What do others' reactions to news on internet portal sites tell us? Effects of presentation format and readers' need for cognition on reality perception. In: *Communication Research*, 37, 2010, pp. 825-846. doi: 10.1177/0093650210376189
- LEE, J. Y.; S. S. SUNDAR: To tweet or to retweet? That is the question for health professionals on Twitter. In: *Health Communication*, 28(5), 2013, pp. 509-524
- LESKOVEC, J.; L. BACKSTROM; J. KLEINBERG: Meme-Tracking and the Dynamics of the News Cycle. In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '09*. New York, NY, USA: ACM, 2009, pp. 497-506. <https://doi.org/10.1145/1557019.1557077>
- LESKOVEC, J.; K. LANG; A. DASGUPTA; M. MAHONEY: Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. In: *Internet Mathematics*, 6(1), 2009, pp. 29-123
- LESKOVEC, J.; A. RAJARAMAN; J. D. ULLMAN: *Mining of massive datasets* (2. Edition). Cambridge [Cambridge University Press] 2014
- LOWE, R.: *OpenAI's GPT-2: the model, the hype, and the controversy*. 2019. <https://towardsdatascience.com/openais-gpt-2-the-model-the-hype-and-the-controversy-1109f4bfd5e8>

- LIN, X.; P. R. SPENCE; K. A. LACHLAN: Social media and credibility indicators: The effect of influence cues. In: *Computers in Human Behavior*, 63, 2016, pp. 264-271
- MCCOMBS, M.: *Setting the agenda: The mass media and public opinion*. Cambridge/UK [Polity Press] 2004
- MESSING, S.; S. J. WESTWOOD: Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. In: *Communication Research*, 41(8), 2014, pp. 1042-1063
- MURTHY, D.; A. B. POWELL; R. TINATI; N. ANSTEAD; L. CARR; S. J. HALFORD; M. WEAL: Automation, algorithms, and politics| bots and political influence: A sociotechnical investigation of social network capital. In: *International Journal of Communication*, 10, 2016, pp. 4952-4971
- NAVARRO, G.: A guided tour to approximate string matching. In: *ACM Computing Surveys*, 33(1), 2001, pp. 31-88. <https://doi.org/10.1145/375360.375365>
- NASIM, M.; A. NGUYEN; N. LOTHIAN; R. COPE; L. MITCHELL: *Real-time Detection of Content Polluters in Partially Observable Twitter Networks*. 2018. <https://arxiv.org/pdf/1804.01235.pdf>
- NEUBAUM, G.; N. C. KRÄMER: Monitoring the opinion of the crowd: Psychological mechanisms underlying public opinion perceptions on social media. In: *Media Psychology*, 20(3), 2017, pp. 502-531
- NEUDERT, L.-M.: Computational Propaganda in Germany: A Cautionary Tale. In: *Computational Propaganda Research Project Working Paper*, 2017(7), 2017. [http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Germany\\_LMNeditv3.pdf](http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Germany_LMNeditv3.pdf)
- NEUDERT, L.-M.: Future elections may be swayed by intelligent, weaponized chatbots: The AI advances that brought you Alexa are teaching propaganda how to talk. In: *MIT Technology Review Magazine*, 121(5), 2018. <https://www.technologyreview.com/s/611832/future-elections-may-be-swayed-by-intelligent-weaponized-chatbots/> amp/
- NEWMAN, M. E. J.: The Structure and Function of Complex Networks. In: *SIAM Review*, 45(2), 2003, pp. 167-256. <https://doi.org/10.1137/S003614450342480>
- NEWMAN, N.; R. FLETCHER; A. KALOGEROPOULOS; D. LEVY; R. K. NIELSEN (Eds.): *Reuters Institute Digital News Report 2018*, 2018. <http://media>.

digitalnewsreport.org/wp-content/uploads/2018/06/digital-  
news-report-2018.pdf?x89475

- NOELLE-NEUMANN, E.; H. M. KEPPLINGER; W. DONSBACH: *Kampa: Meinungsklima und Medienwirkung im Bundestagswahlkampf 1998*. Freiburg i. Br. [Alber] 1999
- O'BRIEN, M.: *How does the Facebook News Feed work? An interview with Dan Zigmond, head of Facebook News Feed analytics*. 2018. <https://milesobrien.com/how-does-the-facebook-news-feed-work-an-interview-with-dan-zigmond-head-of-facebook-news-feed-analytics/>
- OENTARYO, R. J.; A. MURDOPO; P. PRASETYO: *On profiling bots in social media* (Vol. 10046). Cham [Springer International Publishing] 2016. [https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=4650&context=sis\\_research](https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=4650&context=sis_research)
- PAPACHARISSI, Z.: *Affective publics: Sentiment, technology, and politics*. Oxford [Oxford University Press] 2015
- PORTEN-CHEÉ, P.; J. HASSLER; P. JOST; C. EILDERS; M. MAURER: Popularity cues in online media: Theoretical and methodological perspectives in political communication research. In: *Studies in Communication and Media*, 7, 2018, pp. 208-230
- PORTEN-CHEÉ, P.; C. EILDERS: The effects of likes on public opinion perception and personal opinion. In: *Communications*, 2019. <https://www.degruyter.com/view/j/comm.ahead-of-print/commun-2019-2030/commun-2019-2030.xml>
- RADFORD, A.; J. WU; R. CHILD; D. LUAN; D. AMODEI; I. SUTSKEVER: *Language Models are Unsupervised Multitask Learners*. 2019. [https://d4mucfpsywv.cloudfront.net/better-language-models/language\\_models\\_are\\_unsupervised\\_multitask\\_learners.pdf](https://d4mucfpsywv.cloudfront.net/better-language-models/language_models_are_unsupervised_multitask_learners.pdf)
- RATKIEWICZ, J.; M. D. CONOVER; M. MEISS; B. GONCALVES; A. FLAMMINI; F. MENCZER: Detecting and tracking political abuse in social media. In: *Proceedings of the 5th AAAI International Conference on Weblogs and Social Media (ICWSM'11)*, 2011, pp. 297-304
- ROSS, B.; L. PILZ; B. CABRERA; F. BRACHTEN; G. NEUBAUM; S. STIEGLITZ: Are social bots a real threat? An agent-based model of the spiral of silence to analyse the impact of manipulative actors in social networks. In: *European Journal of Information Systems*, 2019, pp. 1-19
- SANDIM, H.; D. AZEVEDO; A. P. C. DA SILVA; M. M. MORO: *The Role of Social Capital in Information Diffusion over Twitter: a Study Case over Brazilian posts*. 2018. <http://ceur-ws.org/Vol-2247/poster11.pdf>



- SCHMITT-BECK, R.: *Politische Kommunikation und Wählerverhalten. Ein internationaler Vergleich*. Wiesbaden [vs Verlag für Sozialwissenschaften] 2000
- SCHONFELD, E.: *Zuckerberg: »We Are Building A Web Where The Default Is Social«*. 2010. <https://techcrunch.com/2010/04/21/zuckerbergs-buildin-web-default-social/>
- SHAO, C.; G. L. CIAMPAGLIA; O. VAROL; K. YANG; A. FLAMMINI; F. MENCZER: *The spread of low-credibility content by social bots*. 2017. <http://arxiv.org/pdf/1707.07592v4>
- SIMON, H. A.: Bandwagon and underdog effects and the possibility of election predictions. In: *Public Opinion Quarterly*, 18(3), 1954, pp. 245-253
- SOFFER, O.; G. GORDONI: Opinion expression via user comments on news websites: Analysis through the perspective of the spiral of silence. In: *Information, communication & society*, 21(3), 2018, pp. 388-403. <https://doi.org/10.1080/1369118X.2017.1281991>
- SOUVATZIS, V.: *Fraudulent social media users detection by a supervised machine learning technique*. 2018. <https://ikee.lib.auth.gr/record/297303/files/GRI-2018-21332.pdf>
- STARK, B.; M. MAGIN; P. JÜRGENS: *Ganz meine Meinung? Informationsintermediäre und Meinungsbildung – eine Mehrmethodenstudie am Beispiel von Facebook*. LfM-Dokumentation: Band 55. Düsseldorf: Landesanstalt für Medien Nordrhein-Westfalen (LfM). 2017
- STELLA, M.; E. FERRARA; M. DE DOMENICO: Bots increase exposure to negative and inflammatory content in online social systems. In: *Proceedings of the National Academy of Sciences*, 115(49), 2018, pp. 12435-12440
- STIEGLITZ, S.; F. BRACHTEN; D. BERTHELE; M. SCHLAUS; C. VENETOPOULOU; D. VEUTGEN (Eds.): *Do Social Bots (Still) Act Different to Humans? – Comparing Metrics of Social Bots with Those of Humans*. 2017. <https://www.springerprofessional.de/do-social-bots-still-act-different-to-humans-comparing-metrics-0/12284366>
- STIEGLITZ, S.; F. BRACHTEN; B. ROSS; A.-K. JUNG: *Do Social Bots Dream of Electric Sheep? A Categorisation of Social Media Bot Accounts*. 2017. <http://arxiv.org/pdf/1710.04044>
- STUKAL, D.; S. SANOVICH; R. BONNEAU; J. A. TUCKER: Detecting Bots on Russian Political Twitter. In: *Big Data*, 5(4), 2017, pp. 310-324. <https://doi.org/10.1089/big.2017.0038>

- SUBRAHMANIAN, V. S.; A. AZARIA; S. DURST; V. KAGAN; A. GALSTYAN; K. LERMAN; L. ZHU; E. FERRARA; A. FLAMMINI; F. MENCZER: The DARPA Twitter Bot Challenge. In: *Computer*, 49(6), 2016, pp. 38-46. <https://doi.org/10.1109/MC.2016.183>
- SUNDAR, S. S.: The MAIN model: A heuristic approach to understanding technology effects on credibility. In: METZGER, M. J.; A. J. FLANAGIN (Eds.): *Digital media, youth, and credibility*. Cambridge/MA [The MIT Press] 2008, pp. 73-100
- SUNDAR, S. S.; S. KNOBLOCH-WESTERWICK; M. R. HASTALL: News cues: Information scent and cognitive heuristics. In: *Journal of the American Society for Information Science and Technology*, 58(3), 2007, pp. 366-378
- TANDOC, E. C. J.; T. P. VOS: The journalist is marketing the news. Social media in the gatekeeping process. In: *Journalism Practice*, 10(8), 2015, pp. 950-966. <https://doi.org/10.1080/17512786.2015.1087811>
- THIELTGES, A.; O. PAPAKYRIAKOPOULOS; J. C. M. SERRANO; S. HEGELICH: *Effects of Social Bots in the Iran-Debate on Twitter*. 2018. arXiv preprint arXiv:1805.10105
- TUCKER, J. A.; A. GUESS; P. BARBERÁ; C. VACCARI; A. SIEGEL; S. SANOVICH; D. STUKAL; B. NYHAN: *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature*. 2018
- TURING, A.: Computing Machinery and Intelligence. In: *Mind*, 59(236), 1950, pp. 433-460. <http://phil415.pbworks.com/f/TuringComputing.pdf>
- VAROL, O.; E. FERRARA; C. A. DAVIS; F. MENCZER; A. FLAMMINI: *Online Human-Bot Interactions: Detection, Estimation, and Characterization*. 2017. <http://arxiv.org/pdf/1703.03107v2>
- WANG, G.; M. MOHANLAL; C. WILSON; X. WANG; M. METZGER; H. ZHENG; B. Y. ZHAO: *Social Turing Tests: Crowdsourcing Sybil Detection*. 2012. <http://arxiv.org/pdf/1205.3856v2>
- WEEDON, J.; W. NULAND; A. STAMOS: *Information Operations and Facebook*. 2017. <https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf>
- WOOLLEY, S. C.: Automating power: Social bot interference in global politics. In: *First Monday*, 21(4), 2016. <https://doi.org/10.5210/fm.v21i4.6161>
- WOOLLEY, S. C.; P. N. HOWARD: Social Media, Revolution, and the Rise of the Political Bot. In: ROBINSON, P.; P. SEIB; R. FRÖHLICH (Eds.): *Routledge*

- handbooks. Routledge handbook of media, conflict and security*. London, New York [Routledge] 2017, pp. 282-292
- WOOLLEY, S. C.; P. N. HOWARD: *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*. Oxford [Oxford University Press] 2018
- YANG, K. C.; O. VAROL; C. A. DAVIS; E. FERRARA; A. FLAMMINI; F. MENCZER: Arming the public with artificial intelligence to counter social bots. In: *Human Behavior and Emerging Technologies*. 2019. <https://onlinelibrary.wiley.com/doi/full/10.1002/hbe2.115>
- ZIEGELE, M.; M. WEBER; O. QUIRING; T. BREINER: The dynamics of online news discussions: Effects of news articles and reader comments on users' involvement, willingness to participate, and the civility of their contributions. In: *Information, Communication & Society*, 21, 2018, pp. 1419-1435. <https://doi.org/10.1080/1369118X.2017.1324505>
- ZWEIG, K. A.: *Network analysis literacy: a practical approach to the analysis of networks*. Wien [Springer] 2016

DANIELA STOLTENBERG / DANIEL MAIER

## Descriptive Methods for Investigating Dynamics in Online Networks

*Abstract:* The importance of communication networks and their dynamics has become more obvious with the rise of digital technologies. Still, network analysis methods have only recently become more common in the field of communication research and most of this work treats the networks as static systems. However, static network snapshots permit only limited insights into the dynamics that drive networks. Our essay provides a selective overview of methods that are useful in describing dynamic communication networks. The type of communication network (i.e., instantaneous contact networks vs. more enduring interval networks) and the temporal research focus (i.e., long-term dynamics vs. short-term processes running on the network) influence the choice of network analytical methods. Some methods and measures from the standard network analysis toolkit can (cautiously) be used in combination with time-series analysis for dynamic networks. Others are unique to dynamic network analysis and specifically require a fine-grained time-ordered data structure. A dynamic framework of time-ordered networks allows researchers to correctly portray, e.g., how content travels through a network.

*Keywords:* social network analysis, dynamic network, network evolution, networked process

## 1. Introduction

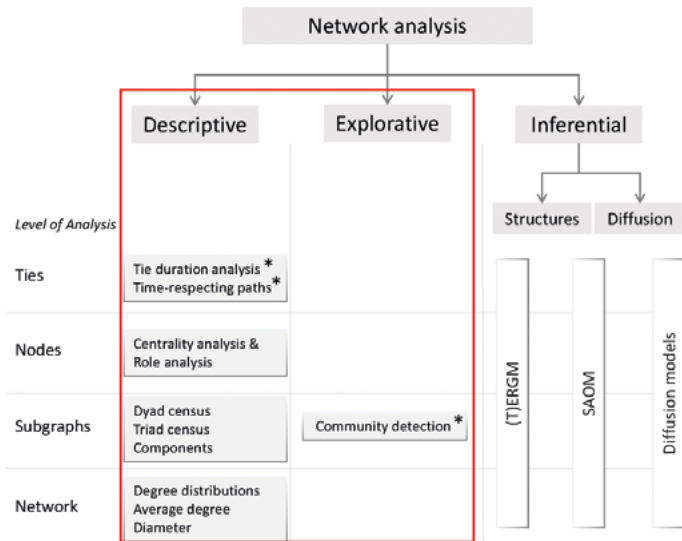
Communication networks are by their very nature dynamic systems (SNIJDERS/VAN DE BUNT/STEGLICH 2010), which constitute the arenas for processes such as agenda setting or the formation of public opinion (HOLME/SARAMÄKI 2012; PANZARASA/OPSAHL/CARLEY 2009). From the perspective of networked public spheres (BENKLER 2006; NEUBERGER 2009) on the web, for example, actors and issues may be in the spotlight of public attention for some time, then disappear, only to gain salience again later. Similarly, actors will not always be equally well-connected to one another. Network analysis has become fairly ingrained in communication science for the investigation of these phenomena (e.g., ADAMIC/GLANCE 2005; BENNETT/SEGERBERG 2013; HIMELBOIM/MCCREERY/SMITH 2013; HINDMAN 2008; NUERNBERGK 2013; SHUMATE 2012). However, much of the time the networks underlying these dynamic processes are studied as if they were static (HOLME/SARAMÄKI 2012; PANZARASA et al. 2009).

Although interesting, such investigations of singular network *snapshots* usually do not offer insights into either the generation or the consequences of network structures because different processes »can give rise to the same static graph« (BERGER-WOLF/SAIA 2006: 523). Furthermore, studying dynamic processes by using static graphs implies critical assumptions, including a fixed network topology, i.e., a fixed set of actors communicating with the same others all the time (BLONDER et al. 2012: 958f.). This implicit assumption is problematic, particularly online, where ever-faster, more interactive means of communication on the social web result in highly dynamic networks. In their literature review, Nuernbergk and Neubarth (2014: 276) identified dynamics as one of the central issues that network analysis for social scientific online research faces. But while computer science and related disciplines have made substantial progress in the study of dynamic networks in recent years, little of this work has made it into the social sciences thus far. Our aims are therefore to explicate what constitutes dynamic communication networks online and to provide an overview of the available methods and tools as well as the challenges of analysis.

A general distinction can be made between descriptive and explorative techniques of network analysis and those aimed at causal inference (see Fig. 1). The aim of descriptive methods and measures, on the one hand, is, of course, to describe a network's structure or – in the case of dynamic network analysis more particularly – to describe how network properties change over time.

This category contains analytical techniques for time-series analysis over metrics that are also used for the analysis of static networks, as well as dynamic measures, such as contact/tie duration and time-respecting paths, i. e., paths that allow for flow in a dynamic network. It must be stressed that, from our perspective, some techniques still fall into the category of descriptive or explorative methods despite relying on inferential calculations, e.g., trend analysis (univariate time-series analysis) or comparisons of means across time. As their aim is to describe whether and how the network structure changes, but not to uncover causes for or effects of such changes. Therefore, we refer to them as instances of descriptive network analysis.

FIGURE 1  
Systematization of dynamic network analysis procedures



Note: Descriptive and explorative analyses refer to procedures aiming to characterize changes in the network empirically. Inferential techniques may still be used here, e.g., univariate time-series analyses or comparisons of means. Techniques marked with an asterisk (\*) are unique to dynamic network analysis, while the other procedures are combinations of methods from (static) network analysis with time-series analysis. Regarding community detection, techniques exist that are unique to dynamic networks. However, methods for static networks have also been applied to dynamic networks. With inferential network analysis techniques we refer to techniques aimed at causal inference, i.e., either identifying the causes for or consequences of changes in the network structure over time. This latter category is not discussed in more detail in this essay (for further information see references in Footnote 1).

Inferential techniques, on the other hand, encompass a range of statistical models for hypothesis testing (CRANMER et al. 2017). Within this category, two branches of models can be differentiated. The first branch of models tackles the challenge of investigating the formation of networks, i.e., what are the underlying processes that lead to the empirically observed relational structure of a network? Put differently, these models focus on social selection processes (FRIEMEL 2015). For the analysis of dynamic networks, the most prominently used model in the social sciences is the Stochastic Actor Oriented Model (SAOM) as proposed by Tom Snijders (1996; see also SNIJDERS/VAN DE BUNT/STEGELICH 2010). Recently, longitudinal extensions for the Exponential Random Graph family of models (ERGM) were developed (HANNEKE/FU/XING 2010; LEIFELD/CRANMER/DESMARAIS 2017; KRIVITSKY/HANDCOCK 2014). In contrast, the second branch of models concentrates on social influence processes in networks, such as the diffusion of behavior, attitudes, or any other type of actor attributes (FRIEMEL 2010).

Interestingly, the majority of dynamic network analysis applications in the social sciences literature currently rely on inferential techniques while descriptive measures and methods are leading a relatively obscure existence. Although inferential techniques are at the forefront of methodological development,<sup>1</sup> little has been written and summarized about descriptive methods of dynamic network analysis. We believe that this blind spot in the literature is worth illuminating. The descriptive methods are important for gaining a fundamental understanding of the dynamics of networked communication on its own. Additionally these methods may be regarded as the first steps in dynamic network analysis. Further inferential analyses may have to take into account what has been found with descriptive analysis in the first place. We believe that even research that ultimately aims for causal inferences has to describe its networks carefully. Therefore, it is our aim here to lay out an overview of these so far neglected descriptive techniques of dynamic network analysis.

The remainder of this essay therefore proceeds by discussing predominantly the descriptive set of methods in more detail, pointing out usage

1 Cranmer et al. (2017) provide a brilliant overview of inferential methods for the analysis of social networks; Shumate and Pallazzo (2010) introduce the Exponential Random Graph Models to the analysis of communication networks. For a deep and thorough understanding and a guideline for application we point readers to the book by Lusher, Koskinen, and Robins (2013).

contexts and available measures and techniques. First, however, it is necessary to explicate what we regard as a dynamic (communication) network and which different types of dynamic networks exist. For the sake of brevity and clearness, we will mostly explicate our considerations on Twitter network examples, although the properties under discussion can be generalized to other types of online networks, too.<sup>2</sup> It is worth pointing out that some basic knowledge about social network analysis is helpful for the reading of this essay. However, when we refer to terms and concepts which are unique to network analysis, short descriptions and/or suggestions for further reading are provided.

## 2. What constitutes a dynamic network data set?

while the question of what a dynamic network is seems intuitive at first glance, the answer is anything but trivial. Due to the interdisciplinary nature of network analysis, a multitude of different labels have been used to describe dynamic networks, including longitudinal networks, temporal networks, graphs over time, evolving graphs, and time-varying graphs (CASTEIGTS et al. 2012; HOLME/SARAMÄKI 2012). The cause for this variation lies in different strands of literature from different disciplinary contexts, but all broadly describe the same phenomena. Beyond semantics, different types of dynamics must be distinguished.

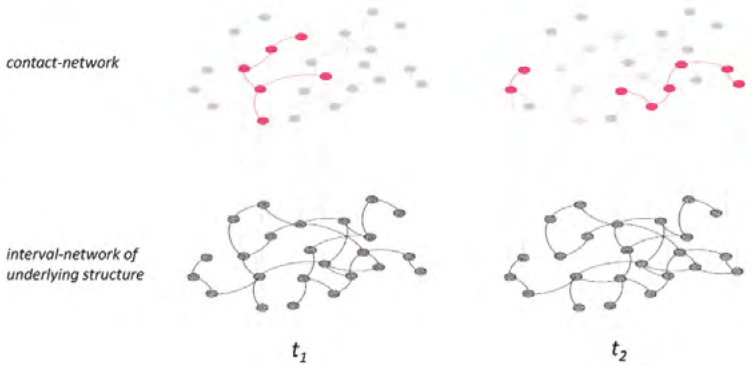
From an abstract analytical point of view, a network is constituted by a set of nodes or actors and the relations or ties of interest between them (WASSERMAN/FAUST 2009: 5). Additionally, actors may have metrically or categorically defined attribute variables attached to them, such as socio-demographic or behavioral information. Depending on the nature of the network under study, each element (i.e., actors, attributes, and relations) may be prone to change or persistence over time. For instance, little change of the set of actors may be expected in the friendship network of students from one school class, while the set of Twitter users that is involved in a public debate network (e.g., all users who tweet under a common hashtag) can be fluctuating massively over time. Actor attributes like gender or re-

2 Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 290045248 - SFB 1265.



ligious affiliation may be much more stable over time than, e.g., political attitudes. Most importantly, the type of relation under study may be either relatively instantaneous or durable (HOLME/SARAMÄKI 2012). A handshake among two people or one Twitter user liking the tweet of another are examples for instantaneous relations, while friendships or follower relationships on Twitter, for example, last for longer periods. We will focus on the variation of different types of relations over time here.

FIGURE 2  
**Contact network and underlying interval network structure**



Note: The figure displays a fictitious (multiplex) network for two time steps. While the networks in the lower row represent the periodically lasting interval networks of, e.g., friendship/following relationships, the upper row networks indicate the instantaneous contact connections, e.g., mention and retweet-connections between users. The interval network remains similar over the course of two time steps, while the contact networks of  $t_1$  and  $t_2$  are likely completely different.

A first important difference is that between *contact graphs* and *interval graphs* (HOLME/SARAMÄKI 2012). Contact graphs, on the one hand, are networks that can be represented as a sequence of instantaneous contacts with a negligible duration (HOLME/SARAMÄKI 2012). Most networks where relations represent instances of communication fall into this category, including e-mail networks, instant messaging networks, or retweet/reply/mention networks on Twitter. These relations can be denoted as a set of  $C$  contacts consisting of triples  $(i, j, t)$  where  $i$  and  $j$  represent the actors of the network and  $t$  is the precise time stamp of the contact. Interval graphs, on the other hand, are networks that can be represented as a sequence of

intervals representing periods of a relation's activity. Here, the duration of contact is substantial and cannot be treated as instantaneous (HOLME/SARAMÄKI 2012). Hyperlink networks or social media friendship/follower networks fall into this category. Instead of timestamps, time period intervals indicate activity  $T_e = \{(t_1, t'_1), \dots, (t_n, t'_n)\}$  where  $T_e$  is the set of activity intervals of relation  $e$ . The tuple elements of the set in parentheses  $(t_n, t'_n)$  indicate the start and end of the  $n$ th period of activity.

Some communication networks, predominantly in the realm of social media, feature observable contact and interval relation sets simultaneously (see Fig. 2). Such networks are called multiplex networks, i.e., they describe »the extent to which two actors are linked together by more than one relationship« (MONGE/CONTRACTOR 2003: 31). On Twitter, there may be a set of follower relations that structure communication. Follower relations are usually active for substantial periods of time because the formation and dissolution of these relations occur comparatively rarely (KWAK/CHUN/MOON 2011). The topology of follower relations strongly guides the flow of messages between users. In this example, the instantaneous flow of messages can be traced by focusing on communication events between users, i.e., when one user retweets, mentions or replies to the message of another user. Instantaneous communication between users can be reflected in a contact network.<sup>3</sup> Both, the topology of the interval and the contact network, where the latter reflects flow phenomena, are subject to dynamic changes, albeit at different rates (GHOSH/LERMAN 2012; LESKOVEC/KLEINBERG/FALOUTSOS 2005). Both are interdependent and feedback mechanisms may exist (BLONDER et al. 2012: 960). On Twitter, for example, one user may follow another after seeing retweets of their messages.

The distinction between contact and interval networks is embedded in the nature of underlying relations themselves. With this in mind, we can now focus on differences in data collection and preprocessing strategies. Regarding data collection, we must differentiate between continuous and snapshot approaches. In data preprocessing, a basic distinction can be drawn between time-ordered and time-aggregated networks.

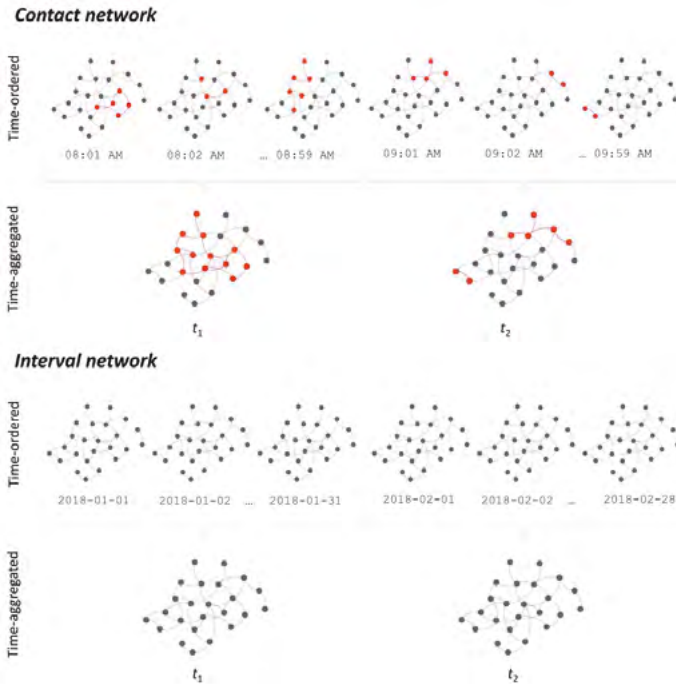
3 The term »contact network« merely denotes that the relation under study is instantaneous. Contact, in the common sense of the term, between Twitter users can already be assumed if one user reads another one's messages. However, not every contact leads to an instance of flow, i.e., the behavior of actor  $i$  influencing the behavior of actor  $j$ .

First, let us turn to data collection. Ideally, researchers would want to collect data *continuously* over the whole investigation period, that is to say, observe for each tie exactly when it is activated and – in the case of interval networks – deactivated. Such an approach will usually be necessary to collect valid representations of contact networks and in the case of social media platforms can oftentimes be achieved via the application programming interfaces (APIs; e.g., Twitter’s Streaming API delivers continuous representations of communication instances). Continuous network data is usually harder to acquire for interval networks, such as hyperlink and follower networks. Instead, such networks can more feasibly be collected as *snapshots*, i.e., data is collected at multiple discrete points in time equidistantly, for example, once per day for social media networks (GONG et al. 2012 for Google+; KWAK/CHUN/MOON 2011 for Twitter). While researchers may miss some data on ties that are created and immediately dissolved again between two consecutive snapshots and lack some information on the exact timing of changes, this approach generally creates valid data for rather slowly changing networks. In contrast, taking a snapshot of a contact network would likely result in sparse network data, as it would capture only the interactions taking place at the exact moment of data retrieval. Here, only a continuous data collection approach will yield valid data.

Secondly, different strategies are possible regarding data preprocessing. Assuming researchers had an ideal longitudinal dataset of a network that was collected continuously over time, how may they analyze it? Broadly speaking, one could either retain all details of the temporal structure which can be intricate, or one could choose a more manageable aggregate approach which comes at the cost of losing some information about network dynamics. BLONDER et al. (2012) differentiate between the frameworks of *time-ordered* vs. *time-aggregated* networks (Fig. 3). Time-ordered networks describe the data as a sequence of interactions, retaining the complete information on ordering, duration, and timing of events: »Thus, the time-ordered network is a complete record of all observed interactions at all time points« (BLONDER et al. 2012: 960). These networks allow us to pinpoint and investigate communication dynamics precisely, e.g., how fast an information originating from one Twitter user is retweeted by another who does not have a direct tie to the originator. Analyzing time-ordered networks requires elaborately collected datasets and analytic strategies that go beyond the regular network analysis software toolkits. The most suitable software for managing and analyzing such networks is the *R*

package *networkDynamic* (BUTTS et al. 2016) that is fully compatible with the *Rstatnet* suit of packages, including *sna* (BUTTS 2016), *tsna* (BENDER-DE MOLL/MORRIS 2016) and *network* (BUTTS 2008). However, even today, larger networks of more than a few hundred actors and several time points (e.g., more than ten) are hardly manageable, even on a well-equipped computer.

FIGURE 3  
Time-ordered and time-aggregated networks



Note: The figure displays a fictitious time-ordered contact network (top, e.g., a retweet network) and interval network (bottom, e.g., a follower network) over the course of several time steps. The lower rows for each network represent time-aggregated projections of the upper row network time-series. While the time-aggregated projections of the interval network largely resemble the topology of each of the time-ordered time slices, the time-aggregated projections of the contact network feature a highly dissimilar topology compared to each of its time slices.

That is one reason why the second analysis framework for dynamic networks, time-aggregated networks, is far more common. Here, the dynamic

network is decomposed, i.e., »the time-ordered data [is collapsed] into a series of static networks« (BLONDER et al. 2012: 965). A time-aggregated approach treats an evolving network as a time sequence of networks, where each network could be studied independently (BORGNAT et al. 2008: 199). As mentioned earlier, this may come at the cost of losing temporal order within each aggregation. It allows researchers, however, to apply all those methods and measures that can be used on static networks to describe the evolution of dynamic networks.

For such networks, the question of the proper length of aggregation periods is critical. Unfortunately, there is no consensus on how to best define aggregation periods (time frames) (BLONDER et al. 2012; BORGNAT et al. 2008). Instead, the choice of time-frame granularity largely depends on the specific dynamics of the network under study and on the research interest (CASTEIGTS et al. 2012). For instance, appropriately describing the spread of a hashtag will usually require a more fine-grained time resolution than investigating the evolution of follower networks. Therefore, prior knowledge about the processes or experimenting with network statistics at different granularity levels can be fruitful for determining an appropriate scale (BLONDER et al. 2012). In practice, for slowly developing networks, such as academic citation graphs, monthly or even yearly periods are typical (HUANG et al. 2008; LESKOVEC/KLEINBERG/FALOUTSOS 2005; PALLA/BARABÁSI/VICSEK 2007). For online social networks or phone call networks, daily snapshots or at most snapshots encompassing a few weeks are more common (GONG et al. 2012; KWAK/CHUN/MOON 2011; PALLA/BARABÁSI/VICSEK 2007; PANZARASA/OPSAHL/CARLEY 2009). Instructively, Panzarasa et al. (2009) investigated the network of a university's instant messaging platform at different aggregation time windows, namely two, three, and six weeks. Analyzing a variety of metrics, including average degrees, reciprocity, and clustering coefficients over time, they found that time-series distribution curves become increasingly smoother for longer time frames; however, they generally follow the same trajectory.

Another question concerns overlap between consecutive slices of time-aggregated contact networks. While in most studies, slices are consecutive, e.g., weekly aggregations always encompassing Monday through Sunday, some researchers choose overlapping data (e.g., PANZARASA/OPSAHL/CARLEY 2009). Each daily time step then contains, for example, all data reaching back two weeks. In either case, aggregation slices implicitly assume dissolving ties: »[W]hen social relationships do not last forever,

past events lose relevance for the current state of the system, depending on when they took place« (PANZARASA/OPSAHL/CARLEY 2009: 917). Basically, if a communication tie between two nodes dissolves the fact that they once communicated may no longer be relevant for the current system; a lifespan is imposed on ties. Alternatively, a cumulative approach could be applied where each time step of a series of networks includes all relations/contacts that took place at some point during the investigation period (HOLME/SARAMÄKI 2012; KUMAR/NOVAK/TOMKINS 2010; PANZARASA/OPSAHL/CARLEY 2009). Each time a new interaction occurs, a tie is added. Dissolution of relations is not considered in this approach. For many network metrics, including degrees or clustering coefficients, such an approach implies growth throughout the observation period (*ibid.*).

### 3. Descriptive Analysis of Dynamic Networks

The distinction between time-ordered and time-aggregated networks has major implications for the descriptive analysis of dynamic networks. On the one hand, time-ordered networks differ in structure from static networks and therefore allow other kinds of insights, but also require different analytical procedures. For time-aggregated networks, on the other hand, each aggregate can be analyzed like a static network. Afterwards, a sequence of time slices can be compared regarding the extent and nature of changes over time. This latter approach is much more common because it allows for easy generalization of available procedures. Researchers should carefully consider whether their questions may be answered through this approach or whether the added level of precision makes a more difficult and computationally expensive analysis on time-ordered networks worthwhile (HOLME/SARAMÄKI 2012). Oftentimes, if researchers want to describe how network structures evolve over time, an aggregation approach will serve them well. However, if they want to learn about processes, such as flow, taking place on the network structure, genuinely dynamic measures are required. In principle, many measures can be applied to both types. They only differ in the level of achievable granularity in results. For instance, we may learn how long ties remain activated by counting the number of consecutive aggregates they appear in. We may also use timestamp data on the ties to calculate their exact duration to the second.

Metrics for describing networks can be categorized along the analytic level of units under consideration (SCHUNTER/LIEBAU 2014), as well as regarding their focus on nodes (actors) or ties (relations). On the lowest level for the former are metrics focusing on individual nodes, such as centrality measures (e.g., in- and out-degree centrality). Secondly, we may focus on dyadic (two actors) and triadic (three actors) subgraph configurations in order to measure the share of reciprocated ties and closed (transitive and cyclical) triads over time. Higher order subgraph configurations are not usually considered. On the next level are measures concerned with larger areas of the network, e.g., cohesive subgroups/communities and connected components. Finally, actor- and relations-counts, the density or cohesiveness and degree-distributions of the whole network may also be compared over time. Focusing more specifically on ties, we can take into consideration both individual ties (e.g., their duration) and longer tie sequences, i.e., paths. We will discuss these options and how they apply to dynamic networks in more detail below. For a comprehensive introduction into different network metrics, we recommend consulting Wasserman and Faust (2009) or Borgatti, Everett, and Johnson (2013).

With some caution most measures known from the analysis of static networks can be applied to multiple consecutive aggregates of the same network (BORGAT ET AL. 2008). This is equally true for snapshots of interval networks, such as a Twitter follower network, and networks representing aggregated contacts, such as mentions, replies, or retweets. After calculating measures for each time slice, we can apply standard time-series analysis to the data, visually inspect plots of developments, or calculate statistical properties such as means, standard deviations, and trends (BLONDER ET AL. 2012; BORGAT ET AL. 2008; CASTEIGTS ET AL. 2012). Again, these measures are descriptive/explorative in the sense that their aim is to describe/uncover what changes occur in the network from one time-slice to the next. However, the use of inferential calculations is necessary to achieve this. We will discuss these options in more detail below, based on specific applications. Tie-centered metrics, however, often have different demands compared to their application in static networks. That is because researchers often look at them not to describe long-term network evolution, but short-term communication and flow processes, in which the precise order of events is of great importance. We will now proceed by discussing metrics at different analytical levels of units under consideration, pointing out where time-series analysis of common measures is appropriate and where dynamic measures are necessary or add analytical value.

### *Individual nodes*

For *individual nodes*, we can easily investigate their *in- and out-degrees* over time. A node's degree denotes the number of other actors it is connected to (WASSERMAN/FAUST 2009: 100ff.). In a directed network, such as a follower network (user A may follow user B, but not vice versa), the in-degree describes the number of actors initiating connections to user A, while the out-degree is the number of users A initiates ties with. These measures can also be aggregated to the network level by calculating average in- and out-degrees over time (e.g., HUANG et al. 2008; PANZARASA et al. 2009). This tells us whether actors gain or lose relations over time or whether they predominantly maintain the same amount of relations they form early on.

Degree measures can also be leveraged for the investigation of *activity dynamics* in the network. For example, Wu et al. (2013) argue that nodes are either active or inactive (i.e., last login not within a specified time period). They investigated the influence of the absolute number and fraction of active friends, the number of inactive friends, and the number of inactive friends who left the network recently on the activity status of a focal actor. They found that while joining the network is best predicted by the number of friends who also joined recently, departure dynamics work differently. The fraction of active friends, that is, the neighborhood activity level, rather than the absolute number of active friends affected the decision for a user to leave (WU et al. 2013: 233). Put differently, when a substantial share of an actor's contacts becomes inactive, they are likely to become inactive themselves.

Besides their degree, we may also be interested in learning how central individual actors are in connecting the whole network. *Centrality*, as Wasserman and Faust (2009: 169) point out, is a way of defining an actors' importance through their (strategic) location in a network. There are different ways to operationalize centrality, such as degree centrality (high number of ties from A to other actors) and betweenness centrality (high number of paths between non-adjacent actors running through A). These centrality measures can in principle also be calculated over time. However, depending on the research goal, some caution is warranted. Applied to snapshots of slowly evolving interval graphs, regular centrality measures are appropriate. The evolution of betweenness centralities, for instance, may be used to identify those actors that become brokers over time (BURT 1999). Brokers play an important role, as they build up connect-



ing pathways between different, otherwise unconnected (or very loosely connected) subgraphs of a network. In contact networks, however, due to the time dependency and instantaneous nature of paths, actor centrality in a dynamic process may also fluctuate widely from one point in time to the next (HOLME/SARAMÄKI 2012). This fluctuation limits the conclusions we can draw from such measures.

If we want to identify central actors in a communication process, there is an additional problem to consider. As Borgatti (2005) and Ghosh and Lerman (2012) point out, the specific nature of the communication process influences who plays a central role in its course. Importantly, a process may be of type ›transfer‹ or ›broadcast‹ (BORGATTI 2005). In transfer processes, a resource moves from one node to another without replication (BORGATTI 2005). For example, surfers in a hyperlink network behave this way, as they can only be on one single website at any given point. As no replication takes place, the flow of type ›transfer‹ can be modelled using a random walk strategy (e.g., betweenness centrality, PageRank algorithm; GHOSH/LERMAN 2012). Online communication, however, often resembles a radio broadcast, as nodes can pass information along to many others simultaneously (BORGATTI 2005). Ghosh and Lerman (2012) demonstrated that for online message diffusion, broadcast centrality measures, such as Alpha-Centrality, are more suitable, as they identified those individuals as central whose messages empirically received the most attention.

### *Dyads and triads*

Turning next to the analysis of local subgraph configurations, on the level of *dyads*, changes in *reciprocity or mutuality* may be of theoretical importance (MONGE/CONTRACTOR 2003; for a deep theoretical background on reciprocity see STEGBAUER 2011). Reciprocity describes the strength of ›the tendency for one actor to ›choose‹ another, if the second actor chooses the first‹ (WASSERMAN/FAUST 2009: 507). That is, it enables researchers to check whether the relation from actor *i* to actor *j* prompts a relation back from actor *j* to actor *i*. Changes in the absolute number or the share of reciprocated dyads can easily be observed over several snapshots (e.g., GONG et al. 2012; HUANG et al. 2008; PANZARASA et al. 2009). Empirical studies suggest that in online social networks, depending on the phase of development, reciprocity may both increase (KWAK et al. 2011) and decrease over time (GONG et al. 2012), making it worthwhile to investigate this question further.

For *triads*, the evolution of *transitivity* – the degree to which nodes  $i$  and  $j$  form a relationship to one another because of their common friend  $k$  (WASSERMAN/FAUST 2009: 150) – can easily be analyzed over time. Such data serve researchers well if they are interested in triads as a feature for the whole network. The (global and local) *clustering coefficient* (WATTS/STROGATZ 1998; NEWMAN 2010) is calculated as the overall share or the mean of closed triples (e.g., GONG et al. 2012; PANZARASA et al. 2009). While network theories, such as activity-focus theory (FELD 1981) or balance theory (CARTWRIGHT/HARARY 1956; HUMMON/DOREIAN 2003), suggest an increase of transitivity over time, no general trend for the clustering coefficient's development can be abstracted from the empirical literature and its development also depends considerably on the temporal resolution of the data at hand (PANZARASA et al. 2009).

However, comparing reciprocity (mutuality) or transitivity over time by investigating consecutive network aggregates has limits. Such data sets are sufficient if researchers are interested in dyadic or triadic developments of the whole network, i.e., to learn whether overall reciprocity measures or clustering coefficients increase or decrease over time. If researchers want to investigate the exact local developments, however, they should employ continuous, time-stamped data. Local developments are the focus of interest when researchers study when actors reciprocate a directed tie, what delay there is between a tie first emerging and being reciprocated (KUMAR/NOVAK/TOMKINS 2010), or when triadic closure emerges. That is because time-series of consecutive snapshots always omit information about what happens in-between snapshots or obscure the order of events.

### *Groups within the network*

For groups within the networks some difficulties arise when investigating their development over time. Typically, such groups are identified as so-called *cohesive subgroups* or *communities*, i.e., »subsets of actors among whom there are relatively strong, direct, intense, frequent, or positive ties« (WASSERMAN/FAUST 2009: 249). There are a number of ways communities may evolve over time. They can grow larger, contract (lose members), merge with other communities, or split up into separate communities. Furthermore, new communities may emerge (birth) and others (temporarily) perish (death) (CAZABET/AMBLARD 2014; PALLA/BARABÁSI/VICSEK 2007).

For large networks, such as online networks tend to be, cohesive subgroups are typically discovered using algorithmic procedures, mostly opti-

mizing a global quality function like modularity (NEWMAN/GIRVAN, 2004; for an overview of community detection algorithms, see FORTUNATO, 2010). Because these algorithms optimize some overarching measure, they are often »very sensitive to small modifications of the topology« (BORGNAT et al. 2008: 200; CAZABET/AMBLARD 2014: 407), such as we would expect to occur from one snapshot to the next. Although it is possible to apply any community detection algorithm to consecutive snapshots of a network and compare the results (CAZABET/AMBLARD 2014), using for example the Rand index (RAND 1971),<sup>4</sup> this may not yield satisfactory insights due to this sensitivity. Minor changes in the network topology can disrupt the community structure and lead to entirely different solutions from one time step to the next.

While a huge body of literature tackling this problem has emerged in computer science and theoretical physics (see AGGARWAL/SUBBIAN 2014 for an extensive review), readily implemented, easy-to-use solutions are still lacking. Generally speaking, one key challenge is matching what is considered to be the same community between two snapshots. Another is to balance the two objectives of accurately reflecting data at the current time step (>consistency<) and obtaining clusters that closely resemble those at previous time steps (>smoothness<; AGGARWAL/SUBBIAN 2014). Proposed solutions that incorporate both properties include an evolutionary spectral clustering approach (CHI et al. 2009) and the ENetClus approach (GUPTA et al. 2012). The latter creates a probabilistic assignment for each snapshot that is then used as the initialization point for the consecutive snapshot, thereby ensuring smoothness (for a similar approach see LANCICHINETTI/FORTUNATO/KERTESZ 2009). But, again, to our knowledge no ready-to-use implementations are available for these procedures.

One relatively straightforward and widely used approach to community detection that has been generalized for dynamic network analysis via consecutive snapshots is the *clique percolation method* (PALLA et al. 2005). This procedure is based on finding unions of fully connected subgraphs (cliques, see WASSERMAN/FAUST 2009: 253f.) that can reach each other through a series of adjacent cliques. As this procedure has a clear-cut, local definition of what constitutes a cohesive subgroup, any changes occurring between

4 The Rand index can be used to evaluate the similarity of two community solutions proposed by an algorithm. It compares the dyad-cluster-assignments in one solution to the assignment in another solution.

two snapshots will only affect the local network area. Palla, Barabási, and Vicsek (2007) suggest the creation of an intermittent network between snapshots, consisting of all nodes and ties present in both  $t$  and  $t+1$ . Communities in this joint graph can only grow, merge, or remain identical. This enables a clear matching of communities between time steps. The authors successfully apply this approach to some real-world networks, including one mobile phone communication network (PALLA/BARABÁSI/VICSEK 2007).

Avoiding community detection altogether, another approach to studying the evolution of groups in online networks is to focus on *components*. While communities are densely connected network areas, a component is a subgraph where all nodes can be reached through some path, i.e., are connected through some path (WASSERMAN/FAUST 2009: 109). Here, we can identify instances of previously disconnected subgraphs establishing connections and merging (HUANG et al. 2008; KUMAR et al. 2010; PANZARASA et al. 2009). Particularly for online social networks, this approach has uncovered a consistent pattern, where initially the largest component grows through merging with isolates and middle-sized components. After some time, an equilibrium is reached and the shares of nodes in the giant and middle-sized components remain steady, despite overall network growth (KUMAR et al. 2010; PANZARASA et al. 2009).

### *Entire networks*

On the level of the whole network, particularly the evolution of density, diameters or average distances are of interest. Again, these measures can be used on a series of network slices, but they should be interpreted with some caution. For instance, we must take into account that the *density* metric, which measures the number of existent relations divided by the number of all possible relations (WASSERMAN/FAUST 2009: 101f.), is sensitive to changes in the network's number of nodes. Thus, density is not a valid measure for the evolution of cohesiveness in dynamic networks that exhibit high variance. The network's *average degree* and the network's *average local clustering coefficient* (i.e., how close are local subgraph configurations to being fully connected cliques, see WATTS/STROGATZ 1998) are more suitable in such a case, as these measures are not as much associated with network size. However, given a relatively stable network size, density can be investigated over time (e.g., GONG et al. 2012; KUMAR et al. 2010). Positive trends in density indicate that a network as a whole is growing more connected,

while negative trends indicate disentanglement. Generally speaking, most online networks tend to grow denser over time (KUMAR et al. 2010; KWAK et al. 2011), as the number of edges grow super-linearly in the number of nodes (LESKOVEC et al. 2005).

The *diameter* of a network is the length of the longest geodesic distance between two nodes of a network (i.e., the longest shortest path between two nodes, see WASSERMAN/FAUST 2009: 111f.). Because this measure is sensitive to outliers, researchers frequently use *average distances* (average length of the shortest paths between two nodes) and *effective diameters* (90th percentile of the shortest path lengths between all pairs of nodes) instead (GONG et al. 2012; KUMAR et al. 2010; LESKOVEC et al. 2005). If researchers are interested in the evolution of the overall connectivity of a graph, calculating these measures for multiple consecutive snapshots is a valid approximation. Overall, results indicate that, despite network growth, paths in online networks tend to shrink over time (LESKOVEC et al. 2005; PANZARASA et al. 2009) which is consistent with the small-world phenomenon (WATTS/STROGATZ 1998). Researchers may also draw interesting conclusions from fluctuations in path length. Abrupt changes will often point to unusual events in a network (AGGARWAL/SUBBIAN 2014). Consistently high diameters point to relatively separate groups, such as ideological factions, while fluctuation may indicate emerging dialogue between these previously separated groups (BLONDER et al. 2012).

Still, particularly regarding contact networks, researchers must be cautious when interpreting diameters and average distances calculated on time-aggregated networks. That is because results strongly depend on the length of the aggregation period. A path in a snapshot of a contact network that covers all contacts within a month or a year, may not indicate that information could actually have travelled this way between  $i$  and  $j$ . That is because the order of contacts within the snapshot determines that possibility. Put simply, a message can only be transmitted from actor  $A$  to  $B$  via  $C$  if  $C$  talks to  $B$  after talking to  $A$  first (time-respecting paths, see below). That means that calculating diameters or average distances for time-aggregated networks allows us to draw conclusions about how the network's overall connectivity develops over time, but not necessarily about the reach of individual instances of communication within the network.

*Degree distributions* over time might reveal changes in the structure of communication roles and the overall network topology over time (e.g., GONG et al. 2012; PANZARASA et al. 2009). For instance, for the study of communication networks we may find that initially relatively many moderate

hubs (as indicated by a high out-degree) and authority actors (as indicated by high in-degree) participate in a discussion. Two scenarios are plausible. First, as actors get to know each other better, they may become more comfortable interacting with others. Consequently, degree distributions may become more equalized in the long run (see also AGGARWAL/SUBBIAN 2014). Conversely, in line with the Matthew effect (MERTON 1968; PERC 2014) that predicts »the rich get richer«, actors who already have a high reputation or authority may become even mightier authorities over time and hubs will accumulate ever more relations. Concentration of degree distributions would then increase. This tendency can ultimately lead to the so-called power law or other heavy-tailed degree distributions (BARABASI/ALBERT 1999), a widely acknowledged topological feature of the World Wide Web and many other networks, such as scientific citation networks. These distributions are highly right-skewed, with the bulk of observations occurring at fairly small sizes (e.g., nodes with very few ties) and a small number of observations with much higher than average values (e.g., hubs with very many ties; see NEWMAN 2005 for a more detailed discussion).

### *Tie duration and time-respecting paths*

The metrics discussed above have mostly been actor-centric, that is, aiming to describe how network properties surrounding a particular actor or a set of actors – be it dyads, triads, communities, or whole actor networks – evolve over time. As we have seen, most network properties can be studied by focusing on a series of consecutive snapshots or aggregates. While this approach enables usage of the whole network analysis toolkit and is therefore an efficient framework for studying the evolution of network topology, »some properties cannot be directly observed in this [snapshot] framework« (BORGNAT et al. 2008: 200). This is particularly true for measures centered on tie dynamics, for instance to investigate patterns of contacts within a network.

For individual ties, *contact duration*, i.e., the (average) amount of time between the formation and dissolution of ties, is a central measure (BORGNAT et al. 2008; CHAINTREAU et al. 2006). This metric only makes sense for interval networks with non-negligible tie durations. For instance, Kwak et al. (2011) studied the dissolution of follower ties on Twitter, i.e., instances of people unfollowing others. They compared daily snapshots to identify vanishing connections. Interestingly, they found that older ties are less likely to break up.

For contact networks, e.g., retweet networks, contact duration is not a sensible measure, as contacts are instantaneous. Here, however, we may be interested in *inter-contact time*, that is, »the time between two transfer-opportunities« (CHAINTREAU et al. 2006). We should point out that the level of accuracy when applying these measures to time-aggregated networks is once again limited by the length of aggregation time windows. If we need highly precise information, we require time-stamped ties or must calculate measures based on time-ordered networks.

As discussed, the temporal order of ties becomes critical when studying possible flow processes on the network structure, rather than the long-term evolution of network topology. Questions that can arise in this context are as follows: Which ways can information, behavior, or opinions travel through a network? How can a message be transmitted from one actor to another? What set of people can a certain actor address within a specified time window? How long will it take for a message to be transferred between different parts of a network?

To avoid erroneous conclusions, the precise temporal order of events is important. A simple example can illustrate this (Fig. 4). Say we are investigating an instant messaging network and would like to know how a message *X* originating from actor *A* reached actor *B*. The relevant nodal attribute can be thought of as »knows about *X*«. We might only have aggregated contact data available, e.g., daily snapshots, and observe that *A* talked to *C*, *C* talked to *D*, and *D* talked to *B*. Still, we may be wrong to assume that *C* and *D* passed the message along to *B* as our aggregated network would hide that the order of communication instances would not allow this flow of information (HOLME/SARAMÄKI 2012: 98f.). That may be the case if *C* talked to *D* before having talked to *A*, making it impossible for them to have passed the information from *A* along, a fact that is obscured by aggregation. Even fine-grained aggregation becomes error-prone, so that there might be a path between the two in every footprint and yet information could not possibly have passed through it (CASTEIGTS et al. 2012: 402). As researchers need precise information about the temporal order of ties, the analysis of time-ordered (continuous) networks is required (BLONDER et al. 2012).

The central concept for these investigations is the *time-respecting path* (KEMPE/KLEINBERG/KUMAR 2002). »[I]n relation to dynamical processes, paths must necessarily be constrained to sequences of link activation that follow one another in time« (HOLME/SARAMÄKI 2012: 105). Time-respecting paths are therefore sequences of contacts with non-decreasing times that

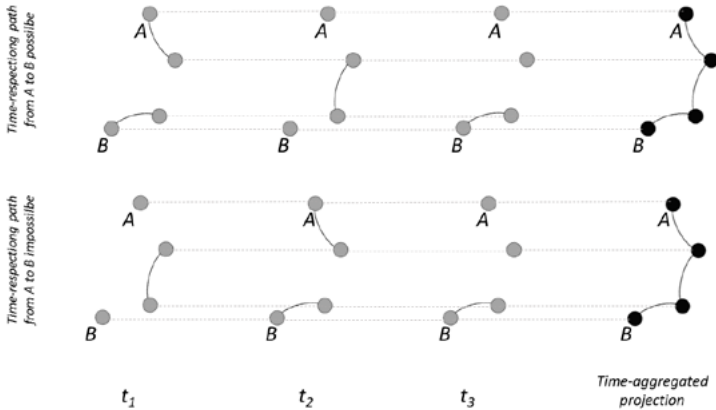
connect sets of vertices (HOLME/SARAMÄKI 2012). They are also referred to as »journeys« (CASTEIGTS et al. 2012). Time-respecting paths can be calculated in  $R$  using the *timeordered* package (BLONDER 2015) or the aforementioned *tsna* package (BENDER-DE MOLL/MORRIS 2016). Time-respecting paths have some nontrivial implications. Even in undirected networks, they are not symmetrical: it is possible that there is a path from  $A$  to  $B$  but not vice versa »because the time dimension creates its own level of direction« (CASTEIGTS et al. 2012: 394; s. also Fig. 4). Time-respecting paths also have two distinct lengths, a *topological* and a *temporal length* (CASTEIGTS et al. 2012: 394), each of which can be specified based on time-ordered network data. Further, a time-respecting path's *latency* gives the shortest time within which one node can reach another (HOLME/SARAMÄKI 2012). Latency can be considered the temporal equivalent of the shortest path length (geodesic distance). Time-respecting paths between nodes vary over time, both regarding their topological and temporal length.

Based on the concept of time-respecting paths, which tells us if, when, and how quickly a resource can travel between two nodes, some additional concepts can be abstracted. By averaging them, durations and latencies can characterize the overall *velocity* of communication within a network (HOLME/SARAMÄKI 2012), that is, how quickly messages can travel overall. Furthermore, we can define the *set of influence* of a node  $i$  that consists of all vertices that can be reached from  $i$  through time-respecting paths (HOLME/SARAMÄKI 2012). Thus, the set of influence describes the maximum number of nodes that can potentially be influenced by a specified actor within a specified time-frame. This is an interesting measure to evaluate the empirical reach some message within the system had. Complementary to this, a *source set* consists of all nodes that can reach  $i$  from anywhere in the network through time-respecting paths (HOLME/SARAMÄKI 2012) and, therefore, all nodes that can potentially exert influence on or pass information to  $i$ .

Finally, we might be interested in the long-term evolution of contact patterns within a graph. Do contact durations evolve, for instance? Is the temporal length of time-respecting paths growing shorter as more users enter a network? Does the network enhance or decrease in terms of its velocity? For these questions, Casteigts et al. (2012) suggest the creation of a sequence of shorter temporal subgraphs for successive time windows. It is then possible to plot the coarse-grain evolution of temporal measures over long periods of time.



FIGURE 4  
 Time-respecting paths and non-time-respecting paths may look the same in aggregation.



Note: The figure illustrates the logic of time-respecting paths and the issue induced by time-aggregated projection. Both rows display time-ordered networks each consisting of four nodes over the course of three time steps. While it is possible that a message is transmitted from node A to B in the example at the top of the figure, it is not possible in the example at the bottom. However, the time-aggregated projections of the two examples are identical.

#### 4. Outlook

Taken together, there is a wide array of metrics and analytical procedures that can be used for the descriptive analysis of dynamic online networks. However, one must be aware of the specifics of the data source one is dealing with and which techniques are reasonable to apply. We believe that if researchers are in doubt about what procedures are appropriate for the description of a particular dynamic network, it is first important to answer some basic questions in this vein: Are ties in the network periodically lasting intervals, i.e., do they have a substantial duration? Or is it an instantaneous contact network where the contact duration is negligible? What is known about the temporal patterns of ties (e.g., how long is a typical follower relation on Twitter, what is the lifespan during which a given tweet still draws new interactions)? These questions are instrumental to answer in order to decide whether researchers can validly employ quasi-static time-aggregated networks and what an appropriate aggregation period may be.

Besides the characteristics of the network relations, a second point to clarify before choosing methods is the research aim. Is the process under study short-term or long-term? Do researchers want to identify flow or diffusion processes on the network or rather describe how tie patterns, subgraph configurations, or network density evolve over time? These questions will be instructive as to whether a more computationally expensive time-ordered approach is necessary. Generally speaking, the study of flow demands the consideration of rigorous temporal order and therefore the application of measures such as contact duration, inter-contact time, and time-respecting paths. Longer-term network-evolutionary questions may be more validly answered through a time-series approach to multiple snapshots.

However, as we have pointed out in the discussion of individual metrics, more caution is necessary in the application of some techniques from static network analysis than with others. For example, community detection and centrality measures require more careful data handling and consideration in order to avoid methodological bias. With community detection, local definitions of communities, such as in the clique percolation method, may be more validly applied than those procedures that rely on the optimization of global metrics, where minor changes in the network may drastically alter the solution. With centrality methods, changes in the size of the node set pose a threat to validity. Other metrics, like degrees, mutuality, transitivity, and local clustering coefficients can more easily be abstracted to the analysis of dynamic networks. Even here, however, we advise researchers to carefully consider whether this abstraction is theoretically appropriate and whether characteristics of the network's dynamics may affect the validity of the constructs under study.

Generally speaking, empirical analyses of dynamic online networks indicate that these networks »travel through distinct stages of growth, characterized by specific behavior in terms of density, diameter, and regularity of component structure« (KUMAR et al. 2010: 617; see also GONG et al. 2012; PANZARASA et al. 2009). The observation of substantive changes over time makes the study of dynamic online networks worthwhile, particularly as there has been little effort so far to connect these observations with substantive social scientific theories.

Our short overview does not fully cover all possibilities regarding descriptive analysis of dynamic networks, of course. For instance, we have not discussed at length what the implications of a volatile node set are and how it may affect various metrics. This may become an issue in networks

where either the overall size of the node set fluctuates or where large parts of the node set change in a short period of time. More specifically, this can be a problem regarding measures that are sensitive to network size, such as density, because comparisons over time can become problematic. Evaluating, for instance, the (dis-)similarity of communities at different time steps is complicated if the underlying set of actors assigned to such groups is not identical.

Furthermore, we largely excluded the dynamics of actor attributes, such as Twitter users' location or the topics they write about. Incorporating such variables into evolution analysis would surely be a fruitful advancement for descriptive network analysis. The simultaneous evolution of content or user attributes and network structure could also aid with explaining when and why network structure changes, e.g., whether users' network positions change as their online interests evolve (AGGARWAL/SUBBIAN 2014; GONG et al. 2012).

Since our focus was on descriptive analysis of dynamic networks here, we have not discussed methodological efforts that aim at explaining what factors affect network evolution, e.g., longitudinal extensions of Exponential Random Graph Models (HANNEKE et al. 2010; LEIFELD/CRANMER/DESMARAIS 2017; KRIVITSKY/HANDCOCK 2014) and the Stochastic Actor Oriented Model (SNIJDERS 1996; SNIJDERS/VAN DE BUNT/STEGELICH 2010). We have also excluded inferential efforts at understanding how the network structure influences diffusion dynamics, through the use of diffusion or contagion models (FRIEMEL 2010; VALENTE 1995).

Connecting inferential techniques with descriptive measures of dynamic networks will likely prove a fruitful avenue in future research. It will facilitate the exploration of underlying social causes for the nature and evolution of empirical communication networks and determination of what role these networks play in diffusion processes.

## References

- ADAMIC, L. A.; N. GLANCE: The political blogosphere and the 2004 US election: divided they blog. In: *Proceedings of the 3rd international workshop on Link discovery*. ACM, 2005, pp. 36-43
- AGGARWAL, C.; K. SUBBIAN: Evolutionary network analysis: A survey. In: *ACM Computing Surveys (CSUR)*, 47(1), 2014, p. 10

- BARABÁSI, A. L.; R. ALBERT: Emergence of scaling in random networks. In: *Science*, 286(5439), 1999, pp. 509-512. doi: 10.1126/science.286.5439.509
- BENDER-DE MOLL, S.; M. MORRIS: *tsna: Tools for Temporal Social Network Analysis*. R package version 0.2.0, 2016. <https://CRAN.R-project.org/package=tsna>
- BENKLER, Y.: *The Wealth of Networks. How Social Production Transforms Markets and Freedom*. New Haven, London [Yale University Press] 2006
- BENNETT, W. L.; A. SEGERBERG: *The logic of connective action: Digital media and the personalization of contentious politics*. New York City/NY [Cambridge University Press] 2013
- BERGER-WOLF, T. Y.; J. SAIA: A framework for analysis of dynamic social networks. In: *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2006, pp. 523-528. doi: 10.1145/1150402.1150462
- BLONDER, B.: *Timeordered. Time-ordered and time-aggregated network analyses*. R package version 0.9.8, 2015. <https://cran.r-project.org/web/packages/timeordered/index.html>
- BLONDER, B.; T. W. WEY; A. DORNHAUS; R. JAMES; A. SIH: Temporal dynamics and network analysis. In: *Methods in Ecology and Evolution*, 3(6), 2012, pp. 958-972. doi: 10.1111/j.2041-210X.2012.00236.x
- BORGATTI, S. P.: Centrality and network flow. In: *Social networks*, 27(1), 2005, pp. 55-71. doi: 10.1016/j.socnet.2004.11.008
- BORGATTI, S. P.; M. G. EVERETT; J. C. JOHNSON: *Analyzing Social Networks*. Thousand Oaks/CA [Sage] 2013
- BORGNAT, P.; E. FLEURY; J.-L. GUILLAUME; C. MAGNIEN; C. ROBARDET; A. SCHERRER: Evolving networks. In: FOGELMAN-SOULIÉ, F.; D. PERROTTA; J. PISKROSKI; R. STEINBERGER (Eds.): *Mining Massive Data Sets for Security*. Amsterdam [IOS Press] 2008, pp. 198-203. doi: 10.3233/978-1-58603-898-4-198
- BURT, R. S.: The social capital of opinion leaders. In: *The Annals of the American Academy of Political and Social Science*, 566(1), 1999, pp. 37-54. doi: 10.1177/000271629956600104
- BUTTS, C. T.: Network: A Package for Managing Relational Data in R. In: *Journal of Statistical Software*, 24(2), 2008, pp. 1-36. doi: 10.18637/jss.v024.i02
- BUTTS, C. T.: *sna: Tools for Social Network Analysis*. R package version 2.4, 2016. <https://CRAN.R-project.org/package=sna>

- BUTTS, C. T.; A. LESLIE-COOK; P. N. KRIVITSKY; S. BENDER-DE MOLL: *networkDynamic: Dynamic Extensions for Network Objects*. R package version 0.9.0, 2016. <https://CRAN.R-project.org/package=networkDynamic>
- CARTWRIGHT, D.; F. HARARY: Structural balance: a generalization of Heider's theory. In: *Psychological review*, 63(5), 1956, p. 277. doi: 10.1037/h0046049
- CASTEIGTS, A.; P. FLOCCINI; W. QUATTROCIOCHI; N. SANTORO: Time-varying graphs and dynamic networks. In: *International Journal of Parallel, Emergent and Distributed Systems*, 27(5), 2012, pp. 387-408. doi: 10.1080/17445760.2012.668546
- CAZABET, R.; F. AMBLARD: Dynamic community detection. In: ALHAJJ, R.; J. ROKNE (Eds.): *Encyclopedia of Social Network Analysis and Mining*. New York/NY [Springer] 2014, pp. 404-414
- CHARENTREAU, A.; P. HUI; J. CROWCROFT; C. DIOT; R. GASS; J. SCOTT: Impact of human mobility on opportunistic forwarding algorithms. In: *IEEE Transactions on Mobile Computing*, 6(6), 2007, pp. 606-620
- CHI, Y.; X. SONG; D. ZHOU; K. HINO; B. L. TSENG: On evolutionary spectral clustering. In: *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 3(4), 2009, pp. 17
- CRANMER, S. J.; P. LEIFELD; S. D. MCCLURG; M. ROLFE: Navigating the range of statistical tools for inferential network analysis. In: *American Journal of Political Science*, 61(1), 2017, pp. 237-251. doi: 10.1111/ajps.12263
- FELD, S. L.: The focused organization of social ties. In: *American Journal of Sociology*, 86(5), 1981, pp. 1015-1035. doi: 10.1086/227352
- FORTUNATO, S.: Community detection in graphs. In: *Physics Reports*, 486 (3-5), 2010, pp. 75-174. doi: 10.1016/j.physrep.2009.11.002
- FRIEMEL, T. N.: Diffusionsforschung. In: STEGBAUER, C.; R. HÄUSSLING (Eds.): *Handbuch Netzwerkforschung*. Wiesbaden [vs] 2010, pp. 825-833
- FRIEMEL, T. N.: Influence Versus Selection: A Network Perspective on Opinion Leadership. In: *International Journal of Communication*, 9, 2015, pp. 1002-1022
- GHOSH, R.; K. LERMAN: *Rethinking centrality: the role of dynamical processes in social network analysis*. 2012. arXiv:1209.4616
- GONG, N. Z.; W. XU; L. HUANG; P. MITTAL; E. STEFANOV; V. SEKAR; D. SONG: Evolution of social-attribute networks: measurements, modeling, and implications using google+. In: *Proceedings of the 2012 Internet Measurement Conference*. ACM, 2012, pp. 131-144

- GUPTA, M.; J. GAO; Y. SUN; J. HAN: Community trend outlier detection using soft temporal pattern mining. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Berlin [Springer] 2012, pp. 692-708. doi: 10.1007/978-3-642-33486-3\_44
- HANNEKE, S.; W. FU; E. P. XING: Discrete temporal models of social networks. In: *Electronic Journal of Statistics*, 4, 2010, pp. 585-605. doi: 10.1214/09-EJS548
- HIMELBOIM, I.; S. MCCREERY; M. SMITH: Birds of a feather tweet together: Integrating network and content analyses to examine cross-ideology exposure on Twitter. In: *Journal of Computer-Mediated Communication*, 18(2), 2013, pp. 40-60. doi: 10.1111/jcc4.12001
- HINDMAN, M. S.: *The myth of digital democracy*. Princeton [Princeton University Press] 2008
- HOLME, P.; J. SARAMÄKI: Temporal networks. In: *Physics reports*, 519(3), 2012, pp. 97-125. doi: 10.1016/j.physrep.2012.03.001
- HUANG, J.; Z. ZHUANG; J. LI; C. L. GILES: Collaboration over time: characterizing and modeling network evolution. In: *Proceedings of the 2008 international conference on web search and data mining*. ACM, 2008, pp. 107-116. doi: 10.1145/1341531.1341548
- HUMMON, N. P.; P. DOREIAN: Some dynamics of social balance processes: bringing Heider back into balance theory. In: *Social Networks*, 25(1), 2003, pp. 17-49. doi: 10.1016/S0378-8733(02)00019-9
- KEMPE, D.; J. KLEINBERG; A. KUMAR: Connectivity and inference problems for temporal networks. In: *Journal of Computer and System Sciences*, 64(4), 2002, pp. 820-842. doi: 10.1006/jcss.2002.1829
- KRIVITSKY, P. N.; M. S. HANDCOCK: A separable model for dynamic networks. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1), 2014, pp. 29-46. doi: 10.1111/rssb.12014
- KUMAR, R.; J. NOVAK; A. TOMKINS: Structure and evolution of online social networks. In: YU, P.; J. HAN; C. FALOUTSOS (Eds.): *Link mining: models, algorithms, and applications*. New York City/NY [Springer vs] 2010, pp. 337-357. doi: 10.1007/978-1-4419-6515-8\_13
- KWAK, H.; H. CHUN; S. MOON: Fragile online relationship: a first look at unfollow dynamics in twitter. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011, pp. 1091-1100
- LANCICHINETTI, A.; S. FORTUNATO; J. KERTESZ: Detecting the overlapping and hierarchical community structure in complex networks. In: *New Journal of Physics*, 11(3), 2009, 033015

- LEIFELD, P.; S. J. CRANMER; B. A. DESMARAIS: Temporal Exponential Random Graph Models with btergm: Estimation and Bootstrap Confidence Intervals. In: *Journal of Statistical Software*, 83(6), 2017, pp. 1-36. doi: 10.18637/jss.v083.i06
- LESKOVEC, J.; J. KLEINBERG; C. FALOUTSOS: Graphs over time: densification laws, shrinking diameters and possible explanations. In: *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery in data mining*. ACM, 2005, pp. 177-187
- LUSHER, D.; J. KOSKINEN; G. L. ROBINS (Eds.): *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge [Cambridge University Press] 2013
- MERTON, R. K.: The Matthew effect in science: The reward and communication systems of science are considered. In: *Science*, 159(3810), 1968, pp. 56-63. doi: 10.1126/science.159.3810.56
- MONGE, P. R.; N. S. CONTRACTOR: *Theories of Communication Networks*. New York City/NY [Oxford University Press] 2003
- NEUBERGER, C.: Internet, Journalismus und Öffentlichkeit. Analyse des Medienumbruchs. In: NEUBERGER, C.; C. NUERNBERGK; M. RISCHKE (Eds.): *Journalismus im Internet. Profession – Partizipation – Technisierung*. Wiesbaden [Springer vs] 2009, pp. 19-105. doi: 10.1007/978-3-531-91562-3
- NEWMAN, M. E. J.: Power laws, Pareto distributions and Zipf's law. In: *Contemporary physics*, 46(5), 2005, pp. 323-351. doi: 10.1080/00107510500052444
- NEWMAN, M. E. J.: *Networks: an introduction*. Oxford [Oxford University Press] 2010
- NEWMAN, M. E. J.; M. GIRVAN: Finding and evaluating community structure in networks. In: *Physical review E*, 69(2), 2004, p. 026113. doi: 10.1103/PhysRevE.69.026113
- NUERNBERGK, C.: *Anschlusskommunikation in der Netzwerköffentlichkeit: Ein inhalts- und netzwerkanalytischer Vergleich der Kommunikation im ›Social Web‹ zum G8-Gipfel von Heiligendamm*. Baden-Baden [Nomos] 2013. doi: 10.5771/9783845246529
- NUERNBERGK, C.; J. NEUBARTH: Netzwerkanalysen in der sozialwissenschaftlichen Online-Forschung. In: WELKER, M.; M. TADDICKEN; J.-H. SCHMIDT; N. JACKOB (Eds.): *Handbuch Online-Forschung. Sozialwissenschaftliche Datengewinnung und -auswertung in digitalen Netzen*. Köln [Herbert von Halem] 2014

- PALLA, G.; A. L. BARABÁSI; T. VICSEK: Quantifying social group evolution. In: *Nature*, 446(7136), 2007, p. 664. doi: 10.1038/nature05670
- PALLA, G.; I. DERÉNYI; I. FARKAS; T. VICSEK: Uncovering the overlapping community structure of complex networks in nature and society. In: *Nature*, 435(7043), 2005, pp. 814-818. doi: 10.1038/nature03607
- PANZARASA, P.; T. OPSAHL; K. M. CARLEY: Patterns and dynamics of users' behavior and interaction: Network analysis of an online community. In: *Journal of the Association for Information Science and Technology*, 60(5), 2009, pp. 911-932. doi: 10.1002/asi.21015
- PERC, M.: The Matthew effect in empirical data. In: *Journal of the Royal Society, Interface*, 11(98), 2014, p. 20140378. doi: 10.1098/rsif.2014.0378
- RAND, W. M.: Objective Criteria for the Evaluation of Clustering Methods. In: *Journal of the American Statistical Association*, 66(336), 1971, pp. 846-850. doi: 10.1080/01621459.1971.10482356
- SCHUNTER, J.; M. LIEBAU: Menschliche Beziehungsnetzwerke verstehen!? Morenos Werk aus der Sicht der sozialen Netzwerkanalyse. In: *Zeitschrift für Psychodrama und Soziometrie*, 13(1), 2014, pp. 151-165. doi: 10.1007/s11620-014-0232-2
- SHUMATE, M.: The evolution of the HIV/AIDS NGO hyperlink network. In: *Journal of Computer-Mediated Communication*, 17(2), 2012, pp. 120-134. doi: 10.1111/j.1083-6101.2011.01569.x
- SHUMATE, M.; E. T. PALAZZOLO: Exponential random graph ( $p^*$ ) models as a method for social network analysis in communication research. In: *Communication Methods and Measures*, 4(4), 2010, pp. 341-371. doi: 10.1080/19312458.2010.527869
- SNIJDERS, T. A.: Stochastic actor-oriented models for network change. In: *Journal of mathematical sociology*, 21(1-2), 1996, pp. 149-172. doi: 10.1080/0022250X.1996.9990178
- SNIJDERS, T. A.; G. G. VAN DE BUNT; C. E. STEGLICH: Introduction to stochastic actor-based models for network dynamics. In: *Social networks*, 32(1), 2010, pp. 44-60. doi: 10.1016/j.socnet.2009.02.004
- STEGBAUER, C.: *Reziprozität: Einführung in soziale Formen der Gegenseitigkeit* (2nd ed.). Wiesbaden [vs Verlag für Sozialwissenschaften] 2011
- VALENTE, T. W.: *Network Models of the Diffusion of Innovations*. Cresskill/NJ [Hampton Press] 1995
- WASSERMAN, S.; K. FAUST: *Social Network Analysis: Methods and Applications* (18th sed.). Cambridge [Cambridge University Press] 2009



- WATTS, D. J.; S. H. STROGATZ: Collective dynamics of ›small-world‹ networks. In: *Nature*, 393(6684), 1998, pp. 440-442. doi: 10.1038/30918
- WU, S.; A. DAS SARMA; A. FABRIKANT; S. LATTANZI; A. TOMKINS: Arrival and departure dynamics in social networks. In: *Proceedings of the sixth ACM international conference on Web search and data mining*. ACM, 2013, pp. 233-242

PHILIPP K. MASUR

## Capturing Situational Dynamics and Processes: Strengths and Pitfalls of the Experience Sampling Method

*Abstract:* Communication research has often either used self-report questionnaires or experimental designs to study communication phenomena. Scholars thereby either draw inferences about these phenomena through people's retrospection in surveys or by isolating particular factors and behaviors to identify causal mechanisms. Although both approaches have strengths, they nonetheless fail to capture dynamics and processes that result from different situationally-varying environmental and personal factors. This chapter therefore proposes a situational framework that allows to investigate non-situational and situational antecedents of perceptions or behaviors. It further describes the merits of using the experience sampling method (ESM) for testing such processes. Different types of data collection techniques and sampling strategies are presented and discussed. Using a real-world data set of 164 participants who completed 1,104 situational questionnaires, this chapter further illustrates how data obtained in ESM studies can be analyzed using a multilevel approach. Several aspects of this analytical approach such as power issues, centering decisions, random intercept and random slope models will be presented and discussed. Finally, an analytical strategy to test whether the potential burden of having to complete repeated measurements in ESM designs produces biases in participant's response patterns will be presented.

*Keywords:* Situational processes, experience sampling method, multilevel modeling, sampling procedures, bias in response patterns

## 1. Introduction

What we know about communication phenomena is often based on what people have told us in standardized, retrospective surveys. With the aim of measuring people's feelings, thoughts, attitudes, or behaviors, we often ask participants to remember past experiences and to aggregate them into a single set of coherent responses in a self-report questionnaire (cf. BARRETT/BARRETT 2001: 175). For example, scholars have asked people to indicate what they expect from using a certain medium and have tried to correlate these expected gratifications with participants' self-reported usage frequency (cf. uses and gratification approach; e.g., RUBIN 1983; LEUNG/WIE 2000; SMOCK et al. 2011). Similarly, another field of research has studied the frequency of online communication and its relation with measures of overall well-being (e.g., KRAUT et al. 1998, 2002; VALENZUELA/PARK/KEE 2009; DIENLIN/MASUR/TREPTE 2017). Survey designs are particularly adopted whenever scholars aim at identifying differences in personality factors as the source for differences in perceptions or behaviors (e.g., MOORE/MCELROY 2012; SEIDMAN 2013). In all fields of research, communication scholars have embraced and used survey methods because they are comparably easy to draft and implement, yet allow for interesting insights into interpersonal differences. Although findings from these studies are invaluable, there is an inherent caveat: Whenever we want to measure a certain behavior (that may vary across different situations), we have no choice but to ask participants to estimate their general behavior in retrospective surveys (typical survey questions such as »How often do you usually ...?«, »On average, how willing are you to ...?«, »Please estimate how many times you use ... per day?«). Such questions require individuals to aggregate several, potentially varying experiences of different situations (in which they remember having engaged in that particular behavior) into a coherent set of responses to predefined items. These aggregated measures flatten situational variance into one broad measure that may be indicative of a general tendency (although potentially biased due to difficulties in estimation and recollection), but at the same time do not contain any information about potential deviations that may result from situational factors.

With regard to the examples mentioned above, we could ask: In how far can findings obtained in the tradition of the uses and gratification approach predict whether a person will watch TV in the evening on a particular day? We potentially know what this person generally expects from watching TV,

but we also need to know what she currently feels and in particular which of the many needs that may be satisfied by watching TV are currently not satisfied. Going further, we might realize that we also would need to know if other persons are present, if the person routinely watches the show that runs on that particular evening, how she currently feels and so on. All these latter factors are inherently situational and vary from evening to evening, from day to day, or even from one situation to the next. With regard to the second example, the rationale is no different: Well-being (as many other concepts) has even been theorized as having state and trait components (DIENER et al. 1999). It is easily imagined that there are situations in which a lot of communication does not contribute to our well-being and others in which too little communication likewise negatively affects our well-being (e.g., whether communication positively influences our well-being surely depends on who we are communicating with).

Of course, not all research uses survey designs: Scholars have also studied communication phenomena by systematically varying the characteristics of a situation and observing changes in a person's affective, cognitive, or behavioral reactions. Experiments are thus valuable for measuring situational behavior, particularly because they allow to prove causality. However, they likewise have several caveats (e.g., CZIENSKOWSKI 1996; BROSIUS/HAAS/KOSCHEL 2016): Experiments are designed to maximize internal validity by isolating single effects and thereby ruling out alternative explanations. But they also reduce the external validity and thereby the generalizability of the findings. Experiments in laboratory settings generally represent highly artificial situations. More importantly, they only capture a small amount of possible intraindividual variance and cannot provide information about the complexity and diversity of situational dynamics. Experiments (whether between- or within-designs) mostly just represent one to three situation(s) among a multitude of similar or diverging situations encountered in daily life. Given that an individual is always influenced by a changing immediate external and social environment (cf. STONE et al. 2007: 7), this may result in findings that are difficult to generalize to everyday life situations.

We hence know that several environmental and personal factors are essential in determining whether a person shows a particular behavior or experiences a certain feeling (e.g., decides to use a certain medium, feels happy or sad ...). In light of this, I argue that it is important to adopt a situational perspective in order to gain a more comprehensive understanding of many communication phenomena. That is not to say, however, that

we should neglect dispositional factors. Instead, I suggest that we should acknowledge and consider dispositional (or trait-like personal factors), contextual, and situational factors when trying to understand behavioral processes. Indeed, some environmental and personal factors are stable and thus non-situational, others are consistent across some, but not all situations, and some factors are different in each situation.

That being said, it currently seems that we are tempted to explain exclusively in dispositional terms (i.e. non-situational person-related factors) what ought to be understood largely in situational terms. In light of this, this paper discusses the experience sampling method (ESM), which can be regarded as an extensive longitudinal research methodology, designed to capture people's momentary feelings, thoughts, behaviors across different situations and as they are encountered in participant's daily lives. It thereby allows researchers to sample a broad range of variables in different situations (BARRETT/BARRETT 2001: 176). This chapter thus represents a primer, illustrating not only the promise and opportunities of ESMs in capturing situational dynamics, but also the challenges and difficulties associated with conducting such research. In the following, I will first provide a situational framework that allows to capture both non-situational and situational antecedents of perceptions or behaviors. I will then provide a more nuanced rationale for the experience sampling method. Third, I will describe different types of experience sampling methods in more detail. Fourth, I will discuss aspects of collecting and analyzing experience sampling data. Multilevel modelling is presented as a standard analytical procedure using a data set that was collected to investigate both personal and environmental factors of self-disclosure in mobile communication situations (MASUR 2018). In doing so, I will also discuss whether the somewhat intensive burden of responding to several prompts per day could have an influence on the data quality. Finally, I will summarize the promises and pitfalls of the method and provide some perspectives on how future research may benefit from adopting a situational perspective and implementing experience sampling designs.

## 2. Adopting a situational perspective and sampling experiences

Combining the two predominant approaches towards analyzing communication phenomena (drawing inferences from retrospective survey

measures or studying causal mechanisms with experiments) requires an in-depth examination of the situation as the unit of analysis. Although the literature is full of attempts to define the situation, we must at least temporarily settle on a working definition that allows us to make the manifold variety of situations amenable to theoretical and empirical investigation.<sup>1</sup>

For the purpose of this article, it is important to note that there is a certain agreement between scholars of different backgrounds that both stable and situationally activated personal factors and environmental cues determine a person's psychological representation of a situation (e.g., LEWIN 1951; MAGNUSSON 1981; SAUCIER/BEL-BAHAR/FERNANDEZ 2007; ROSS/NISBETT 2011; RAUTHMANN/SHERMAN/FUNDER 2015a). The term »situation« thereby has been used to represent those parts of the »total world« that an individual can experience through sensory perception, and interprets as having reference to himself and his behavior (MAGNUSSON 1981: 15). These combinations of environmental and personal factors can be positive (or driving) forces and negative (or inhibiting) forces (ROSS/NISBETT 2011: 14) of situationally exhibited behavior.

Based on this work, I argue that we can understand a situation »as the entirety of circumstances that affect people's thoughts, feelings, and behaviors at a given time« (MASUR 2018, chapter 7.1). These circumstances include both personal and environmental factors (cf. MASUR 2018, chapter 7.1.3) that vary with regard to their stability. Personal factors, on the one hand, refer to characteristics of the person itself, such as personality traits and trait-like qualities (e.g., opinions, attitudes, skills, long-term goals ...) that are rather stable and thus do not vary across situations. On the other hand, they also refer to internal affective and cognitive processes (e.g., perceived duties, motives, goals, or feelings) that are situationally activated and thus less stable. Environmental factors can be differentiated into interpersonal (e.g., assessments of other persons) and external factors (physical or virtual cues of the environment) that generally differ across situations. That being said, they sometimes also show a certain stability across contexts. For example, in the context of my family, my assessments of those persons that are typically present (e.g., my parents or siblings) will be fairly similar across all potential situations that may

1 The literature is too vast to be summarized here. For comprehensive overviews and reviews, see e.g., MAGNUSSON 1981; RAUTHMANN/SHERMAN/FUNDER 2015a, 2015b; MASUR 2018.

occur within this context. Similarly, and unequally more important to studying behavior in media environments, certain characteristics of a medium (e.g., norms in certain media environments) are environmental factors that should to some extent always influence a person's behavior when he or she uses this particular medium. They thus have a certain stability across some situations, yet they are not as stable as person-related factors (e.g., a person can use different media). There is an on-going debate about how long a situation lasts and how the stability of factors can be conceptualized (e.g., RAUTHMANN et al. 2015b). For example, scholars have argued that a situational framework needs to distinguish between (1) occurrences, (2) situations, (3) episodes, (4) life events, (5) typical situations, and (6) contexts. As Rauthmann and colleagues (2015b: 5) denote, »one may think of these concepts as being ›nested‹ within each other like layers of an onion«. For our purpose, it is sufficient to accept that a person experiences a ›new‹ situation as soon as any of the situational factors changes. With regard to the stability of certain factors, a broad differentiation between non-situational and situational factors may often be sufficient. However, if environmental factors are studied that are clearly related to contexts, certain media, or cultures, their stability should be appropriately addressed (see also section 4).

In sum, this situational framework allows to theoretically examine general behavioral patterns while still acknowledging situational variance. Specifically, we can investigate the influence of stable person characteristics on situational behavior while controlling for situational influences both of personal or environmental nature. Additionally, we can study the influence of situational personal and environmental factors on situational behavior while controlling for trait characteristics. Finally, we can also investigate in how far situational influences are affected by more stable context or person characteristics. In many contexts, the effects of situational factors on behavior might differ between persons. We can hence analyze what person-related characteristics moderate potential situational processes. Yet, to investigate this situational framework empirically, we cannot rely on traditional survey methods or experimental designs. Instead, we have to assess situational factors across a variety of differing situations and additionally measure more stable person characteristics. In the following, I will show how a combination of experience sampling and survey methods allows to capture such situational processes in more detail.

### 3. Sampling from everyday life: Assessing actual experiences

The experience sampling method (ESM) is an intensive longitudinal data collection methodology that can be understood as a specific form of ecological momentary assessment (EMA). The basic idea of such methods is to measure thoughts, opinions, feelings, or behavior directly or slightly after they occurred. Other forms of EMA include diary studies, behavioral observations, continuous physiological assessments, or self-monitoring. Within ESM studies, participants are prompted to answer several short self-report questionnaires per day over a longer period of time. While the exact methods, technologies, and sampling strategies can differ, the overall goal is to collect data about momentary states in real-time, in natural environments, with multiple repeated assessments over time (SCOLLON/KIM-PRIETO/DIENER 2003: 5).

Although several groups of scholars developed techniques to measure behavior and momentary states in situ independently, the method that closest resembles its current form was developed in the 1970s by psychologists from the University of Chicago (CSIKSZENTMIHALYI/KUBEY 1981; CSIKSZENTMIHALYI/LARSON 1987; KUBEY/LARSON/CSIKSZENTMIHALYI 1996; HEKTNER/SCHMIDT/CSIKSZENTMIHALYI 2007; LARSON/CSIKSZENTMIHALYI 2014). These researchers aimed at measuring »internal (how people think and feel) as well as external (time, location, and social context) dimensions of experience« (cf. VANDEWATER/LEE 2009: 8). Since then, it has become an integral part of psychological methodology (SCOLLON/KIM-PRIETO/DIENER 2003: 5). In communication science, however, the method has not gained as much popularity so far. Few scholars, for example, have applied the ESM to investigate TV use (CSIKSZENTMIHALYI/KUBEY 1981; KUBEY/LARSON 1990; VANDEWATER/LEE 2009; KARNOWSKI/VON PAPE 2009), video games (KUBEY/LARSON 1990), music listening (SLOBODA/O'NEILL/IVALDI 2001; GREASLEY/LAMONT 2011; RANDALL/RICKARD 2013), gratifications of media use (SCHLÜTZ 2002; SCHERER/SCHLÜTZ 2002), media use for procrastination (REINECKE/HOFMANN 2016), mood and situational Facebook usage (BAYER et al. 2017), and privacy and self-disclosure processes (SCHENK et al. 2012; MASUR 2018). These studies all obtained fine-grained – and often surprising – insights in the studied phenomena. Yet, what is the precise rationale for conducting ESM studies?



### 3.1 Rationale for the experience sampling method

In general, three advantages of the *ESM* can be brought forward (cf. *STONE et al. 2007: 3*): (1) *ESM* minimizes the recalling of information and thus reduces potential bias in self-reports; (2) *ESM* has a high ecological validity as it collects data of momentary states in real world situations, and (3) it achieves a high temporal resolution by repeated assessments and thereby allows for the analysis of dynamic processes over time. In light of the proposed necessity to focus more on situational dynamics of communication phenomena, particularly the second and the third characteristic make a compelling case for using the *ESM*. In the following, I will briefly discuss each rationale.

#### 3.1.1 Momentary data: Reducing recall bias

Traditional surveys provide reliable measures of variables of interest when these are reasonably stable over a longer period of time, represent historical information, or refer to rare or at least regular behaviors (*SCHWARTZ 2007: 12*; *BROSIUS/HAAS/KOSCHEL 2016: 133*). Only in those cases, participants are able to provide accurate information.<sup>2</sup> If the variables of interest refer to more frequent, yet irregular and mundane behaviors or indications of intensity in past experience, retrospective self-reports are likely to be biased (*SCHWARZ 2007: 12*). A growing body of research has shown that the recollection of such information is influenced by many different factors including the salience of recent experiences, heuristics, implicit theories, or survey-related motives such as social desirability (e.g., *BRADBURN/RIPS/SHEVELL 1987*; *NISBETT/ROSS 1980*; *TOURANGEAU/RIPS/RASINSKI 2000*). Human beings are generally not very good at retrieving certain information and thus rather try to reconstruct what must have happened based on heuristics and other mental-shortcuts (*GORIN/STONE 2002*).

*ESMs* reduce the methodological disadvantages of surveys because instead of having to remember past experiences, people are able to report their current experiences and observations directly while or shortly after they occurred. It

2 Depending on the type of information that has to be recalled, survey measures of historical information or rare behavior can still be biased. However, experience sampling methods offer no advantage towards measuring these types of constructs.

has to be noted, however, that depending on the type of sampling strategy (see section 2.2.1), self-reports in ESM studies can also be biased. That being said, the time between the actual experiences and the assessment is often considerably shorter compared to traditional survey designs.

### 3.1.2 Ecological validity: Assessing situational factors in real-world environments

As noted already earlier, we are often interested in individuals' behavior under certain conditions or in specific environments. In other words, we want to investigate within-person processes. In survey studies, environmental influences are barely investigable and can only be achieved by confronting participants with scenarios (also called vignettes) in order to compare their responses after they are prompted to imagine being in these particular scenarios. In laboratory settings, researchers generally try to model real-world scenarios, but findings from such artificial settings may not necessarily transfer into the real world (STONE et al. 2007: 7).

By implementing the ESM, participants answer questionnaires in real-world environments and during or shortly after situations of interest. Researchers are hence able to assess environmental influences on their behavior while still controlling for person-related factors. In context of psychotherapeutical assessments, for example, researchers are often interested in what causes pain. Using an ESM design allows to assess pain whenever it occurs while simultaneously enquiring about potential situational causes. In communication science, assessing communication whenever it occurs allows to investigate what aspects of the environment (e.g., who is communicating, the type of relationship between the communicators, the location ...) as well as which internal personal processes (e.g., motives, feelings, perceived norms or duties ...) influence how much people disclose of themselves, what type of language they use, how they react to certain messages, or any other aspects relevant to the research goal respectively.

### 3.1.3 Temporal dynamics: Analyzing processes over time

Finally, ESM provide longitudinal data that can be used to analyze temporal processes. By sampling several situations (whether time-based, ran-

dom, or event-based, see further below) throughout a person's day for one or two weeks, researchers have the possibility to study intra-individual change of behaviors, moods, or other states over time and independent of situational factors. Research on media use, for example, could benefit heavily from *ESM* because they could try to identify a variety of temporal patterns in media use (e.g., are there diurnal or weekly patterns in TV use? Does Internet use differ between weekdays and weekends?). Although such question can sometimes also be studied with time budget or diary studies, *ESM* often provides a higher temporal resolution of the studied processes.

For example, Hasebrink and colleagues (2013, 2017) argued that we need to investigate people's media repertoires. Assessing which media people use over the course of each day over several weeks might not only provide better assessments of people's media repertoires, but also add a temporal resolution to the analysis and thus providing insights into how media repertoires are connected to different times of the day and how they evolve over time as situational conditions change.

### 3.2 Classification of experience sampling methods

*ESMs* can be differentiated according to their data collection technique and their respective sampling strategy. Each particular method has its own strengths and weaknesses which should be evaluated with regard to its fit with the aims and objectives of respective studies. Particularly the sampling strategy profoundly affects in what situations participants will be prompted to answer the short questionnaires and thus how close these assessments will be to the relevant real-world experiences. Investigators have to consider their research questions and hypotheses, their knowledge about the studied phenomenon, practical and logistical considerations, and also need to develop a priori defined plans for the data analysis in order to find the appropriate data collection and sampling technique (cf. SHIFFMAN 2007: 27).

#### 3.2.1 Data collection techniques

In early *ESM* studies, participants were given small hand-held devices such as pagers or alarm watches (personal digital assistants = PDA) that signaled

them when to answer short paper-pencil questionnaires. Today, three types of data collection approaches can be distinguished (HOFMANN/PATEL 2015: 238): (1) the original PDA approach, which gives the researcher much freedom in programming the necessary sampling strategy, but requires participants to carry extra devices at all times; (2) the short message service (SMS) survey distribution approach, which refers to sending links to mobile versions of a web-based survey via SMS to individuals' smartphones and thus requires that participants' devices have Internet access and (3) the local application approach which requires participants to install a specifically-programmed application on their own smartphones that both signals and collects data and only requires Internet access in order to upload data every once in a while. The proliferation of smartphones in recent years as well as constant improvements in their technical capabilities have made the SMS survey distribution and the local app approach much more feasible. Furthermore, the time needed to answer ESM questionnaires is thereby significantly reduced. As much contemporary communication practices occur via smartphones, sampling smartphone-based communication situations or smartphone usage episodes becomes even more unobtrusive. The specific benefits of the mobile experience sampling method (M-ESM; KARNOWSKI/DOEDENS 2010), however, will be discussed further below.

### 3.2.2 Sampling designs

As the goal of any ESM study is to capture experiences in real-world settings using multiple repeated assessments, researchers have to carefully evaluate what kind of data collection arrangement they want to implement. Most importantly, they have to ask themselves what situations they want to sample and how they will be sampled and measured. Most studies will be either focusing on assessing the flow of behavior and experience over time (and thereby receive reliable assessments of varying variables with minimal retrospective bias) or on assessing particular events in which situational processes will be analyzed (e.g., measuring a variety of situational factors including personal and environmental characteristics that may be used to predict situationally varying behavior). A first question is hence whether a study aims at sampling discrete situations or rather a continuous flow of experiences. Whereas the former requires to create a sampling scheme that allows to get as close to the relevant situations as possible, the latter can only be accomplished by intensive

time-based or random sampling procedures. Table 1 provides an overview of typical sampling designs and combinations (cf. also BARRETT/BARRET 2001: 176; SCOLLON/KIM-PRIETO/DIENER 2003: 7f.; SHIFFMAN 2007: 30f.).

We can generally distinguish between (1) time-based sampling, (2) random sampling, (3) event-based sampling, and (4) combinations thereof. Time-based sampling refers to all sampling designs that prompt participants to answer questionnaires at regular or irregular yet set times throughout the day. We can further distinguish between daily sampling, interval sampling, and intensive sampling. Daily sampling is particularly valuable, if the investigator is interested in summative measures for each day. Similar to panel studies, such designs allow to investigate day-to-day changes or how day-specific characteristics influence aggregative measures for each day. Interval sampling refers to prompting participants at pre-defined times per day (both at set intervals or at systematically varying intervals). It is hence particularly valuable when the study focuses on phenomena that vary considerably in intensity or character at different times during the day, e.g., feelings of hunger or tiredness (SHIFFMAN 2007: 35). However, if the phenomenon under investigation is considered to be continuously present (yet with varying intensity, e.g., measures of well-being or feelings) and a continuous analysis is wanted, interval sampling may not provide a representative sample of all possible experiences or situations. In this case, intensive sampling (or a random sampling procedure; see further below) can provide higher data quality. Intensive sampling refers to prompting people in short intervals (e.g., every 30 minutes or every hour). That being said, such a procedure is very burdensome for participants and might even interfere and influence the phenomenon under investigation.

Random sampling generally refers to prompting participants to answer questionnaires at irregular, random intervals. In comparison to time-based interval sampling designs, such a procedure is often implemented to avoid potential bias and ensure a representative sample of a person's overall experience (SHIFFMAN 2007: 39). Most *ESM* studies implemented such a design in order to obtain a representative sample of all situations their participants experienced during the time of the study.

Finally, event-based sampling designs are most important for studying situational dynamics and processes. The general idea is to prompt participants to answer a questionnaire whenever the studied phenomena occur (e.g., the person watches TV, feels pain, is happy, has communicated, uses the smartphone...). Event-based sampling strategies can be categorized into

TABLE 1  
Overview of different sampling designs

Sampling-Approach	Definition	Target phenomena	Advantages	Issues
<b>Time-based sampling</b>				
Daily sampling	Participants are prompted to answer the questionnaire once a day (usually at the end of the day)	Summary of day's events or experiences (e.g., media consumption per day ...)	Coverage of whole day via recall and summary	Recall-based assessment may lead to bias; Time of completion may influence responses
Interval sampling	Participants are prompted multiple, pre-defined times a day (e.g., 8.00 am, 2.00 pm, and 8.00 pm)	Summary of certain episodes throughout the day (e.g., phase-related variables such as stress)	Coverage of whole day via recall and summary of smaller episodes	Recall-based assessment may lead to bias; Time of completion may influence responses
Intensive sampling	Frequent prompts (e.g., once per hour or even more)	Fast-changing or continuous phenomena (e.g., feelings, stress ...)	Almost continuous assessment of the phenomena of interest	Significant subject burden due to intensive prompting
Random sampling	Participants are prompted to answer the questionnaire at random, variable intervals (e.g., five times a day)	Continuous phenomena (e.g., levels of well-being, stress ...)	Aims to achieve a representative sample of occasions; allows for stratified or weighted sampling schemes	If certain events are of interest (e.g., media use), measures might be recall-based; requires devices for prompting and implantation of sampling strategy
<b>Event-based sampling</b>				
Self-sampling	Participants identify the pre-defined events or episodes and answer the questionnaire autonomously	Pre-defined events or episodes that are easily accessible to individuals' perception, but impossible to track (e.g., eating, smoking ...)	Multiple events can be tracked; no extra signaling devices necessary (e.g., by using printed questionnaires)	Events must be well defined; compliance per questionnaire hard to document

Sampling-Approach	Definition	Target phenomena	Advantages	Issues
Automatic sampling	Participants are prompted to answer the questionnaire whenever a pre-defined event or episode occurs (requires the trackability of the event)	Pre-defined events or episodes that can be tracked (e.g., TV use, application use, smartphone use, using a particular website, watching a particular show ...)	Multiple events can be tracked; compliance can be documented	Tracking and prompting device necessary (e.g., mobile phone); requires a complex sampling strategy (particularly when the pre-defined events occur often); willingness to participate may be low due to privacy concerns
<b>Combinations</b>				
Event-based, time-based or random sampling and daily sampling	Combining within-day event or time-based sampling with end of the day summary	Daily summary after within-day assessments (e.g., assessing evaluations of daily media use and assessing situational factors of actual media use)	Daily and within-day assessments can be sampled independently	Daily assessment is recall-biased and should be analyzed separately
Event-based and random sampling	Combining event-based and random sampling	Correlates or antecedents of events (e.g., motives to watching TV and gratifications while or after watching TV)	Analyses of retrospective or prospective event antecedents	Overlaps between event-based and randomly sampled events possible

self-sampling and automatic sampling. Self-sampling relies on participants to identify target events themselves. Such a design hence requires a clear a priori definition of the events that should prompt participants to answer a questionnaire. If the event is objectively measurable (e.g., every time a person has watched tv), such a sampling procedure can be very useful. However, if the event is defined as a subjective experience (e.g., pain), idiosyncratic biases may arise. Furthermore, compliance in event-based self-sampling designs is hard to document. If the goal is to measure the frequency of certain events, such a design may provide less valid results. Automatic event-based sampling refers to all procedures that allow to trigger a questionnaire automatically after a certain event occurred. Schenk et al. (2012), for example, provided participants with a browser extension that logged participants Internet use. Every time a person engaged in specific Facebook activities (e.g., posted a status update) and a corresponding link was saved as log datum, the software automatically triggered a questionnaire that popped up as a new window in the browser (a similar approach was also implemented by BAYER et al. 2017). Masur (2018) used the local application approach. In his two-week study, participants were required to install the application *movisensXS* on their smartphones which tracked their application use and provided the possibility to automatically trigger situational questionnaires after participants had used certain pre-defined applications (e.g., WhatsApp, Twitter, Facebook ...). The advantage of such a sampling design is that it allows to sample directly from pre-defined situations. If the goal of a study is to sample such specific, clearly defined situations, time-based or random sampling procedures are less useful as they sample independent of whether the event of interest occurred. Although there is the chance that the event of interest is sampled, the probability of missing it is very high. In those cases, we can only ask participants to remember the last time the event occurred which again introduces bias due to difficulties in recollection.

Finally, it is important to note that all sampling designs may be combined to achieve an even greater temporal resolution. Particularly combining time-based, random, or event-based sampling schedules with daily assessments can be useful to get a more detailed picture of the studied phenomena. It should be noted that the literature provides no single recommendation on how many times participants should be prompted per day. According to Christensen et al. (2003: 61), decisions in this regard (independent of the used sampling method) should be based on four criteria: (a) the number of observations needed for stable estimations on the within-person level (see also power



consideration in section 4.1), (b) the naturalistic occurrence of the studied situations, (c) the burden to participants, and (d) anticipated compliance.

### 3.2.3 Focus: The mobile experience sampling method

Some scholars have recently introduced the term mobile experience sampling method (M-ESM) for research designs in which mobile devices are used to collect data (KARNOWSKI/DOEDENS 2010; KARNOWSKI 2013; RANDALL/RICKARD 2013). In light of the growing proliferation of smartphones as daily companions of individuals around the world, communication scholars thereby have the unique opportunity to study mediated communication on individuals' own devices (GÖRLAND 2017: 23). This has several beneficial implications for conducting mobile communication research (cf. KARNOWSKI/DOEDENS 2010: 214f.): First, participants do not have to carry an extra device (e.g., electronic pager) which makes sampling less obtrusive and less burdensome. Second, participants do not have to carry printed questionnaires as web- or application-based surveys can be specifically programmed and assessed right from participants' smartphones. M-ESM often allow to record time stamps both for the trigger and for the completion of the survey. Researchers hence obtain data on whether and when a questionnaire was answered and thereby document compliance more easily.

With continuously growing technical abilities of regular smartphones, M-ESM studies may furthermore allow to integrate other methods of data and content collection (e.g., multimedia content, smartphone sensor data such as GPS, etc.). This way, the local application approach allows to implement automatic event-based sampling procedures (e.g., using movisensXS) and participants' smartphone use can be objectively assessed by recording log files. Finally, data can be collected, saved, and uploaded during the study reducing the probability of data loss.

## 4. Collecting and analyzing experience sampling data

When conducting an ESM study, the researcher is confronted with many decisions that involve (1) deciding for an appropriate strategy to recruit participants (including power analysis and incentive strategies), (2) investi-

gating which of the sampling strategies above is most appropriate to study the phenomenon of interest (including decisions with regard to filtering and potential sampling breaks), (3) evaluating the necessity of conducting several larger pretests, (4) deciding about necessary data preparation steps (including missing value analyses and centering decisions), (5) choosing an appropriate data analysis method, and (6) evaluating the quality of the data. At any rate, researchers should clearly describe and report their design and decisions (for first reporting guidelines, see STONE/SHIFFMAN 2007: 371-386).

I will exemplify these decisions with the help of a multi-method study on privacy and self-disclosure processes in smartphone-based communication situations (MASUR 2018)<sup>3</sup>. This study combined survey (to assess non-situational personal factors), tracking (logging smartphone and application use), and experience sampling methods (to assess situational personal and environmental factors). Participants answered in an initial online survey and subsequently installed a specifically programmed application (movisensXS) on their own smartphone that tracked their smartphone use for 14 days. Following an automatic event-based sampling design, the application automatically triggered situational questionnaires (roughly a maximum of 6 questionnaires per day depending on the individual applications use) after participants used pre-defined applications (i.e. whenever the log data revealed that participants used communication applications such as WhatsApp, Threema, e-mail, Facebook, Google+, Twitter, Instagram ...). Overall,  $N = 164$  smartphone users took part in the study ( $M = 30$  years, 67% female). Overall,  $N = 3,802$  questionnaires were prompted. The number of prompts per day per person varied considerably due to individual differences in application use over the course of the study ( $M = 20.56$  per person on average,  $SE = 15.04$ ; range = 1-87). The following analyses are based on a subset of  $n = 1,104$  situations in which participants indicated having communicated to one or several other persons.

3 For more information about the methods used in the overall study, see Part II in MASUR (2018). For the subset analyzed in this paper and the corresponding R code, see the online supplement: <https://osf.io/jjpdvc/>

#### 4.1 Preparation and data collection

It is important to note that conducting ESM studies is disproportionately more complex than data collection methods with one or only a limited number of repeated observations.<sup>4</sup> First, several pretests are recommended for different reasons. On the one hand, investigators should take care to develop reliable short scales or items for the situational assessments. Larger scales should be avoided due to the repeated measurements and the related burden for participants. A preliminary survey can help to identify those items of a larger scale that best represent the latent variable (e.g., the item with the highest factor loading) or to develop more generic items (e.g., by correlating these new items with the latent construct of an existing scale). On the other hand, the technical side of ESM designs should be tested carefully prior to the actual data collection. Final simulations of the whole design should be undertaken with several subjects in order to evaluate the burden for future participants and to identify potential problems that may arise during the study. It is furthermore important to ensure that participants are able to contact the researcher at any time during the study.

A second a priori consideration refers to estimating the power needed to test situational and/or personal effects. In ESM studies, it is important to note that effects of non-situational, person-related characteristics on behavior are estimated as interpersonal effects and are thus based on the number of participants. Effects of environmental or situationally varying personal factors on behavior, in contrast, are tested based on the number of situations obtained by the respective sampling strategy. For example, if we want to test small non-situational effects (i.e. an effect size of  $r = .10$ ) while assuming an alpha-error of 5 percent, we would need to recruit  $N = 779$  participants to achieve a power of 80 percent. In the example study ( $N = 164$  participants), we are only able to test non-situational effects with moderate effect sizes ( $r > .21$ ) with a power of 80 percent. On the situational level, however, it is much easier to achieve enough power to test small effects. If  $N = 164$  participants provide on average  $M = 6$  situational questionnaires ( $= 164 * 6 = 984$  situations), we can test situational effects with a power of

4 Of course, there are also very complex surveys and experimental designs. Yet, compared to most of such studies, ESM studies are nonetheless rather complex and involve many unusual decisions.

88 percent. In the present study, participants on average disclosed private information in  $M=6.78$  ( $SD=5.34$ ) situations during the time of the study (total number of situations that can be used in the analysis:  $N=1,104$ ). Small situational effects can hence be tested with a power of 92 percent.

Finally, it should be noted that *ESM* studies require participants to be engaged with the study for a longer period of time (usually 1 to 2 weeks) and they involve being prompted to answer questionnaires several times a day. Thus, the willingness to participate is considerably lower compared to traditional survey studies. If additional data is collected (e.g., log data), privacy concerns might also influence the willingness to participate. Researchers aiming at recruiting large sample sizes (particularly on the person level) should consider using higher incentives for participation.

## 4.2 Analyzing the data

Analyzing data obtained from *ESM* studies can be tricky. On the one hand, such data are generally organized in a hierarchical structure that we have to account for in our analyses. On the other hand, they often require preparative steps before we can actually analyze them with statistical models. Of course, the specific types of data preparations and the analytical approach depends on the research questions or hypotheses. Depending on the study's goals, methods may include aggregative approaches, time series analysis, structural equation modelling, or multilevel approaches. However, when studying situational dynamics of human behavior as outlined above, multilevel modelling often represents the most valuable approach. In the following, I will discuss important aspects and potential problems that scholars typically encounter when adopting such a data analytical approach.

### 4.2.1 Hierarchical data structures

In *ESM* studies, it is important to remember that we sample data in two consecutive steps. First, we take a sample of individuals, and next we take a sample of situations that these individuals experience in their daily lives (cf. *HOX* 2010: 2). The data structure is called hierarchical because several observations are nested within persons (or: several repeated measurements are clustered within individuals). Sometimes, it could be meaningful to think about other

nested structures in such data sets as well. For example, situations could also be nested in media environments (e.g., situations that occurred while using Facebook compared to situations that occurred while using WhatsApp), contexts (family, friends ...) or certain time-based structures (days, weeks, mornings, evenings ...). Even the participants could be nested in countries or cultures. Whether to account for these hierarchies or not depends on the variance in the respective factor and the research question at hand. In the following, I will only discuss the typical data structure that mostly results from conducting *ESM* studies: measurements nested in persons.

Depending on the study's design, variables may be defined both at the personal (Level 2) and at the situational level (Level 1). In the example study, I measured socio-demographics, trait-like characteristics such as personality and other non-situational factors such as privacy concerns directly at the personal level (in a preliminary survey). Additionally, I measured situationally varying factors such as interpersonal assessments and people's behavior (e.g., the depth of self-disclosure) at the situational level (in the subsequent *ESM* study). A trivial, yet important first step is to merge both types of data sets. For being able to do so, it is important to provide each of the data sets with unique person identifiers (IDs). Only with such IDs, one is able to determine which situational measures belong to which personal measures. There are generally two ways of merging these two data sets which will be discussed in the next section. To exemplify the procedure, I used a subset of the original data set which only includes two variables on the person level (age and sex) and three variables on the situational level (app used, interpersonal trust, and self-disclosure).

#### 4.2.1.1 Data formats

We can generally distinguish between a so-called wide and a long format. In the wide format, each row refers to one subject. A subject's repeated responses is hence in a single row and each response is in a separate column (see Tab. 2). Repeated measures are indicated by the respective time stamp in front of the variable name (e.g., T1\_, T2\_, etc.). The wide format allows to conduct descriptive analyses of person-related variables. Similar to data formats obtained from classical survey studies, for example, we can estimate population means. Using simple descriptive analyses, we can conclude that our example sample of participants is  $M = 30$  years ( $SD = 10$  years) in average and 67 percent of them are female.

In the long format, each row refers to one situation in which the situational variables were assessed (see Tab. 3). Each response per situational questionnaire is hence in a separate column. In contrast to the wide format, several rows now belong to one person (as indicated by the ID variable).

TABLE 2  
Example ESM data (long format)

id	age	sex	time	it	sd
25	24	1	T1	9,7	5,8
25	24	1	T2	9,7	7,2
35	24	1	T1	7,6	8,3
35	24	1	T2	9,3	8,7
35	24	1	T3	9,1	6,6
35	24	1	T4	5,8	3,6
35	24	1	T5	9,4	0,5
...	...	...	...	...	...

Note: Each row represents measures in one situational questionnaire.  
it = interpersonal trust, sd = self-disclosure

TABLE 3  
Example ESM data (wide format)

id	age	sex	T1_it	T1_sd	T2_it	T2_sd	...
25	24	1	9,7	5,8	9,7	7,2	...
35	24	1	7,6	8,3	9,3	8,7	...
37	40	1	7	6,1	8,1	6,4	...
41	47	1	6,2	3,2	6,6	5,8	...
48	33	1	8,2	6,4	8,5	5,6	...
49	27	1	4,9	3,2	9,6	0,2	...
54	26	1	9,8	1,1	9,9	5,8	...
...	...	...	...	...	...	...	...

Note: Each row represents one participant.  
it = interpersonal trust, sd = self-disclosure

The long format is typically used to estimate multilevel models (see further below). It is generally possible to move a variable from the situational level to the personal level by aggregation (HOX 2010: 2). For example,

I could assign to each person her or his mean self-disclosure score. Disaggregation, in contrast, refers to moving a variable from a higher to a lower level. In the wide format, age and sex are person-related characteristics that occupy one column each. In the long format, we need to assign such person-related characteristics to each situation (similar to assigning the mean score of situational variables to each situation). As Hox (2010: 3) denoted, multilevel problems have historically led to approaches that moved all variables by aggregation or disaggregation to one single level. The resulting data set was then used to estimate ordinary multiple regression models, to conduct analysis of variance, or to use other standard analysis methods. However, using aggregation or disaggregation in this manner, i.e. moving variables that belong to different levels to one single level (in ESM designs, this was often the personal level), is inadequate for several reasons (Hox 2010: 3): (1) Aggregation results in a loss of information and also in a loss of power. (2) Disaggregation is similarly problematic as »using the larger number of disaggregated cases for the sample size leads to significance tests that reject the null-hypothesis far more often than the nominal alpha level suggests« (Hox 2010: 3). (3) Aggregation and disaggregation may easily lead to wrong conclusions (e.g., when disaggregated data is wrongly interpreted on the situational level).

Multilevel models avoid these statistical problems and treat each variable appropriately with regard to its generic level of measurement. As we will see, however, aggregation and disaggregation is nonetheless useful if appropriately implemented.

#### 4.2.1.2 Missing values

In data analyses of traditional survey or experimental data, we often treat missing values resulting from item nonresponse with listwise deletion. Particularly if the amount of missing values is small, such an approach can be appropriate. In ESM studies, in contrast, missing values can occur on both levels and thereby amount to a cause of concern. A first type of missing values results from participants not providing the same number of situations. Depending on the type of sampling strategy and participants' compliance, participants can differ significantly regarding the number of completed situational questionnaires. Although systematic missingness may result in bias, a varying number of measurements per person is not necessarily worrisome for the data analysis because even one or two obser-

variations per person can be enough to fit a multilevel model (HOX 2010: 106). That being said, the (random) intercepts may not be estimated precisely for persons with few observations. Yet, even few observations per person can still provide partial information for the estimation of the regression coefficients and variance parameters on both levels (GELMAN/HILL 2009: 276).

A second and more concerning type of missing values refers to participants' item nonresponse both in the preliminary survey and in the situational questionnaires. If we implement listwise deletion, nonresponse to one item in a situational questionnaire would cause the deletion of that situation and would thus only leave that particular participant with fewer situational measurements. Missing values in the preliminary survey, however, are more worrisome as they would eliminate the participant from the analysis. Given that it is more difficult to recruit participants for ESM studies, it can be disastrous if a person who provided high quality data in each situational questionnaire is deleted only due to one missing item in the preliminary survey. In sum, treating missing values with case-wise deletion may severely decrease the power to detect effects on both levels. Hence, I highly recommend conducting a missing value analysis before engaging in the actual data analysis and consider imputation strategies if the missing values can be considered missing at random (for an overview, see SCHAFER/GRAHAM 2002).

#### 4.2.1.3 Data preparation and centering decisions

A final step of data preparation refers to centering. Many communication phenomena and particularly psychological constructs are expressed on arbitrary metrics (e.g., Likert-type scales) that lack a clearly interpretable or meaningful zero point (BLANTON/JACCARD 2006). Centering can be used to establish more meaningful zero points on these scales (e.g., the population's mean) resulting in interpretable intercepts. In the context of multi-level modelling, however, centering decisions are more complex as situational measures (Level 1) can be centered around the grand mean (i.e., the population's mean) or they can be centered around a person's individual mean (i.e., the mean of a person's repeated measurements). Whereas the first has been termed centering at the grand mean (CGM) or grand-mean centering, the latter is referred to as centering within cluster (CWC) or group-mean centering (ENDERS/TOFIGHI 2007: 121). Although both centering approaches can be used to establish meaningful zero points on level 1



variables, it is important to note that they produce different parameters that must be interpreted differently (ibid.: 121). It can be reasonable to use both in the context of one study when different research questions need to be answered (ENDERS/TOFIGHI 2007: 136). Enders and Tofighi (2007) generally propose the following guidelines: (1) CWC is appropriate when one is primarily interested in relationships between two Level 1 variables, (2) CGM is appropriate when one is interested in Level 2 predictors while simultaneously controlling for Level 1 covariates; and (3) CWC is preferable when one is interested in investigating cross-level interactions and interactions that involve a pair of Level 1 variables, while CGM is appropriate for interactions between Level 2 variables (ibid.: 136).

In the context of ESM studies which aim at investigating situational processes, the following procedure, however, is recommended: First, all level 2 predictors that lack a meaningful zero point should be centered using CGM. This way, we account for (2) and (3) of the guidelines and get an interpretable intercept for Level 2 predictors (the value of the dependent variable when the predictor equals the population's mean). Second, level 1 predictors should be separated into between-person (non-situational) and within-person (situational) variance. To do so, one first needs to aggregate the situational measure to the (higher) person level by estimating the mean per person (e.g., estimating the average perception of interpersonal trust for each participant in the study, see online supplement). Including this aggregate measure allows to identify influences of stable variance parts in situational measures on the stable variance parts in the dependent variable. Second, the situational measure should be centered around this aggregated person's mean thus accounting for (1) and (3) of the guidelines. This centered situational measure represents situational deviations from a person's average ESM measure and hence allows to identify truly situational effects. In the following multilevel analysis, we will include both the person's mean (the aggregated measure) and the person-mean centered situational variable in the model.

#### 4.2.2 Multilevel analysis

The multilevel regression model (also known as random coefficients model, variance component model, hierarchical linear model, or mixed-effects model) generally assumes a hierarchical data set with a single dependent variable at the lowest level (HOX 2010: 11). In our example, we will use a

multilevel model to predict the depth of self-disclosure (variable code = sd) using both Level 1 (interpersonal trust; variable code = it) and Level 2 predictors (age and sex).<sup>5</sup>

We could choose to ignore the multilevel structure of the data and estimate a simple regression model (no pooling approach) which would result in a moderate, positive relationship;  $b = 0.20$ ,  $se = 0.03$ ,  $b = .17$ . Or we could decide to estimate a regression model for each person (complete pooling approach). In this case, the mean of all intrapersonal effects even suggests a negative relationship;  $Mb = -0.18$ ,  $SDb = 6.06$  (for more information on no pooling and complete pooling approaches and their caveats see online supplementary material). However, complete pooling ignores variation between persons and no pooling overstates it (cf. GELMAN/HILL 2009: 253). Only in very limited cases (i.e. when there is very little between-person variance), the multilevel regression model is similar to the simple regression model. The multilevel model is designed as a compromise between these two extremes. It is thus often referred to as a form of partial pooling (GELMAN/HILL 2009: 258). It can be thought of as a generalization of the linear regression model in which intercepts and possibly also slopes are allowed to vary per group (in the case of *ESM* data: per person).<sup>6</sup> The aim is to estimate effects that reflect both the relative information we have about within-person variance and between-person variance (ibid.: 253). We therefore estimate fixed and random effects. Fixed effects can be thought of as weighted estimates of the effects that pertain to all persons. Random effects represent how much each person deviates from these fixed effects.

The most basic multilevel model is the random- or varying-intercept model represented in the following formula:

$$y_i = \alpha_{j[i]} + \beta x_i + \varepsilon_i$$

In this model, we predict a situational outcome variable (with  $i = 1, \dots, n$  referring to situations as the lowest level of measurement) with another situational (or disaggregated personal) variable while allowing the intercept  $\alpha$  to vary across  $j = 1, \dots, J$  persons. We hence assume a fixed effect for the relationship between  $x$  and  $y$ , but acknowledge that each person's intercept may deviate from the fixed intercept (see also Fig. 3).

5 Please note that all analytical steps described in the following sections can be reproduced with the online supplementary material: <https://osf.io/jpdvc/>

6 Formal statistical description can be found in de Leeuw and Meijer (2007), Gelman and Hill (2009), and Hox (2010).

The varying-intercept and varying-slope model further allows slopes to vary between persons.<sup>7</sup> It is represented with the following formula:

$$y_i = \alpha_{j[i]} + \beta x_{j[i]} + \varepsilon_i$$

In this model, we still estimate fixed effects for the intercept and the slope, yet both are allowed to vary. We get additional random effects for both the intercept and the slopes which tell us how much persons deviate from the weighted fixed effects.

#### 4.2.2.1 Assessing personal and situational variance

A first step in conducting multilevel analyses of *ESM* data should be the estimation of a varying-intercept model that includes no predictors. This null-model provides one fixed effect that is equivalent to weighed mean across all situations and persons (i.e., persons with more situational measures provide more information than persons with less situational measures). We further get random effects, which, in this case, represent the variance that is attributable to between-person differences ( $\delta^2_{\text{person|intercept}} = 1.86$ ) and the unexplained residual variance (i.e. situational and error variance;  $\delta^2_{\text{residual}} = 1.45$ ). This is helpful, because it allows us to compute the intraclass correlation coefficient (*ICC*) with the following formula:

$$ICC = \frac{\sigma^2_{\text{person|intercept}}}{\sigma^2_{\text{person|intercept}} + \sigma^2_{\text{residual}}}$$

In our example, the *ICC* for depth of self-disclosure is .29, meaning that 29 percent of the variance in depth of self-disclosure is attributable to stable between-person differences. In other words, person-related, non-situational factors can explain one third of the variance of the depth of self-disclosure. Conversely, this suggests that a larger part of the variance may be due to varying situational factors. Hence, self-disclosure (as a communication behavior) is inherently situational.

#### 4.2.2.2 Estimating fixed and random effects

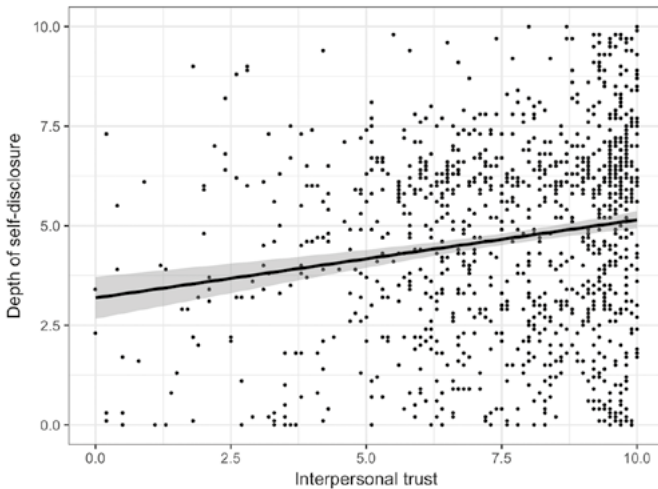
In a next step, we add predictors to the varying-intercept model. Table 4 shows an overview of all estimated models. Model *M*<sub>0</sub> represents the null-

<sup>7</sup> Although random slope models are theoretically possible, they are seldom found in reality. For this reason, I refrained from discussing them here.

model described above. Model M1 includes the situational predictor interpersonal trust (both the person's mean (it.m) and the person-mean centered situational deviance (it.c)). The results reveal that both stable variance ( $b = .23$ ,  $se = .11$ ) and situational variance components ( $b = .23$ ,  $se = 0.03$ ) of interpersonal trust are positively related to depth of self-disclosure. This means, that people, who generally perceive other recipients in smartphone-based communication situations as trustworthy, are also generally more likely to self-disclose. More importantly, the results also suggest that if a person perceives receivers of their disclosures as more trustworthy than normally, she or he will also be more inclined to self-disclose more in that situation. Both findings are visualized in Figure 3 which shows that the same slope was estimated for each person, but intercepts were allowed to differ.

FIGURE 1  
**Complete pooling approach**

A linear regression model is estimated on the basis of all situational measurements. This approach ignores between-person variance and thereby violates the assumption of independent observations.



In model M2, age and sex were included as Level 2 predictors. Only age emerged as a positive predictor ( $b = 0.03$ ,  $se = 0.01$ ) indicating that older people are generally more willing to self-disclose, independent of situa-

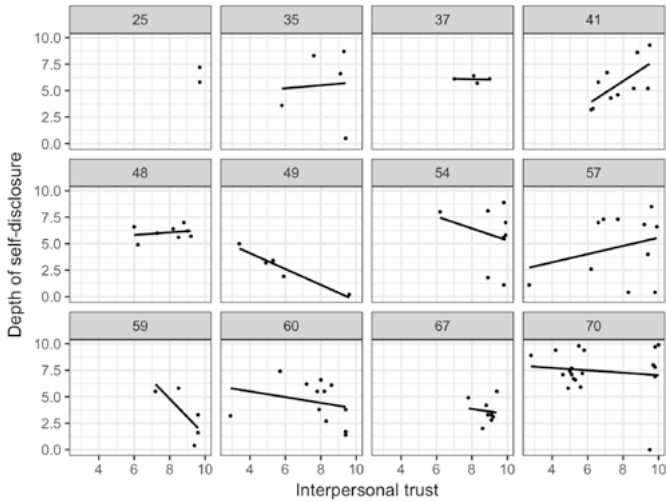
tional trust evaluations. Interestingly, the small effect of trait trust on self-disclosure thereby became non-significant. In model M3, the situational predictor was allowed to vary across situations. We are hence estimating a varying intercept, varying slope model. Although the fixed estimate for the relationship between situational trust and self-disclosure barely changes ( $b=0.22$ ,  $se=0.05$ ), we now also get random effect estimates (i.e., the variance of the deviations from this fixed effect:  $\delta^2_{\text{persontrust}}=0.12$ ). In general, allowing situational effects to vary across persons should result in a better model fit. However, if the increase in model fit is negligible, we could stick with the varying-intercept model which is easier to interpret. As a general rule, it is recommended to estimate both models and compare their model fits. If the increase in model fit is not significant, one can interpret the varying-intercept model appropriately. As the models are nested, they can be compared using the likelihood ratio test (SNIJDERS/BOSKER 2011: 97). This test is based on a comparison of the models' deviances. Models with a lower deviance show a better fit than models with a higher deviance (HOX 2010: 47). As the difference in deviances has a  $\chi^2$ -distribution, we can use a standard  $\chi^2$ -test to assess whether varying slopes significantly improved model fit. In our case, inclusion of random effects for the situational predictor variable did increase model fit ( $\chi^2(2)=38.35$ ,  $p<.001$ ). Looking at the distribution of the random effects (Fig. 4), we can see that some persons deviate considerably from the overall fixed effect. For some individuals, the relationship between situational trust and depth of self-disclosure is even negative.

Finally, model M4 includes cross-level interactions. Whenever a situational effect varies, cross-level interactions can help to understand why they vary. In our example, we analyze whether including the interaction between (1) age and situational trust and (2) sex and situational trust improves the model fit. It is important to bear in mind that after inclusion of interaction terms, the effects of the moderated relationships become conditional. Due to our centering decision, the effect of situational trust on self-disclosure only applies to men who have the population's average age. However, results show that age does not moderate the situational effect of trust on self-disclosure. The relationship between trust and self-disclosure is only slightly stronger for women compared to men. However, the interaction effect is not significant ( $b=0.12$ ,  $se=0.11$ ) and thus does not result in an improved model fit ( $\chi^2(2)=1.23$ ,  $p=.540$ ).

FIGURE 2

## No pooling approach

Linear regression models are estimated for each person. This approach overestimates the between-person variance.



Based on the likelihood ratio test, we reached the best model fit already with model M3. Yet, we do not have any information about its absolute fit. Computing the amount of explained variance is not as straightforward as in standard linear regression models. The simplest approach is to compute how much the model with predictors reduces the residual variance on both levels compared to the null-model. This variance reduction approach can only be used with random-intercept models. In our example, it reveals that model M2 explains 5 percent of the variance on Level 1, but no variance on Level 2. This is not too surprising as the non-situational variables (trait trust, age, and sex) barely influence self-disclosure. Nakagawa and colleagues (2013) recently proposed a marginal and a conditional that are simple measures of the variance explained by fixed effects only and by both fixed and random factors. Johnson (2014) further extended this approach to random-slope models. In this regard, model M3 explains 5 percent ( $R^2_{\text{marginal}} = .049$ ) by fixed effects only, but 42 percent ( $R^2_{\text{conditional}} = .417$ ) by fixed and random effects.

TABLE 4  
**Results from the multilevel models predicting self-disclosure**

	M0	M1	M2	M3	M4
	Fixed Effects				
(Intercept)	4.65 (0.13) ***	2.91 (0.85) ***	3.21 (0.88) ***	3.34 (0.88) ***	3.32 (0.88) ***
Level 1					
Interpersonal trust (CWC)		0.23 (0.03) ***	0.23 (0.03) ***	0.22 (0.05) ***	0.14 (0.09)
Average interpersonal trust (GMC)		0.23 (0.11) *	0.17 (0.11)	0.16 (0.11)	0.16 (0.11)
Level 2					
Age (GMC)			0.03 (0.01) *	0.03 (0.01) *	0.03 (0.01) *
Sex (1 = female)			0.15 (0.28)	0.11 (0.28)	0.13 (0.28)
Cross-level interactions					
Age x Interpersonal trust					-0.00 (0.01)
Sex x Interpersonal trust					0.12 (0.11)
	Random Effects				
	1.86	1.87	1.88	1.97	1.97
	4.45	4.23	4.22	3.76	3.76
				0.12	0.12
	Goodness-of-Fit Statistics				
AIC	4982.42	4936.22	4934.77	4900.42	4903.19
BIC	4997.44	4961.25	4969.82	4945.48	4958.26
log-likelihood	-2488.21	-2463.11	-2460.38	-2441.21	-2440.59
deviance	4976.42	4926.22	4920.77	4882.42	4881.19
D deviance		50.20	5.45	38.35	1.23
df		2	2	2	2
p		< .001	.066	< .001	.540

Note: Fixed effects are estimated by restricted maximum likelihood (REML). Numbers in parentheses represent standard errors.

Basis:  $n_{\text{participants}} = 164$ ,  $n_{\text{situations}} = 1,104$

\*\*\*  $p < .001$ , \*\*  $p < .01$ ,  $p < .05$

### 4.2.3 Assessing potential intervention effects

An important, yet seemingly neglected aspect of analyzing ESM data refers to possible biases resulting from the considerable burden that participants might experience during the study. If individuals are prompted to answer questionnaires several times a day for a longer period of time, it seems likely that this somehow affects how they respond. Stone, Kessler, and Haythomthwatte (1991) denoted that diary studies (and thus ESM studies as well due to their similar assessment) require respondents to maintain interest over a considerable period of time. Otherwise, responses in situational questionnaires might be of poor quality because people might answer more quickly and with less reflection over the course of the study. To the best of my knowledge, there is almost no literature on such potential intervention effects in ESM study. Nonetheless, I believe that several effects can be postulated and should be investigated.

First, we can simply investigate whether the repetition of the same questions somehow affects our variables of interest. We can test this assumption by computing a variable that indicates the current number of repeated measures for each situational questionnaire (i.e., the first questionnaire per person received the number 1, the second number 2, and so on). This variable reflects the temporal position of each questionnaire within the study. We can now use this variable to predict our variable of interest using multilevel modeling. In our example, depth of self-disclosure increased by 0.05 points with each additional questionnaire. However, the effect was not significant ( $b=0.05, se=0.03$ ).

Second, we could assume that participants increasingly deviate from the scale's midpoint. In this particular study, the situational measures were assessed using a slider. A pretest had shown that this answering tool has a high usability. Yet, a slider could also lead to less thoughtful answers by just sliding to right or left extremes. I hence computed another variable, that measures the deviation of the response to the scale's midpoint in each situational questionnaire. Results again suggest that people did not deviate more from the scale's midpoint over the course of the study ( $b=-0.02, se=0.01$ ).

Third, we could assume that the situational variance decreases with each additional situational assessment). However, the difference between two consecutive situations did not decrease significantly over the course of the study ( $b=-0.01, se=0.02$ ).



Finally, we could also test whether the between-person variance decreases over time as response patterns could become more consistent. Therefore, we only have to compute the variance of self-disclosure per situational questionnaire (across all persons). Results suggest that between-person variance did decrease over time ( $b = -0.13$ ,  $se = 0.01$ ).

In this study, these analyses suggest that only small biases exist that can be considered negligible (see also Fig. 5 in the online supplement). From a general perspective, this can be regarded as good news. At least in this study, it seems that the sampling design did not affect the way people responded. However, as the last analysis has shown, ESM designs seem to be able to influence the balance between situational and personal variance in the situationally measured variables. In light of this influence on people's response patterns, particular intensive longitudinal ESM studies should test whether such biases exist and if so, control for them in the main analyses.

## 5. Summary and future perspectives

In this chapter, I have argued that we should often adopt a situational perspective when investigating communication phenomena. Many theories in communication science provide assumptions about situational processes that have been neglected in empirical research. In many fields of research, it seems that we too often overestimate the power of dispositional factors to predict human behavior and simultaneously neglect situational factors. Much of contemporary research either draws inferences about communication phenomena through people's retrospection in surveys or by isolating particular factors and behaviors to identify causal mechanisms. It thereby misses the opportunity to bridge both perspectives. If the goal is to arrive at a comprehensive understanding of human behavior (and I believe it should be), we should not ignore the situation and its potential to alter behavior or behavioral intentions. Based on a long tradition of investigating person-situation interactions, I first proposed a general situational framework which allows to investigate both non-situational and situational factors and their interactions as antecedents of behavior (cf. MASUR 2018).

Second, I have introduced the experience sampling method (ESM) as a promising data collection method to test such a situational framework. The appeal of ESM thereby rests upon three general rationales (cf. STONE et al. 2007: 3): (1) ESM studies avoid the disadvantages of retrospective

self-reports; (2) they increase the ecological validity of self-reports as people are surveyed in their natural environments; and (3) they allow researchers to investigate temporal patterns in people's daily experiences. In sum, ESM studies provide the opportunity to study people's behavior immediately when it occurs and while participants are in their real-world settings. This way, analyzing the influences of the environment and other situational factors becomes possible. With a variety of sampling designs, it provides a promising data collection approach for a variety of research questions.

Third, I have introduced multilevel modelling as a best practice for analyzing situational processes with ESM data. The motivation for conducting multilevel analysis is that it accounts for both situational and non-situational variation in estimating individual-level and situation-level regression coefficients. In contrast to complete pooling or no pooling approaches, it does not under- or overestimate between-person variance. That being said, a drawback of implementing multilevel modelling is its complexity. Several a priori decisions concerning data transformations and centering decisions have to be made and researchers have to carefully plan both the data collection process and the corresponding data analyses. With this chapter, I have outlined several important steps in this regard. It is my hope that this chapter (and its corresponding online supplement) will help scholars, who are interested in adopting a situational perspective and in conducting ESM studies, to overcome potential methodological issues.

In the future, ESM studies will undoubtedly contribute to existing communication research. Particularly the multifunctionality of smartphones and their embeddedness in the daily lives of individuals provides communication scholars with novel opportunities to capture mediated communication and how it is enacted in real-world settings. With the increasing data processing power of handheld devices, scholars should aim at combining ESM data with other (potentially automatic) data collection approaches. For example, ESM data in combination with physiological data obtained from smartwatches can help to study unconscious reactions in communication situations or when people engage with different media content. Combining log data of people's media use with ESM sampling strategies can provide more information about what situations have been sampled and in how far the sampling strategy resulted in a representative sample of a person's experiences. Within the ESM sampling framework, scholars can also ask participants to capture their immediate environments by taking pictures or recording videos and thereby providing an actual window

into their world. In this regard, the ESM could also bring interesting data for qualitative approaches. With the growing mobility of devices, scholars become more and more interested in examining mobile media use and mobile communication. The ESM provides a possibility to study these situations. Combining with data obtained from the smartphone's ability to use GPS and geo-positioning technologies, it is even possible to sample »on the go«-situations automatically and representatively.

In sum, we can conclude that the ESM is a useful and innovative data collection approach that can be used for a variety of research questions. Combined with the situational perspective outline in this chapter, I strongly believe that it can provide novel insights into long-researched as well as new communication phenomena. Yet, as Scollon and colleagues (2003) denote: »Like any other tool, [the ESM's] utility can only be measured by the care with which the researcher plans and conducts the research, analyzes the findings, and interprets the results« (ibid.: 28). If done with care, the ESM can be a powerful tool that provides researchers with fine-grained and rare insights into people's behavior – or, more generally, into people's lives.

## References

- BARRETT, L. F.; D. J. BARRETT: An introduction to computerized experience sampling in psychology. In: *Social Science Computer Review*, 19(2), 2001, pp. 175-185
- BAYER, J.; N. B. ELLISON; S. SCHOENEBECK; E. BRADY; E. B. FALK: Facebook in context(s): Measuring emotional responses across time and space. In: *New Media & Society*, 20(3), 2017, pp. 1047-1067
- BLANTON, H.; J. JACCARD: Arbitrary metrics in psychology. In: *American Psychologist*, 61, 2006, pp. 27-41
- BRADBURN, N. M.; L. RIPS; S. SHEVELL: Answering autobiographical questions: The impact of memory and inference on surveys. In: *Science*, 236(4798), 1987, pp. 157-161
- BROSIUS, H.; A. HAAS; F. KOSCHEL: *Methoden der empirischen Kommunikationsforschung: Eine Einführung* [Methods of empirical communication science: An introduction]. Wiesbaden [Springer] 2016
- CHRISTENSEN, T. C.; L. F. BARRETT; E. BLISS-MOREAU; K. LEBO; C. KASCHUB: A practical guide to experience-sampling procedures. In: *Journal of Happiness Studies*, 4, 2003, pp. 53-78

- CSIKSZENTMIHALYI, M.; R. KUBEY: Television and the rest of life: A systematic comparison of subjective experience. In: *Public Opinion Quarterly*, 45(3), 1981, pp. 317-328
- CSIKSZENTMIHALYI, M.; R. LARSON: Validity and reliability of the Experience-Sampling Method. In: *The Journal of Nervous and Mental Disease*, 175(9), 1987, pp. 526-536
- CZIENSKOWSKI, U.: *Wissenschaftliche Experimente: Planung, Auswertung, Interpretation* [Scientific experiments: Planning, analysis, interpretation]. Weinheim [Beltz, Psychologie-Verl.-Union] 1996
- DE LEEUW, J.; E. MEYER: *Handbook of multilevel analysis*. New York [Springer] 2007
- DIENER, E.; E. M. SUH; R. E. LUCAS; H. L. SMITH: Subjective well-being: Three decades of progress. In: *Psychological Bulletin*, 125(2), 1999, pp. 276-302
- DIENLIN, T.; P. K. MASUR; S. TREPTE: Displacement or reinforcement? The reciprocity of FtF, IM, and SNS communication and their effects on loneliness and life satisfaction. In: *Journal of Computer-Mediated Communication*, 22(2), 2017, pp. 71-87
- ELLISON, N. B.; C. STEINFELD; C. LAMPE: The benefits of Facebook ›friends‹: Social capital and college students' use of online social network sites. In: *Journal of Computer-Mediated Communication*, 12(4), 2007, pp. 1143-1168
- ENDERS, C. K.; D. TOFIGHI: Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. In: *Psychological Methods*, 12(2), 2007, pp. 121-138
- GELMAN, A.; J. HILL: *Data analysis using regression and multilevel/hierarchical models: Analytical methods for social research* (11th ed.). Cambridge [Cambridge University Press] 2009
- GÖRLAND, S.: Surveys go mobile: Experience Sampling Method als methodischer Ansatz für mobile Rezeptionssituationen [Surveys go mobile: Experience sampling method as methodological approach to mobile reception situations]. In: HOOFFACKER, G.; C. WOLF (Eds.): *Technische Innovationen – Medieninnovationen?* Wiesbaden [Springer] 2017, pp. 247-258
- GORIN, A. A.; A. A. STONE: Recall biases and cognitive errors in retrospective self-reports a call for momentary assessments. In: BAUM, A.; T. REVENSON; J. SINGER (Eds.): *Handbook of Health Psychology*. Mahwah/NJ [Lawrence Erlbaum Associates] 2002, pp. 405-413

- GREASLEY, A. E.; A. LAMONT: Exploring engagement with music in everyday life using experience sampling methodology. In: *Musicae Scientiae*, 15(1), 2011, pp. 45-71
- HASEBRINK, U.; A. HEPP: How to research cross-media practices? Investigating media repertoires and media ensembles. In: *Convergence*, 23(4), 2017, pp. 362-377
- HASEBRINK, U.; J. SCHMIDT: Medienübergreifende Informationsrepertoires. In: *Media Perspektiven*, 1, 2013, pp. 2-12
- HEKTNER, J. M.; J. A. SCHMIDT; M. CSIKSZENTMIHALYI: *Experience sampling method: Measuring the quality of everyday life*. Thousand Oaks/CA [Sage Publications] 2007
- HOFMANN, W.; P. V. PATEL: SurveySignal: A convenient solution for experience sampling research using participants' own smartphones. In: *Social Science Computer Review*, 33(2), 2015, pp. 235-253
- HOX, J. J.: *Multilevel analysis: Techniques and applications* (2nd ed.). New York [Routledge] 2010
- JOHNSON, P. C. D.: Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. In: *Methods in Ecology and Evolution*, 5(9), 2014, pp. 944-946. <https://doi.org/10.1111/2041-210X.12225>
- KARNOWSKI, V.: Befragung in situ: Die Mobile Experience Sampling Method (MESM) [Survey in situ: The mobile experience sampling method (MESM)]. In: MÖHRING, W.; D. SCHLÜTZ (Eds.): *Handbuch standardisierte Erhebungsverfahren in der Kommunikationswissenschaft*. Wiesbaden [Springer] 2013, pp. 235-247
- KARNOWSKI, V.; S. DOEDENS: Mobile Experience Sampling: Eine Methode zur Untersuchung mobilen Mediennutzungsverhaltens [Mobile experience sampling: A method to investigate mobile media use]. In: JACKOB, N.; T. ZERBACK; O. JANDURA; M. MAURER (Eds.): *Das Internet als Forschungsinstrument und -gegenstand in der Kommunikationswissenschaft* (6th ed.). Köln [Herbert von Halem] 2010, pp. 235-247
- KARNOWSKI, V.; T. VON PAPE: Mobile TV im Alltag der Nutzer: Ergebnisse einer dreimonatigen Panelstudie [Mobile TV in the everyday life of the user: Results from a three-month panel study]. In: KRONE, J. (Ed.): *Fernsehen im Wandel*. Baden-Baden [Nomos] 2009, pp. 241-255
- KRAUT, R.; S. KIESLER; B. BONEVA; J. CUMMINGS; V. HELGESON; A. CRAWFORD: Internet paradox revisited. In: *Journal of Social Issues*, 58(1), 2002, pp. 49-74

- KRAUT, R.; M. PATTERSON; V. LUNDMARK; S. KIESLER; T. MUKOPADHYAY;  
W. SCHERLIS: Internet paradox: A social technology that reduces  
social involvement and psychological well-being? In: *American  
Psychologist*, 53(9), 1998, pp. 1017-1031
- KUBEY, R.; R. LARSON: The use and experience of the new video media  
among children and young adolescents. In: *Communication Research*,  
17(1), 1990, pp. 107-130
- KUBEY, R.; R. LARSON; M. CSIKSZENTMIHALYI: Experience sampling  
method applications to communication research questions.  
In: *Journal of Communication*, 46(2), 1996, pp. 99-120
- LARSON, R.; M. CSIKSZENTMIHALYI: The experience sampling method.  
In: CSIKSZENTMIHALYI, M. (Ed.): *Flow and the foundations of positive  
psychology: The collected works of Mihaly Csikszentmihalyi*. Dordrecht  
[Springer] 2014, pp. 21-34
- LEUNG, L.; R. WIE: More than just talk on the move: Uses and gratifications  
of the cellular phone. In: *J&MC Quarterly*, 77(2), 2000, pp. 308-320
- LEWIN, K.: *Field theory in social science: Selected theoretical papers*. New York  
[Harper & Brothers] 1951
- MAGNUSSON, D.: Problems in environmental analyses – An introduction.  
In: MAGNUSSON, D. (Ed.): *Toward a psychology of situations:  
An international perspective*. Hillsdale/NJ [Erlbaum] 1981, pp. 3-7
- MASUR, P. K.: *Situational privacy and self-disclosure: Communication processes  
in online environments*. Cham/Switzerland [Springer International  
Publishing] 2018
- MOORE, K.; J. C. MCELROY: The influence of personality on Facebook  
usage, wall postings, and regret. In: *Computers in Human Behavior*, 28,  
2012, pp. 267-274
- NISBETT, R. E.; L. ROSS: *Human inference: Strategies and shortcomings of social  
judgment*. Englewood Cliffs/NJ [Prentice-Hall, Inc.] 1980
- RANDALL, W. M.; N. S. RICKARD: Development and trial of a mobile  
experience sampling method (m-ESM) for personal music listening.  
In: *Music Perception: An Interdisciplinary Journal*, 31(2), 2013, pp. 157-170
- RAUTHMANN, J. F.; R. A. SHERMAN; D. C. FUNDER: Principles of situation  
research: Towards a better understanding of psychological situations.  
In: *European Journal of Personality*, 29(3), 2015a, pp. 363-381
- RAUTHMANN, J. F.; R. A. SHERMAN; D. C. FUNDER: New Horizons in  
Research on Psychological Situations and Environments. In: *European  
Journal of Personality*, 29(3), 2015b, pp. 382-432

- REINECKE, L.; W. HOFMANN: Slacking off or winding down? An experience sampling study on the drivers and consequences of media use for recovery versus procrastination. In: *Human Communication Research*, 43(3), 2010, pp. 441-461
- ROSS, L.: The intuitive psychologist and its shortcomings. In: BERKOWITZ, L. (Ed.): *Advances in experimental social psychology*. New York [Academic Press] 1977, pp. 173-220
- ROSS, L.; R. E. NISBETT: *The person and the situation: Perspectives of social psychology*. London [Pinter & Martin] 2011
- RUBIN, A. M.: Television uses and gratifications: The interactions of viewing patterns and motivations. In: *Journal of Broadcasting*, 27(1), 1983, pp. 37-47
- SAUCIER, G.; T. BEL-BAHAR; C. FERNANDEZ: What modifies the expression of personality tendencies? Defining basic domains of situation variables. In: *Journal of Personality*, 75(3), 2007, pp. 479-503
- SCHAFFER, J. L.; J. W. GRAHAM: Missing data: Our view of the state of the art. In: *Psychological Methods*, 7(2), 2002, pp. 147-177
- SCHENK, M.; J. NIEMANN; G. REINMANN; A. ROSSNAGEL: *Digitale Privatsphäre: Heranwachsende und Datenschutz auf sozialen Netzwerkplattformen* [Digital privacy: Adolescents and data protection on social network sites]. Berlin [Vistas Verlag] 2012
- SCHERER, H.; D. SCHLÜTZ: Gratifikation à la minute: Die zeitnahe Erfassung von Gratifikationen [Gratification à la minute: The realtime measurement of gratifications]. In: RÖSSLER, P.; S. KUBISCH; V. GEHRAU (Eds.): *Empirische Perspektiven der Rezeptionsforschung* (Vol. 23). München [Fischer] 2002, pp. 133-152
- SCHLÜTZ, D.: *Bildschirmspiele und ihre Faszination: Zuwendungsmotive, Gratifikationen und Erleben interaktiver Medienangebote* [Video games and their fascination: Motives, gratifications, and experiencing interactive media content]. München [Fischer] 2002
- SCHWARTZ, N.: Retrospective and concurrent self-reports: A rationale for real-time data capture. In: STONE, A. A.; S. SHIFFMAN; A. A. ATIENZA; L. NEBELING (Eds.): *The science of real-time data capture*. Oxford [Oxford University Press] 2007, pp. 11-16
- SCOLLON, C. V. N.; C. KIM-PRIETO; E. DIENER: Experience sampling: Promises and pitfalls, strengths and weaknesses. In: *Journal of Happiness Studies*, 4(1), 2003, pp. 5-34

- SEIDMAN, G.: Self-presentation and belonging on Facebook:  
How personality influences social media use and motivations.  
In: *Personality and Individual Differences*, 54, 2013, pp. 402-407
- SHIFFMAN, S.: Designing protocols for ecological momentary assessment.  
In: STONE, A. A.; S. SHIFFMAN; A. A. ATIENZA; L. NEBELING (Eds.): *The science of real-time data capture*. Oxford [Oxford University Press] 2007, pp. 27-53
- SLOBODA, J. A.; S. A. O'NEILL; A. IVALDI: Functions of music in everyday life: An exploratory study using the experience sampling method.  
In: *Musicae Scientiae*, 5(1), 2001, pp. 9-32
- SMOCK, A. D.; N. B. ELLISON; C. LAMPE; D. Y. WOHN: Facebook as a toolkit: A uses and gratification approach to unbundling feature use. In: *Computers in Human Behavior*, 27(6), 2011, pp. 2322-2329. <https://doi.org/10.1016/j.chb.2011.07.011>
- SNIJDERS, T. A. B.; R. J. BOSKER: *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Los Angeles/CA [Sage Publications] 2011
- STONE, A. A.; R. C. KESSLER; J. A. HAYTHOMTHWATTE: Measuring daily events and experiences: Decisions for the researcher. In: *Journal of Personality*, 59(3), 1991, pp. 575-607
- STONE, A. A.; S. SHIFFMAN: Appendix: Capturing momentary, self-report data: A proposal for reporting guidelines. In: STONE, A. A.; S. SHIFFMAN; A. A. ATIENZA; L. NEBELING (Eds.): *The science of real-time data capture*. Oxford [Oxford University Press] 2007, pp. 371-386
- STONE, A. A.; S. SHIFFMAN; A. A. ATIENZA; L. NEBELING: Historical roots and rationale of ecological momentary assessments (EMA). In: STONE, A. A.; S. SHIFFMAN; A. A. ATIENZA; L. NEBELING (Eds.): *The science of real-time data capture*. Oxford [Oxford University Press] 2007, pp. 3-10
- TOURANGEAU, R.; L. J. RIPS; K. RASINSKI: *The psychology of survey response*. Cambridge [Cambridge University Press] 2000
- VALENZUELA, S.; N. PARK; K. F. KEE: Is there social capital in a social networking site? Facebook use and college students' life satisfaction, trust, and participation. In: *Journal of Computer-Mediated Communication*, 14, 2009, pp. 875-901
- VANDEWATER, E. A.; S. LEE: Measuring children's media use in the digital age: Issues and challenges. In: *The American Behavioral Scientist*, 52(8), 2009, pp. 1152-1176



ANNA SOPHIE KÜMPEL

## Dynamik im Blick: Die qualitative Beobachtung mit Post-Exposure-Walkthrough als Verfahren für die Rekonstruktion individueller Navigations- und Selektionshandlungen auf sozialen Netzwerkseiten (SNS)

*Abstract:* Die Untersuchung der Nutzung hochgradig personalisierter und sich dynamisch wandelnder Online-Anwendungen – allen voran soziale Netzwerkseiten (SNS) – stellt die Kommunikationswissenschaft vor methodische Herausforderungen. Während quantitative Analysen digitaler Verhaltens- und Spurendaten Aufschluss über generelle Nutzungstendenzen geben, liefern sie in aller Regel keine Erkenntnisse über die subjektiven Beweggründe, die SNS-Nutzer dazu veranlassen, sich einzelnen (Medien-) Angeboten zuzuwenden, diese auszuwählen, zu bewerten oder weiterzuleiten. Blind sind solche Analysen insbesondere für die Frage, warum die überwiegende Zahl an Angeboten keinerlei Auseinandersetzung provoziert. Qualitativ-rekonstruierende Verfahren bieten indes die Möglichkeit, personalisierte Informationsumgebungen und damit verknüpftes Navigations- und Selektionshandeln unmittelbar zu erfassen sowie die Motive für (nicht) ausgeführte Handlungen zu ergründen. Der Beitrag diskutiert ein solches Verfahren – die qualitative Beobachtung mit Post-Exposure-Walkthrough – und erläutert an einer Studie zur Nachrichtennutzung auf Facebook beispielhaft den Ablauf von Datenerhebung und -auswertung. Darüber hinaus werden die grundsätzlichen Einsatzmöglichkeiten, die Abgrenzung zu vergleichbaren methodischen Ansätzen sowie die Vor- und Nachteile des Verfahrens erörtert. Der Beitrag endet mit einem Plädoyer für den Einsatz von Mehrmethodendesigns, die insbesondere im Kontext

von Online- und SNS-Rezeptionsforschung einen ganzheitlichen Blick auf Nutzungspraktiken erlauben.

*Schlüsselwörter:* Qualitative Methoden, Beobachtung, Post-Exposure-Walkthrough, soziale Netzwerkseiten (SNS), Facebook, Nachrichtennutzung

## 1. Einleitung

Soziale Netzwerkseiten (SNS), insbesondere Facebook, sind zu einer zentralen Schnittstelle für die Auseinandersetzung mit Nachrichteninhalten avanciert (NEWMAN et al. 2017; VAN EIMEREN/KOCH 2016) und daher in letzter Zeit (noch) stärker in das Blickfeld kommunikationswissenschaftlicher Forschung gerückt. Der SNS-Präsentationslogik entsprechend liegen Nachrichten auf Twitter oder Facebook jedoch nicht als journalistisch gebündeltes »Komplettpaket« vor, sondern sind in Form *einzelner* Posts Teil eines »ständig aktualisierten und personalisierten Strom[s] von Neuigkeiten« (SCHMIDT 2017: 53) und mit vielfältigen sozialen Kontextinformationen verknüpft. Bedingt durch im- und explizite Formen der Personalisierung ist jeder SNS-Nutzer dabei mit einem einzigartigen und hochgradig dynamischen Angebot an sozialen, werblichen und nachrichtlichen Informationen konfrontiert.

Will man vor diesem Hintergrund die Auswahl von (Nachrichten-) Inhalten sowie den weiteren Umgang mit diesen rekonstruieren, benötigt man einen methodischen Zugang, der einerseits in der Lage ist, individuelle Informationsumgebungen und damit zusammenhängende Nutzungspraktiken zu erfassen, andererseits den Nutzern Raum gibt, die Entscheidungen, die bestimmten Navigations- bzw. Selektionshandlungen zugrunde liegen, zu verbalisieren und einzuordnen. Eine solche Möglichkeit wird durch methodische Zugänge eröffnet, die Messungen des Navigations- und Selektionshandelns mit der Erhebung subjektiver Bedeutungszuweisungen verknüpfen. In diesem Beitrag wird ein solches Verfahren – *die qualitative Beobachtung mit Post-Exposure-Walkthrough* – anhand einer konkreten kommunikationswissenschaftlichen Fragestellung illustriert. Die für den Beitrag als Fallbeispiel dienende Studie verfolgte angesichts der eingangs beschriebenen Ausgangssituation das Ziel, zu untersuchen, wie verschiedene theoretisch identifizierte Einflussfaktoren den nutzerseitigen Umgang mit auf SNS entdeckten Nachrichteninhalten beeinflussen. Konkret ging es dabei um die Frage, unter welchen Umstän-

den die beiläufige Konfrontation mit Nachrichten auf Facebook in tiefergehende Auseinandersetzung mündet.

Neben einer generellen Charakterisierung des Verfahrens sowie der beispielhaften Illustration von Erhebungssituation und Auswertungsstrategie wird im Rahmen des Beitrags eine (kritische) Einordnung des methodischen Zugangs vorgenommen, indem einerseits eine Abgrenzung zu vergleichbaren Ansätzen erfolgt, andererseits die Vor- und Nachteile des Vorgehens diskutiert werden. Der Beitrag endet mit einem Plädoyer für den Einsatz qualitativ-rekonstruierender Verfahren in der SNS-Rezeptionsforschung, die insbesondere als Ergänzung zu digitalen Verhaltens- und Spurendaten zu tiefschärferen Befunden beitragen können.

## 2. Problemstellung: Die Rekonstruktion von Navigations- und Selektionshandlungen bei der SNS-Nachrichtenrezeption

Die Untersuchung der SNS-(Nachrichten-)Nutzung stellt die Kommunikationswissenschaft angesichts der Personalisierung sowie der stetigen und dynamischen Anpassung der SNS-Informationsumgebung vor methodische Herausforderungen. Während quantitative Analysen digitaler Verhaltens- und Spurendaten Aufschluss über generelle Nutzungstendenzen geben können, liefern sie in aller Regel keine Erkenntnisse über die subjektiven Beweggründe, die SNS-Nutzer dazu veranlassen, sich einzelnen (Nachrichten-)Inhalten zuzuwenden, diese auszuwählen, zu bewerten oder weiterzuleiten. Blind sind solche Analysen insbesondere für die Frage, warum die überwiegende Zahl an Angeboten *keinerlei* Auseinandersetzung provoziert. So fokussiert die Mehrheit quantitativer Analysen auf Erfolgsfaktoren und die Frage, welche Inhalte viel geklickt, gelikt oder geteilt werden (siehe z. B. GARCÍA-PERDOMO et al. 2018; TRILLING/TOLOCHKO/BURSCHER 2017). Missachtet wird bei solchen Untersuchungen jedoch nicht nur die inhärent subjektive Qualität inhaltlicher Wahrnehmungen, sondern auch, inwiefern SNS-spezifische Kontextfaktoren (Wurde der Beitrag von einem Freund empfohlen? Wie ist er überhaupt in den Newsfeed des Nutzers gelangt? etc.) die Auseinandersetzung mit Nachrichten und anderen medialen Inhalten beeinflussen können.

Mit Blick auf diese Forschungslücken sollte im Rahmen der hier vorgestellten Studie untersucht werden, wie verschiedene theoretisch identifizierte Einflussfaktoren die nutzerseitige Auswahl von beiläufig auf Facebook ent-

deckten Nachrichteninhalten sowie den weiteren Umgang mit diesen beeinflussen. Auf Basis des dynamisch-transaktionalen Ansatzes (DTA, siehe FRÜH/SCHÖNBACH 1982, 2005; SCHÖNBACH/FRÜH 1984) wurde dafür zunächst ein theoretisches Modell entwickelt, das die zentralen Rahmenbedingungen und Einflussfaktoren im Prozess der SNS-Nachrichtennutzung systematisiert und die Auswahl von Nachrichteninhalten bei Facebook (bzw. den Umgang mit diesen) als Resultat von imaginären oder realen Interaktionsprozessen (*Inter-Transaktionen*) sowie innerhalb des Nutzers zu verortenden *Intra-Transaktionen* konzeptualisiert. Das auf der Mikroebene zu verortende Modell fokussiert auf den auf einen Nachrichtinhalt aufmerksam werdenden Nutzer (bezeichnet als *Nachrichtenempfänger*) und ermöglicht durch die Berücksichtigung 1) der (wahrgenommenen) Merkmale des *Nachrichtenanbieters* (*Von welchem Anbieter wurde der Beitrag ursprünglich gepostet?*), 2) der Inhalte des Nachrichtenposts-/links (*Worum geht es in dem verlinkten Beitrag? Wie sind die Inhalte aufbereitet?*), 3) des Nachrichtenvermittlers (*Welcher Facebook-Kontakt hat den Beitrag geteilt/empfohlen/steht mit diesem in Verbindung?*), 4) der Art des Nachrichtenerfahrens (*Wie genau ist der Nachrichtenempfänger auf den Post/Beitrag aufmerksam geworden?*) sowie 5) weiterführender Eigenschaften des Nachrichtenempfängers einen ganzheitlichen Blick auf die Faktoren, die die während der SNS-Nachrichtennutzung ablaufenden Wirkungsprozesse beeinflussen können.

Um die identifizierten Einflussfaktoren in einer möglichst natürlichen Nutzungssituation sowie in ihrem Zusammen- und Wechselspiel untersuchen zu können, schienen quantitative Ansätze – nicht zuletzt aufgrund Facebook-seitiger Zugangsrestriktionen – für die vorliegende Studie ungeeignet. Qualitativ-rekonstruierende Verfahren bieten indes die Möglichkeit, personalisierte Informationsumgebungen und damit verknüpft Navigations- und Selektionshandeln unmittelbar zu erfassen sowie die Motive für (nicht) ausgeführte Handlungen umfassend zu ergründen.

### 3. Methodischer Zugang: Qualitative Beobachtung mit Post-Exposure-Walkthrough

#### 3.1 Charakterisierung des Verfahrens

Angesichts des Forschungsinteresses wurde mit einem Verfahren gearbeitet, das hier als qualitative Beobachtung mit Post-Exposure-Walkthrough bezeichnet wird. Dabei handelt es sich um einen methodischen Zugang, bei

dem eine offene, direkte, künstliche und passiv-teilnehmende qualitative Beobachtung (siehe MEYEN et al. 2011: 124) mit Verfahren der retrospektiven Verbalisierung individueller Handlungen verknüpft wird. Im vorliegenden Fall liegt der Fokus auf der Facebook-Nutzung der Teilnehmer, die beobachtet sowie mittels Screen-Capture-Software aufgezeichnet wird. Die Verbalisierungen werden schließlich entlang dieser Aufzeichnung erhoben. Individuelle Handlungen werden also gemeinsam mit dem Teilnehmer *nach* der eigentlichen Ausführung noch einmal schrittweise und in voller Länge nachvollzogen (= Post-Exposure-Walkthrough, Begriff von SCHWEIGER 2010: 195). Der Forscher hat dabei die Gelegenheit, die subjektiven Beweggründe für bestimmte (nicht) ausgeführte Handlungen zu erfragen und die gemachten Beobachtungen auf Basis der Einordnungen des Teilnehmers zu kontextualisieren.

Zunächst lässt sich das Verfahren über den Teilaspekt der *Beobachtung* charakterisieren. Dabei handelt es sich im (kommunikations-)wissenschaftlichen Sinne um »die selektive und systematische Erfassung und Protokollierung von sinnlich wahrnehmbaren Aspekten prinzipiell sichtbaren menschlichen Verhaltens« (BROSIOUS/HAAS/KOSCHEL 2016: 185; siehe auch GEHRAU 2002: 26). Auch wenn die spezifischen Vor- und Nachteile der Beobachtung stark mit deren konkreter Ausrichtung in Zusammenhang stehen, lassen sich einige prinzipielle Stärken und Schwächen der Methode benennen (BROSIOUS/HAAS/KOSCHEL 2016: 185ff.; DÖRING/BORTZ 2016: 323ff.; MEYEN et al. 2011: 121ff.; QUANDT 2011: 294; VICARI 2016: 300): So bieten Beobachtungen einerseits einen (relativ) direkten und unmittelbaren Blick auf den Untersuchungsgegenstand und können bei dem Fokus auf Mediennutzer auch solche Handlungen sichtbar machen, die in Befragungen unerwähnt bleiben, weil sie den Teilnehmern belanglos vorkommen, habitualisierten und automatisierten Mustern unterliegen oder ob ihrer Komplexität retrospektiv nur schwer verbalisiert werden können. Schwächen zeigen sich hingegen darin, dass Beobachtungen in der Regel sehr aufwendig sind und darum mit deutlich geringeren Fallzahlen arbeiten (müssen) als Befragungen. Darüber hinaus stellt sich das Problem – insbesondere, wenn es um die Erfassung subjektiver Bedeutungszuweisungen geht –, dass Beobachtungen allein nichts über den »praktischen Sinn [sagen], den Menschen mit Handlungen und Strukturen verbinden« (MEYEN et al. 2011: 122). In der Praxis der rezeptionsorientierten Forschung wird daher, wie auch hier, die Beobachtung in der Regel mit Befragungen verknüpft, die einerseits durch gezielte Nachfragen dabei helfen,

die Beobachtung zu kontextualisieren (im vorliegenden Fall durch den Post-Exposure-Walkthrough), andererseits genutzt werden, um generelle Nutzungsmuster oder soziodemografische Merkmale zu erfragen. Meyen und Kollegen (2011: 123f.) regen an, qualitative Beobachtungen entlang der Kriterien ›Transparenz‹, ›Anwesenheit‹, ›Eingriff‹ und ›Teilnahme‹ zu charakterisieren. Wie eingangs erwähnt, lässt sich die hier eingesetzte Variante in diesem Sinne als *offen* (Teilnehmer wissen um die Beobachtung), *direkt* (Forscher ist bei der Beobachtung dabei), *künstlich* (Teilnehmern wird im Rahmen der Beobachtung eine Aufgabe gestellt) und *passiv-teilnehmend* (Forscher als Forscher statt Teil des ›normalen‹ Settings) beschreiben. In Anlehnung an Döring und Bortz (2016: 334) ließe sich zudem von einer qualitativen Beobachtung »mit geringem Komplexitätsgrad« sprechen, da – im Gegensatz etwa zur ethnografischen Forschung – von vornherein nur bestimmte interessierende Handlungen und Verhaltensweisen in den Blick genommen werden (siehe Kapitel 3.2).

Der hier als *Post-Exposure-Walkthrough* beschriebene zweite Teilaspekt des Verfahrens rekuriert auf die Idee, die beobachteten (und aufgezeichneten) Handlungen der Teilnehmer unmittelbar *nach* der eigentlichen Ausführung (→ Post-Exposure) erneut schrittweise (→ Walkthrough) nachzuvollziehen und dabei die subjektiven Beweggründe für bestimmte (nicht) ausgeführte Handlungen von den Teilnehmern verbalisieren zu lassen. Dieses Vorgehen erinnert an Ansätze wie das ›nachträgliches Lautes Denken‹ (BILANDZIC 2012; BILANDZIC/TRAPP 2000) oder das vor allem in der Psychologie eingesetzte ›Selbstkonfrontationsinterview‹ (BREUER 1995; *self-confrontation interview* bei LIM 2002), unterscheidet sich von diesen aber mit Blick auf 1) die Direktivität der Anweisungen sowie 2) die Tiefe der interpretativen Einordnung der gewonnenen Daten. So werden die Teilnehmer im Gegensatz zur Methode des nachträgliches Lauten Denkens nicht angewiesen, alles zu verbalisieren, was ihnen ›durch den Kopf geht‹, sondern spezifisch um die Kontextualisierung bestimmter Aktivitäten gebeten. Auch wenn durch diese Lenkung die Gefahr besteht, dass die tatsächlichen Denkprozesse von den Teilnehmern stärker rationalisiert werden (BILANDZIC/TRAPP 2000: 190), schien es für die vorliegende Untersuchung sinnvoll, mittels konkreter Erzählaufforderungen (*reflection prompts*, KONRAD 2010: 479) die Ausführungen vorrangig auf interessierende Teilbereiche zu fokussieren.

Vom Selbstkonfrontationsinterview unterscheidet sich das hier eingesetzte Verfahren vor allem hinsichtlich des Umgangs mit den gewonnenen Daten. Angesichts des häufigen Einsatzes in der Psychotherapie werden bei

diesem methodischen Zugang vorrangig die Divergenzen zwischen aufgezeichneter Handlung (beschrieben als Handlung 1) und Aussagen der Teilnehmer über die Handlungen (beschrieben als Handlung 2) betrachtet und etwa »Sprünge« [...] des Explikations-Umfangs« (BREUER 1995: 176f.) als Hinweise für stattfindende innerpsychische Prozesse verstanden. Eine solche Differenzierung und Interpretationstiefe scheint angesichts des Forschungsinteresses der vorliegenden Untersuchung – und für kommunikationswissenschaftliche Fragestellungen generell – allerdings wenig zweckmäßig.

Bei einer ganzheitlichen Betrachtung der qualitativen Beobachtung mit Post-Exposure-Walkthrough offenbaren sich somit spezifische Vor- und Nachteile. Stärken liegen vor allem in der großen Prozessbezogenheit, die das Erfassen dynamischer und individueller Navigations- und Selektionsverläufe im Sinne des DTA ermöglicht. Im Gegensatz zu klassischen Befragungen, die im Allgemeinen zeitlich und affektiv von der interessierenden Rezeptionssituation abgekoppelt sind, können nicht nur allgemeine Navigations- und Selektionsstrategien, sondern auch situative Inter-Transaktionen zwischen Nutzer und (Nachrichten-)Inhalt eruiert werden. Da die Teilnehmer im vorliegenden Fall mit ihrem *persönlichen* Facebook-Newsfeed konfrontiert werden, lassen sich etwa Wechselwirkungen zwischen der inhaltlichen Wahrnehmung eines entdeckten Beitrags (Ist das Thema spannend/relevant?) und der Einschätzung des empfehlenden Freundes (Ist dieser ein Experte? Wie eng ist die Freundschaftsbeziehung?) beleuchten: So mag auch ein eigentlich interessant anmutender Artikel nicht angeklickt werden, da der mit diesem assoziierte Facebook-Kontakt negativ evaluiert wird und sich diese Einschätzung auf die inhaltliche Wahrnehmung überträgt. Durch die Entscheidung, die Kognitionen der Teilnehmer erst nach den Handlungen entlang der Aufzeichnung von diesen verbalisieren und kontextualisieren zu lassen, wird einerseits Reaktivität vermieden, andererseits eröffnet sich die Möglichkeit, die Aufnahme anzuhalten, zurückzuspulen oder zu verlangsamen, um auch subtilere Handlungen oder wiederkehrende Selektionsroutinen mit den Teilnehmern zu erörtern. Durch die unaufdringliche Aufzeichnung des Bildschirms mittels Screen-Capture-Software (siehe Kapitel 3.2) und den Verzicht auf Videoaufnahmen der Teilnehmer wird zudem Befangenheit reduziert und die Nutzungssituation nicht übermäßig artifiziell. Als Schwächen lassen sich aufgrund der Anwesenheit des Forschers während der Nutzungsepisode zum einen reguläre Interviewer-Effekte benennen (siehe dazu BROSIUS/HAAS/KOSCHEL 2016: 127f.), andererseits der verhält-

nismäßig hohe Aufwand sowohl für die Erhebung selbst als auch für die Auswertung des umfangreichen Datenmaterials. Dieser Aufwand lohnt sich jedoch insbesondere dann, wenn der Schwerpunkt der Untersuchung auf dynamischen Nutzungsverläufen sowie subjektiven Bedeutungszuweisungen und Interpretationen liegt.

### 3.2 Ablauf der Erhebung

#### *Konstruktion des Beobachtungsprotokolls*

Der eigentlichen Erhebung ging die Erstellung eines Beobachtungsprotokolls voraus, in dem neben Beobachtungsfeld, -objekt und -fall (GEHRAU 2002: 65ff.) auch Situationen definiert wurden, die stets mit einer Protokollierung einhergehen mussten (s. unten). Als *Beobachtungsfeld* – für dessen Festlegung eine räumliche und zeitliche Einschränkung vorgenommen werden muss – wurde die persönliche Facebook-Seite der Teilnehmer definiert (>Raum<). Die Handlungen der Teilnehmer in diesem Raum wiederum wurden für eine Dauer von fünf Minuten (>Zeit<) beobachtet. Als *Beobachtungsobjekt* wurde folglich der jeweilige Teilnehmer definiert, während die Handlungs- und Reaktionsmuster der Teilnehmer als *Beobachtungsfälle* beschrieben werden können.

Im Sinne der von Gehrau (2002: 37) definierten Arten der Protokollierung lässt sich die hier eingesetzte als weitgehend unstrukturiert beschreiben, was immer dann von Vorteil ist, »wenn über den Forschungsgegenstand wenig bekannt ist, wenn man eher explorativ arbeitet oder den Untersuchungsgegenstand möglichst ganzheitlich erfassen möchte« (ebd.: 38). Trotz der gebotenen Offenheit wurden im Vorfeld sieben Handlungs- bzw. Reaktionsmuster definiert, die bei Vorkommen im Post-Exposure-Walkthrough grundsätzlich thematisiert werden sollten (s. Tab. 1).

Das ausgefüllte Beobachtungsprotokoll wurde schließlich als eine Art Leitfaden<sup>1</sup> für die Gesprächsstrukturierung im Post-Exposure-Walkthrough

1 Daneben wurde im Rahmen der Studie noch ein im Vorfeld konzipierter Leitfaden eingesetzt, der genutzt wurde, um im Anschluss an den Post-Exposure-Walkthrough generelle Tendenzen der Nachrichten- und Facebook-Nutzung der Teilnehmer zu erfragen. Angesichts des Fokus des Beitrags wird an dieser Stelle jedoch auf eine Darstellung der Konstruktion dieses Leitfadens verzichtet.



genutzt. Wurde beispielsweise notiert, dass ein Teilnehmer lange bei einem Post der *Tagesschau* verweilte, dem Link zum Artikel aber (dennoch) nicht gefolgt ist, konnte dies thematisiert werden. Zumeist genügte in solchen Situationen eine unspezifische Feststellung des Beobachteten (z. B. »Bei dem Post bist du kurz hängengeblieben«), um die Teilnehmer zu einer Kontextualisierung zu motivieren. In Abhängigkeit von der Ausführlichkeit der Aussagen und dem Forschungsinteresse konnten dann konkrete Nachfragen angeschlossen werden – etwa, wenn ein Post mit einem Nachrichtenvermittler in Verbindung stand (z. B. »Welche Rolle hat es gespielt, dass [Name] den Beitrag gelikt hat?«) oder der Teilnehmer auf eine spezifische Art und Weise auf einen Post aufmerksam geworden ist (z. B. »Unter dem Post wurdest du markiert. Was hat das für Gedanken bei dir ausgelöst?«).

TABELLE 1

### Definierte Handlungs- und Reaktionsmuster im Beobachtungsprotokoll

Handlungs-/ Reaktionsmuster	Beschreibung
Verweildauer	Nutzer verweilt auffällig lange bei einem (Nachrichten-)Post
(Non-)Verbale Reaktion	Nutzer zeigt auffällige verbale oder nonverbale Reaktion auf einen (Nachrichten-)Post
Selektion	Nutzer folgt dem in einem (Nachrichten-)Post hinterlegten Link
Passive Auseinandersetzung	Nutzer setzt sich passiv mit (Nachrichten-)Post auseinander (z. B. Klick auf Kommentare, um diese zu rezipieren)
Aktive Auseinandersetzung	Nutzer setzt sich aktiv mit (Nachrichten-)Post auseinander (z. B. Liken, Kommentieren, Sharen, Facebook-Kontakte markieren)
Rückkehr	Nutzer kehrt zu einem bestimmten (Nachrichten-)Post zurück, den er zuvor bereits länger betrachtet und/oder selektiert hat
Aktivitäten außerhalb des Neuigkeiten-Bereichs	Nutzer verlässt den Neuigkeitenbereich und wendet sich spezifischen Seiten/Profilen oder Sektionen (z. B. private Nachrichten, Veranstaltungen) zu

### *Auswahl der Teilnehmer*

Obschon Repräsentativität »für qualitative Stichproben kein sinnvolles Kriterium [ist]« (HELFFERICH 2011: 172), sind auch die Befunde qualitativer Forschung in hohem Maße von der Auswahl der Teilnehmer abhängig, weswegen die für eine (Beobachtungs-)Studie gewählte Samplingstrategie nicht nur in Abhängigkeit des Forschungsinteresses durchdacht, sondern auch angemessen dokumentiert werden muss (MEYEN et al. 2011: 67f.; dazu auch MISOCH 2015: 185f.). Angesichts des besonderen Interesses am Umgang mit Nachrichteninhalten sowie des Fokus auf Facebook wurde die Auswahl der Teilnehmer im vorliegenden Fall durch folgende Kriterien bestimmt:

- *Alter*: Die tatsächliche Altersverteilung unter deutschen SNS-Nutzern zum Rekrutierungszeitpunkt reflektierend (KOCH/FREES 2016: 428), wurde bei der Auswahl der Teilnehmer ein besonderer Fokus auf die Kernnutzergruppe der 14- bis 29-Jährigen gelegt.
- *Geschlecht*: Ebenfalls an der Verteilung unter deutschen SNS-Nutzern zum Rekrutierungszeitpunkt orientiert (KOCH/FREES 2016: 435), wurde mit Blick auf das Geschlecht der Teilnehmer darauf geachtet, dass für die Studie mehr Frauen als Männer rekrutiert werden.
- *Bildung*: Angesichts des Forschungsinteresses schien es sinnvoll, ein Auswahlkriterium zu wählen, das mutmaßlich mit der Wahrscheinlichkeit in Zusammenhang steht, auf Facebook (auch) mit nachrichtlichen Inhalten konfrontiert zu werden. In Anlehnung an das vor allem aus der quantitativen Forschung bekannte Konzentrationsprinzip (BROSIUS/HAAS/KOSCHEL 2016: 74) – ein bewusstes Auswahlverfahren, bei dem sich das Sampling auf jenen Teil der Grundgesamtheit konzentriert, in dem der überwiegende Teil der interessierenden Elemente vermutet wird – wurde als Bedingung für die Studienteilnahme die allgemeine Hochschulreife (bzw. bei jüngeren Teilnehmern die laufende gymnasiale Ausbildung) definiert. Unter den 14- bis 49-Jährigen nutzt die Gruppe der Abiturienten das Internet häufiger für Nachrichten als jene mit niedrigeren Bildungsabschlüssen (VAN EIMEREN/KOCH 2016: 280), weswegen erwartet werden konnte, dass es für Facebook-Nutzer mit Abitur wahrscheinlicher ist, auch dort mit Nachrichten konfrontiert zu werden: Einerseits, weil das Ausmaß persönlicher Kuratation (z. B. durch Abonnements der Seiten von Nachrichten Anbietern) größer sein dürfe. Andererseits, weil Menschen eher mit Personen befreundet sind, die ähnliche soziodemografische und Persönlichkeitsmerkmale haben

wie sie selbst und somit auch bei diesen eine erhöhte Nachrichtennutzung vermutet werden kann (LÖNNQVIST/ITKONEN 2016; MCPHERSON/SMITH-LOVIN/COOK 2001). Teilen jene Nutzer dann Beiträge auf Facebook, so erhöht sich prinzipiell auch die Wahrscheinlichkeit, dass ein Studienteilnehmer mit – in diesem Fall sozial kuratierten – Nachrichteninhalten konfrontiert wird. Durch den ›Umweg‹ über das Kriterium Bildung konnte zudem vermieden werden, eine nachrichtenbezogene Variable (z. B. Interesse oder Nutzungshäufigkeit) zum Auswahlkriterium zu machen: Die genaue Kenntnis des Forschungsinteresses hätte die Teilnehmer ggf. dazu veranlasst, sich während der Beobachtung verstärkt Nachrichteninhalten zuzuwenden.

Basierend auf den Empfehlungen von Meyen und Kollegen (2011: 71-73) wurde entlang dieser Kriterien ein Quotenplan erstellt, in dem Größenordnungen (Mindestangaben) für bestimmte Ausprägungen der drei Kriterien definiert wurden. Insgesamt wurden 16 Beobachtungen durchgeführt (6 Männer, 10 Frauen, zwischen 16 und 47 Jahren), die inklusive Post-Exposure-Walkthrough und Leitfadeninterview zwischen 30 und 55 Minuten dauerten. Unerlässlich bei allen Studien dieser Art ist es, das Einverständnis über die Aufzeichnung (hier: der Facebook-Nutzung) und Verwendung der Daten bereits im Rahmen der Rekrutierung einzuholen (s. zu forschungsethischen Implikationen auch Kapitel 4).

### *Verlauf einer typischen Erhebungsepisode*

Eine typische Erhebungsepisode lief in der Studie wie folgt ab: Der Teilnehmer wurde – nach erneuter Aufklärung über das Prozedere – gebeten, sich mit seinem persönlichen Facebook-Account an einem vorbereiteten Laptop einzuwählen, auf dem eine Screen-Capture-Software<sup>2</sup> (*Microsoft Expression Encoder*) die folgende, auf fünf Minuten festgelegte, Nutzungsepisode inklusiver aller Mausbewegungen und Klicks aufzeichnete. Daneben wurde zudem eine Tonaufnahme angefertigt, da auf Basis eines Pretests

2 Kostenlose Screen-Capture-Software gibt es auch für Smartphones, für Android-Geräte z. B. *DV Recorder* oder *AZ Screen Recorder*. Aktuelle ios-Geräte wie iPhone und iPad haben eine Bildschirmaufnahmefunktion sogar standardmäßig installiert. So können auch Online-Anwendungen, die ausschließlich oder schwerpunktmäßig für die mobile Nutzung optimiert sind (z. B. Instagram), im Rahmen von Post-Exposure-Walkthrough-Studien untersucht werden.

anzunehmen war, dass manche Teilnehmer bereits während der Handlung Gedanken zum Geschehen verbalisieren würden.<sup>3</sup>

Zu Beginn der Beobachtung stellte sich die Forscherin den Teilnehmern vor und erläuterte noch einmal kurz das Ziel der Studie. Um die Teilnehmer bzw. ihr Verhalten während der Beobachtung nicht übermäßig zu beeinflussen, wurde das Erkenntnisinteresse zu Beginn eher vage formuliert und angegeben, dass es darum ginge, »herauszufinden, wie sich die Facebook-Nutzung verschiedener Nutzer charakterisieren lässt«. Das tatsächliche Erkenntnisinteresse wurde erst zu Beginn des Leitfadeninterviews spezifiziert, da es in diesem schließlich schwerpunktmäßig um die (Facebook-bezogene) Nachrichtennutzung der Teilnehmer ging. Trotz vollständiger Aufklärung bei der Terminvereinbarung, wurden die Teilnehmer unmittelbar vor der Beobachtung erneut darüber informiert, dass ihr Surfverhalten sowie ihre verbalen Äußerungen aufgezeichnet und für die Datenauswertung dokumentiert, d. h. transkribiert und protokolliert werden. Jedem Teilnehmer wurde darüber hinaus zugesichert, dass alle im Ergebnisbericht verwendeten Daten und Aussagen derart angepasst werden, dass sie sich nicht mehr mit seiner Person in Verbindung bringen lassen.

Die Forscherin saß während der fünfminütigen Nutzungsepisode leicht versetzt neben dem Teilnehmer und notierte dabei die im Beobachtungsprotokoll definierten Handlungs- und Reaktionsmuster sowie nonverbale Reaktionen (Mimik, Gestik), auf die im Rahmen des Post-Exposure-Walkthrough Bezug genommen werden sollte. Die Teilnehmer wurden im Vorfeld darauf hingewiesen, dass ihnen während des Surfens keinerlei Fragen gestellt werden und zudem explizit darum gebeten, »so bei Facebook zu surfen, wie Du es sonst auch tun würdest«. Der anschließende Befragungsanteil der Studie gliederte sich wie oben beschrieben in zwei Teile, wobei im ersten Teil durch Erzählaufforderungen die Gründe für die beobachteten Navigations- und Selektionshandlungen des Teilnehmers eruiert wurden (→ Post-Exposure-Walkthrough, s. dazu ausführlich das Fallbeispiel in Kapitel 3.4), während im zweiten Teil generelle Tendenzen der Facebook- und Nachrichtennutzung im Mittelpunkt standen (→ Leitfadeninterview).

3 Diese Annahme erwies sich als gerechtfertigt: Drei der Teilnehmer verbalisierten bereits vor dem Post-Exposure-Walkthrough ausgiebig (d. h. über vereinzelte Kommentare hinaus) ihre Handlungen und Gedanken.

### 3.3 Auswertungsstrategie

Für die Auswertung wurden zwei Dokumente angefertigt: Das erste enthielt das Beobachtungsprotokoll und die Transkripte von Post-Exposure-Walk-through und Leitfadeninterview, ergänzend wurden in einem zweiten Dokument jeweils die ersten zehn Posts im Neuigkeitenbereich des Teilnehmers aufgeschlüsselt sowie mit Blick auf untersuchungsrelevante Aspekte analysiert (= Postanalyse). Auch wenn naturgemäß nicht alle Posts Nachrichtenposts waren, erwies sich diese Analyse als hilfreich, um ein Gefühl für die Zusammensetzung des Neuigkeitenbereichs des Teilnehmers sowie dessen generellen Umgang mit Facebook-Posts zu erlangen (Wie hoch ist der Anteil von Medieninhalten? Wie häufig wird auf ›Gefällt mir‹ geklickt?). Beide Dokumente wurden mit Screenshots der Posts illustriert, um die Auswertung bei Bedarf ohne Konsultation der Videodateien durchführen zu können.<sup>4</sup> Die Kombination von schriftlichen und aufgezeichneten Beobachtungsdaten (Bildschirm- und Tonaufnahme) mit den expliziten Verbalisierungen der Teilnehmer ermöglichte zusammengekommen einen vertieften Einblick in die (subjektiven) Prozesse, die die SNS-Nachrichtennutzung determinieren. Insbesondere die für den DTA zentralen, bei quantitativen Verfahren oder reinen Beobachtungen jedoch im Verborgenen bleibenden Inter-Transaktionen zwischen Teilnehmer und einem spezifischen Nachrichteninhalt sowie innerhalb des Teilnehmers zu verortende Intra-Transaktionen (Aktivation, Wissen) konnten so nachvollzogen werden.

Zur Analyse des Materials wurde schließlich auf eine Form der qualitativen Inhaltsanalyse zurückgegriffen (MAYRING 2015; die Anwendung auf kommunikationswissenschaftliche Fragestellungen diskutieren auch MEYEN et al. 2011: 171ff.; NAWRATIL/SCHÖNHAGEN 2009; FÜRST/JECKER/SCHÖNHAGEN 2016). Dabei handelt es sich um ein systematisches Verfahren zur Analyse von Texten, das sich durch die Verknüpfung von deduktiver und induktiver Kategorienbildung auszeichnet. So wird einerseits auf ein sich aus den theoretischen Vorannahmen abgeleitetes »Kriterien-

4 Zusätzlich bietet sich die Nutzung von Zeitstempeln an, die eine schnelle Identifikation der Posts im Originalmaterial ermöglichen. Bei der Arbeit mit *MAXQDA Plus 12* können Codierungen auch direkt am Videomaterial vorgenommen werden: Ein Klick auf die Codierungen ermöglicht dann eine direkte Wiedergabe des entsprechenden Abschnitts im programmeigenen Multimedia-Browser.

raster« (NAWRATIL/SCHÖNHAGEN 2009: 335) zurückgegriffen, das als erstes Ordnungssystem die Analyse anleitet. Andererseits werden die durch das Kriterienraster festgelegten Kategorien durch mehrfaches, intensives Lesen des Untersuchungsmaterials um neue Kategorien ergänzt, die im ursprünglichen Kriterienraster nicht oder nicht ausreichend expliziert sind. Es geht also nicht (nur) darum, im Vorfeld festgelegte Aspekte im Material ausfindig zu machen. Vielmehr hat die qualitative Inhaltsanalyse einen stark explorativen Charakter und erlaubt so eine weitgehend offene und flexible Bearbeitung des Untersuchungsmaterials (vgl. FÜRST/JECKER/SCHÖNHAGEN 2016: 216; MEYEN et al. 2011: 171; NAWRATIL/SCHÖNHAGEN 2009: 335).

Die Verwaltung und Sichtung der Auswertungsdokumente (Beobachtungsprotokoll und Transkripte, Postanalysen), die Entwicklung des Kategoriensystems sowie die eigentliche Codierung erfolgten mithilfe der für die qualitative Datenanalyse optimierten QDA-Software *MAXQDA Plus 12*.

### 3.4 Fallbeispiel

Um ein Gefühl für das Potenzial des Verfahrens zu erhalten, wird auf eine überblicksartige Beschreibung der Ergebnisse verzichtet und stattdessen am Beispiel einer der durchgeführten Erhebungen erläutert, wie die qualitative Beobachtung mit Post-Exposure-Walkthrough dabei helfen kann, die individuellen Prozesse zu rekonstruieren, die dem nachrichtenbezogenen Navigations- und Selektionshandeln auf Facebook zugrunde liegen. Auszugsweise beschrieben wird die Ende August 2016 durchgeführte Beobachtung von Carolin<sup>5</sup>, 27 Jahre alt, zum Zeitpunkt der Untersuchung Studentin in einem sozialwissenschaftlichen Masterstudiengang.

Nach einem kurzen Blick auf den ersten Post in ihrem Neuigkeitenbereich – ein Videobeitrag, der von einer Facebook-Freundin kommentiert wurde, aber keinerlei Relevanz für Carolin hat (»[D]as interessiert mich nicht«) – scrollt sie zunächst zum fünften Post, kehrt dann aber zum dritten zurück – ein Post von *Spiegel Online* über Veganismus –, um kurz den

5 Die Namen aller Teilnehmer wurden für Auswertung und Ergebnispräsentation pseudonymisiert, d. h. »möglichst gleichwertig i.S. sozialwissenschaftlicher Nützlichkeit« (GEBEL et al. 2015) ersetzt.

Teaser zu lesen, dann aber doch wieder nach unten zu scrollen. Erneut hält sie bei einem Post von *Spiegel Online* (dieses Mal über eine nachgebildete Raumstation auf Hawaii) inne, entscheidet sich aber wiederum gegen einen Klick auf den Artikel. Carolin erläutert dazu: »Ich bin schon an dem Spiegel-Online-Post hängengeblieben, weil Beiträge von Spiegel Online mich schon interessieren. Das ist auch meine Startseite, spiegel.de. Das ist so mein Medium, mit dem ich mich informiere [...] Aber dann hab ich schnell gesehen: Oh, nee, das interessiert mich nicht [*Carolin lacht*]. Das Thema hat mich nicht interessiert.« Hier und im weiteren Verlauf der Beobachtung wird deutlich, dass der Name bzw. die Marke eines Nachrichtenanbieters als Heuristik ihre initiale Zuwendung zu Nachrichtenposts sichtlich beeinflusst, jedoch nicht der alleinige Auslöser für Auswahlentscheidungen ist. Im zitierten Beispiel überwiegt etwa das (fehlende) thematische Interesse die eigentlich aufmerksamkeitsfördernde Wirkung des Anbieters. Nach weiterem Scrolling und einem kurzen Blick auf einen genuin sozialen Beitrag (eine Freundin verkündet eine Reise in die USA), setzt sich Carolin erneut länger mit dem Teaser zu einem DEUTSCHLANDFUNK-Artikel auseinander. Auslöser dafür war eine der Facebook-Nutzungsperiode vorausgehende Nachrichtenrezeption: »Ich hab gestern Abend die Tagesschau gesehen [...] Und da war ein kurzer Ausschnitt aus diesem ARD-Sommerinterview. Und da hab ich gesehen, dass der Post sich auch darum dreht – ob Merkel jetzt kandidiert und welche Rolle die CSU spielt.« Allerdings hält Carolin auch in diesem Fall der vom Teaser angekündigte Fokus – versprochen wird eine Einordnung des ehemaligen Bild-Chefredakteurs Hans-Hermann Tiedje – von einer Rezeption des verlinkten Artikels ab: »Was der dazu sagt, das interessiert mich dann doch nicht so. Aber das Thema hat mich per se schon interessiert.« Dieses grundlegende Interesse bestätigt eine kurz darauffolgend stattfindende Interaktion mit einem *Handelsblatt*-Post zum Sommerinterview, die in einer tatsächlichen Auseinandersetzung mit dem verlinkten Artikel mündet. Hier wurden laut Carolin schließlich die Informationen geboten, die sie sich bereits bei der Auseinandersetzung mit dem DEUTSCHLANDFUNK-Post erhofft hatte. Bei der Rückkehr in den Neuigkeitenbereich zeigt sich schließlich, wie auch Facebook-interne Features das (nachrichtenbezogene) Navigations- und Selektionshandeln lenken können. Die »Related Articles«-Funktion sorgt dafür, dass Carolin nach der Rückkehr von der *Handelsblatt*-Webseite drei weitere illustrierte Teaser zu Nachrichtenbeiträgen angezeigt bekommt, die thematisch jenem ähneln, den sie zuvor rezipiert hat. Tatsächlich führt die Sichtung

dieser Teaser dazu, dass sie noch einen weiteren Beitrag anklickt, der sich mit Merkels Kanzlerkandidatur beschäftigt und besonderes Augenmerk auf den Einfluss der CSU legt.

Angestoßen durch die Konfrontation im Post-Exposure-Walkthrough reflektiert Carolin jedoch nicht nur die Auseinandersetzung mit einzelnen Posts, sondern auch ihre generelle Facebook-Nutzung sowie die Zusammensetzung ihres Neuigkeitenbereichs: »Ich hab auch – das fällt mir jetzt erst so richtig auf – viele Medien bei Facebook gelikt. Deutschlandfunk, Spiegel Online, Handelsblatt, die Süddeutsche Zeitung – also Informationsmedien, über die ich mich auch sonst informieren würde, aber ich nutze meine Facebook-Timeline quasi so als kleine, eigene zusammengebastelte Zeitung.« Vor diesem Hintergrund überrascht es nicht, dass Carolin fast ausschließlich deshalb auf Nachrichtenbeiträge aufmerksam wird, weil sie ihre Informationsumgebung durch persönliche Kuration bewusst »nachrichtenbetont« gestaltet hat. Im Gegensatz zu vielen anderen Teilnehmern spielt die soziale Kuration von Nachrichtenposts/-beiträgen bei ihr eine untergeordnete Rolle. Im Interview wird deutlich, dass in ihrem Freundeskreis als interessant empfundene Nachrichtenbeiträge primär über private Nachrichten ausgetauscht werden, da so mit Blick auf die »so heterogen[e]« Gruppe an Facebook-Freunden sehr zielgenau Empfehlungen ausgesprochen werden könnten.

Was der Post-Exposure-Walkthrough ebenfalls offenbart – gerade im Vergleich mit den anderen Teilnehmern –, ist die generelle Herangehensweise an den Umgang mit externen Inhalten. Während manche Teilnehmer nahezu jedem (Nachrichten-)Link folgen, dem sie begegnen, setzen sich andere, wie auch Carolin, im Vorfeld recht ausführlich mit den Teasern auseinander und entscheiden erst auf Basis inhaltlicher Vorüberlegungen über den weiteren Umgang. Durch die Nähe zur Rezeptionssituation sowie die Kombination von Beobachtungsdaten und verbalen Einordnungen dieser Daten durch die Teilnehmer werden Nutzungsdynamiken sicht- und thematisierbar, die reinen Beobachtungen oder Befragungen mit großer Wahrscheinlichkeit entgangen wären. Im vorliegenden Fall etwa die durch eine vorherige TV-Nachrichtenrezeption angestoßene Aufmerksamkeit für einen Nachrichtenpost (→ *Aktivierung*), die zwar nicht in eine Rezeption des Artikels mündete, dafür aber den Anstoß für die Interaktion mit zwei weiteren Beiträgen geliefert hat, die den Informationsbedürfnissen besser entsprochen haben und folglich für eine Rezeption ausgewählt wurden. Solche dynamischen Nutzungsketten sind schließlich nur vor dem indivi-



duellen Erfahrungshorizont der Teilnehmer zu verstehen, der durch den Post-Exposure-Walkthrough weitgehend organisch zur Sprache kommt bzw. gebracht werden kann.

#### 4. Diskussion

Der vorgestellte methodische Ansatz ist grundsätzlich nicht neu und wurde, wie oben geschildert, in Abhängigkeit von Disziplin und inhaltlichem Fokus bereits unter Begriffen wie »gestütztes« nachträgliches Lautes Denken« (BILANDZIC/TRAPP 2000), »Selbstkonfrontationsinterview« (BREUER 1995; LIM 2002) oder »Stimulated Recall« (CALDERHEAD 1981; MESSMER 2014) diskutiert, wobei mit jedem Begriff auch unterschiedliche Schwerpunkte in Datenerhebung und -auswertung einhergehen. Die hier genutzte Bezeichnung findet Anwendung, da sie den Aspekt der *Beobachtung* sowie die *schrittweise* Rekonstruktion einzelner Nutzungsvorgänge in den Mittelpunkt stellt. Alle Ansätze tragen jedoch gleichermaßen dem Umstand Rechnung, 1) dass Beobachtungen allein nicht ausreichen, um den subjektiven Sinn von Handlungen zu erfassen und 2) dass eine »nachträgliche Verbalisierung von Kognitionen« (SCHWEIGER 2010: 195) sinnvoll erscheint, um die mit handlungsbegleitenden Teilnehmerkommentaren einhergehende Reaktivität (z. B. Verlangsamung/Veränderung der Handlung) einzuschränken, die etwa beim simultanen Lauten Denken ein zentrales Problem darstellt.

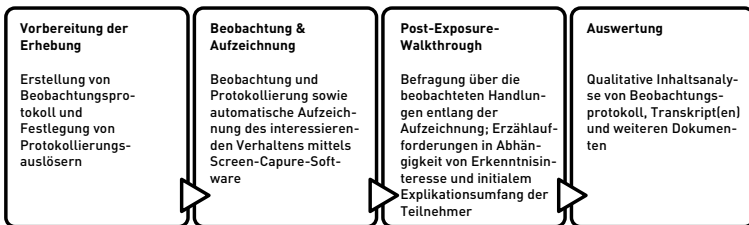
Eine fachliche Diskussion des hier umrissenen, streng genommen »alten« Ansatzes rechtfertigt sich dennoch und nicht zuletzt aufgrund der gerade in der SNS-Rezeptionsforschung zunehmend zu beobachtenden »Schieflage«, die dadurch entsteht, dass »wir zwar auf der einen Seite Unmengen von Verhaltens- und Kommunikationsdaten erheben können, über Urheber und die Entstehungssituation aber vergleichsweise wenig wissen« (BROSIUS 2016: 369). Insbesondere bei der Arbeit mit prozess- und subjektorientierten Theorien wie dem DTA sollten daher methodische Ansätze in Erwägung gezogen werden, die neben individuellen Informationsumgebungen auch damit zusammenhängende, zum Teil nur durch Einordnungen der Teilnehmer zu verstehende Nutzungspraktiken und -dynamiken in die Analyse integrieren. Daneben bietet der Versuchsaufbau die Möglichkeit, ohne erheblichen Mehraufwand auch Verfahren wie das Mouse-Tracking (CHEN/ANDERSON/SOHN 2001; DIEKAMP/SCHWEIGER 2001; FREEMAN/AMBADY 2010) zu integrieren, das zwar nicht die gleiche

Präzision wie Messungen via Eye-Tracking aufweist, aber bei vielen Fragestellungen dennoch als ökonomischer Proxy für Aufmerksamkeitsprozesse fungieren kann.

Abbildung 1 fasst den grundsätzlichen Ablauf der qualitativen Beobachtung mit Post-Exposure-Walkthrough noch einmal überblicksartig zusammen und eröffnet so auch die Möglichkeit, die in den jeweiligen Phasen zu verortenden Leistungen und Grenzen des Verfahrens zu diskutieren.

ABBILDUNG 1

## Die qualitative Beobachtung mit Post-Exposure-Walkthrough im Überblick



Während sich Phase 1 (*Vorbereitung der Erhebung*) grundsätzlich nicht von anderen Formen der Beobachtung unterscheidet, lassen sich für Phase 2 (*Beobachtung & Aufzeichnung*) spezifische Vor- und Nachteile benennen. So entlastet die Aufzeichnung einerseits den Beobachtungsprozess, indem sich durch den anschließenden Post-Exposure-Walkthrough die Möglichkeit eröffnet, auch jene Auffälligkeiten zu thematisieren, die dem Forscher bei der ›Live-Beobachtung‹ entgangen sind, andererseits erfordern Aufzeichnungen – insbesondere von hochgradig individualisierten Online-Umgebungen wie dem Neuigkeitenbereich auf Facebook<sup>6</sup> – zwingend die informierte Einwilligung der Teilnehmer und somit schon im Vorfeld eine exakte Vorstellung über Aus- und Verwertung sowie Archivierung der aufgezeichneten Daten. Daneben ist in dieser Phase aufgrund der Anwe-

6 Auch wenn das Interesse im vorliegenden Fall auf den Umgang mit Nachrichteninhalten beschränkt war, ist unvermeidlich, dass auch private oder gar intime Inhalte im Neuigkeitenbereich der Teilnehmer beobachtet und aufgezeichnet werden. Generell bietet es sich daher an, alle Inhalte, die für das Forschungsinteresse irrelevant sind, zu schwärzen oder von vornherein aus den Auswertungsdokumenten auszuschließen.

senheit des Forschers mit klassischen Formen der Reaktivität zu rechnen, die zu einem veränderten Nutzungsverhalten führen können. Der Großteil der Vorteile lässt sich schließlich in Phase 3 (*Post-Exposure-Walkthrough*) verorten. Durch die zeitliche und affektive Nähe zur beobachteten Nutzungsepisode können situative Wechselwirkungen zwischen Nutzern und genutzten Inhalten explizit berücksichtigt und (eigen-)dynamische Prozesse nachgezeichnet werden. Als besonders lohnenswert hat sich während des Post-Exposure-Walkthrough zudem die Möglichkeit zum Anhalten, Verlangsamens und Zurückspulens der Aufzeichnungen herausgestellt, da so auch nuancierte Handlungen oder wiederkehrende Routinen mit den Teilnehmern erörtert werden konnten. Problematisch hingegen ist – wie bei Aufforderungen zur Kontextualisierung des eigenen Verhaltens generell –, dass nur bewusste Prozesse artikuliert werden können und das eigene Handeln von den Teilnehmern unter Umständen (stärker) rationalisiert wird. In Phase 4 (*Auswertung*) schließlich zeigen sich angesichts der Fülle an Auswertungsdaten und -dokumenten primär Herausforderungen, die sich mit einem strukturierten Analyseplan jedoch gut bewältigen lassen.

Die SNS-Rezeptionsforschung droht durch immer bessere Möglichkeiten zur automatisierten und großflächigen Erhebung von Verhaltens- und Spurendaten den Blick auf die Genese dieser Daten sowie die für eine Interpretation der Befunde nötige Tiefenschärfe zu verlieren. Qualitativ-rekonstruierende Verfahren wie das hier beschriebene können dazu beitragen, quantitative Befunde zu kontextualisieren und sinnvoll einzuordnen. Vor allem als Ergänzung dürfte – und *sollte* – ihnen dabei auch künftig eine gewichtige Rolle bei der Analyse von (SNS-)Nutzungsprozessen zukommen. Methodische Ansätze, bei denen die Teilnehmer ihre Aktivitäten oder Gedanken aus einer rezenten Erinnerung herleiten müssen, empfehlen sich hier in besonderem Maße – neben der qualitativen Beobachtung mit Post-Exposure-Walkthrough etwa Spielarten der Tagebuchmethode oder der Mobile Experience Sampling Method (MESM, FAHR/KARNOWSKI 2017; KARNOWSKI 2013).

## Literatur

- BEAM, M. A.; M. J. HUTCHENS; J. D. HMIELOWSKI: Clicking vs. sharing: The relationship between online news behaviors and political knowledge. In: *Computers in Human Behavior*, 59, 2016, S. 215-220.  
<https://doi.org/10.1016/j.chb.2016.02.013>

- BILANDZIC, H.: Wie Selektion und Interpretation zusammenhängen: Kombination von Lautem Denken und Beobachtung.  
In: LOOSEN, W.; A. SCHOLL (Hrsg.): *Methodenkombinationen in der Kommunikationswissenschaft. Methodologische Herausforderungen und empirische Praxis*. Köln [Herbert von Halem] 2012, S. 195-209
- BILANDZIC, H.; B. TRAPP: Die Methode des lauten Denkens: Grundlagen des Verfahrens und die Anwendung bei der Untersuchung selektiver Fernsehnutzung bei Jugendlichen. In: PAUS-HAASE, I.; B. SCHORB (Hrsg.): *Qualitative Kinder- und Jugend-Medienforschung. Theorie und Methoden: Ein Arbeitsbuch*. München [KoPäd] 2000, S. 183-209
- BODE, L.: Political news in the news feed: Learning politics from social media. In: *Mass Communication and Society*, 19(1), 2016, S. 24-48.  
<https://doi.org/10.1080/15205436.2015.1045149>
- BREUER, F.: Das Selbstkonfrontations-Interview als Forschungsmethode.  
In: KÖNIG, E.; P. ZEDLER (Hrsg.): *Bilanz qualitativer Forschung. Band 11: Methoden*. Weinheim [Deutscher Studien Verlag] 1995, S. 159-180
- BROSIUS, H.-B.: Warum Kommunikation im Internet öffentlich ist. In: *Publizistik*, 61(4), 2016, S. 363-372. <https://doi.org/10.1007/s11616-016-0304-6>
- BROSIUS, H.-B.; A. HAAS; F. KOSCHEL: *Methoden der empirischen Kommunikationsforschung. Eine Einführung* (7., überarbeitete und aktualisierte Auflage). Wiesbaden [Springer vs] 2016
- CALDERHEAD, J.: Stimulated recall: A method for research on teaching.  
In: *British Journal of Educational Psychology*, 51(2), 1981, S. 211-217. <https://doi.org/10.1111/j.2044-8279.1981.tb02474.x>
- CHEN, M. C.; J. R. ANDERSON; M. H. SOHN: What can a mouse cursor tell us more? Correlation of eye/mouse movements on web browsing.  
In: *CHI '01 Extended Abstracts on Human Factors in Computing Systems*. New York/NY [ACM] 2001, S. 281f. <https://doi.org/10.1145/634067.634234>
- DIEKAMP, O.; W. SCHWEIGER: Zur visuellen Wahrnehmung von Webseiten – Ergebnisse einer Rezeptionsstudie mit dem Mouse-Tracking-Verfahren. In: BECK, K.; W. SCHWEIGER (Hrsg.): *Attention please! Online-Kommunikation und Aufmerksamkeit*. München [Reinhard Fischer] 2001, S. 197-214
- DÖRING, N.; J. BORTZ: *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. vollständig überarbeitete, aktualisierte und erweiterte Auflage). Berlin, Heidelberg [Springer] 2016

- FAHR, A.; V. KARNOWSKI: Die mobile Onlinebefragung. In: MIKOS, L.; C. WEGENER (Hrsg.): *Qualitative Medienforschung. Ein Handbuch* (2., völlig überarbeitete und erweiterte Auflage). Konstanz [UVK] 2017, S. 340-346
- FREEMAN, J. B.; N. AMBADY: MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. In: *Behavior Research Methods*, 42(1), 2010, S. 226-241. <https://doi.org/10.3758/BRM.42.1.226>
- FRÜH, W.; K. SCHÖNBACH: Der dynamisch-transaktionale Ansatz. Ein neues Paradigma der Medienwirkungen. In: *Publizistik*, 27(1), 1982, S. 74-88
- FRÜH, W.; K. SCHÖNBACH: Der dynamisch-transaktionale Ansatz III. Eine Zwischenbilanz. In: *Publizistik*, 50(1), 2005, S. 4-20. <https://doi.org/10.1007/s11616-005-0115-7>
- FÜRST, S.; C. JECKER; P. SCHÖNHAGEN: Die qualitative Inhaltsanalyse in der Kommunikationswissenschaft. In: AVERBECK-LIETZ, S.; M. MEYEN (Hrsg.): *Handbuch nicht standardisierte Methoden in der Kommunikationswissenschaft*. Wiesbaden [Springer] 2016, S. 209-225
- GARCÍA-PERDOMO, V.; R. SALAVERRÍA; D. K. KILGO; S. HARLOW: To share or not to share. The influence of news values and topics on popular social media content in the United States, Brazil, and Argentina. In: *Journalism Studies*, 19(8), 2018, S. 1180-1201. <https://doi.org/10.1080/1461670X.2016.1265896>
- GEBEL, T.; M. GRENZER; J. KREUSCH; S. LIEBIG; H. SCHUSTER; R. TSCHERWINKA; O. WATELER; A. WITZEL: Verboten ist, was nicht ausdrücklich erlaubt ist: Datenschutz in qualitativen Interviews. In: *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 16(2), 2015. <http://nbn-resolving.de/urn:nbn:de:0114-fqs1502279>
- GEHRAU, V.: *Die Beobachtung in der Kommunikationswissenschaft*. Konstanz [UVK] 2002
- HELFFERICH, C.: *Die Qualität qualitativer Daten. Manual für die Durchführung qualitativer Interviews* (4. Auflage). Wiesbaden [vs Verlag] 2011
- KARNOWSKI, V.: Befragung in situ: Die Mobile Experience Sampling Method (MESM). In: MÖHRING, W.; D. SCHLÜTZ (Hrsg.): *Handbuch standardisierte Erhebungsverfahren in der Kommunikationswissenschaft*. Wiesbaden [Springer] 2013, S. 235-247
- KIM, Y.; H.-T. CHEN; H. GIL DE ZÚÑIGA: Stumbling upon news on the Internet: Effects of incidental news exposure and relative

- entertainment use on political engagement. In: *Computers in Human Behavior*, 29, 2013, S. 2607-2614. <https://doi.org/10.1016/j.chb.2013.06.005>
- KOCH, W.; B. FREES: Dynamische Entwicklung bei mobiler Internetnutzung sowie Audios und Videos. In: *Media Perspektiven*, (9), 2016, S. 418-437
- KONRAD, K.: Lautes Denken. In: MEY, G.; K. MRUCK (Hrsg.): *Handbuch Qualitative Forschung in der Psychologie*. Wiesbaden [vs Verlag] 2010, S. 476-490
- KUCKARTZ, U.: *Einführung in die computergestützte Analyse qualitativer Daten* (3., aktualisierte Auflage). Wiesbaden [vs Verlag] 2010
- LEE, J. K.; E. KIM: Incidental exposure to news: Predictors in the social media setting and effects on information gain online. In: *Computers in Human Behavior*, 75, 2017, S. 1008-1015. <https://doi.org/10.1016/j.chb.2017.02.018>
- LIM, S.: The self-confrontation interview. Towards an enhanced understanding of human factors in web-based interaction for improved website usability. In: *Journal of Electronic Commerce Research*, 3(3), 2002, S. 162-173
- LÖNNQVIST, J.-E.; J. V. A. ITKONEN: Homogeneity of personal values and personality traits in Facebook social networks. In: *Journal of Research in Personality*, 60, 2016, S. 24-35. <https://doi.org/10.1016/j.jrp.2015.11.001>
- MAYRING, P.: *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (12., überarbeitete Auflage). Weinheim, Basel [Beltz] 2015
- MCPHERSON, M.; L. SMITH-LOVIN; J. M. COOK: Birds of a feather: Homophily in social networks. In: *Annual Review of Sociology*, 27(1), 2001, S. 415-444. <https://doi.org/10.1146/annurev.soc.27.1.415>
- MESSMER, R.: Stimulated recall as a focused approach to action and thought processes of teachers. In: *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 16(1), 2014. <http://www.qualitative-research.net/index.php/fqs/article/view/2051>
- MEYEN, M.; M. LÖBLICH; S. PFAFF-RÜDIGER; C. RIESMEYER (Hrsg.): *Qualitative Forschung in der Kommunikationswissenschaft. Eine praxisorientierte Einführung*. Wiesbaden [vs Verlag] 2011
- MISOCH, S.: *Qualitative Interviews*. Berlin, München, Boston [De Gruyter] 2015
- NAWRATIL, U.; P. SCHÖNHAGEN: Die qualitative Inhaltsanalyse: Rekonstruktion der Kommunikationswirklichkeit. In: WAGNER, H.

- (Hrsg.): *Qualitative Methoden in der Kommunikationswissenschaft. Ein Lehr- und Studienbuch* (Vollständig überarbeitete, erweiterte und ergänzte Neuauflage). Baden-Baden [Nomos] 2009, S. 333-346
- NEWMAN, N.; R. FLETCHER; A. KALOGEROPOULOS; D. A. L. LEVY; R. K. NIELSEN: *Reuters Institute Digital News Report 2017*. Oxford/UK [Reuters Institute for the Study of Journalism] 2017
- QUANDT, T.: Journalisten unter Beobachtung. Grundlagen, Möglichkeiten und Grenzen der Beobachtung als Methode der Journalismusforschung. In: JANDURA, O.; T. QUANDT; J. VOGELGESANG (Hrsg.): *Methoden der Journalismusforschung*. Wiesbaden [vs Verlag] 2011, S. 277-297
- SCHMIDT, J.-H.: *Social Media* (2., aktualisierte und erweiterte Auflage). Wiesbaden [Springer vs] 2017
- SCHÖNBACH, K.; W. FRÜH: Der dynamisch-transaktionale Ansatz II. Konsequenzen. In: *Rundfunk und Fernsehen*, 32(3), 1984, S. 314-329
- SCHWEIGER, W.: Informationsnutzung online: Informationssuche, Selektion, Rezeption und Usability von Online-Medien. In: SCHWEIGER, W.; K. BECK (Hrsg.): *Handbuch Online-Kommunikation*. Wiesbaden [vs Verlag] 2010, S. 184-210
- TRILLING, D.; P. TOLOCHKO; B. BURSCHER: From newsworthiness to shareworthiness. How to predict news sharing based on article characteristics. In: *Journalism & Mass Communication Quarterly*, 94(1), 2017, S. 38-60. <https://doi.org/10.1177/1077699016654682>
- VALERIANI, A.; C. VACCARI: Accidental exposure to politics on social media as online participation equalizer in Germany, Italy, and the United Kingdom. In: *New Media & Society*, 18(9), 2016, S. 1857-1874. <https://doi.org/10.1177/1461444815616223>
- VAN EIMEREN, B.; W. KOCH: Nachrichtenkonsum im Netz steigt an – auch klassische Medien profitieren. In: *Media Perspektiven*, (5), 2016, S. 277-285
- VICARI, J.: Beobachtung in der Kommunikationswissenschaft. In: AVERBECK-LIETZ, S.; M. MEYEN (Hrsg.): *Handbuch nicht standardisierte Methoden in der Kommunikationswissenschaft*. Wiesbaden [Springer] 2016, S. 289-301

MARTIN WETTSTEIN / ANDREAS FAHR /  
ALEXANDER ORT

## Beyond Eyeballing: Automatische Mustererkennung in Prozessdaten

*Abstract:* Die Kommunikationswissenschaft hat zunehmend Zugriff auf umfangreiche und feingliedrige Beobachtungsdaten, mit denen unter anderem Prozesse der Mediennutzung, -rezeption und -wirkung beschrieben werden. Neue Erhebungsmethoden erlauben es etwa, das Online-Verhalten von MediennutzerInnen (z. B. die Kommunikation auf Twitter), die physiologischen Reaktionen (z. B. Augenbewegungen von Probanden) oder das Nutzungsverhalten auf Websites detailliert und in Echtzeit zu beobachten. Um diese – teilweise sehr umfangreichen – Daten quantitativ wie qualitativ auswerten zu können, müssen Muster in den Daten erkannt und verlässlich charakterisiert werden.

In diesem Beitrag wird anhand zweier Beispielstudien gezeigt, wo die Kommunikationswissenschaft mit diesen Prozessdaten in Berührung kommt und welche quantitativen Methoden sich anbieten, um Muster in den Daten verlässlich zu finden. Zwei neue Verfahren zur Identifikation bekannter Verlaufsmuster und repetitiver Sequenzen werden vorgestellt und in einem frei verfügbaren Statistikprogramm implementiert. Anhand der Beispielstudien wird die Anwendung der Verfahren demonstriert, die wertvolle Informationen zu den Prozessen liefern und zur Beantwortung von Forschungsfragen herangezogen werden können.



## 1. Einleitung

Die Kommunikationswissenschaft hat zunehmend Zugriff auf umfangreiche und feingliedrige Beobachtungsdaten, mit denen unter anderem Prozesse der Mediennutzung, -rezeption und -wirkung beschrieben werden. Neue Erhebungsmethoden erlauben es etwa, das Online-Verhalten von MediennutzerInnen (z. B. Nutzung von Online-News, Like- und Share-Verhalten oder Online-Diskussionen), die physiologischen Reaktionen (z. B. Augenbewegungen und körperliche Erregung von Probanden) oder das Nutzungsverhalten auf Websites detailliert und in Echtzeit zu beobachten. Obschon Längsschnittdaten im Fach bereits aus Inhaltsanalysen und Tagebuchstudien bekannt sind, stellen der große Umfang und die Komplexität solcher Prozessdaten das Fach vor neue Möglichkeiten, aber auch methodische Herausforderungen.

Eine dieser Herausforderungen besteht darin, Umfang und Komplexität der erhobenen Zeitreihen so zu reduzieren, dass sie quantitativ und qualitativ ausgewertet, sinnvoll interpretiert und schließlich zur Beantwortung von Forschungsfragen und Hypothesen herangezogen werden können. Dieser Herausforderung kann man begegnen, indem man versucht, bedeutungshaltige Muster in Prozessdaten zu identifizieren und zu isolieren. Dadurch werden relevante Ereignisse in den Daten zum einen zähl- und charakterisierbar, was eine quantitative Auswertung ermöglicht. Darüber hinaus werden Prozessdaten dadurch in isolierte Muster zerlegt, die ihrerseits auch qualitativ beschrieben werden können.

Bei der Identifikation von Mustern in Datenreihen gibt es traditionell zwei mögliche Herangehensweisen: Einerseits verfügt der Mensch über die Fähigkeit, Muster und Regeln in Daten zu erkennen und diese visuell unmittelbar zu erfassen und zu beschreiben. Wird eine Zeitreihe in geeigneter Weise dargestellt, kann ein Codierer bestimmte Muster erkennen und benennen. Dieses ›Eyeballing‹ stößt jedoch dort an seine Grenzen, wo Daten zu umfangreich werden oder die Identifikation nicht augenfällig ist und nach klaren und objektiven Kriterien erfolgen muss, um eine quantitative Auswertung zu ermöglichen. Hier sind andererseits empirische Methoden erforderlich, die Muster verlässlich automatisch identifizieren.

Zwei solcher empirischen Methoden zur Mustererkennung in Prozessdaten werden in diesem Beitrag eingeführt und anhand von Beispielstudien illustriert. Die Anwendung der Methoden erfolgt vor dem Hintergrund zweier unterschiedlicher Datenformate. Das erste Datenformat sind Verlaufsdaten,

in denen die Veränderung eines Merkmals über die Zeit beobachtet und festgehalten wird. Diese Art von Längsschnittdaten fallen insbesondere bei physiologischen Messungen an, bei denen Hautleitwiderstand, Pupillengröße, Muskelbewegungen oder der Herzschlag hochaufgelöst gemessen werden. Für dieses Datenformat wird eine Methode vorgestellt, mit welcher a priori bekannte Muster verlässlich identifiziert werden können. Das zweite Datenformat sind Sequenzdaten, bei denen eine Sequenz von Ereignissen zeitlich festgehalten wird. Diese Daten fallen an, wenn das Online-Verhalten, die Augenbewegungen oder einzelne Handlungen von Probanden untersucht und als Sequenz festgehalten wird. An diesem Datenformat wird eine Methode zur induktiven Mustererkennung vorgestellt.

## 2. Vordefinierte Muster in Verlaufsdaten

Die erste Methode, die in diesem Beitrag behandelt wird, ist die Mustererkennung in längsschnittlichen Verlaufsdaten. Diese Muster sind in der Kommunikationswissenschaft vor allem aus der Forschung zu Themenkarrieren oder Nachrichtenzyklen (DOWNS 1972) bekannt. Hier wurden etwa in longitudinalen Inhaltsanalysen wiederholt prototypische Muster ansteigender und absinkender Themenaufmerksamkeit aufgezeigt. Während sich diese Muster in einer Inhaltsanalyse über einen überschaubaren Zeitraum relativ leicht visuell identifizieren lassen, gestaltet sich eine manuelle Codierung von Mustern in umfangreicheren Datensätzen, wie sie in den letzten Jahren generiert werden (z. B. durch psychophysiologische Messungen oder automatische Inhaltsanalysen über mehrere Jahre an Berichterstattung) deutlich schwieriger. Je umfangreicher die Daten und je komplexer die zu identifizierenden Muster sind, desto größer wird das Bedürfnis nach quantitativen Methoden zu deren objektiver Erfassung und Charakterisierung.

In der Vergangenheit wurden prototypische Muster in Verläufen hauptsächlich deduktiv bestimmt und charakterisiert. So werden in Studien, in denen psychophysiologische Parameter erhoben werden, Stimuli eingesetzt, um zu einem festgelegten Zeitpunkt eine Reaktion auszulösen, die anschließend untersucht wird (MEITZ et al. 2017). In der Nachrichtenzyklenforschung werden konkrete Ereignisse gesucht und die darauf folgende Berichterstattung untersucht (DOWNS 1972); oder sie werden in Simulationsstudien zu einem festgelegten Zeitpunkt initiiert (WALDHERR 2014). In wenigen Fällen wird zudem visuell nach Mustern gesucht, um

sie qualitativ zu beschreiben (WETTSTEIN 2015). Eine quantitative Erfassung von prototypischen Mustern in Verlaufsdaten hat bis anhin in der Kommunikationswissenschaft noch keine Tradition. Das Bedürfnis nach geeigneten Methoden ist jedoch angesichts der wachsenden Menge an Verlaufsdaten nicht von der Hand zu weisen.

## 2.1 Methodischer Zugang

Bei der quantitativen Erfassung von Verlaufsmustern ist zu bedenken, dass die Muster selten eine klar definierte und sich exakt wiederholende Form aufweisen, sondern vielmehr in ihrer Intensität, Frequenz und Dauer variieren. So kann die Dauer eines Nachrichtenzyklus – vom Ereignis bis zum Abflachen der Berichterstattung – eine Woche, aber genauso gut drei Wochen dauern. Darüber hinaus sind die Anzahl und Länge der Artikel von Fall zu Fall verschieden. Im Hinblick auf physiologische Daten kann beispielsweise eine kurzfristige Erregung, die sich in einer Veränderung der Hautleitfähigkeit zeigt, in Dauer und Intensität schwanken. Ebenso können Orientierungsreaktionen (OR) zwischen drei und acht Sekunden andauern (LANG 1990) und in ihrer Intensität variieren.

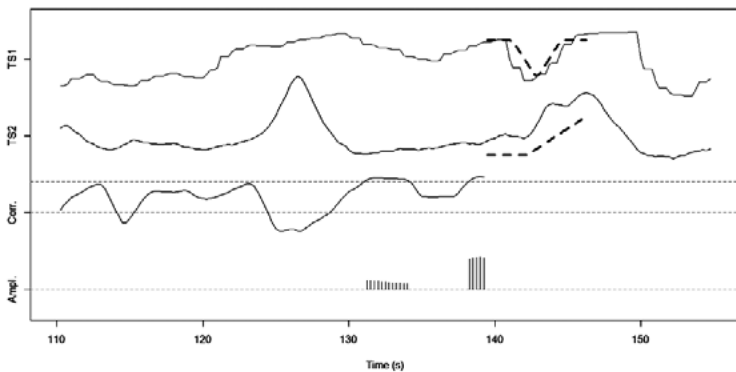
Schließlich ist in allen Fällen der Grundzustand oder die Baseline (z. B. die Anzahl von Artikeln außerhalb des Nachrichtenzyklus oder die Herzfrequenz im Ruhezustand) von Fall zu Fall verschieden. Die quantitative Erfassung dieser Muster muss also zunächst unabhängig von absoluten Werten, Frequenz und Amplituden des Musters gelingen und in der Lage sein, kürzere und längere Muster gleichermaßen zu identifizieren.

Die Grundidee der quantitativen Mustererkennung in Zeitreihendaten, die in diesem Beitrag aufgegriffen wird, basiert auf der Pearson Korrelation zwischen zwei Zeitreihen gleicher Länge, die unabhängig von Amplitude und absoluten Werten der Zeitreihen ist. Die Korrelation ist damit ein sinnvolles Maß, um prototypische Muster mit Zeitreihen unbekannter Amplitude und Varianz zu vergleichen. Voraussetzung dafür ist allerdings, dass beide Zeitreihen dieselbe Länge haben und möglichst keine fehlenden Werte aufweisen. Da die Startposition eines Musters nicht bekannt ist und seine Länge variieren kann (z. B. Lebenszyklen von Themen oder die Dauer einer elektrodermalen Reaktion), muss die Möglichkeit geschaffen werden, ein prototypisches Muster mit unterschiedlich positionierten und unterschiedlich langen Abschnitten der gemessenen Zeitreihe zu vergleichen.

Um diesem Problem zu begegnen, wird in unserem Vorschlag die Korrelation mit einer brute-force Abtastung der Verlaufsdaten kombiniert, um die Korrelation zwischen einem prototypischen Muster und den erhobenen Daten an jedem Punkt einer Verlaufskurve mit jeder theoretisch möglichen Länge zu berechnen. Konkret wird ein gleitendes Fenster mit variabler Länge verwendet, das die gesamte Zeitreihe nach Übereinstimmungen, also hohen Korrelationen, mit einem vorgegebenen Muster durchsucht. Wird eine >hohe< Korrelation der Messdaten mit dem theoretisch vorgegebenen Muster gefunden, wird die Stelle in den Verlaufsdaten markiert. Um schwaches und zufälliges Rauschen von tatsächlich vorhandenen Mustern zu unterscheiden, ist es notwendig, dass das Verfahren nicht nur die Korrelationen ausgibt, sondern weitere Informationen bereitstellt. So werden Informationen über die Streuung der Verlaufsdaten an der gefundenen Stelle sowie die Länge des Fensters, in welchem das Muster gefunden wurde, benötigt, um Aufschluss über die Intensität und die Dauer eines Musters in der Messreihe zu geben. Theoretische Annahmen zu den Verlaufsmustern sowie Informationen über zu erwartende Messfehler bei der Erhebung können dabei helfen, Mindestgrenzen für die Identifikation eines Musters zu definieren und Rauschen als solches zu erkennen.

ABBILDUNG 1

### Parallele Abtastung zweier Zeitreihen (TS1 und TS2) mit zwei prototypischen Mustern (gestrichelt)



Die schwarze Kurve (Corr.) gibt für jeden Punkt der Kurven die höchste Korrelation aller gleitenden Fenster ab diesem Punkt an. Übersteigt die Korrelation 0.8 (obere gestrichelte Linie), wird auch die Amplitude (Ampl.) an dieser Stelle ausgegeben. Die Korrelation und Amplitude werden jeweils dem Startpunkt des Musters zugewiesen.

Falls sich ein Phänomen in mehr als einer Messreihe gleichzeitig als Muster zeigt, z. B. eine Ressourcenallokation, die sich durch eine Orientierungsreaktion (identifizierbar durch ein typisches Muster des Verlaufs der Herzfrequenz, GRAHAM/CLIFTON 1966) sowie einer gleichzeitigen Veränderung der Hautleitfähigkeit äußert, kann das Verfahren auch auf mehrere Zeitreihen und mehrere theoretische Muster erweitert werden. Dabei wird für jedes mögliche Zeitfenster in den Daten die Korrelation zweier Muster mit den jeweiligen Messreihen berechnet und deren Mittelwert als Korrelation mit den parallelen Mustern ausgewiesen. Als Indikator für die Streuung der Messpunkte kann entweder der arithmetische (für ähnlich skalierte Messreihen) oder der geometrische (für unterschiedlich skalierte Messreihen) Mittelwert berechnet werden.

## 2.2 Anwendungsbeispiel: Orientierungsreaktionen während eines Filmes

Ein Anwendungsfeld, in welchem eine quantitative Erfassung von Verlaufsmustern besonders hilfreich sein kann, ist die mikropsychologisch orientierte Rezeptionsforschung. Hier werden Probanden bei der Rezeption von Medieninhalten mithilfe von Sensoren überwacht, um physiologische Parameter wie Herzfrequenz, Hautleitfähigkeit, Muskelbewegungen oder Hirnaktivitäten aufzuzeichnen. Die Verläufe der Messungen geben Aufschluss über Reaktionen der Probanden auf die Medieninhalte, die mit Befragungen nicht ohne Weiteres erfassbar sind (FAHR 2013).

Während zwar langfristige Veränderungen in Hautleitfähigkeit und Herzfrequenz qualitativ dem Augenschein nach oder quantitativ mittels Trendanalysen identifiziert werden können, gibt es auch kurzfristige – oft komplexere – Muster, deren manuelle Codierung sich insbesondere bei langen Untersuchungszeiträumen, wie sie für die Rezeptionsforschung oft typisch sind, schwieriger gestaltet. Ein Beispiel für eine solche typische – und theoretisch bekannte – Reaktion ist die Orientierungsreaktion, die bei überraschenden abrupten Veränderungen, beispielsweise in der direkten Umgebung oder im rezipierten Medieninhalt, hervorgerufen werden und sich unter anderem durch charakteristische Ausschläge im Verlauf der Herzfrequenz und der Hautleitfähigkeit bemerkbar machen. Der Organismus stellt hier kognitive Ressourcen bereit, um die Veränderung in der Umwelt zu verarbeiten (vgl. in der Übersicht FAHR 2013; POTTER/BOLLS 2012).

In der hier vorgestellten Beispielstudie wurden 245 Probanden bei der Rezeption eines Filmes von acht Minuten Länge beobachtet. Der Film zeigte eine Dokumentation mit Werbeunterbrechung, in welcher einer der Werbespots – eine Kampagnenwerbung für Safer Sex – experimentell manipuliert wurde, um fröhlich, neutral oder bedrohlich zu wirken. Neben der mittelfristigen Wirkung des Spots, die mithilfe einer Befragung nach der Rezeption erhoben wurde, interessierte bei dieser Studie auch die kurzfristige Wirkung, die sich an Orientierungsreaktionen während des Spots zeigt. Es sollte untersucht werden, auf welche Elemente des Spots die Probanden in den unterschiedlichen Gruppen besonders stark reagieren – im Sinne kognitiver Informationsverarbeitung.

### 2.2.1 Methode

#### *Datenbereinigung*

Für alle 245 Probanden wurden die Herzfrequenz und die Hautleitfähigkeit über die gesamte Zeit in einer Auflösung von 40 Hertz gemessen. Bei der Datenbereinigung wurden alle Zeitmarken in den Daten mit der Zeitmessung des Stimulus synchronisiert und anschließend Herzfrequenz sowie Hautleitfähigkeit in eine neue Zeitreihe mit einer Frequenz von 4 Hertz transformiert. Grund für die Aggregation ist die benötigte Rechenleistung der brute-force Abtastung, welche mit einer Aggregation um den Faktor 10 um das 100-fache verringert wird.

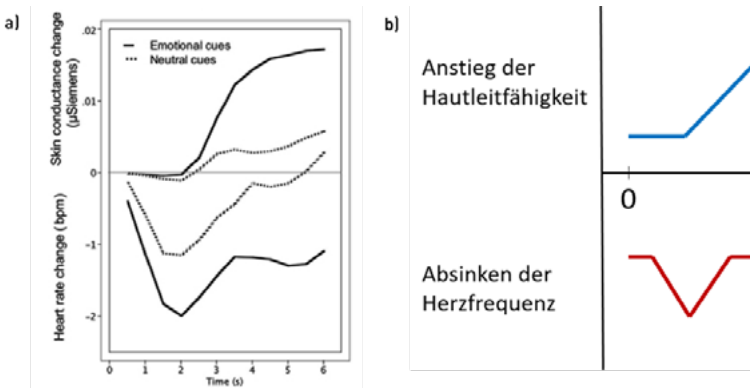
#### *Prototypische Verlaufsmuster*

Die Orientierungsreaktion bezüglich der Herzfrequenz wird üblicherweise in einem Fenster von acht Sekunden identifiziert (MACKAY-BRANDT 2011), in der Regel setzt spätestens nach drei Sekunden eine Zunahme der Hautleitfähigkeit ein (BOUCSEIN et al. 2012). Wie alle physiologischen Daten, so besitzt auch dieses spezifische Muster im Detail einen individuellen Charakter und ist von Person zu Person verschieden. Das bedeutet, dass die Dauer des Rückgangs der Herzfrequenz und das Einsetzen der Veränderung der Hautleitfähigkeit variabel sind und die Zeit der gesamten Reaktion zwischen drei und acht Sekunden betragen kann (LANG 1990). Ebenso hängt die Intensität der Reaktion (also der Anstieg der Hautleitfähigkeit

und die Verringerung der Herzfrequenz) von der Person und der Stärke des Reizes ab, der die Orientierungsreaktion auslöst.

Ausgehend von diesen bekannten Mustern und ihrer Varianz wurden zwei prototypische Muster für die Orientierungsfunktion definiert, die in den Verlaufsdaten gefunden werden sollen. Diese Muster sind für die Herzfrequenz durch ein Abfallen und anschließenden Wiederanstieg in Form einer V-Kurve und für die Hautleitfähigkeit durch einen Anstieg, der leicht verzögert zum Muster der Herzfrequenz einsetzt, repräsentiert (Abb. 2). Als Minimaldauer einer OR wurden 3 Sekunden (= 12 Messpunkte), als Maximaldauer 8 Sekunden (= 32 Messpunkte) festgelegt.

ABBILDUNG 2  
Typische Orientierungsreaktion



(a) (Quelle: Bradley/Keil/Lang 2012). Das prototypische Verlaufsmuster für die Herzfrequenz- und Hautleitfähigkeitskurven (b).

### Auswertung

Für jeden Probanden und jeden Messzeitpunkt wurden die folgenden 12 bis 32 Werte beider Datenreihen mit den prototypischen Mustern verglichen. Dazu wurde die Korrelation beider Abschnitte mit ihrem jeweiligen Muster berechnet und der Mittelwert der beiden Korrelationen als Ergebnis ausgegeben. Die höchste Korrelation, die für diesen Messzeitpunkt erreicht wurde, wurde in eine neue Datenreihe geschrieben. In zwei weitere Datenreihen wurde die Länge des Fensters, in dem diese Korrelation gefunden wurde, und das arith-

metische Mittel der Spannweite der Messwerte in diesem Fenster ausgegeben<sup>1</sup>. Dadurch entstanden, parallel zu den Messwerten für jeden Probanden, drei neue Datenreihen, an welchen das Vorkommen prototypischer Muster und deren Amplitude direkt abgelesen werden kann.

Eine Reaktion wird in dieser Anwendung dann als bedeutsam betrachtet, wenn die Korrelation zwischen Messreihe und Muster >außergewöhnlich< hoch wird. Eine erste Auswertung zeigte, dass Korrelationen nur in 2,2 Prozent der Fälle einen Wert von  $r > 0.80$  übersteigen, was diesen zu einem sinnvollen Grenzwert für eine gute Übereinstimmung zwischen Muster und Zeitreihe macht. Dieser Grenzwert, während er sich für diese Auswertung als sinnvoll erweist, kann je nach Daten und Forschungsinteresse variieren und sollte immer über deskriptive Analysen der Korrelationen ermittelt werden.

Um Messfehler und zufälliges Rauschen aus der Analyse auszuschließen, wurde zudem ein unterer Grenzwert für die Amplitude der Reaktion festgelegt. Zu erwarten wären bei einer Orientierungsreaktion ein Abfall der Herzfrequenz um 2 bis 4 Schläge pro Minute und ein Anstieg der Hautleitfähigkeit um 0.05 bis 0.1  $\mu\text{S}$ . Das geometrische Mittel dieser beiden Werte beläuft sich auf 0.3 bis 0.7. Die Messfehler liegen bei einer 40Hz-Abtastung der Herzfrequenz bei 1 Schlag pro Minute und sollten bei der Hautleitfähigkeit unter 0.01  $\mu\text{S}$  liegen. Ausschläge mit einem geometrischen Mittel von unter 0.1 sollten deswegen nicht als tatsächliche Muster interpretiert werden, sondern sind auf Rauschen zurückzuführen.

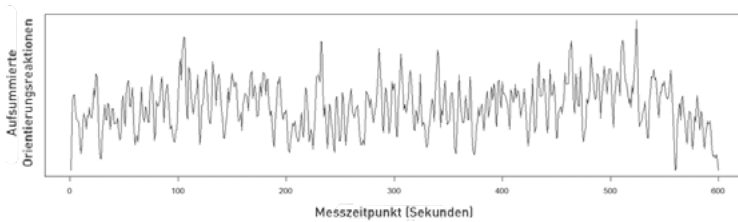
Übereinstimmungen des prototypischen Musters mit den physiologischen Verlaufsdaten wurden für alle Probanden automatisch gesucht und markiert. Anschließend wurden diese >Treffer< für alle Probanden einer Gruppe aufsummiert und zu einer neuen Zeitreihe mit einer Auflösung von 1 Hertz verdichtet. Das Ergebnis ist damit (für jede Experimentalbedingung) eine Zeitreihe, die für jede Sekunde des Stimulus angibt, bei wie vielen Probanden an dieser Stelle eine Orientierungsreaktion begann. Besonders starke Ausschläge weisen darauf hin, dass die Mehrheit der Probanden dem Medieninhalt in dieser Sekunde kognitive Verarbeitungskapazität zur Verfügung gestellt hat (Abb. 3).

1 Für die Identifikation der Muster wurde ein Python-Skript verwendet, das in Nogrod (WETTSTEIN 2018) eingebunden ist. <https://github.com/Tarlan/Nogrod>



## ABBILDUNG 3

## Aggregierte Orientierungsreaktionen aller 245 Probanden



## 2.2.2 Validierung der Ergebnisse

Ein Abgleich der Verteilung der gefundenen Orientierungsreaktionen mit den Szenen des Films zeigt, dass die theoretisch erwarteten Muster jeweils zwei bis vier Sekunden vor einem Szenenwechsel im Film und vor dem Einsetzen eines Werbespots gefunden werden. Da die Muster drei bis acht Sekunden lang sind und die niedrigste Herzfrequenz in ihrer Mitte liegt, entsprechen die gefundenen Stellen genau den erwarteten und von anderen Autoren dokumentierten Orientierungsreaktionen bei harten Schnitten (z. B. POTTER/LANG/BOLLS 2008). Die Rezipientinnen und Rezipienten nutzen diese Zeit zur Konsolidierung der Wahrnehmung sowie zur Herausbildung von Erwartungen gegenüber Neuem. Die Mustererkennung erweist sich damit als valide Methode, um Orientierungsreaktionen automatisch in physiologischen Daten zu identifizieren.

Für die Fragestellung der Studie sind aber jene Orientierungsreaktionen zentral, die während des experimentell variierten Werbespots der »Love Live«-Kampagne (BAG 2016) aufgetreten sind. Hier lassen sich zwischen den Experimentalgruppen drei (bedrohlich, fröhlich, neutral) deutliche Unterschiede festmachen. Erstens reagierten die Probanden scheinbar stark auf die musikalische Untermalung der Spots, die sich in den drei Gruppen unterschied. Es zeigen sich vermehrt OR vor dem Einsetzen eines neuen Instruments oder einer Stimme. Gleichzeitig sind in der bedrohlichen Version die OR bei Szenen mit verdeckten Gesichtern stärker ausgeprägt und in der unterhaltsamen Version die OR bei Szenen mit unbekleideten Darstellern. Der klarste Unterschied ist jedoch am Ende der Spots erkenn-

bar. Dort wurden allen Probanden dieselben Fakten zu sexuell übertragbaren Krankheiten als kurze Texte präsentiert. Probanden, die zuvor die bedrohliche Version des Spots rezipiert hatten, hatten deutlich stärkere OR beim Lesen dieser Fakten.

## 2.3 Diskussion

Die Ergebnisse weisen insgesamt darauf hin, dass OR während der Rezeption eines Filmes durch ganz unterschiedliche Reize ausgelöst werden können. Besonders verlässlich ist die Wirkung von harten Schnitten und Szenenwechseln in einem Film. Dies führt bei allen Probanden zu starken physiologischen Reaktionen. Zudem zeigt die Reaktion auf die unterschiedlichen Versionen des experimentell variierten Werbespots, dass auch die musikalische Untermalung und bestimmte Bildinhalte eine OR hervorrufen können. Die bedrohliche Stimmung, die in der zugehörigen Version erzeugt wurde, hatte eine stark aktivierende Wirkung auf die Probanden, die im Anschluss auf die eingeblendeten Texte mit einer klaren OR reagiert haben. Probanden, die einen unterhaltsamen Spot gesehen hatten, reagierten weniger ausgeprägt auf die präsentierten Informationen, was auf eine eher oberflächliche Rezeption der Informationen hinweist.

Die automatische Erkennung von Mustern in Zeitreihen hat in diesem Anwendungsbeispiel dazu beigetragen, die Reaktionen von Probanden auf Filmmaterial besser zu verstehen und nachvollziehbar zu quantifizieren. Generell erwies sich die Korrelation in gleitenden Fenstern, die in diesem Kapitel vorgestellt wurde, als valide und sinnvolle Möglichkeit, Muster in Verlaufsdaten zu identifizieren.

Wir sind der Meinung, dass sich dieses Verfahren nicht nur eignet, um in der Psychophysiologie das Auftreten von bestimmten Reaktionsmustern erkennen zu können, sondern sich darüber hinaus auch problemlos auf andere Forschungsfelder in der Kommunikationswissenschaft anwenden lässt. So wurden etwa in der Medieninhaltsforschung durch Simulationsstudien und theoretische Überlegungen prototypische Verläufe von Nachrichtenzyklen und Aufmerksamkeitsschüben skizziert (WALDHERR 2014; WETTSTEIN 2015). Auch diese Muster ließen sich mittels Korrelation in Längsschnittdaten automatisch nachweisen und charakterisieren.

### 3. Wiederkehrende Muster in Sequenzdaten

Eine weitere Form von Prozessdaten, die durch die Beobachtung von Probanden oder inhaltliche Auswertungen entstehen können, sind Sequenzdaten. Anders als bei den oben beschriebenen Verlaufsdaten wird hier nicht die Veränderung eines Merkmals über die Zeit, sondern die Abfolge von Ereignissen in einer Sequenz erfasst. In der Kommunikationswissenschaft sind solche Daten aus der Programmanalyse oder der Beschreibung der Zeitungsrezeption bekannt, wo die Abfolge von gesendeten oder rezipierten Inhalten festgehalten wird (BUCHER/SCHUMACHER 2007; HOHLFELD 1998). Da die Datenmenge bei diesen Analysen überschaubar ist, sind hier vertiefte und qualitative Analysen der Abfolgen durchaus möglich. Dies wird jedoch weniger gangbar, wenn die Datenmengen unübersichtlich groß werden. In der Kommunikationswissenschaft haben technologische Neuerungen und neue Erhebungsmethoden in den letzten Jahren für eine Zunahme solcher großen Datenmengen gesorgt. Besonders hervorzuheben sind zum einen Eye-Tracking-Erhebungen, bei denen die Blickrichtung von Probanden aufgezeichnet wird, um die Reihenfolge fixierter Medieninhalte festzuhalten, und zum anderen Tracking-Erhebungen, mit denen besuchte Websites während der Online-Nutzung aufgezeichnet werden.

Bei beiden Erhebungsmethoden entstehen Daten, die nicht nur um mehrere Größenordnungen umfangreicher sind als die Sequenz einer Zeitungslektüre oder eines TV-Programms, sondern auch weniger systematisch. Zudem zeichnen sie sich durch unbedeutende Ausrutscher wie unwillkürliche Augenbewegungen oder unbeabsichtigte Besuche von Internetseiten aus, die bei der Auswertung ignoriert werden können. Diese Eigenheiten müssen bei der Suche nach Mustern in Sequenzdaten berücksichtigt werden.

Prinzipiell gibt es bei der Mustererkennung in Sequenzdaten zwei unterschiedliche Arten von Fragestellungen: Einerseits kann nach vorbestimmten und klar definierten Mustern in der Abfolge von Ereignissen gesucht werden, andererseits kann induktiv nach unbekanntem Mustern gesucht werden, die sich häufig wiederholen. Die erste Problemstellung hat eine lange Tradition in der Bioinformatik, wo Abschnitte bekannter Sequenzen in umfangreichen Datenmengen (z.B. DNS- und Proteinsequenzen) gesucht werden. Entsprechend gibt es für diese Problemstellung eine Vielzahl von Lösungen und einsatzbereiten Hilfsmitteln. Die zweite Problemstellung ist etwas voraussetzungsreicher, da die Muster *a priori* nicht bekannt sind und in einem ersten Schritt automatisch erkannt werden

müssen. Gerade diese Identifikation von regelhaften Mustern in scheinbar unsystematischen Sequenzen ist jedoch interessant für Fragestellungen, die sich beispielsweise auf Gewohnheiten und feste Routinen in der Online-Nutzung oder auf häufig auftretende Blickpfade beim Lesen eines Medieninhalts beziehen.

### 3.1 Methodischer Zugang

Um wiederkehrende Muster in Sequenzdaten zu finden, wird in der Kommunikationswissenschaft seit einigen Jahren die T-Pattern-Analyse diskutiert (vgl. SCHWAB/UNZ 2008). Diese Analysemethode sucht nach Elementen, die unter Annahme einer Gleichverteilung über die Zeit überzufällig häufig nacheinander auftreten (vgl. SU/BRDICZKA/BEGOLE 2013; UNZ/SCHWAB 2005). Weil die Grundannahme dieser Methode aber eine gleichmäßige und zufällige Verteilung aller Elemente ist und gegen diese Nullhypothese getestet wird (MAGNUSSON 2000), eignet sie sich nur bedingt für Beobachtungsdaten menschlichen Verhaltens.

Anders als in der T-Pattern-Analyse angenommen wird, ist menschliches Verhalten nämlich vielmehr durch Repetition und schubweise Handlungen geprägt (BARABÁSI 2005). Das bedeutet, dass Menschen sich lange Zeit monoton verhalten und dann in kurzer Zeit mehrere Handlungen vollziehen, bevor die Aktivität wieder nachlässt. Diese schubweisen Aktivitäten sind zudem häufig repetitiv, also zyklisch. So kann zum Beispiel ein Online-Nutzer längere Zeit auf einer Seite verweilen, um sie zu lesen. Nach der Lektüre werden mehrere Seiten geöffnet und mehrfach zwischen Google und einzelnen Ergebnissen hin und her gewechselt, bis sich wieder eine ruhigere Phase der Lektüre einstellt. Ähnlich schweifen die Augen nach dem Lesen eines Artikels auf einer Zeitungsseite über mehrere Texte und Bilder und wechseln zwischen benachbarten Elementen, bevor der nächste Artikel gelesen wird.

Eine T-Pattern-Analyse eines solchen schubartigen und repetitiven Verhaltens führt unweigerlich dazu, dass der Wechsel zwischen Elementen, die mehrfach kurz nacheinander vorkommen, als dominantes Muster erkannt wird, da die Wahrscheinlichkeit einer schnellen Repetition in einer gleichmäßigen Verteilung extrem gering ist. Dieser Effekt wird umso prägnanter, je länger die untersuchte Sequenz ist. Bei Anwendungen mit Mediennutzungsdaten muss deswegen der zulässige Alpha-Fehler auf ex-

trem tiefe Werte (z. B.:  $p < 0.0005$ ; UNZ/SCHWAB 2005;  $p < 0.000001$ : SCHWAB/UNZ 2008) gesetzt werden, um überhaupt ein Resultat zu erhalten. Dieses weist aber nur einzelne besonders kurzfristige Repetitionen aus und ignoriert die Komplexität der restlichen Sequenz.

Als Alternative zur T-Pattern-Analyse wird daher in diesem Beitrag eine neu entwickelte und speziell auf Rezeptionsdaten angepasste Adaption der Motiv-Extraktion vorgestellt. Das zweistufige Verfahren reduziert in einem ersten Schritt die Komplexität der Sequenz, indem repetitive Phasen in der Sequenz zusammengefasst und durch einen Platzhalter ersetzt werden. In einem zweiten Schritt folgt eine Motiv-Extraktion, die an den ADIOS-Algorithmus (SOLAN et al. 2005) angelehnt ist. Der Algorithmus stammt aus der Computerlinguistik und übersetzt die Gesamtheit aller Sequenzen in einen nicht-linearen Graphen, in welchem besonders häufige und möglichst lange Routen identifiziert werden. Das Ergebnis ist eine Sammlung von Sequenzmustern, die auch den Kontext repetitiver Wechsel zwischen Elementen ausweist.

Entwickelt wurde ADIOS für die induktive Extraktion grammatikalischer Regeln aus großen Textmengen und für die Ermittlung häufiger Abfolgen von Wörtern sowie deren Rolle in einem Satz (MCFALL 2007). Mit geringfügigen Anpassungen kann der Algorithmus jedoch dazu verwendet werden, Sequenzen von Nutzungsdaten nach häufig vorkommenden und möglicherweise repetitiven Abfolgen von Elementen bzw. Ereignissen zu durchsuchen. Sind diese Teilsequenzen gefunden, lassen sie sich durch Sequenzabgleiche in den ursprünglichen Daten finden und quantifizieren. Der angepasste Algorithmus wurde in Python implementiert und ist in Nogrod (WETTSTEIN 2018) als Auswertungsmethode für Sequenzdaten verfügbar.

### 3.2 Anwendungsbeispiel: Komplexe Lesemuster in Eye-Tracking-Daten

Gerade bei Eye-Tracking-Erhebungen fallen große Mengen an Sequenzdaten für jeden Probanden an, in denen die Abfolge von Fixationen verzeichnet ist. Betrachtet ein Proband einen Medieninhalt, so schweift der Blick über mehrere Bestandteile des Inhalts, die als isolierte Bereiche, sogenannte »Areas of Interest« (AOI), definiert werden. Eine AOI kann beispielsweise ein Bild, ein Textabschnitt, ein Titel oder ein bestimmtes Wort sein. Während traditionell

eher die Verweildauer auf einer bestimmten AOI, die Zeit bis zur ersten Fixierung einer AOI oder die Anzahl von Fixationen quantitativ ausgewertet werden (HOFER/MAYERHOFER 2010; RÖTTING 2001), wird der eigentliche Prozess, der in diesen Sequenzdaten aufgezeichnet ist, meist nur qualitativ betrachtet. So kann der Blickpfad visuell nachvollzogen werden, um visuell Muster in der Abfolge für einzelne oder mehrere Probanden zu finden.

Werden jedoch mehrere Probanden über eine längere Zeit bei der Mediennutzung beobachtet, ist eine visuelle Analyse der Blickpfade kaum mehr möglich. Besonders wenn sich eine Forschungsfrage explizit auf die Blickpfade und mögliche Unterschiede zwischen Experimentalgruppen bezieht und eine objektive und reliable Identifikation wiederkehrender Muster erforderlich ist, sind automatische und quantitative Erhebungen angebracht.

### 3.2.1 Methode

Im folgenden Anwendungsbeispiel wurde Eye-Tracking in einer Studie mit 48 Probanden eingesetzt, in der diese einen Zeitungsartikel lesen sollten. Der Zeitungsartikel bestand aus fünf Textteilen (einen Titel, einen Lead, einen Text, eine Grafik und eine Box mit Meinungen zum Thema), die als AOI definiert wurden. Zudem wurde der Header des Textes mit Datum und Name der Zeitung als weitere und vom Artikel getrennte AOI abgegrenzt.

Der Artikel behandelte eine neue Technologie im Bereich der Fortpflanzungsmedizin. Je nach Stimulusgruppe nahm der Artikel entweder eine zustimmende oder ablehnende Haltung gegenüber der thematisierten Entwicklung ein. Die enthaltene Grafik veranschaulichte das Ergebnis einer repräsentativen Umfrage zur Zustimmung in der Bevölkerung, die je nach Experimentalbedingung mehrheitlich (70%) für oder gegen die Technologie ausfiel. Die in einer Box dargestellten Zitate mit Einzelmeinungen zum Thema konnten sich ebenfalls mehrheitlich (drei von vier) entweder für oder gegen die Technologie aussprechen. Die Probanden wurden in vier Experimentalgruppen eingeteilt, die mit jeweils einer der vier Versionen des Artikels konfrontiert wurden:

- Konsonanter Contra-Artikel: Der Artikel, die Statistik und die Leser sind gegen die Technik.
- Konsonanter Pro-Artikel: Der Artikel, die Statistik und die Leser sind für die Technik.

- Dissonanter Pro-Artikel: Der Artikel ist für die Technik, die Statistik und Leser sind aber dagegen.
- Dissonanter Contra-Artikel: Der Artikel ist gegen die Technik, die Statistik und die Leser sind aber dafür.

Unabhängig von der Wirkung, die ein solcher Stimulus auf die Einstellung von Probanden haben könnte, wird vermutet, dass die Rezeption je nach Art des Artikels einer anderen Dynamik folgt. Ausgehend von der Überlegung, dass Probanden bemüht sein werden, entstehende Dissonanz aufzulösen, wird bei dissonanten Texten eine intensivere und komplexere Rezeption vermutet. Dabei dürfte mehrfach zwischen den widersprüchlichen Teilen hin und her gewechselt werden. Um diese Hypothese zu testen, müssen zunächst Lesemuster in den einzelnen Gruppen erkannt und quantifiziert werden. Hierfür wurde der entwickelte Algorithmus eingesetzt.

### 3.2.2 Auswertung und Ergebnisse

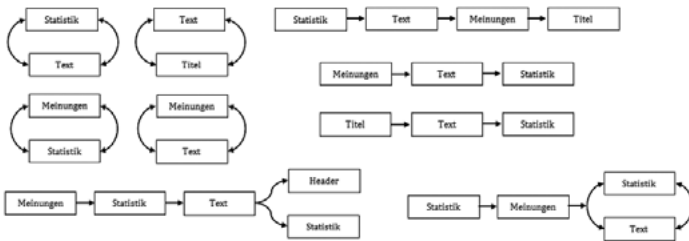
In einem ersten Schritt wurden repetitive Teilsequenzen oder n-gramme in den Sequenzen gesucht und durch Platzhalter ersetzt. Als oberer Schwellenwert wurden drei aufeinanderfolgende Elemente gesetzt, wodurch die wiederholte Fixation eines einzelnen sowie die Oszillation zwischen zwei oder drei AOI als Schlaufe markiert wurde. In der Sequenz würde damit zum Beispiel die Repetition des Musters *ABABABAB* durch den Platzhalter *(AB)\** ersetzt. Ein Platzhalter kann für eine beliebig lange Oszillation zwischen Elementen stehen, die Länge wird jedoch registriert und ausgegeben. Die gefundenen repetitiven Muster waren Bi- und Trigramme, die zwei- bis dreimal wiederholt wurden und einzelne Elemente, die bis zu 680 aufeinanderfolgende Fixationen auf sich vereinten. Für die Motiv-Extraktion wurde ADIOS mit einem Schwellenwert von  $\eta = 0.9$  verwendet. Eine Sequenz wurde also als vollständig betrachtet, wenn eine Verlängerung um das häufigste folgende Element die Auftretenswahrscheinlichkeit um den Faktor 0.9 reduzieren würde (SOLAN et al. 2005).

Die durchschnittliche Lesedauer betrug 143.7 Sekunden ( $SD=68.9$ ) und generierte eine Sequenz von 7356.5 Elementen ( $SD=3458.3$ ). Die Analyse erkannte über alle Probanden hinweg insgesamt 30 wiederkehrende Muster. Die Muster wurden in den Originalsequenzen eingesetzt, wodurch diese auf eine Länge von 11.18 ( $SD=5.77$ ) Elemente reduziert werden konnten. Die Komplexe Information von 48 Blickpfaden mit jeweils über 7.000 Fixatio-

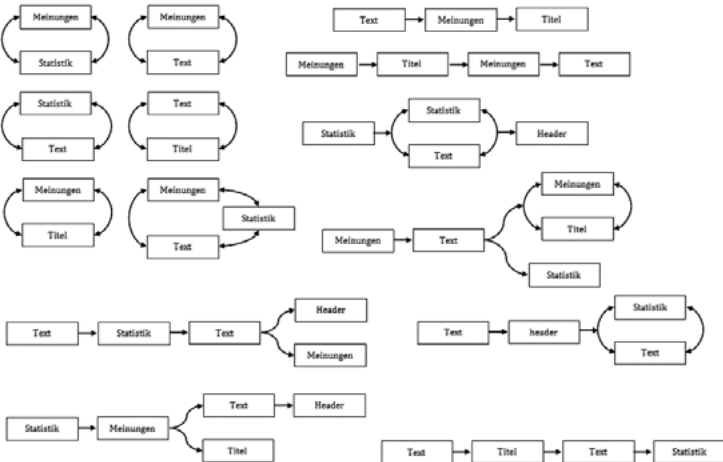
nen konnte damit auf eine Sammlung von 30 wiederkehrenden Mustern und 48 Sequenzen mit jeweils 3 bis 21 Abfolgen dieser Muster und einzelner Fixationen reduziert werden. Sowohl der Inhalt und die Komplexität der wiederkehrenden Muster als auch die einzelnen Abfolgen lassen sich nun qualitativ und quantitativ auswerten.

ABBILDUNG 4  
**Wiederkehrende Blickmuster beim Lesen dissonanter (a) und konsonanter (b) Artikeln**

**a) Dissonante Bedingung**



**b) Konsonante Bedingung**



Elemente, die mit bidirektionalen Pfeilen verbunden sind, stellen Schleifen dar, die oft repetiert wurden. Abzweigungen zeigen an, dass beide abzweigenden Pfade wahrscheinlicher waren als alle Alternativen.



Für eine inhaltliche Auswertung wurden die Probanden in die Experimentalbedingung (konsonant vs. dissonant) eingeteilt, um die Komplexität der Muster zu vergleichen. Die Hypothese, dass Probanden bei einem dissonanten Artikel stärker dazu tendieren, zwischen den Artikelteilen zu wechseln, konnte sich jedoch nicht halten. In Abbildung 4 sind die Muster für beide Experimentalbedingungen schematisch aufgeführt. Anders als angenommen traten bei der Lektüre dissonanter Texte geradlinigere und einfachere Muster auf, wogegen sich bei konsonanten Texten komplexere Muster mit mehreren Schleifen bei mehreren Probanden wiederholten. Dieser Befund wird dadurch unterstrichen, dass die Lektüre des Fließtexts in der konsonanten Bedingung öfter unterbrochen wurde und dass dort mehr repetitive Wechsel zwischen dem Text und der dargestellten Statistik verzeichnet wurden. In der dissonanten Bedingung wurden mehr aufeinanderfolgende Fixationen auf den Text, die Statistik und die Meinungen gemessen. Dissonante Artikel führten folglich nicht dazu, dass die Artikelteile abwechselnd betrachtet wurden, sondern dass der Text, die Statistik und die Meinungen konzentrierter und länger studiert wurden.

### 3.3 Diskussion

Durch die Untersuchung wiederkehrender Muster in Sequenzdaten in der Blickverlaufsstudie wurde getestet, ob die Komplexität dieser Verläufe vom Inhalt (konsonant vs. dissonant) eines textbasierten Stimulus beeinflusst wird. Wir konnten zeigen, dass sich die Gruppen in ihrem Leseverhalten unterscheiden, wenn ihnen ein Artikel mit konsonanten oder dissonanten Informationen präsentiert wird. Darüber hinaus wurde deutlich, dass sich mitunter komplexe Muster in der Rezeption wiederholen. Anders als angenommen wurden Artikel, bei denen sich die Tonalität des Artikels von den Randinformationen unterschied, konzentrierter und geradliniger gelesen – wohingegen in konsonanten Artikeln komplexere Rezeptionsmuster auftraten.

Die Motiv-Extraktion hat zudem die Komplexität der Daten deutlich reduziert und sie für inhaltliche Betrachtungen zugänglich gemacht. Aus einzelnen Blickpfaden mit jeweils über 7.000 Fixationen wurden 30, teilweise stark repetitive Muster extrahiert, um die Blickpfade auf Sequenzen von durchschnittlich elf dieser Muster zu vereinfachen. Dies ermöglicht nicht nur die qualitative Auswertung der wiederkehrenden Muster, sondern macht auch die Sequenzen für manuelle Analysen zugänglich.

Diese Methode erweist sich daher als vielversprechender Ansatz, um etwas über repetitive Muster in Sequenzdaten zu erfahren und die Komplexität dieser Daten erheblich zu reduzieren. Die Anwendung ist dabei nicht auf Eye-Tracking-Daten beschränkt, sondern kann überall eingesetzt werden, wo Sequenzen von Handlungen oder Inhalten erhoben werden. So können Tracking-Daten, die das Online-Surfverhalten oder die Smartphone-Nutzung aufzeichnen, auf wiederkehrende Nutzungsmuster untersucht werden. Gleichsam können Narrationsmuster in Nachrichtensendungen oder feste Gewohnheiten in Tagebuchstudien induktiv erfasst und ausgewertet werden.

#### 4. Allgemeine Diskussion

Die Kommunikationswissenschaft hat mit neuen Erhebungsmethoden und der zunehmenden Automatisierung von Inhaltsanalysen und Beobachtungen vermehrt Zugriff auf umfangreiche Verlaufsdaten. Dies können Tracking-Daten aus der Online-Nutzungsforschung, psychophysiologische Messreihen, Eye-Trackingprotokolle oder Ergebnisse automatischer oder manueller Längsschnittinhaltsanalysen sein. In Anbetracht des Umfangs der Daten und der Ausdifferenzierung spezifischer Forschungsfragen zu den exakten Abläufen und Prozessen erscheint es heute nicht mehr zeitgemäß, Muster in diesen Daten via ›Eyeballing‹ zu bestimmen. Vielmehr ist es hilfreich – und deutlich reliabler – empirische Methoden zu entwickeln, mit denen sich diese Prozessdaten automatisch analysieren und auf vordefinierte oder repetitive Muster durchsuchen bzw. prüfen lassen.

Im vorliegenden Beitrag wurden als Beispiel zwei Methoden zur automatischen Mustererkennung in Prozessdaten vorgestellt. Mit der biserialen Korrelation in gleitenden Fenstern lassen sich a priori bekannte – prototypische – Verlaufsmuster in Zeitreihendaten finden. Diese Methode ist also besonders dort sinnvoll, wo spezifische Muster (z. B. Orientierungsreaktionen) identifiziert und gezählt werden sollen oder wo Muster unterschiedliche Formen annehmen können (z. B. Aufmerksamkeitsschübe während der Medienrezeption), die eine automatische Identifikation notwendig machen.

Mit der Motiv-Extraktion wurde eine Methode vorgestellt, die repetitive Muster in Sequenzen entdecken kann. Im Anwendungsbeispiel wurde dieses Verfahren verwendet, um Muster in Blickpfaden einer Eye-Tracking-

Studie zu finden, sie kann jedoch auch auf Tracking-Daten, Beobachtungsdaten oder andere Zeitreihen angewandt werden.

Beide Vorgehensweisen sind nicht neu, sondern Adaptionen bereits existierender Verfahren aus anderen Wissenschaftszweigen, die sich schon länger mit feingliedrigen Längsschnittdaten auseinandersetzen. Die biseriale Korrelation etwa hat in der Archäologie mit dendrochronologischen Untersuchungen eine lange Tradition und die Motiv-Extraktion wird in der Computerlinguistik ebenfalls seit Jahren eingesetzt. Auch für die Identifikation bekannter Muster in Sequenzen wurde in diesem Beitrag auf die bereits bestehende Forschungstradition in der Bioinformatik verwiesen.

Für die Kommunikationswissenschaft eröffnen sich folglich mit neuen Datenquellen Chancen und Herausforderungen, aber auch neue Analyse-möglichkeiten. Durch die Integration von Verfahren, die in anderen Forschungsfeldern bereits seit Jahren eingesetzt werden, soll dieser Beitrag dazu animieren, über den Tellerrand der eigenen Disziplin zu blicken und sich mit dem Methodeninventar anderer Bereiche auseinanderzusetzen. In der Vergangenheit wurden bereits Simulationen magnetischer Momente in ionischen Gasen oder chemische Gleichgewichtsmodelle erfolgreich auf soziale Phänomene übertragen (CASTELLANO/FORTUNATO/LORETO 2009). Ähnlich könnten auch Modelle aus der Fluidodynamik, Meteorologie, Paläontologie oder Linguistik Probleme lösen, die sich unserem Fach erst jetzt stellen, da sich die Kommunikationswissenschaft mit neuen Datenquellen und -mengen im digitalen Zeitalter befassen muss. Die Kommunikationswissenschaft muss Computational Science nicht neu erfinden, sondern kann auf Jahrzehnten der Forschung zu Computermodellen unserer Welt aufbauen.

## Literatur

- BAG: *Informationskampagne LOVE LIFE*. 2016. <https://www.bag.admin.ch/bag/de/home/das-bag/aktuell/news/neue-love-life-praeventionskampagne.html>
- BARABÁSI, A.-L.: The origin of bursts and heavy tails in human dynamics. In: *Nature*, 435(7039), 2005, S. 207-211. <https://doi.org/10.1038/nature03459>
- BOUCSEIN, W.; D. C. FOWLES; S. GRIMNES; G. BEN-SHAKHAR; W. T. ROTH; M. E. DAWSON; D. L. FILION: Publication recommendations for

- electrodermal measurements. In: *Psychophysiology*, 49(8), 2012, S. 1017-1034. <https://doi.org/10.1111/j.1469-8986.2012.01384.x>
- BRADLEY, M. M.; A. KEIL; P. J. LANG: Orienting and emotional perception: Facilitation, attenuation, and interference. In: *Frontiers in Psychology*, 3, 2012, S. 493. <https://doi.org/10.3389/fpsyg.2012.00493>
- BUCHER, H.-J.; P. SCHUMACHER: Tabloid versus Broadsheet: Wie Zeitungsformate gelesen werden. In: *Media Perspektiven*, 10, 2007, S. 514-528
- CASTELLANO, C.; S. FORTUNATO; V. LORETO: Statistical Physics of Social Dynamics. In: *Reviews of Modern Physics*, 81(2), 2009, S. 591-646
- DOWNES, A.: Up and Down with Ecology: The »Issue-Attention Cycle«. In: *Public Interest*, 28, 1972, S. 38-50
- FAHR, A.: Physiologische Ansätze der Wirkungsmessung. In: SCHWEIGER, W.; A. FAHR (Hrsg.): *Handbuch Medienwirkungsforschung*. Wiesbaden [Springer VS] 2013, S. 601-626
- GRAHAM, F. K.; R. K. CLIFTON: Heart-rate change as a component of the orienting response. In: *Psychological Bulletin*, 65(5), 1966, S. 305-320. <http://dx.doi.org/10.1037/h0023258>
- HOFER, N.; W. MAYERHOFER: Die Blickregistrierung in der Werbewirkungsforschung: Grundlagen und Ergebnisse. In: *Der Markt*, 49(3-4), 2010, S. 143-169. <https://doi.org/10.1007/s12642-010-0039-2>
- HOHLFELD, R.: Fernsehprogrammanalyse: Formen, Einsatzmöglichkeiten und Reichweite. In: KLINGLER, W.; G. ROTERS; O. ZÖLLNER (Hrsg.): *Südwestrundfunk Schriftenreihe Medienforschung: Bd. 1. Fernsehforschung in Deutschland: Themen, Akteure, Methoden* (1. Auflage). Baden-Baden [Nomos] 1998, S. 197-224
- LANG, A.: Involuntary Attention and Physiological Arousal Evoked by Structural Features and Emotional Content in TV Commercials. In: *Communication Research*, 17(3), 1990, S. 275-299. <https://doi.org/10.1177/009365090017003001>
- MACKAY-BRANDT, A.: Orienting Response. In: KREUTZER, J. S.; J. DELUCA; B. CAPLAN (Hrsg.): *Encyclopedia of Clinical Neuropsychology*. New York/ NY [Springer Science+Business Media LLC] 2011, S. 183of. [https://doi.org/10.1007/978-0-387-79948-3\\_1317](https://doi.org/10.1007/978-0-387-79948-3_1317)
- MAGNUSSON, M. S.: Discovering hidden time patterns in behavior: T-patterns and their detection. In: *Behavior Research Methods, Instruments & Computers*, 32(1), 2000, S. 93-110

- MCFALL, J. D.: *Using the ADIOS Algorithm for Grammar Induction and Psycholinguistic Comparison*. Master Thesis, University of Athens/GA, Athens/GA, 2007
- MEITZ, T. G. K.; M. HUFF; A. ORT; A. FAHR: *Does heart rate detect event boundaries? Orienting responses as implicit measures for changes in event models*. Paper presented at the Psychonomic Society 58th Annual Meeting, Vancouver/BC, 2017
- POTTER, R. F.; P. BOLLS: *Psychophysiological Measurement and Meaning: Cognitive and Emotional Processing of Media*. [Routledge Communication Series: Taylor & Francis] 2012. <https://books.google.ch/books?id=lgPJBQAAQBAJ>
- POTTER, R. F.; A. LANG; P. D. BOLLS: Identifying Structural Features of Audio. In: *Journal of Media Psychology*, 20(4), 2008, S. 168-177. <https://doi.org/10.1027/1864-1105.20.4.168>
- RÖTTING, M.: *Parametersystematik der Augen- und Blickbewegungen für arbeitswissenschaftliche Untersuchungen*. Zugl.: Aachen, Techn. Hochsch., Diss., 2001. Schriftenreihe Rationalisierung und Humanisierung; Vol. 34. Aachen [Shaker] 2001
- SCHWAB, F.; D. UNZ: Die T-Pattern-Analyse: Eine Methode zur Untersuchung von zeitlichen Strukturen in (Ereignis-)Daten. In: MATTHES, J.; W. WIRTH; G. DASCHMANN; A. FAHR (Hrsg.): *Die Brücke zwischen Theorie und Empirie: Operationalisierung, Messung und Validierung in der Kommunikationswissenschaft*. Köln [Herbert von Halem] 2008, S. 257-270
- SOLAN, Z.; D. HORN; E. RUPPIN; S. EDELMAN: Unsupervised learning of natural languages. In: *Proceedings of the National Academy of Sciences of the United States of America*, 102(33), 2005, S. 11629-11634. <https://doi.org/10.1073/pnas.0409746102>
- SU, N. M.; O. BRDIZKA; B. BEGOLE: The Routineness of Routines: Measuring Rhythms of Media Interaction. In: *Human-Computer Interaction*, 28(4), 2013, S. 287-334. <https://doi.org/10.1080/07370024.2012.697026>
- UNZ, D.; F. SCHWAB: Viewers Viewed: Facial Expression Patterns while Watching TV News. In: ANOLLI, L.; S. DUNCAN JR.; M. S. MAGNUSSON; G. RIVA (Hrsg.): *The Hidden Structure of Interaction: From Neurons to Culture Patterns*. Amsterdam [IOS Press] 2005, S. 254-262
- WALDHERR, A.: Emergence of News Waves: A Social Simulation Approach. In: *Journal of Communication*, 64(5), 2014, S. 852-873. <https://doi.org/10.1111/jcom.12117>

- WETTSTEIN, M.: Quantitative Ursachenbestimmung medialer Aufmerksamkeitsschübe. In: *Publizistik*, 60(3), 2015, S. 325-343. <https://doi.org/10.1007/s11616-015-0238-4>
- WETTSTEIN, M.: *Nogrod 1.0: Quick Tutorial*, 2018. [https://www.ikmz.uzh.ch/dam/jcr:fd26fbd3-2b3f-4421-96f8-4c2750417551/Nogrod\\_1-1.pdf](https://www.ikmz.uzh.ch/dam/jcr:fd26fbd3-2b3f-4421-96f8-4c2750417551/Nogrod_1-1.pdf)

CHRISTIAN SCHEMER / STEFAN GEISS /  
PHILIPP MÜLLER

## Applying the Reinforcing Spirals Model to Dynamic Communication Phenomena: Conceptual and Statistical Pitfalls

*Abstract:* The reinforcing spirals model (SLATER 2007) is considered an important conceptual model in research on both media use and media effects. Researchers relied on this conceptual model to explain reciprocal effects of media use and media effects in various domains of research. Although Slater (2015) refined the model by considering boundary conditions, moderators, and alternative processes, empirical research has lagged behind this theoretical advancement. We discuss how the use of inappropriate statistical modeling techniques can produce misleading findings about the presumed occurrence of reinforcing spirals processes and what can be done about that. Conceptually, extant research has focused most on escalating spiral dynamics while such processes may occur very rarely. At the same time, research may have ignored more common dynamic processes, e.g., homeostasis, wear-out, depolarization. Therefore, we urge researchers to think about communication dynamics from different angles and to model such processes accordingly.

*Keywords:* Reinforcing Spirals Model, Media Effects, Media Use, Dynamics

### 1. Introduction

The reinforcing spirals model (RSM; SLATER 2007, 2015, 2017) has received considerable attention in communication research. Without doubt, this

model has inspired a lot of research across various domains of communication and has clearly advanced scientific understanding of the dynamic relationships between media uses and effects. Its unique contribution is to model the development of communication processes over time in a way that facilitates understanding of the underlying causes and consequences rather than only fitting the data or making accurate predictions. The principal idea is that there is a reciprocal relation over time between exposure to particular kinds of messages and cognitive, attitudinal, or behavioral variables (which are thus both outcomes and preconditions of message exposure). This approach formally integrates media effects research (which is seeking to explain change in the outcome variables through media consumption) and research on selective exposure (which is aiming to explain how message selection is influenced by a variety of exogenous factors). Selective exposure is a phenomenon that plays into almost every theory or approach in media effects research, though to varying degrees and in varying shapes. However, not all media effects approaches explicitly discuss their relationship to selective exposure.

This is where the RSM taps in. The model is not explicitly linked or limited to any single media effects theory, but is presented by Slater (2007, 2015, 2017) as a general mechanism, which resonates with the omnipresence of selective exposure across many communication effects phenomena. The only condition necessary for applying the RSM empirically is that a study focuses on media exposure and an outcome variable related to media exposure. This limits the scope of the RSM to individual-level media effects analyses predicting that exposure to media content results in media effects; the RSM adds the prediction (if the theory does not posit it itself) that the respective media effect will impact the level of exposure.

The unifying nature of the RSM across theoretical approaches may come at the price of a lack of specificity, meaning that the general idea or mechanisms of the RSM need to be implemented in a context-sensitive fashion to fit the state of the art in the theory it is applied to. It should also be noted that the RSM is first and foremost a conceptual model. This model does not tell us what statistical model should be applied to test the RSM. As a consequence, previous research relied on different statistical modeling strategies. The graphical display of the RSM – reminiscent of a cross-lagged path model – may have contributed to equating the RSM with a cross-lagged panel model or simplex model (see Fig. 1). Besides its unifying nature and its resonance with the state of the art in many theories of media effects, its



focus on spiraling or escalating processes in the outcome variable – often undesirable outcomes, such as smoking behavior, racist attitudes, or political polarization – has certainly contributed to the attractiveness of the RSM. A certain degree of sensationalism revolves around stating, e.g. that political radicalization is driven by a reinforcing spiral, even though it is certainly possible that this model predicts no change in the outcome variable at all.

The present article discusses two issues related to the RSM that we argue deserve more scholarly attention. First, not only are different statistical modeling strategies used in extant research on reinforcing spiral processes, but they also tend to produce inconsistent results: Many studies relying on *cross-lagged path analysis* provide evidence of reinforcing spirals processes, whereas much research using *growth curve modeling* fails to find supporting evidence. We argue and show that cross-lagged path analysis cannot unequivocally uncover reinforcing spirals processes while growth curve modeling can. Second and more generally, the focus on escalation processes has impeded research to look for other than escalating dynamics of communication phenomena, e.g., depolarization instead of polarization, habituation or wear-out instead of escalation, or dynamic equilibria in which two variables influence each other in a way that leads them to remain largely stable. Even if depolarization, habituation, wear-out effects and equilibria may seem less spectacular, research suggests that these phenomena may be more frequent than polarization, escalation, or radicalization (see for instance, research on self-regulatory systems, CARVER/SCHEIER 1998; DEBOEK/BOKER 2010). Therefore, the present article urges researchers to consider reinforcing spirals as special cases. Additionally, researchers need to justify why they expect these to occur (or under which conditions they expect them) and rather than other >less exciting< alternative communication dynamics. In terms of modeling, it might not be >either, or<; rather, alternative types of dynamics may be included in a single study's statistical model.

In the following, we will first review the formulation of the RSM and its basic assumptions; this provides the background for critically reviewing empirical studies as to their conceptual and statistical underpinnings. We will then have a closer look at the statistical modeling of previous research and discuss which statistical strategies should be applied in order to account for the RSM logic as it has been suggested by Slater (2007, 2015, 2017). Finally, we will outline alternative approaches for modeling communication dynamics that account for varying contexts and accom-

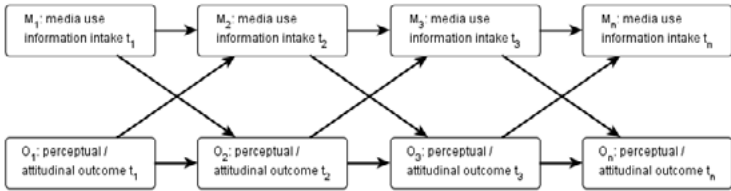
panying influences more explicitly than the predominant ways in which researchers have applied RSM. Finally, we will conclude with a number of recommendations for future empirical attempts to capture the dynamics between media exposure and different outcome variables.

## 2. Assumptions of the RSM

The RSM takes a holistic view on media uses and effects by integrating both perspectives in a dynamic process over time. Slater (2007, 2015, 2017) has repeatedly outlined that two basic premises have been instructive for the development of the model. First, individuals' media use should not only be conceptualized as an exogenous (i.e., input) variable, as media effects research typically does, but also as an endogenous (i.e., outcome) variable which is affected by a number of situational and dispositional factors. Second, the RSM adopts a dynamic perspective, i.e., individuals' media use and their beliefs, attitudes, and behaviors affect each other over time. The result is a theoretical model in which media use of an individual or a population ( $M_1$  to  $M_n$ ) and the experience of certain outcomes in this individual or in the population ( $O_1$  to  $O_n$ ) affect each other over time ( $M_{1-n} \rightarrow O_{2-O_n}$  or  $O_{1-n} \rightarrow M_2-M_n$ ; s. Fig. 1). It is important to note, that Slater (2007) conceptualizes these processes of mutual reinforcement on the individual level. That is, he assumes media exposure and outcome variables to continuously affect each other within one and the same media user. For instance, an individual who is exposed to media content changes her or his attitude in response to the media content received. This attitude, in turn, can predispose the same individual to expose herself or himself even more to the kind of media content that has shaped the attitude in the first place. An example for this is attitude polarization in response to attitudinally congruent news (BEAM/HUTCHENS/HMIELOWSKI 2018).

These general assumptions of the RSM can be read as suggesting that both media use decisions as well as beliefs, attitudes, and behaviors can reinforce each other over time and can thus fuel an escalating process. Specifically, Slater (2007) has argued that the RSM can explain polarization processes. Many empirical studies referring to RSM have interpreted the RSM logic that way (e.g., BEAM/HUTCHENS/HMIELOWSKI 2018; DVIR-GVIRSMAN et al. 2014; HUTCHENS/HMIELOWSKI/BEAM 2019; MOELLER/DE VREESE 2015; SCHEMER 2012).

FIGURE 1  
Reinforcing Spirals Model



Source: Adapted from Slater (2007: 284)

While mentioning some limits to the escalation predictions suggested by the basic outline of the model already in the early paper (SLATER 2007), these points became more salient and more elaborate as the model developed and matured in response to the empirical evidence (SLATER 2015). For instance, Slater (2007) acknowledged that reinforcing spirals would require a continuous unidirectional impact of media use on outcome variables and vice versa. Moreover, there must only be positive and no negative feedback in response to media exposure (BAUMGARTNER/JONES 2002), and no ceiling or floor effects may occur. Additionally and most importantly, the model does not allow for counter-acting influences to exist such as cross-cutting exposure or fading of topic careers in media coverage. Such processes would dampen self-reinforcing escalating processes.

Slater (2007: 288) has argued that reinforcing spirals spinning out of control can only be expected in »a perfectly closed system (free from the effects of competing social, psychological, or environmental influences)«. Just to add that »in the social world, there are always competing social, psychological, and environmental influences present«. So what the RSM actually suggests is that individuals' media use and their beliefs, attitudes, and behaviors *could* interact in a constant over-time escalation *if* individuals were no social beings embedded in a societal environment in which buffering influences almost always occur, either (a) just incidentally or (b) as a consequence of the reinforcing spiral process itself.

Slater continues to explicate which circumstances need to be given to increase the likelihood of an escalating reinforcing spirals process. He argues that (1.) the individual's need to continuously receive positive reinforcement through communication exposure is strongest when the beliefs, attitudes, and behaviors which are to be reinforced relate to individuals'

personal and social identity (SLATER 2007: 290). However, this would still allow for floor or ceiling effects to occur when a certain level of identity affirmation is reached. Then, media use on the one hand and beliefs, attitudes, and behaviors on the other would not find themselves in an escalating spiral process, but rather in a state of homeostasis, i.e., a dynamic equilibrium in which the overall level of all variables remains relatively stable through continuous mutual reinforcement over time (and if that reinforcement would discontinue, regression towards more moderate levels would result). However, (2.) external influences such as social change, conflict, or economic pressures can induce increased identity threat that might lead to a higher demand for identity reinforcement within an individual (SLATER 2015). For this to lead to an escalation of reinforcement (3.) system closure is required in addition. As long as there is an openness to communication flows from other topic areas or with attitudinally inconsistent viewpoints, or there are competing social identifications within the individual, escalation is not to be expected (SLATER 2015). That is why a true positive feedback loop can only be found in »highly atypical cases such as fundamentalists and zealots of various stripes; terrorists of course come to mind« (SLATER 2015: 375).

These constraints are important to notice. Taken together, if a reinforcing relationship between media use and other variables is to be found, these other variables will relate to social identity, and homeostasis rather than mutual escalation will most likely be the resulting pattern. Moreover, if escalation occurs it will occur only for a certain period and certain individuals. Instead of reinforcing spirals processes researchers are more likely to encounter homeostatic processes.

Several additional arguments decrease the likelihood of empirically detecting escalating reinforcing spirals. Most importantly, as Slater (2007, 2015, 2017) has repeatedly argued, there is a need for moderator variables that help distinguish individuals for whom a certain set of beliefs, attitudes, and behaviors is considered part of their personal or social identity from those for which it is not. The latter are very unlikely to enter a reinforcing spirals process in this very domain, even one that ends in homeostasis. Therefore, reinforcing spirals should only be hypothesized for specific populations or subgroups of more general populations.

Another issue refers to satisficing or time restrictions (SLATER 2015). People have only a limited amount of time for media use and these time restrictions will involuntarily make people stop using media once other

tasks (concerning school, work, family, and so on) become more attractive or more pressing. This limits the extent to which reinforcing spirals can occur resulting in natural ceiling effects that make homeostasis much more likely than escalation.

Finally, exposure to information or communication that is inconsistent with predispositions or extant attitudes can dampen or reverse self-reinforcing spirals processes (SLATER 2015). This kind of information is unlikely to produce more extreme attitudes. Conversely, attitude- or identity-inconsistent information or communication can even moderate or depolarize people's predispositions (SLATER 2015). Thus, as long as individuals are not caught in a filter bubble (PARISER 2011) in which they are exclusively exposed to attitude-consistent information, continuous depolarization or homeostasis is more likely to occur than escalating self-reinforcing spirals processes (GARRETT 2009). However, even for algorithmically organized online environments filter-bubble like information repertoires can hardly be found empirically (see, e.g., FLAXMAN et al. 2016; HAIM et al. 2018). This further supports the notion that the occurrence of escalating reinforcing spirals processes should be regarded an exception rather than the rule. Interestingly, however, most studies that refer to the seminal article by Slater (2007) focus on exactly this exceptional case of escalating reinforcing spirals processes without considering conditioning variables and without discussing other processes, such as homeostasis, depolarization, or wear-out effects. We will review them in more detail to illustrate the point.

### 3. Extant empirical research on the RSM

Previous research has investigated reinforcing spirals processes in various domains, e.g., political news use and political learning (MOELLER/DE VREESE 2015), violent media use and aggressiveness (SLATER et al. 2003), music television use and smoking (SLATER et al., 2010), sexual media use and sexual behavior (HENNESSY et al. 2009), attention to political advertisement and negative intergroup emotions (SCHEMER 2012), exposure to political violence in media and aggression and post-traumatic stress (DVIR-GVIRSMAN et al. 2014), and Facebook political information use and affective polarization (BEAM/HUTCHENS/HMIELOWSKI 2018), interpersonal discussion and attitudinal polarization (HUTCHENS/HMIELOWSKI/BEAM 2019), or media skepticism and information seeking (HUTCHENS et al. 2016). Most of these

studies are in search of escalating reinforcing spirals processes.<sup>1</sup> Some of these studies claim to provide evidence of such escalating processes. This research is presented first. Subsequent sections refer to studies that found no indication of co-escalation of processes.

For instance, Hutchens and colleagues (2019) found that affective polarization at the beginning of an election campaign increased interpersonal political discussion with like-minded people at a later point in time. This increase in attitude-consistent political discussion resulted in more affective polarization at the end of the campaign. However, there was no effect of like-minded interpersonal discussion at the beginning of the campaign on affective polarization in the mid of the campaign nor an effect of affective polarization in the mid of the campaign on interpersonal discussion at the end. In this sense, the authors found only half of a reinforcing spirals process. Another study by Hutchens et al. (2016) demonstrates that media skepticism increased information seeking over time, which, in turn, increased skepticism at a later point in time. The authors conclude that »skepticism and political information seeking may work in a reinforcing process in which each variable positively reinforces the other variable« (HUTCHENS et al. 2016: 11f.).

In a study on media violence, Dvir-Gvirsman and colleagues (2014) found that exposure to mediated violence in a first panel wave produced aggressiveness or post-traumatic stress disorder at a subsequent measurement occasion. Albeit weaker, the authors demonstrated that maladjusted people were more likely to use violent media content at a later point in time. Similar to the study by Hutchens et al. (2019) this can only be interpreted as partial evidence of reinforcing spirals process. Finally, Beam et al. (2018) found that less affective polarization in wave one increased news exposure on Facebook in wave two which, in turn, reduced affective polarization in wave three. So, this finding – a »modest over-time depolarization spiral through reading news on Facebook« (ibid.: 11) – is contrary to escalating processes in the RSM framework. Put differently, exposure to attitude-consistent communication and affective polarization are related in a homeostatic fashion. Overall, much previous research has claimed to provide evidence of reinforcing spirals processes over time. However, in

1 In the following sections, we only consider studies that relied on at least three waves of panel data to test the RSM. Having only two panel waves or cross-sectional data tells us very little about change that unfolds over time. Two-wave designs are thus by no means sufficient empirical grounds to argue within the RSM framework.

most of these studies this claim is based on significant cross-lagged paths between communication exposure and attitudinal outcomes in a cross-lagged panel model – which is not unproblematic as we will explain in the section on statistical modeling.

Other research – although in other topic domains – has failed to find evidence of reinforcing spirals processes. Möller and de Vreese (2015) found that political knowledge affected news use over time. However, there was only weak evidence for the effect of news use on learning over time. The authors concluded that this spiral is not a self-reinforcing one since intra-individual over-time change of learning is not related to intra-individual change of media use (MOELLER/DE VREESE 2015). The findings by Schemer (2012) point into a similar direction. This study demonstrated cross-lagged effects of attention to political advertising on negative intergroup emotions and selective exposure effects of intergroup emotions on attention to political advertising. However, intra-individual over-time change of advertising attention was unrelated to intra-individual over-time change of negative intergroup emotions. Similarly, a number of other studies also failed to provide evidence of correlated intra-individual change processes and concluded not being able to detect escalating reinforcing spirals processes (HENNESSY et al. 2009; SLATER/HAYES 2010; SLATER et al. 2003).

As we have summarized in the previous section, Slater (2015) has already raised a number of issues that may have caused findings that are inconsistent with the assumption of escalating reinforcing spirals processes, e.g., different topic domains, measures of variables that differ in stability, failure to consider vulnerable sub-populations, counter-acting influences, homeostasis, ceiling or floor effects. However, what is missing from this discussion is the choice of the statistical models that are used to put RSM to an empirical test. Put simply, the choice of the statistical approach that is used to model reinforcing spirals processes can also determine the findings and their interpretation. In the remainder of this article, we thus focus on what statistical models researchers use to test RSM and how this choice of models can affect the findings and conclusions that researchers draw from this evidence.

#### 4. Statistical models to test the RSM

Previous studies used different modeling strategies to study reinforcing spirals processes. Some studies relied on simplex models that analyze

cross-lagged effects over time while controlling for autoregressive effects (so-called simplex model: BEAM et al. 2018; DVIR-GVIRSMAN et al. 2014; HUTCHENS et al. 2019; HUTCHENS et al. 2016). Other researchers relied on growth curve modeling by using structural equation modeling (MOELLER/DE VREESE 2015; SCHEMER 2014) or hierarchical linear modeling (HENNESSY et al. 2009; SLATER et al. 2003; SLATER/HAYES 2010). This choice of modeling strategy is consequential for the results achieved. In hindsight, it appears as if studies that use the simplex approach have a higher likelihood of detecting traces of escalating reinforcing spirals processes over time. A likely reason for this is that simplex models confound within-person change with between-person change while growth curve models disentangle intra-individual from inter-individual processes. Therefore, empirical findings can vary as a function of the model that is chosen.

#### 4.1 Using cross-lagged panel models to test the RSM

By and large, most studies relying on simplex models claim to have found evidence or partial evidence of escalating (or de-escalating) reinforcing spirals processes over time, while research that used growth curve modeling claims that there were none. Interestingly, however, in nearly all studies conducted so far cross-lagged effects occurred. Why do some researchers interpret cross-lagged effects as evidence of self-reinforcing spirals processes and others do not? A possible explanation is that the theoretical model of RSM is unspecific with respect to the question what statistical results are an indication of reinforcing spirals processes. In Slater's (2007) seminal paper the theoretical model suggests that reinforcing spirals processes can be estimated with a cross-lagged panel model. Researchers trained in traditional social science models are quite familiar with this model. The problem however is that this model depicts relations of variables over time, i.e., cross-lagged effects. Some authors interpret cross-lagged effects in a path model as evidence of reinforcing spirals processes being at work (BEAM/HUTCHENS/HMIELOWSKI 2018; DVIR-GVIRSMAN et al. 2014; HUTCHENS/HMIELOWSKI/BEAM 2019; HUTCHENS et al. 2016). The validity of this inference hinges upon two conditions.

First, since cross-lagged panel models cannot disentangle *intra-individual change* and *inter-individual difference in change*, observed cross-lagged



effects are always a mixture of both effects. However, escalating reinforcing spirals processes refer to intra-individual processes meaning that an individual experiences a change in a perceptual, attitudinal or behavioral outcome in response to media exposure. This outcome, in turn, predisposes this specific individual to expose herself or himself to media even more, resulting in more extreme outcomes. When cross-lagged effects are a mixture of intra-individual change and inter-individual differences in change then cross-lagged effects are only valid estimates of intra-individual reinforcing spirals processes when inter-individual differences in change did not occur. This is a strong assumption that is hardly ever explicated or tested in extant research. If inter-individual differences in change occur, then cross-lagged effects represent biased estimates of reinforcing spirals processes; given the lack of convergence of findings from simplex and growth models, the bias seems to be to overestimate the extent to which escalating reinforcing spirals occur.

Second, a single cross-lagged effect, e.g.,  $M_1 \rightarrow O_2$ , represents an average period-specific shock-like effect meaning that some part of the population experiences this effect of media use at  $t_1$  on an outcome at  $t_2$ . A subsequent cross-lagged effect  $O_2 \rightarrow M_3$  is also a period-specific effect that occurs for some individuals in the population. The occurrence of subsequent cross-lagged effects can therefore mean different things. On the one hand, this chain of effects may be indicative of reinforcing spirals processes. Put differently, the same individuals who experience the first cross-lagged effect also experience the second one. However, it is also possible that some part of the population experiences the first period-specific effect and another part of the population experiences the second period effect. However, in this case we would not interpret this finding as a reinforcing spirals process since there is no mutual reinforcement at the intra-individual level. Therefore, based solely on findings from cross-lagged panel analysis, we cannot know whether the same individuals experienced the whole effects chain that is indicative of reinforcing spirals processes at the intra-individual level.

These two problems are additive in nature. Put differently, in the worst case researchers observe subsequent cross-lagged effects that are due to changes of inter-individual differences over time for different sub-populations in subsequent periods and interpret this finding erroneously as evidence of self-reinforcing spirals processes.

In reality, repeated cross-lagged effects are likely to be a mix of intra-individual change and inter-individual differences in change processes

in different periods and it makes sense to assume that there is some overlap in these sub-populations. However, the indirect effect  $M_1 \rightarrow O_2 \rightarrow M_3$  that is quantified in some studies (e.g., BEAM/HUTCHENS/HMIELOWSKI 2018; HUTCHENS/HMIELOWSKI/BEAM 2016; HUTCHENS et al. 2019) almost certainly overestimates the true reinforcing spirals process. Otherwise, inter-individual differences in change would have to be zero and overlap between sub-populations experiencing an effect in subsequent periods must be 100 percent. These are strong assumptions that researchers have not tested so far and the probability that these assumptions hold is low.<sup>2</sup>

Since cross-lagged panel models cannot disentangle intra-individual processes of change from inter-individual differences in change researchers have proposed to rely on statistical models that consider these different sources of variance (SLATER et al. 2003; Slater 2015). This is done in growth-curve models (for an overview, see SCHEMER/GEISS 2017).

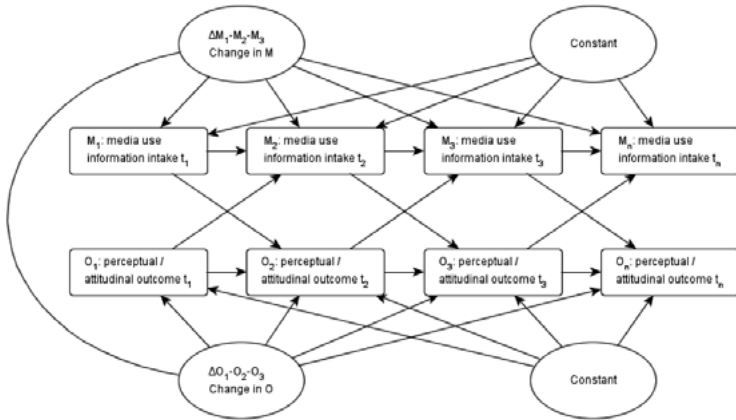
#### 4.2 Using growth-curve models to test the RSM

Slater and colleagues (2003; SLATER 2015) claim that it is necessary to distinguish between-person and within-person effects when studying reinforcing spirals processes. In order to disentangle inter-individual differences of change and intra-individual change researchers estimate one or more average growth factor(s) for the intra-individual over-time change in media use, e.g.,  $\Delta M_{1-2-3}$ , for three measurement occasions. The same is done for an outcome, e.g.,  $\Delta O_{1-2-3}$ . As can be seen in figure 2, growth curve models are an extension of cross-lagged panel models, since they include cross-lagged paths. In addition to potential cross-lagged effects, researchers can test whether growth curves of media use and outcomes are related. A positive correlation of growth factors is evidence of co-occurring processes, i.e., the growth (or decline) of media use is related to an increase (or decrease) of an outcome when the slopes of the growth factors have the same sign. In this model, the researcher clearly can infer what effects are specific to a

2 A third explanation for a cross-lagged effect is that there are omitted variables that are not modeled, but can nevertheless result in a spurious correlation between media use and any outcome variable. As this omission bias is a general bias not specific to the RSM, we do not further consider this issue for now. However, it has to be kept in mind as an additional potential source for false positive results.

period between two measurement occasions and what effects occur over time, i.e., all measurement occasions. However, a correlation between two growth factors cannot be interpreted in a causal fashion.<sup>3</sup> Put differently, we cannot say what comes first or what effect is stronger, media selection or media effect. This can only be inferred from the pattern of cross-lagged effects. The interpretation of these effects is again plagued by the problem, that it is not possible to know for sure that two subsequent period effects are indicative of intra-individual reinforcing spirals processes.

FIGURE 2  
Growth Curve Model to Study Reinforcing Spirals Processes



Note: Correlations among the constants and between constants and growth factors are omitted for the sake of simplicity, but are conventionally estimated in growth curve models. The same holds true for the correlation of disturbances at each measurement occasion.

Möller and de Vreese (2015) used latent growth curve modeling to study the mutual influences of political news media use and political learning over time. The authors found cross-lagged effects indicating that political knowledge affected news use over time. However, there was only weak

3 When the temporal order can be determined in a parallel-process model, e.g., change in media use precedes change in an outcome, then researchers are in a better position to infer causal order of variables. This, however, requires even more measurement occasions.

evidence for the effect of news use on learning. A correlation between the growth factors could not be observed ( $\Delta M_{1-2-3}$  and  $\Delta O_{1-2-3}$ ). In other words, changes in learning and changes in news use did not occur within the same individuals and were not related over time. The authors interpret the effect of political knowledge on news use from wave one to wave two in combination with the effect of news use in wave two on political knowledge in wave three as evidence of an upward spiral. However, they state that this spiral is not a self-reinforcing spiral since intra-individual over-time change of learning is not related to intra-individual change of media use. The findings by Schemer (2012) point into a similar direction. This study demonstrates cross-lagged effects of attention to political advertising on negative intergroup emotions and selective exposure effects of intergroup emotions on attention to political advertising. However, over-time change of advertising attention was unrelated to over-time change of negative intergroup emotions. The take-away from that study is also that there are no reinforcing spirals in which attention to political advertising and negative intergroup emotions grow hand in hand.

Thus, researchers relying on growth curve models produce similar findings as researchers that make use of simplex models. Yet, their interpretation differs since simplex modeling cannot disentangle intra-individual processes from inter-individual processes, but growth curve modeling can and does. It makes clear that some effects are period specific and others unfold over periods. Cross-lagged effects cannot be interpreted as evidence of reinforcing spiral processes without either testing for growth factor correlation or making the strong assumptions referred to above: That is, that there is no inter-individual change over time and that sub-populations experiencing an effect remain completely stable across measurement occasions. However, these assumptions are usually not tested. Thus, cross-lagged effects represent shock-like effects that are specific to a given period. Correlations between growth curves are evidence of parallel processes. Parallel processes, in turn, come closer to the theoretical assumptions of the RSM (SLATER 2007) that reinforcing spirals processes evolve over time and are not just period-specific.

An alternative to parallel process modeling is latent change score analysis. In latent change score analysis, growth processes are dissected into distinct change processes (for an overview, see MCARDLE 2009). This strategy is useful when a change process over time is non-continuous, for instance, if several independent change processes occur in subsequent time periods. For instance, by means of latent change models researchers can

test whether intra-individual change in media use from a first to a second measurement occasion affects intra-individual change in an outcome in a subsequent period ( $\Delta M_{1-2} \rightarrow \Delta O_{2-3}$ ). Thus, scholars can uncover whether a change process over time is driven by media use or by a specific predisposition of the media user, i.e., the outcome. Latent change models and latent growth curve models can also be combined (MCARDLE 2009). There are more modeling options that consider intra-individual change and inter-individual differences in change as sources of variance that can be useful (SCHEMER/GEISS 2017).

Taken together, the RSM as a theoretical model does not inform researchers what statistical models should be used for an empirical test. Some researchers relied on cross-lagged panel models, others on growth curve models. However, implicit assumptions of the RSM do suggest that some statistical approaches are better suited than others. Namely, RSM assumes intra-individual processes. As we have made clear in this section, simplex models can provide misleading findings in this respect because they confound intra-individual and inter-individual processes and repeated cross-lagged effects can have various meanings unrelated to reinforcing spirals. The application of statistical models that enable researchers to investigate intra-individual processes, e.g. latent growth curve models, is not only more appropriate to represent the conceptual ideas of the RSM. The use of more refined statistical models can also broaden the perspective conceptually. When escalating reinforcing spirals do not occur at the individual level researchers need to refine their theoretical models and should consider competing influences, moderating factors of reinforcing spirals processes or even alternative processes. These factors and reinforcing spirals or alternative processes need to be theorized and modeled accordingly. Thus, improving statistical and conceptual underpinnings of empirical RSM applications goes hand in hand.

## 5. Conceptual pathways towards realistic modeling of communication dynamics

The RSM had considerable impact on the field, also because the model is inclusive and provides an overarching framework that resonates well with various theoretical approaches in the research literature. Additionally, it resonates with hopes and apprehensions that particular desirable or undesirable beliefs, attitudes, or behaviors escalate and become more extreme as

media use grows more extreme or more selective as well (e.g., attitudinal polarization, echo chambers). It suggests an interesting explanation for such phenomena as excessive exposure to media portrayals of violence that lead to violent behavior, excessive exposure to extreme religious views and religious fundamentalism, excessive exposure to extreme political views and political extremism. In the literature on religious fundamentalism and religion-motivated terrorism, (self-)radicalization is a frequent topic that fits the concept of reinforcing spirals quite well (MCCAULEY/MOSKALENKO 2008; BORUM 2011).

Despite more nuanced applications, the major explanatory contribution of the RSM is that it can model the co-escalation, mutual reinforcement, or reciprocal relationship of two or more variables, one of them being media exposure. Even though it is possible, or even more plausible, to use the model for making more conservative predictions in the direction of homeostasis, de-escalation, or wear-out-effects, the use of the model in empirical studies has focused on tests of >spectacular< co-escalating processes. The more specific modeling of contingent conditions and counterforces has only infrequently been explicated and incorporated in statistical models, despite the explicit theoretical treatment of such processes in the original literature (SLATER 2015). This may be due to a lack of awareness that there are methods to handle dynamic processes other than coescalation. This may also be due to a lack of training in appropriate methods for the analysis of dynamics in communication research.

The literature review also suggests that in various contexts, data do not show co-escalation processes. And the examples provided above – violent behavior, religious and political fundamentalism – would also suggest that such processes are limited to particular sub-populations or to particular circumstances. Thus, escalating spiral processes are the exception, and not the rule. Therefore, we suggest a number of ways in which empirical studies drawing from the RSM could proceed differently than they have predominantly done in the past, to realign the statistical models used with the expanded theoretical scope of the RSM. There are several approaches to tackle the empirical and theoretical problems discussed so far and to enhance the use of RSM to make more clear-cut judgments as to whether and in which cases reinforcing spirals as individual-psychological processes can be measured empirically.

First, we should be careful in judging what the exact role of a reinforcing spirals component in a theory could be. As the model is applicable to

a wide range of theoretical approaches, the way it can be applied best will most likely vary from theory to theory. The RSM alone cannot substitute theories, but it serves as their conceptual complement that helps to predict how specific processes – according to a given theory – unfold over time or how variables – according to a given theory – interact over time. In other words, the RSM should be used and implemented in an idiosyncratic fashion that matches the theory. This also involves different building blocks on top of reinforcement in terms of escalation or non-escalation. Escalating processes can be an important ›module‹ that allows for modeling co-escalation of media use on the one hand and beliefs, attitudes, and/or behaviors on the other hand. However, there may be other ›modules‹ that model less exceptional developments or consider counter-acting mechanisms that impede, dampen or limit co-escalation processes. Such modules have been outlined before (SLATER 2015), but adding them directly to RSM may water down the concept and blur the distinction between reinforcing processes and other process components. Rather, we opt for defining additional potential model components with which an RSM module could be combined. This would also sensitize researchers to the need for defining process idiosyncratic models that fit the theory at hand rather than schematically applying the same RSM logic across different domains of media effects. Thresholds, floors, and ceilings for potential spiral processes should be important components in any conceptual application of the RSM framework.

Moreover, theories should be reviewed intensely as to whether they specify any subpopulations which are at risk for (co-)escalation, any variables that increase the susceptibility, or any variables that increase resistance or resilience against co-escalation. If, for instance, we wanted to view ›radicalization‹ processes from an RSM perspective, an implementation of RSM would certainly not predict a reinforcing spiral for everyone who is exposed to a high dose of radical messages; rather, group processes could reinforce or counteract a spiral process, socialization agencies would in many cases take action to counteract radicalization etc. For a spiral process to occur, certain psychological vulnerabilities are necessary conditions (SLATER 2015) or in other words, psychological resilience (e.g., resulting from high life satisfaction, functional social networks, high self-confidence) would need to be impeded.

Second, researchers should check their designs and data gathering. Defining the population as well as interesting subpopulations and variables that are predicted to increase or reduce susceptibility to co-escalating processes are important in field studies. After all we know about limiting factors for

the occurrence of escalating self-reinforcing spirals it does not seem advisable to look for such processes within heterogeneous populations without clearly laying out theoretical boundary conditions that allow to narrow down the search to specific subpopulations. To delve more deeply into contingent conditions, it may be worth considering experimental or quasi-experimental approaches that could lead to predictions under which conditions or in which subgroups reinforcing spirals would be expected or not.

In conjunction, intensifying theoretical and methodological rigor in empirical studies can increase the chances for plausibly demonstrating reinforcing spirals considerably – mostly independent of which data analytic approaches are used. For instance, finding a reinforcing spiral in exactly the subpopulation where it is expected but not in several other >control< groups would make the point much better than finding an average reinforcing spiral process in the whole sample; they will most likely not emerge, and one could have expected that in advance.

Third, we should check for additional components in the statistical modeling. This is the data analysis counterpart of many of the arguments outlined above: Besides an RSM module other processes would usually need to be included. In the most basic setup, this could be subgroups (e.g., to separately analyze subgroups with and without risk factors) or conditioning variables that moderate the relationship between the two processes (e.g. >turn on< or >turn off< the spiraling). In more elaborate models, this could be separate processes that also influence the outcome (but are not tied to the RSM component), or linked processes, where the reinforcing spiral process itself serves as input for the linked process.

Again going for the radicalization example, a *separate process* might be that – parallel to searching for radical content on the Internet – individuals may be exposed to radicalization literacy in schools, where a particularly charismatic teacher might succeed in detailing the intricacies and benefits of pluralism, dampening the ideological radicalization. However, note that this dampening effect would have occurred with or without the ongoing radicalization in the background, which makes it a separate process. In contrast, a *linked process* would be a reaction to the ongoing radicalization, e.g. outside interventions by parents, friends, school psychologists, etc., or intra-individual interventions such as conflicts between perceived self, ought, and/or ideal self that induce self-regulation processes. Such linked processes can cause *negative* or *positive feedback* (BAUMGARTNER/JONES 2002). Negative feedback processes result if the outcome variable deviates from a reference state,



pushing to drive it back towards the reference state. If increasing religious extremism leads to interventions by one's social environment attempting to reduce or contain one's extremism or the factors that reinforce it would be negative feedback. A positive feedback process also results if the outcome variable deviates from a reference state, but pushes it further away from the reference state. For instance, persons with increasing religious extremism may start changing their social environment, associating with other people who are religiously extreme themselves. The potential for both positive and negative feedback can recede or deplete over time, e.g. when one's social environment becomes frustrated if interventions to contain the extremization turn out to have no, little, or adverse effects.

Fourth, we should check the modeling strategy of the reinforcing spirals processes themselves. Since the statistical estimation technique determines the interpretation of findings researchers should be explicit about what statistical results are evidence of reinforcing spirals processes. Given that cross-lagged panel models cannot unequivocally uncover reinforcing spirals processes, researchers should rely on statistical models that can disentangle intra-individual change processes from inter-individual differences in change over time, for instance latent growth curve modeling, latent change score analysis, dynamic factor analysis, techniques that rely on differential equations, spectral analysis and many more (e.g., CHOW/FERRER/HSIEH 2010; BOKER/WENGER 2007). If researchers use a sequence of significant cross-lagged path coefficients as evidence of reinforcing spirals processes and others do not, then scientific advancement can become difficult. Only if it is sufficiently clear on what grounds researchers infer the occurrence of escalating reinforcing spirals processes (or their absence) in a particular study, it is possible to fairly and productively discuss the assumptions behind and limitations of the respective study.

## 6. Summary and conclusion

The present article started from the observation that many empirical studies using the RSM (SLATER 2007; 2015; 2017) are looking for co-escalating processes in which levels of exposure and behavioral, attitudinal or cognitive outcomes increase concurrently at the individual level, and not a single study has produced evidence supporting such a reinforcing spiral. A review of studies employing the RSM confirmed that impression, except

for studies using a simplex model which cannot isolate individual-level processes from inter-individual growth processes. Therefore, these studies cannot be regarded direct tests of the RSM. This is because the RSM is always arguing at the level of within-subject effects. It is, of course, also imaginable to use a similar structural pattern as the one described by the RSM to describe mutual growth processes between media exposure and cognitive or emotional outcome variables from a societal macro-perspective. However, this is not how the RSM and studies referring to it are arguing.

Looking for reasons of the widespread failure to empirically detect escalating self-reinforcing spirals processes, we encountered several possibilities which most likely all contribute to this general picture. In its theoretical formulation, the RSM includes important caveats and contingent conditions (SLATER 2007) that have been highlighted and put forth more systematically in more recent writings (SLATER 2015). Yet, empirical studies applying the RSM most frequently overlook the boundary conditions of a reinforcing spiral to occur in full effect. The fact that the label ›reinforcing spirals‹ and its basic formulation in Slater's (2007) seminal paper certainly highlight the co-escalating nature of exposure-outcome relationships, we would opt for reserving the term RSM exactly for this kind of relationship (which empirical studies have practically done anyway). Other dynamics modules (such as thresholds, ceilings and floors, separate co-processes and linked co-processes) should rather be explicitly named and bear labels which are precisely reflecting the statistical modeling that is necessary to empirically assess them. Acknowledging this, researchers should more extensively reflect whether other conceptualizations of the dynamics between media exposure and outcome variables might not be a better fit for the processes under consideration than the RSM logic. RSM needs to be regarded as what it is according to Slater (2015): a very specific case of dynamic communication processes – and certainly not the only conceivable approach to conceptualize such dynamics.

Another important and special property of the RSM is that it is not a theory dealing with any one specific outcome variable, but a conceptual model that can be used in different theoretical contexts which explicitly or implicitly assume reinforcing processes between exposure and outcomes. But this cross-cutting nature of the RSM comes at the price of the necessity to apply it in a context-sensitive idiosyncratic way in each theory. Combining it with other ›dynamics modules‹ seems a neat way to account for the idiosyncrasies of the different outcome variables and theoretical approaches.

Looking at the results so far, researchers might be disenchanted with the RSM and ready to drop it. This would be premature, given that the empirical studies to date used fairly simplified versions of the RSM, both conceptually and statistically. In fact, the RSM has moved beyond modeling reinforcing spirals, and it might be a good idea to formulate different model components (modules) that can be combined to illustrate the complex decisions involved in modeling dynamics. Interestingly, researchers using simplex models seemingly uncovered something like reinforcing spirals that may not only be due to intra-individual processes. Hence psychological explanations are not enough to account for them. There may be also inter-individual processes at work suggesting that processes like critical mass, diffusion, imitation, or contagion play a role. Accounting for this serendipity is beyond the theoretical scope of the RSM, but nevertheless an important avenue for future research.

In this research, scholars should be more explicit about what dynamics are expected. This refers to the univariate rate and shape of change of communication phenomena, e.g., change of media use or attitude polarization in campaigns. But it also refers to how the phenomena under investigation affect each other, e.g., mutual escalation, mutual dampening, whether such effects are linear or non-linear in nature, and whether they are subject to moderating influences of third variables. In order to address these more specific question, communication scholars will need more training in the statistical modeling of dynamic processes.

## References

- BAUMGARTNER, F. R.; B. D. JONES: Positive and negative feedback in politics. In: BAUMGARTNER, F. R.; B. D. JONES (Eds.): *Policy Dynamics*. Chicago [University of Chicago Press] 2002, pp. 3-28
- BEAM, M. A.; M. J. HUTCHENS; J. D. HMIELOWSKI: Facebook news and (de)polarization: reinforcing spirals in the 2016 US election. In: *Information, Communication & Society*, 21(7), 2018, pp. 940-958. doi: 10.1080/1369118X.2018.1444783
- BOKER, S. M.; M. J. WENGER (Eds.): *Data Analytic Techniques for Dynamical Systems*. Mahwah/NJ [Erlbaum] 2007

- BORUM, R.: Radicalization into violent extremism I: A review of social science theories. In: *Journal of Strategic Security*, 4(4), 2011, pp. 7-36. doi: 10.5038/1944-0472.4.4.1
- CARVER, C. S.; M. F. SCHEIER: *On the Self-Regulation of Behavior*. New York [Cambridge University Press] 1998
- CHOW, S.-M.; E. FERRER; F. HSIEH (Eds.): *Statistical Methods for Modeling Human Dynamics*. New York [Routledge] 2010
- DEBOEK, P. R.; S. M. BOKER: Unbiased, smoothing-corrected estimation of oscillators in psychology. In: CHOW, S.-M.; E. FERRER; F. HSIEH (Eds.): *Statistical Methods for Modeling Human Dynamics*. New York [Routledge] 2010, pp. 179-212
- DVIR-GVIRSMAN, S.; L. R. HUESMANN; E. F. DUBOW; S. F. LANDAU; K. SHIKAKI; P. BOXER: The effects of mediated exposure to ethnic political violence of middle East youth's subsequent post-traumatic stress symptoms and aggressive behavior. In: *Communication Research*, 41(7), 2014, pp. 961-990. doi: 10.1177/0093650213510941
- FLAXMAN, S.; S. GOEL; J. M. RAO: Filter bubbles, echo chambers, and online news consumption. In: *Public Opinion Quarterly*, 80(S1), 2016, pp. 298-320. <https://doi.org/10.1093/poq/nfw006>
- GARRETT, R. K.: Echo chambers online? Politically motivated selective exposure among Internet news users. In: *Journal of Computer-Mediated Communication*, 14, 2009, pp. 265-285. doi: 10/fsk58s
- HAIM, M.; A. GRAEFE; H.-B. BROSIUS: Burst of the filter bubble? Effects of personalization on the diversity of Google News. In: *Digital Journalism*, 6(3), 2018, pp. 330-343. <https://doi.org/10.1080/21670811.2017.1338145>
- HENNESSY, M.; A. BLEAKLEY; M. FISHBEIN; A. JORDAN: Estimating the longitudinal association between adolescent sexual behavior and exposure to sexual media content. In: *Journal of Sex Research*, 46(6), 2009, pp. 586-596. doi: 10.1080/00224490902898736
- HUTCHENS, M.; J. HMIELOWSKI; B. E. PINKLETON; M. A. BEAM: A spiral of skepticism? The relationship between citizens' involvement with campaign information to their skepticism and potential knowledge. In: *Journalism & Mass Communication Quarterly*, 93(4), 2016, pp. 1073-1090. doi: 10.1177/1077699016654439
- HUTCHENS, M. J.; J. D. HMIELOWSKI; M. A. BEAM: Reinforcing spirals of political discussion and affective polarization. In: *Communication Monographs*. (Published online) 2019. doi: 10.1080/03637751.2019.1575255

- MCARDLE, J. J.: Latent variable modeling of differences and changes with longitudinal data. In: *Annual Review of Psychology*, 60, 2009, pp. 577-605. doi: 10.1146/annurev.psych.60.110707.163612
- MCCAULEY, C.; S. MOSKALENKO: Mechanisms of political radicalization: Pathways toward terrorism. In: *Terrorism and Political Violence*, 20(3), 2008, pp. 415-433. doi: 10.1080/09546550802073367
- MOELLER, J.; C. H. DE VREESE: Spiral of political learning: The reciprocal relationship of news media use and political knowledge among adolescents. In: *Communication Research*, 2015, pp. 1-17. doi: 10.1177/0093650215605148
- PARISER, E.: *The Filter Bubble*. London [Penguin] 2011
- SCHEMER, C.: Reinforcing spirals of negative affect and selective attention to advertising in a political campaign. In: *Communication Research*, 39(3), 2012, pp. 413-434. doi: 10.1177/0093650211427141
- SCHEMER, C.; S. GEISS: Latent Growth Curve Modeling. In: MATTHES, J.; R. POTTER; C. S. DAVIS (Eds.): *International Encyclopedia of Communication Research Methods*. Hoboken/NJ [Wiley] 2017
- SLATER, M. D.: Reinforcing spirals: The mutual influences of media selectivity and media effects and their impact on individual behavior and social identity. In: *Communication Theory*, 17(2), 2007, pp. 281-303. doi: 10.1111/j.1468-2885.2007.00296.x
- SLATER, M. D.: Reinforcing spirals model: Conceptualizing the relationship between media content exposure and the development and maintenance of attitudes. In: *Media Psychology*, 18(3), 2015, pp. 370-395. doi: 10.1080/15213269.2014.897236
- SLATER, M. D.: Reinforcing spirals model. In: RÖSSLER, P. (Ed.): *The International Encyclopedia of Media Effects*. Hoboken/NJ [Wiley] 2017
- SLATER, M. D.; A. F. HAYES: The influence of music television viewership on changes in cigarette use and association with smoking peers: A social identity, reinforcing spirals perspective. In: *Communication Research*, 37(6), 2010, pp. 751-773. doi: 10.1177/0093650210375953
- SLATER, M. D.; K. L. HENRY; R. C. SWAIM; L. L. ANDERSON: Violent media content and aggressiveness in adolescents. A downward spiral model. In: *Communication Research*, 30(6), 2003, pp. 713-736. doi: 10.1177/0093650203258281
- SONG, H.; H. G. BOOMGAARDEN: Dynamic spirals put ot test: An agent-based model of reinforcing spirals between selective exposure,

- interpersonal networks, and attitude polarization. In: *Journal of Communication*, 67(2), 2017, pp. 256-281. doi: 10.1111/jcom.12288
- WANG, Z.; A. C. MOREY; J. SRIVASTAVA: Motivated selective attention during political ad processing: the dynamic interplay between emotional ad content and candidate evaluation. In: *Communication Research*, 41(1), 2014, pp. 119-156. doi: 10.1177/0093650212441793

MICHAEL SCHARKOW / MARKO BACHL

## Stable Attitudes and Behaviors as Boundary Conditions for Testing Reinforcing Spirals Models

*Abstract:* The reinforcing spirals model (RSM) has quickly become a popular way of thinking about the reciprocal dynamics of media use and effects. One key precondition for testing the model empirically is that both exposure to media content and individual attitudes vary over time – if there is no variation, there cannot be any covariation. In this paper, we discuss the preconditions for testing the predictions of the RSM regarding intra-individual variation and inter-individual differences in media use and political attitudes. We present empirical results on the temporal stability of different phenomena of interest in political communication, such as media use, campaign involvement, or candidate evaluations, drawing on panel data from two national election studies in the USA and Germany. Our results suggest that, after accounting for imperfect measurement, media use and most political orientations and attitudes are very stable over the course of a campaign, making it rather unlikely to detect patterns corresponding to the RSM predictions with meaningful effect sizes. We conclude with reflections on whether and how to refine and empirically test the RSM and similar dynamic media effects models in light of our results.

*Keywords:* Reinforcing Spirals Model, Stability, Reliability, GLES, ANES

## 1. Introduction

Many theories and models of media effects describe dynamic processes in which media selection, media exposure, and attitudes mutually influence one another over time. The reinforcing spirals model (RSM, SLATER 2007, 2015, 2017) has become a popular way of thinking about such reciprocal dynamics of media use and effects: Higher exposure to certain media messages leads to attitude change, which in turn leads to selective media use and higher exposure to similar messages, which in turn reinforces the attitude change, and so on. Moeller and de Vreese (2015), for example, investigated whether learning from the news and political news use reinforce each other over time in a sample of Dutch adolescents. The growth processes of political knowledge and news use were found to be interrelated, indicating a »spiral of political learning« (ibid.: 1). Another example is Schemer (2012), who applied the RSM reasoning to the longitudinal relationship between negative affects toward asylum seekers and attention to political advertisements in a Swiss referendum campaign.

One key precondition to empirically test the predictions of the RSM is that both exposure to media content and individual attitudes vary over time – if there is no variation, there cannot be any covariation. However, the relative stability of media use, attitudes, and behaviors is often ignored in RSM studies. Taking political communication as a prominent example, we present empirical evidence that news media use and many relevant self-report measures were exceedingly stable over the course of typical panel studies during two election campaigns in the USA and Germany. The rank-order of individuals hardly changed and the individual growth processes showed very little variation. This makes it rather unlikely to find any meaningful covariation of the variables over time, let alone evidence for the patterns, which are predicted by the RSM, with substantial effect sizes. We conclude with reflections on whether and how to refine and empirically test the predictions of the RSM and similar dynamic media effects models in light of our results.

## 2. Preconditions for empirical tests of reinforcing spirals models

Empirical tests of predictions of individual-level dynamic models such as the RSM of course require longitudinal designs with repeated measures of



the same constructs from the same individuals. A minimum of three consecutive measurements per person is necessary to establish the complete process as proposed in Slater's (2007, 2015, 2017) conceptual work. The data is most commonly analyzed with cross-lagged panel models (CLPM; see, e.g., Figure 1 in SLATER 2007) or variants of parallel latent growth-curve models (PLGCM; e.g., MOELLER/DE VREESE 2015; SCHEMER 2012; SLATER/HAYES 2010). Specific patterns of (temporal) covariation of media use and relevant attitudes or behaviors (usually multiple cross-lagged paths in CLPM and covariation of slopes in PLGCM) are taken as evidence in favor of reinforcing spirals. A necessary precondition for detecting any kind of covariation – both contemporary or over time – is, of course, that the measures or derived latent variables themselves vary at all. This concerns both CLPM and PLGCM and is especially important if we are interested in causal processes after accounting for previous measures of the same construct (i.e., its stability). If the measures of interest are highly stable, then there is just no way to empirically distinguish between homeostasis (caused by negative feedback loops which occurred in between the waves) and causal independence (no reinforcing process took place). The case under research were then unsuited to test the predictions of the RSM, because a null finding cannot be interpreted unambiguously. We therefore aim to answer two basic research questions to clarify the importance of stable attitudes and behaviors as boundary conditions in empirical tests of RSM predictions:

1. How stable were typical election-study measures when controlling for reliability?
2. How variable were within-person growth trajectories of typical election-study measures?

We took political communication as a suitable example, because the RSM is frequently tested in this domain and because high-quality data sets are publicly available.

### 3. Methods

#### 3.1 Data and measures

We conducted a secondary analysis of two established panel studies: the *2013 Short-term Campaign Panel* from the German Longitudinal Election Study (GLES13; RATTINGER et al. 2016) and the *2008-2009 Panel Study* from the Amer-

ican National Election Study (ANES08; THE AMERICAN NATIONAL ELECTION STUDIES 2010). The data from these studies are not only of very high quality, but also structurally very similar to many published studies that implicitly or explicitly refer to the reinforcing spirals model (MOELLER/DE VREESE 2015; SCHEMER 2012). In line with these studies, we selected  $k=33$  measures to test, including media use variables (television, newspaper, and Internet use, measured with the common days per week items), measures related to attitudes toward parties and candidates, and political self-assessments such as interest in politics. In order to estimate the temporal stability and variability of these measures, at least three repeated measurements are required. This led to the exclusion of some measures that were only measured once or twice during the course of the study. Where available, we selected three equally spaced pre-election waves for each measure, although in some cases unequal time lags were unavoidable. The procedure yielded three-wave panel data with measurement occasions roughly every two months for most of the variables. The average sample sizes across all measures were  $n=3,521$  (GLES13) and  $n=1,257$  (ANES08). Table 1 summarizes the selected measures. Details about design, sampling, and measurements are available from the documentations of the primary researchers, and our transformations and statistical analyses are available in the replication material.

TABLE 1  
Measures used in the secondary analysis

Variable	ANES08	GLES13
Media use	3	3
Evaluation parties	2	2
Opinion differentiation parties	1	1
Evaluation top candidates	2	2
Opinion differentiation top candidates	1	1
Evaluation government	1	1
Left-right classification top candidates	2	2
Perceived polarization top candidates	1	1
Interest in politics	1	1
Interest in the election campaign	N/A	1
Conversations about politics	1	1
Left-right self-classification	1	1

Note: Cells show the number of distinct variables selected.

### 3.2 Analysis

For every variable in the two data sets, we estimated four different measures of temporal stability and reliability. In general, one can distinguish between test-retest reliability, rank-order (or covariance) stability and mean stability over time. Test-retest reliability is commonly defined as the correlation between two repeated measures. However, using this simple correlation as a reliability measure assumes perfect stability, i.e. no true change between measurement occasions. This assumption is not only unfounded in general, but contradicts the theoretical reasoning of any RSM study. Heise (1969) developed a model that avoids this assumption and can separate reliability from stability. The Heise model requires three repeated measures to be identified. Using only the correlations between the measurements, one reliability estimate and two stability estimates can be computed using the following formulas (see ALWIN [2007] for a detailed discussion).

$$Stab_{12} = \frac{r_{13}}{r_{23}}; Stab_{23} = \frac{r_{13}}{r_{12}}; Rel = \frac{r_{12} * r_{23}}{r_{13}}$$

where  $r$  is a correlation and the subscripts denote waves in a three-wave panel survey. We present the results for the stability between waves 1 and 2 in this report and note that the results for  $Stab_{23}$  are substantially identical. The Heise (1969) model can also be specified as a path model, specifically an autoregressive model where prior values predict the subsequent measurements. This autoregressive or simplex model is the major component of the cross-lagged panel model that has historically been used to model reciprocal relationships between two measures over time. It is often assumed that the cross-lagged paths from  $X_1$  to  $Y_2$  and *vice versa* are indicative of mutual reinforcement. This, however, requires that not all variance at wave 2 is explained by the autoregressive path, i.e. that both  $X$  and  $Y$  are not perfectly stable. For example, when  $Stab_y = .90$ , even under perfect circumstances  $X$  can at most explain the 20 percent remaining variance of  $Y$ . This is true for all regression models that use lagged dependent variables, and therefore also for the CLPM.

In classical test theory, the observed score  $X$  is the sum of the true score  $\tau_X$  and random error  $e_X$ , and the correlation between consecutive measures is influenced by both the stability of the true scores and the amount of error variance. A measure with low reliability might show high variance (i.e. low stability) over time, but still cannot be explained by a causal process, because the variation over time is mainly due to random noise. In

other words, without accounting for measurement error, researchers will underestimate the stability of a measure  $Y$  and falsely conclude that there are many rank-order changes that could be explained by  $X$ . Therefore, it is necessary to separate stability and reliability, for example using the Heise (1969) model, and look at how they might influence the results of a CLPM.

Until now, we were only concerned with covariance-based models and rank-order stability. These do not consider changes in means at all, i.e. perfect rank-order stability is consistent not only with mean stability, but also with homogeneous growth or decline. Since the CLPM does not even include parameters for means, this model cannot say anything about whether, for example, interest in the campaign increases or declines over time. In terms of the RSM, it is impossible to distinguish whether there is homeostasis or a spiraling-out-of-control process over time using the CLPM. Therefore, a second model class, the latent growth curve model, has been employed to identify reinforcing spiral processes. Just like the CLPM is composed of two simplex models, the PLGCM is composed of two simple growth curve models for  $X$  and  $Y$ .

The simple growth curve model can be defined as a multilevel regression with random effects of a *time* predictor (GRIMM/RAM/ESTABROOK 2016),

$$y_{t,i} = \beta_{00} + \beta_{10} * time + u_{0i} + u_{1i} * time_{t,i} + e_{t,i}$$

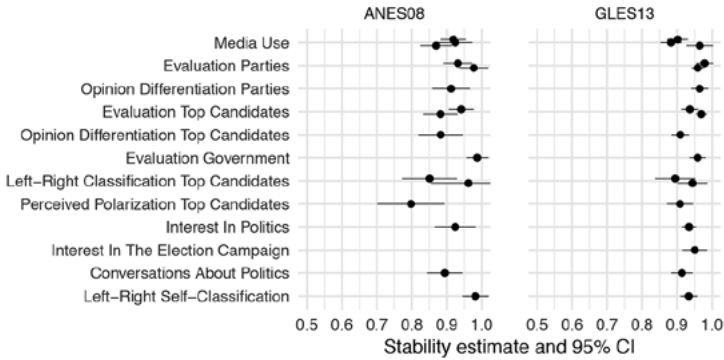
where the  $t$  subscript denotes the measurement occasion (in our model scaled in months) and the  $i$  subscript denotes the individual respondent. Since the selected measures had different response scales, we transformed them before estimating the latent growth curve model. The variables were re-scaled by units of standard deviations, so that the effect sizes for the time predictor can roughly be understood as Cohen's  $d$  growth per month (FEINGOLD 2009). Although absolute growth or decline of the selected measures over time might be of interest for applied research, we are mainly interested in the  $u_{1i}$  term, which quantifies the between-person variation in within-person growth. A relationship between the growth curves in a PLGCM can only be established if the individuals differ in their trajectories. In other words, if there is no variation in growth curves, there is no covariation to observe and explain.

In total, we estimated one simplex model and one latent growth curve model for every selected measure, which yielded four estimates plus their standard errors for our analysis: (1) rank-order stability, (2) reliability, (3) linear growth, and (4) variance in linear growth. We used R (R CORE TEAM 2016), the *lme4* (BATES et al. 2017) and *lavaan* (ROSSEEL 2017) packages for estimation, and *ggplot2* (WICKHAM 2009) for graphical summaries.

4. Results

We first looked at the rank-order stability and test-retest reliability estimates from the Heise (1969) model. The rank order of respondents in both studies was exceedingly stable for most of the attitudes and behaviors we analyzed. The average rank-order stability estimates for the ANES 2008 were  $M = 0.91$  ( $SD = 0.05$ ), for the GLES 2013 the stability was even higher with  $M = 0.94$  ( $SD = 0.03$ ). There were few meaningful differences for media use, political attitudes, and political self-assessments (see Figure 1). Only the assessments of the political positions of the candidates were less stable. This could reflect actual changes in candidates' positions or at least noticeable swings in their perception during the course of a campaign. Overall, a researcher trying to model reinforcing spirals of political attitudes and media exposure with a CLPM would find relatively little variance (about 20 percent on average) left to explain after accounting for rank-order stability and measurement error.

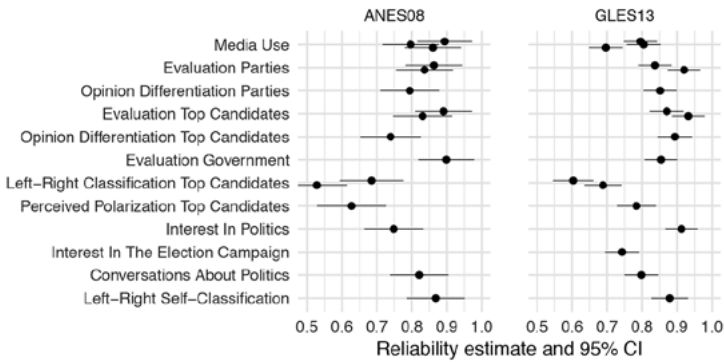
FIGURE 1  
Rank-order stability estimates



Looking at the reliability coefficients from the Heise model in Figure 2, we obtained strongly varying test-retest reliability estimates for both the ANES 2008 ( $M = 0.79$ ,  $SD = 0.10$ ), and the GLES 2013 ( $M = 0.82$ ,  $SD = 0.09$ ). Most media use measures and self-assessments were of sufficient reliability. In contrast, the questions about candidate positions and perceived political

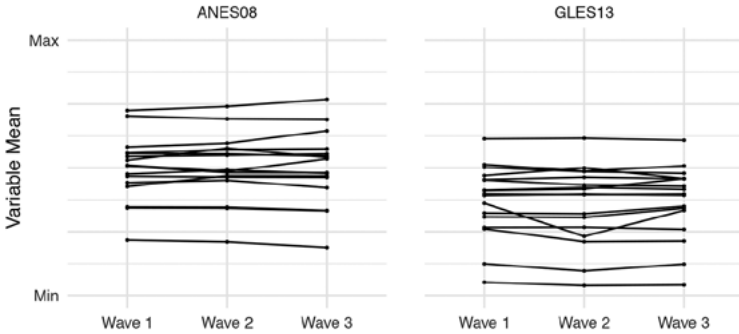
polarization showed greater shares of measurement error. The combination of high stability and imperfect reliability poses serious problems for an adequate application of the CLPM. If one fails to account for measurement error, the magnitude of the autoregressive path will be underestimated, which incorrectly implies that there is sufficient variation left to explain with cross-lagged paths from other predictors. However, such models would mainly try to fit measurement noise instead of true effects, quite likely leading to either null findings or false positive results.

FIGURE 2  
Test-retest reliability estimates



Given the high stability and moderate-to-high reliability estimates from the Heise model, there seems to be little reason to expect notable media or selection effects in CLP models of political attitudes. Following Schemer, Matthes, and Wirth (2009), it can be useful to look at individual growth trajectories in the context of media effects research. However, our results show that not only the overall rank-order was very stable, there was also little indication of mean variation over time. Figure 3 shows descriptively that the overall levels of almost all variables remained largely unchanged over the course of the campaigns. Even average campaign interest or perceived polarization remained stable over time, therefore homogeneous growth or decline as cause for high rank-order stability can be ruled out. By and large, media use and political attitudes in both campaign studies were rather constant over time.

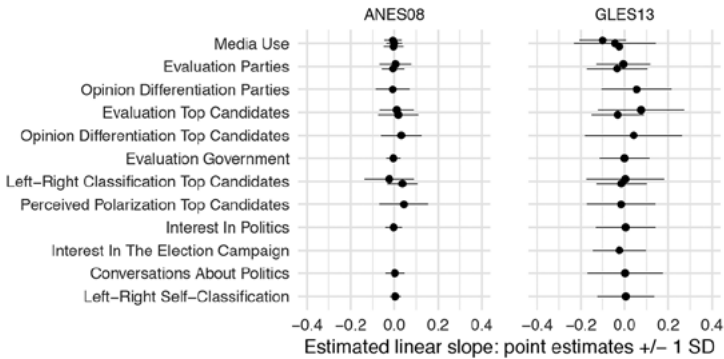
**FIGURE 3**  
**Differences in retrospective accessibility between content types and parties**



Notes: All variables were rescaled for the plot so that 0 is the empirical minimum and 1 is the empirical maximum of each variable across all waves.

This finding is supported by the near-zero population-level point estimates of the growth curves' slopes (Figure 4). The overall average slope was 0. The largest estimated linear growth was  $d=0.08$  per month, the strongest decrease was  $d=-0.10$ , both in the GLES 2013 study. However, there were no systematic differences between the different types of variables. Moreover, there was very little variation in these slopes, as indicated by the width of the intervals for each measure. The standard deviation of the individual slopes rarely exceeded  $d=0.1$  in the GLES 2013 or  $d=0.2$  in the ANES 2008. The aggregate stability was mostly not the result of changes in opposing directions which cancel each other out, but of rather homogeneously flat individual trajectories. The lack of variation was stronger in the ANES 2008 study than in the GLES 2013. Moreover, the relatively more varying slopes in perceived candidate positions are also related to the more pronounced measurement error in these variables (see Figure 2). If the observed variance is inflated per measurement occasion, it is likely that the variance in the observed changes will also be inflated. Considering the relative lack of inter-individual variation in intra-individual trajectories, most PLGCM involving the measures we investigated will therefore likely coincide with weakly correlated processes. And even if one does find substantial correlations between two growth curves, the underlying pattern will likely turn out to be parallel flat trajectories.

FIGURE 4  
**Differences in retrospective accessibility between content types and parties**



Notes: The horizontal lines depict a range of +/- 1 standard deviations of the individual slope estimates around the population point estimate. The individual slopes of approximately two thirds of the respondents fall in this range.

## 5. Discussion

Many typical election study measures (including typical measures of news media use) turned out to be exceedingly stable. Typical panel designs with representative samples of the electorate thus do not appear to be well suited to test the predictions of the RSM, at least if we expect to find covariations and changes with meaningful effect sizes. This applied to both studies and irrespective of whether we follow the analytical logic of the CLPM or the PLGCM. The results have relevant consequences for future research: They highlight the need for making the empirical preconditions explicit when making predictions based on RSM reasoning (see also SLATER 2015), and to consider all sources of variation when looking at possible reciprocal effects. Our first recommendation is that communication scholars pay more attention to measurement, e.g. by using more reliable measures (scales instead of single item measures), estimating the reliability of these measures whenever possible, and use statistical models that explicitly account for measurement error (BACHL/SCHARKOW 2019). In most cases, this will lead to higher stability estimates, and therefore less variance to explain, but it



also makes it more likely to detect media or selection effects at all and with more realistic effect sizes.

Concerning the fundamental questions of appropriate research designs to test the RSM, Slater (2007) prominently noted that »[c]learly, the longitudinal design must fit the phenomenon under study« (ibid.: 286; see also SLATER 2015). As we have demonstrated, the typical election campaign study may not be ideal to uncover reinforcing spirals, for several reasons: Media use, e.g. the number of days per week where one watches the news, does not vary much month-to-month. In order to uncover cumulative effects of media exposure, longer time lags (in the order of years rather than weeks or months) may be necessary. This takes us outside the realm of campaign effects into mid- to long-term cultivation or socialization processes (KRUIKEMEIER/SHEHATA 2017). Alternatively, researchers could design studies with shorter time lags (e.g. daily or even more frequent measurements) to capture short-term effects that are more in line with previous experimental research (OTTO/THOMAS/MAIER 2018). While the amount of daily media exposure might not vary much, either, there will likely be more variation over time when researchers consider (actual or probable) exposure to specific content rather than general media use. Such linkage analyses (SCHARKOW/BACHL 2017) also match Slater's (2015) theoretical notion that it is exposure to media content, rather than general media use, which drives the spiraling process. An interesting addition might be to take an even more detailed perspective by looking into message processing instead of media use or message exposure.

Another obstacle for empirical tests of the RSM's predictions is that not only media use, but also the orientations, attitudes, or behaviors of the respondents need to vary over time. As we have shown, most political orientations and attitudes do not fulfill this condition, at least in the span of a typical election campaign. It might therefore be more promising to apply the RSM to conceptually less stable constructs (e.g., perceptions or emotions) rather than trait-like characteristics such as political interest or party preferences. In addition or alternatively, the RSM can be used to make predictions about populations other than typical voters, e.g. children and adolescents (MOELLER/DE VREESE 2015), or move to more volatile contexts, e.g. referendums (SCHEMER 2012), or political systems. However, it can be debated whether moving the research focus to presumably more dynamic environments is a useful endeavor when much political media effects research is explicitly or implicitly legitimized by the political importance

of elections and election campaigns, which often display rather little dynamics. The question of appropriate constructs of interest may be answered differently for other fields of study, where the RSM is applied, such as media effects in public health (e.g., SLATER/HAYES 2010) or youths' sexual socialization (e.g., PETER/VALKENBERG 2009). Our empirical results may not transfer to these domains, yet our basic concerns of strong stability should be considered there as well.

Finally, and beyond the empirical scope of the present study, we want to note that it is necessary to explicitly specify an appropriate statistical model according to the *theoretical* predictions which one derived from the RSM, regardless of the topic under study. We have addressed problems which follow from two different kinds of stability in the measures under study for the two most common statistical translations of the conceptual RSM, the CLPM and the PLGCM. Both statistical model specifications have been applied with comparatively little justification for choosing one over the other (or even third options). However, since both models make strong assumptions about the causal processes under investigation, and often answer distinctly different questions, it seems prudent to explicitly state which parameters in a statistical model supposedly match the spiral process, and which the reinforcing effects. Moreover, most of the original formulations of the RSM (SLATER 2007, 2015) suggest that the processes under investigation are meant to be intra-individual rather than inter-individual. However, as pointed out by Baumgartner et al. (2017), the CLPMS

»do not disaggregate between-person differences from within-person changes over time. Therefore, common CLPM do not allow one to draw conclusions about individual change over time. If a cross-lagged path is found in a traditional CLPM, it is thus not clear whether this effect is due to longitudinal changes on the individual level, due to changes on the between-person level, or due to both types of effects« (ibid.: 11).

The same argument holds for the PLGCM, since only the covariance of individual trajectories is modeled, which is a between-person estimate. Recently, Hamaker, Kuiper, and Grasman (2015) suggested a modified random intercept CLPM that accounts for both within- and between-person effects, with the cross-lagged effects on the within-person level. We agree with Baumgartner et al. (2017) that the RI-CLPM currently seems the best statistical model to empirically test intra-individual predictions of the RSM, and their results seem to deliver on that promise in terms of model fit and interpretability. However, only few studies in communication research have

used this model, so we have to wait whether this solves the statistical issues in reinforcing spirals research. At the very least, it will force researchers to think about and explicitly state what exactly these mutually reinforcing effects are and how they can be investigated empirically.

## References

- ALWIN, D. F.: *Margins of error: A study of reliability in survey measurement*. Hoboken [Wiley] 2007
- BACHL, M.; M. SCHARKOW: Some suggestions on dealing with measurement error in linkage analyses. In: PETER, C.; T. NAAB; R. KÜHNE (Eds.): *Measuring media use and exposure: Recent developments and challenges*. Köln [Herbert von Halem] 2019, pp. 175-195. <https://doi.org/10.17605/OSF.IO/C5UTA>
- BATES, D.; M. MAECHLER; B. BOLKER; S. WALKER: *Lme4: Linear mixed-effects models using eigen and s4*. 2017. <https://CRAN.R-project.org/package=lme4>
- BAUMGARTNER, S. E.; W. A. VAN DER SCHUUR; J. S. LEMMENS; F. TE POEL: The relationship between media multitasking and attention problems in adolescents: Results of two longitudinal studies. In: *Human Communication Research*, 44(1), 2017, pp. 3-30. doi:10.1111/hcre.12111
- FEINGOLD, A.: Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. In: *Psychological Methods*, 14(1), 2009, pp. 43-53. doi:10.1037/a0014699
- GRIMM, K. J.; N. RAM; R. ESTABROOK: *Growth modeling: Structural equation and multilevel modeling approaches*. New York [Guilford] 2016
- HAMAKER, E. L.; R. M. KUIPER; R. P. GRASMAN: A critique of the cross-lagged panel model. In: *Psychological Methods*, 20(1), 2015, p. 102. doi:10.1037/a0038889
- HEISE, D. R.: Separating reliability and stability in test-retest correlation. In: *American Sociological Review*, 34(1), 1969, pp. 93-101. doi:10.2307/2092790
- KRUIKEMEIER, S.; A. SHEHATA: News media use and political engagement among adolescents: An analysis of virtuous circles using panel data. In: *Political Communication*, 34(2), 2017, pp. 221-242. doi:10.1080/10584609.2016.1174760

- MOELLER, J.; C. DE VREESE: Spiral of political learning. In: *Communication Research*. 2015. doi:10.1177/0093650215605148
- OTTO, L.; F. THOMAS; M. MAIER: Everyday dynamics of media skepticism and credibility. In: OTTO, K.; A. KÖHLER (Eds.): *Trust in media and journalism*. Wiesbaden [Springer vs] 2018, pp. 111-133
- PETER, J.; P. VALKENBERG: Adolescents' exposure to sexually explicit internet material and notions of women as sex objects: Assessing causality and underlying processes. In: *Journal of Communication*, 59(3), 2009, pp. 407-433. doi:10.1111/j.1460-2466.2009.01422.x
- R CORE TEAM: *R: A language and environment for statistical computing*. Vienna/Austria [R Foundation for Statistical Computing] 2016. <https://www.R-project.org/>
- RATTINGER, H.; S. ROSSTEUTSCHER; R. SCHMITT-BECK; B. WESSELS; C. WOLF; T. PLISCHKE; E. WIEGAND: *Short-term Campaign Panel 2013 (GLES)*. GESIS Data Archive, 2016. doi:10.4232/1.12561
- ROSSEEL, Y.: *Lavaan: Latent variable analysis*. 2017. <https://CRAN.R-project.org/package=lavaan>
- SCHARKOW, M.; M. BACHL: How measurement error in content analysis and self-reported media use leads to minimal media effect findings in linkage analyses: A simulation study. In: *Political Communication*, 34(3), 2017, pp. 323-343. doi:10.1080/10584609.2016.1235640
- SCHEMER, C.: Reinforcing spirals of negative affects and selective attention to advertising in a political campaign. In: *Communication Research*, 39(3), 2012, pp. 413-434. doi:10.1177/0093650211427141
- SCHEMER, C.; J. MATTHES; W. WIRTH: Applying latent growth models to the analysis of media effects. In: *Journal of Media Psychology*, 21(2), 2009, pp. 85-89. doi:10.1027/1864-1105.21.2.85
- SLATER, M. D.: Reinforcing spirals: The mutual influence of media selectivity and media effects and their impact on individual behavior and social identity. In: *Communication Theory*, 17(3), 2007, pp. 281-303. doi:10.1111/j.1468-2885.2007.00296.x
- SLATER, M. D.: Reinforcing spirals model: Conceptualizing the relationship between media content exposure and the development and maintenance of attitudes. In: *Media Psychology*, 18(3), 2015, pp. 370-395. doi:10.1080/15213269.2014.897236
- SLATER, M. D.: Reinforcing spirals model. In: RÖSSLER, P.; C. A. HOFFNER; L. VAN ZOONEN (Eds.): *The International Encyclopedia of Media Effects*. Hoboken [Wiley] 2017

SLATER, M. D.; A. F. HAYES: The influence of youth music television viewership on changes in cigarette use and association with smoking peers: A social identity, reinforcing spirals perspective. In: *Communication Research*, 37(6), 2010, pp. 751-773. doi:10.1177/0093650210375953

THE AMERICAN NATIONAL ELECTION STUDIES: 2008-2009 *Panel Study*. Stanford University and the University of Michigan. 2010. www.electionstudies.org

WICKHAM, H.: *Ggplot2: Elegant graphics for data analysis*. New York [Springer] 2009

LUKAS P. OTTO / FABIAN THOMAS

## Experience Sampling und multivariate Wachstumsmodelle zur Erfassung kurzfristiger selbstverstärkender Dynamiken in der Kommunikationswissenschaft

*Abstract:* Innerhalb dieses Kapitels gehen wir auf die Erhebung und Analyse von kurzfristigen dynamischen Prozessen innerhalb der Kommunikationswissenschaft ein. Wir übertragen dabei das Reinforcing-Spirals-Model von Slater (2007), das einen reziproken und selbstverstärkenden Prozess zwischen zwei Variablen annimmt, auf die tägliche Nachrichtenrezeption. Der Schwerpunkt des Kapitels liegt einerseits auf der Vorstellung der Experience-Sampling-Methode zur Erfassung von kurzfristigen Dynamiken, andererseits in der Vorstellung einer geeigneten Analysemethode zur Modellierung dieser Daten mittels Multilevel-Modellen. Als Anwendungsbeispiel wird der reziproke, selbstverstärkende Zusammenhang zwischen emotionalen Reaktionen auf politische Information in den Nachrichten und der selbstberichteten Aufmerksamkeit gegenüber Nachrichten untersucht. In einer Experience-Sampling-Studie berichteten die Probanden ihre emotionalen Zustände und Aufmerksamkeit gegenüber politischer Information direkt nach der Rezeption dieser Nachrichten. Ein solches Längsschnittdesign erlaubt eine genauere Erfassung und detaillierte Analyse von kurzfristigen Mustern und Dynamiken als klassische Panelstudien. Die Ergebnisse weisen auf die erwarteten Zusammenhänge zwischen den Emotionen ›Angst‹, ›Wut‹, ›Freude‹ und der Aufmerksamkeit hin, allerdings zeigen sich keine zeitverzögerten Effekte oder Spiralprozesse. Der Beitrag schließt mit einer Diskussion der Anwendbarkeit von Experience

Sampling und dem Reinforcing-Spirals-Model zur Erfassung und Modellierung kurzfristiger dynamischer Prozesse.

*Schlüsselbegriffe:* Experience-Sampling-Methode, Reinforcing-Spirals-Model, Nachrichtenrezeption, Emotionen, Aufmerksamkeit

## 1. Einleitung

Dynamische Prozesse, insbesondere selbstverstärkende Prozesse, sind in vielen Theorien, Ansätzen und Modellen der Kommunikationswissenschaft enthalten (SCHNEIDER/OTTO/BARTSCH 2017). Dabei kann grob zwischen inhaltlichen Theorien, also solchen Theorien, die den dynamischen Zusammenhang zwischen zwei (oder mehreren) bestimmten Variablen beschreiben, und generellen Theorien und Modellen, die eine Dynamik beschreiben, die sich auf unterschiedliche Variablen, Felder und Domänen der Kommunikationswissenschaft übertragen lassen, unterschieden werden. So beschreibt beispielsweise die Theorie der Schweigespirale die reziproke Verstärkung von wahrgenommener öffentlicher Meinung und der Bereitschaft, sich öffentlich zu äußern (NOELLE-NEUMANN 1974). Auch das Phänomen der Polarisierung in der politischen Kommunikation umschreibt einen selbstverstärkenden dynamischen Prozess aus selektiver Mediennutzung und der Verstärkung von politischen Einstellungen und Ideologie, was wiederum zur selektiven Zuwendung zu politischen Medieninhalten führen kann (BEAM/HUTCHENS/HMIELOWSKI 2018; STROUD 2010). Diese Theorien beschränken sich, wie bereits angedeutet, auf den Zusammenhang zwischen zwei oder mehreren bestimmten Variablen und beschreiben den zeitlichen und kausalen Zusammenhang dieser Variablen.

Neben diesen spezifischen, inhaltlichen Modellen gibt es in der Kommunikationswissenschaft auch die Tradition, (mediale) Kommunikation allgemein als dynamischen Prozess zu beschreiben. Der dynamisch-transaktionale Ansatz (DTA) von Früh und Schönbach (1982) stellt ein solches »allgemeines Kommunikationsparadigma« (FRÜH/SCHÖNBACH 2005: 6) dar und beschreibt einerseits einen Feedbackprozess zwischen Kommunikator und Rezipient (Inter-Transaktionen) und zum anderen einen dynamischen Zusammenhang zwischen Wissen und Motivation des Rezipienten (Intra-Transaktionen). Der DTA ist dabei unabhängig von einer bestimmten Domäne, von bestimmten Variablen oder Prozessen in der

Kommunikationswissenschaft zu sehen und beschreibt allgemeine Kommunikationsprozesse. Ebenso kann das Modell selbstverstärkender Spiralen, das Reinforcing-Spirals-Model (RSM, SLATER 2007, 2015), als allgemeine Beschreibung des Zusammenhangs zwischen Mediennutzung oder -selektion und Medienwirkung gesehen werden. Dieses Modell steht im Fokus dieses Kapitels und soll – im Gegensatz zur bisherigen Forschung – auf kurzfristige Dynamiken angewendet werden.

Das RSM beschreibt einen ›Spiralprozess‹ zwischen zwei Variablen – meist die Medienselektion und Medienwirkung. Diese Spirale ist eine Kausalkette, bei dem sich Mediennutzung bzw. -selektion und Medienwirkungen bedingen und gegenseitig verstärken oder abmildern, sodass eine selbstverstärkende negative oder positive Feedbackschleife entsteht (siehe auch SLATER/HAYES 2010; SLATER et al. 2003). Die oben beschriebenen Schweigespiral- oder Polarisierungsprozesse sind prominente Beispiele für solche selbstverstärkende Spiralen.

In diesem Kapitel stellen wir zunächst die Grundannahmen des RSM dar und zeigen anhand der bisherigen Forschung (methodische) Herausforderungen und offene Fragen in der Erforschung selbstverstärkender Spiralen. Wir gehen dabei insbesondere auf Schwächen der bisherigen Forschung ein und zeigen, wie kurzfristige dynamische Zusammenhänge einerseits gemessen und andererseits statistisch modelliert werden können. Genauer zeigen wir, wie die Experience-Sampling-Methode (ESM) genutzt werden kann, um die alltägliche Mediennutzung, aber auch Rezipientenreaktionen auf diese Mediennutzung zu erfassen.

## 2. Das Reinforcing-Spirals-Model – Grundzüge und Voraussetzungen

Innerhalb der Kommunikationswissenschaft wurden – historisch betrachtet – meist entweder Medienwirkungsprozesse (innerhalb des Paradigmas der starken Medien und passiven Rezipienten) oder Selektionsprozesse auf Seiten des Rezipienten erfasst (im Sinne der wirkungsschwachen Medien und aktiven Rezipienten). Im Gegensatz zu diesen Traditionen geht das RSM davon aus, dass Medienselektion und Medienwirkung gleichzeitig stattfinden können und sich gegenseitig bedingen (SLATER 2007). Anders ausgedrückt wird nicht mehr ein einseitiger, sondern ein wechselseitiger Kausalprozess angenommen – »causality runs both ways« (OTTO/THOMAS/



MAIER 2018: 119). Diese selbstverstärkende Spirale lässt sich anschaulich mit dem Beispiel der selektiven Nutzung politischer Nachrichten und politischen Einstellungen beschreiben. Viele Autoren zeigen, dass sich Rezipienten eher einstellungskongruenten Medieninhalten aussetzen (HART et al. 2009), gleichzeitig können Medieninhalte politische Einstellungen verstärken und polarisieren (PRIOR 2013; STROUD 2008). Die aus der Medienrezeption entstandene extremere, stärkere politische Einstellung kann nun erneut zu selektiver Zuwendung zu politischen Inhalten führen. Dieser selbstverstärkende Prozess ist – zumindest im amerikanischen Kontext – für bestimmte Bevölkerungsgruppen nachweisbar (STROUD 2008, 2010). Eine solche reziproke, selbstverstärkende Beziehung aus Medienwahl und Medienwirkung lässt sich auf viele andere Fragestellungen übertragen.

## 2.1 Voraussetzungen des RSM

Das Beispiel der Polarisierung verdeutlicht drei (methodische) Voraussetzungen, die gegeben sein müssen, damit von einem selbstverstärkenden Spiralprozess, also einer Reinforcing Spiral im Sinne des Modells, gesprochen werden kann<sup>1</sup> (siehe dazu auch SLATER 2007, 2015):

- 1) Die Variablen sind über die Zeit veränderlich. Diese Voraussetzung ist nicht trivial, schließlich sind viele Variablen – gerade in der politischen Kommunikationsforschung – durch eine sehr hohe Stabilität gekennzeichnet (SCHNEIDER et al. 2014; SCHNEIDER/OTTO/BARTSCH 2017). Dies trifft auch für die (Häufigkeit der) Mediennutzung, also eine der wichtigsten Variablen des RSM zu. Bei zeitlich und situativ stabilen Variablen ist selbstverständlich auch keine selbstverstärkende Spirale vorstellbar (SCHARKOW/BACHL 2017).
- 2) Die Variablen weisen ein korreliertes Wachstum (positiv oder negativ) auf. Die alleinige gegenseitige Beeinflussung über die Zeit hinweg muss noch keinen Spiralprozess anzeigen – es kann sich auch um temporäre Zusammenhänge oder ›Schocks‹ handeln (SCHEMER 2012).

1 In diesem Beitrag beschränken wir uns vor allem auf die methodischen Voraussetzungen und Zusammenhänge des RSM. Bezüglich der theoretischen Grundlagen sei auf die Vorstellung des RSM von SLATER (2007, 2015) verwiesen, der die psychologischen Prozesse einer selbstverstärkenden Spirale dabei sehr ausführlich beschreibt.

- 3) Ebenso muss auch ein reziproker Zusammenhang vorliegen, da ein gemeinsames Wachstum auch auf eine Scheinkorrelation hindeuten könnte; eine dritte Variable könnte für das (gleichzeitige) Wachstum von beiden Variablen verantwortlich sein.

Diese Voraussetzungen lassen sich mittels verschiedener Analysemethoden überprüfen und analysieren, in der bisherigen Forschung wurden meist parallele Wachstumskurvenmodelle mit reziproken Effekten (MOELLER/DE VREESE 2015; SCHEMER 2012) oder Mehrebenenmodelle (KRUIKEMEIER/SHEHATA 2017; SLATER et al. 2003) zur Analyse der Prozesse verwendet. Beide Analysemethoden sind in der Lage, Korrelationen, reziproke Effekte (lagged effects) sowie Wachstumsfaktoren (growth factors) zu quantifizieren, wie wir innerhalb dieses Kapitels zeigen werden.<sup>2</sup>

## 2.2 Herausforderungen des RSM

Auf Grundlage des RSM wurden in vielen unterschiedlichen Teildisziplinen der Kommunikationswissenschaft Längsschnittstudien durchgeführt und einige der wichtigsten und intensiv beforschten Forschungsfragen unter diesem Gesichtspunkt neu untersucht: Selbstverstärkende Spiralprozesse wurden bezüglich des Pornografiekonsums und der Sexualisierung von Frauen (PETER/VALKENBURG 2009), bezüglich gewalthaltigen Medieninhalten und aggressivem Verhalten (DVIR GVIRSMAN et al. 2014; SLATER et al. 2003), Nachrichtennutzung und politischem Wissen (MOELLER/DE VREESE 2015), Nachrichtennutzung und politischer Partizipation (KRUIKEMEIER/SHEHATA 2017), Mediennutzung und politischem Skeptizismus (HUTCHENS et al. 2016) sowie dem Interesse an einer politischen Kampagne und Emotionen (SCHEMER 2012) untersucht.

Obgleich diese Studien sehr unterschiedliche Variablen, Theorien und Prozesse verwenden, ist die Untersuchungsanlage meist recht ähnlich: Alle Studien arbeiten mit Daten aus Längsschnittbefragungen. Dies bringt neben einigen Vorteilen, wie den meist sehr großen Stichproben, auch einige Nachteile mit sich, die wir im Folgenden erörtern möchten.

2 Es sei an dieser Stelle darauf hingewiesen, dass Mehrebenenmodelle und Strukturgleichungsmodelle ineinander überführbar bzw. bedeutungsgleich sind, aufgrund der Unterschiede in Anwendbarkeit und Logik halten wir den Hinweis für diesen Text dennoch für sinnvoll (s. dazu CURRAN 2003).

Erstens ist die Abfrage von Mediennutzung innerhalb dieser Studien, wie in vielen repräsentativen Befragungen, sehr breit und bezieht sich lediglich auf die Häufigkeit der Nutzung pro Tag oder pro Woche. Eine solche Messung von Mediennutzung bringt jedoch einige Probleme mit sich (DE VREESE/NEIJENS 2016): Einerseits geht eine solche Erfassung mit einer Verzerrung seitens der Probanden einher (z. B. durch fehlerhafte Erinnerung), andererseits wird – wie schon beschrieben – lediglich die Häufigkeit erfasst, nicht jedoch, welche Medieninhalte genau rezipiert wurden, z. B. welche Zeitungsartikel genau gelesen wurden oder ob eine Nachrichtensendung vollständig rezipiert wurde. Diesen Nachteil von Panelbefragungen kann die ESM beheben, da die Befragung unmittelbar im Anschluss an die Mediennutzung stattfinden kann (KARNOWSKI 2013; KARNOWSKI/DOEDENS 2010).

Zweitens sind die Abstände zwischen den Messzeitpunkten im längsschnittlichen Design von repräsentativen Panelbefragungen meist sehr lang. So werden die Probanden im Abstand von einigen Wochen (SCHEMER 2012), Monaten (DVIR GVIKSMAN et al. 2014) oder Jahren (MOELLER/DE VREESE 2015) befragt. Während ein solches Design für bestimmte Variablen durchaus sinnvoll ist (vor allem zur Untersuchung von Variablen, die sich nur langsam verändern), gibt es jedoch auch deutlich schneller ablaufende Prozesse, die durch solche weit auseinanderliegenden Messzeitpunkte nicht valide erfassbar sind: »Clearly, the longitudinal design must fit the phenomenon under study« (SLATER 2007: 286). ESM stellt somit eine geeignete Ergänzung zu bisherigen längsschnittlichen Designs in der Kommunikationswissenschaft dar, das besonders geeignet ist, kurzfristige Dynamiken und Prozesse abzubilden (siehe auch EVELAND/MOREY 2011).

Ein Beispiel für einen solchen schnell ablaufenden Prozess könnte das Zusammenspiel zwischen emotionalen Reaktionen auf die Medienrezeption auf der einen Seite und Aufmerksamkeitsprozessen auf der anderen Seite darstellen. Einerseits ist es naheliegend, diese Variablen direkt im Anschluss an die Medienrezeption zu erfassen – schließlich sind deutliche Erinnerungsverzerrungen bei der retrospektiven Bewertung von Aufmerksamkeit und emotionalen Zuständen zu erwarten, so werden beispielsweise retrospektive Einschätzungen von emotionalen Zuständen erheblich durch Geschlechterstereotype verzerrt (BARRETT et al. 1998). Darüber hinaus sind affektive Prozesse – per Definition – veränderbare, schnell ablaufende, dynamische Prozesse (FRIJDA 2000) und das Zusammenspiel zwischen emotionalen Prozessen und Aufmerksamkeit ist damit ebenfalls kurzfristig und dynamisch. Mit anderen Worten ist die Analyse

von emotionalen Prozessen und Aufmerksamkeitsprozessen nur mit feinkörnigen und dynamischen Untersuchungsmethoden wie beispielsweise Echtzeitmessungen oder ESM möglich (ARNDT et al. 2017).

### 3. Experience Sampling als Erhebungsmethode in der Kommunikationswissenschaft

Experience Sampling gehört zu den intensiven Längsschnittdesigns und wird mitunter auch Ambulatory Assessment oder Ecological Momentary Assessment genannt. Grob beschrieben handelt es sich dabei um Panelbefragungen, bei denen die Befragten (annähernd) in Echtzeit in der (Rezeptions-)Situation kurze Fragebögen ausfüllen (CONNER et al. 2009; FRALEY/HUDSON 2014; KARNOWSKI 2013; KARNOWSKI/DOEDENS 2010; KUBEY/LARSON/CSIKSZENTMIHALYI 1996). Die ESM ist insbesondere seit der Verbreitung von Smartphones (Mobile Experience Sampling) deutlich ökonomischer geworden und wird daher für die Beantwortung vieler sozialwissenschaftlicher Fragen genutzt (FRALEY/HUDSON 2014; KARNOWSKI 2013). In der Kommunikationswissenschaft wird die ESM vor allem zur Erfassung von Mediennutzung eingesetzt (GREENBERG et al. 2005; OHME/ALBAEK/DEVREESE 2016). Die Vorteile der Experience-Sampling-Methode zur Erfassung von Mediennutzung gegenüber klassischen Befragungsdesigns liegen dabei auf der Hand: 1) Einerseits werden Verzerrungen durch Erinnerung minimiert, schließlich berichten die Probanden über ihre Mediennutzung unmittelbar nach der Rezeption. 2) Weiterhin kann die Mediennutzung deutlich valider und reliabler erfasst werden als in klassischen Befragungen, da die vorangegangene Rezeption inklusive Dauer und Auswahl des exakten Medienitems (z. B. ganze Sendung, Zeitungsartikel, Homepage) besser erinnerbar ist. 3) Schließlich besteht eine Stärke von Experience Sampling darin, die erfassten Längsschnittdaten hinsichtlich zeitlicher Muster oder Dynamiken auszuwerten. Wang und Kollegen (WANG/TCHERNEV 2012; WANG/TCHERNEV/SOLLOWAY 2012) konnten beispielsweise mittels Experience-Sampling-Daten das dynamische Zusammenspiel aus Rezipientenbedürfnissen, Mediennutzung und Gratifikationen untersuchen und waren in der Lage, reziproke und selbstverstärkende Prozesse zu modellieren (WANG/TCHERNEV 2012; WANG/TCHERNEV/SOLLOWAY 2012).

#### 4. Aufmerksamkeit und Emotionen als reziproke Prozesse

Wie bereits erläutert, stellen Aufmerksamkeit und Emotionen geeignete Variablen für die Untersuchung von kurzfristigen dynamischen Prozessen mittels *ESM* dar. Zum einen handelt es sich bei Emotionen per Definition um kurzfristige Zustände und damit veränderliche Variablen (FRIJDA 2000), zum anderen gibt es auch viele Hinweise darauf, dass sich Emotionen und Informationsverarbeitung über die Zeit hinweg beeinflussen (ARNDT et al. 2017; SCHEMER 2012).

Verschiedene Theorien und Modelle beschreiben den Zusammenhang zwischen Emotionen und Informationsverarbeitung. Die *Affective Intelligence Theory* (AIT, MARCUS/MACKUEN/NEUMAN 2011; MARCUS/NEUMAN/MACKUEN 2000) beschreibt die aufmerksamkeitsleitende Funktion von affektiven Zuständen und postuliert, dass negative Emotionen wie Wut und Angst einen positiven Zusammenhang mit Informationssuche, Aufmerksamkeitsallokation und Elaboration aufweisen. Die AIT geht davon aus, dass negative Emotionen als affektive Marker für neue und potenziell bedrohliche Situationen dienen und daher Aufmerksamkeit erhöhen.

Das *Cognitive-Functional-Model* (CFM, NABI 1999) postuliert ebenfalls einen Einfluss von emotionalen Reaktionen auf die Elaboration. Wie bei der AIT nimmt das CFM ebenfalls einen positiven Zusammenhang zwischen Wut oder Ärger und Aufmerksamkeit an, allerdings macht es eine gegensätzliche Annahme zur negativen Emotion ›Angst‹. Wenn der Rezipient Angst verspürt, löst dies eher Fluchtreflexe aus, die eine erhöhte Aufmerksamkeit und tiefe Verarbeitung verhindern (NABI 1999, 2003, 2010). Es wird daher angenommen, dass die Emotion ›Wut‹ mit erhöhter Aufmerksamkeit (Hypothese 1), Angst oder Furcht jedoch mit niedrigerer Aufmerksamkeit einhergeht (Hypothese 2).

Positive Emotionen, insbesondere solche mit niedriger Erregung, wie beispielsweise Zufriedenheit oder Freude, sind laut AIT und CFM nicht oder negativ mit Aufmerksamkeit und tiefer Informationsverarbeitung korreliert (BRADER/MARCUS/MILLER 2011; NABI 1999). Die AIT postuliert für diese Emotionen, dass sie Marker für das Dispositionssystem sind und daher anzeigen, dass keine bedrohliche Situation vorliegt. Daher gehen diese Emotionen mit habituellem Verhalten und niedriger Aufmerksamkeit einher. Wir nehmen daher ebenfalls einen negativen Zusammenhang zwischen der positiven Emotion ›Freude‹ und Aufmerksamkeit gegenüber den rezipierten Medieninhalten an (Hypothese 3).

Scherer (2012) konnte zeigen, dass sich Emotionen und Aufmerksamkeit in reziproker und selbstverstärkender Weise über die Zeit beeinflussen. Innerhalb einer Längsschnittbefragung konnte gezeigt werden, dass negative emotionale Reaktionen auf Kampagnenkommunikation und Interesse an der Kampagne eine selbstverstärkende Spirale im Sinne des RSM bilden. Dies bedeutet, dass emotionale Zustände und Aufmerksamkeit nicht nur korreliert sind, sondern auch reziproke Effekte (lagged effects auf den nächsten Messzeitpunkt) sowie ein korreliertes Wachstum über die Zeit hinweg aufweisen (growth factors).

Die Forschungsfrage im Sinne der selbstverstärkenden Spiralprozesse lautet daher:

FF1: Gibt es einen gleichzeitigen, reziproken und selbstverstärkenden Zusammenhang zwischen emotionalen Reaktionen (Wut, Angst, Freude) auf politische Information und der selbstberichteten Aufmerksamkeit gegenüber diesen Informationen?

## 5. Methode

### 5.1 Datenerhebung

Um die oben aufgestellten Hypothesen und Forschungsfragen zu überprüfen, wurde eine Mobile-Experience-Sampling-Studie durchgeführt. Acht Tage lang nahmen 96 Probanden an dieser Studie teil.

Im Vorfeld der Studie füllten die Probanden einen allgemeinen Fragebogen zu demografischen Angaben und weiteren (Kontroll-)Variablen aus. Die Teilnehmer wurden in etwa zur Hälfte durch das Marktforschungsinstitut KIM (Karlsruhe) rekrutiert. Alle anderen Teilnehmer (44,8%) waren Studierende der Universität Koblenz-Landau. Alle Versuchsteilnehmer wurden für die Teilnahme entlohnt.

Für die Experience-Sampling-Studie wurden die Teilnehmer dazu aufgefordert, direkt im Anschluss an die Rezeption politischer Nachrichten einen kurzen Fragebogen auf der Smartphone-App movisensXS, Version 0.7.4162, der Firma movisens GmbH in Karlsruhe auszufüllen. In diesem wurden verschiedene emotionale Reaktionen auf den Inhalt der Nachrichten und die Aufmerksamkeit der Teilnehmer während der Medienrezeption abgefragt. Durch dieses event-basierte Design, konnte ein möglichst natürliches Rezeptionsverhalten gewährleistet werden, da die Teilnehmer frei

wählen konnten, wann und wie oft sie Nachrichten rezipieren und infolge dessen einen Fragebogen abschicken. Lediglich die Anzahl der täglichen Fragebogen beschränkte sich auf ein individuelles Maximum von fünf. Die Probanden nahmen diese Möglichkeit jedoch eher selten wahr und lagen im Durchschnitt ( $M=10.20$ ,  $SD=8.09$ )<sup>3</sup> deutlich unter der maximal möglichen Anzahl von insgesamt 40 Fragebögen. Dies spiegelt sich auch in den zeitlichen Abständen zwischen der Medienrezeption wider, die mit durchschnittlich 12,84 Stunden ( $SD=16.34$ ) relativ lang sind. Außerdem deutet die hohe Standardabweichung auf große Unterschiede zwischen den Probanden hin. Der gegen Ausreißer robuste Median lässt mit 7,46 Stunden auf kürzere Intervalle zwischen der Medienrezeption schließen.

Um den Stichprobenumfang in allen Modellen konstant zu halten, wurden alle Fälle ausgeschlossen, die fehlende Werte bei den emotionalen Reaktionen, der Aufmerksamkeit oder einer Kontrollvariablen aufwiesen und somit per listenweisem Fallausschluss ohnehin nicht berücksichtigt worden wären. Danach verblieben 76 Probanden, deren Alter von 18 bis 70 Jahre reicht ( $M=32.49$ ,  $SD=14.32$ ). Sie verfügen im Schnitt über einen hohen formalen Bildungsabschluss (59,2% mit Abitur oder höherem Abschluss) und sind überwiegend weiblich (71,1%).

## 5.2 Messung

*Emotionen.* Die emotionalen Reaktionen wurden jeweils mit einem einzelnen Item gemessen. Die Teilnehmer wurden gefragt, inwiefern sie sich während der Rezeption von politischen Nachrichten wütend ( $M=2.70$ ,  $SD=1.54$ ), freudig ( $M=1.99$ ,  $SD=1.20$ ) oder ängstlich ( $M=1.93$ ,  $SD=1.14$ ) fühlten. Alle Emotionen wurden auf einer 6-Punkte-Skala gemessen (1 fühle mich überhaupt nicht »...« 6 fühle mich sehr »...«).

*Aufmerksamkeit.* Die Aufmerksamkeit der Probanden wurde mit zwei Items erhoben (»Während ich den Beitrag gelesen/angeschaut habe, war ich hochkonzentriert«, »Ich war häufig abgelenkt, als ich den Beitrag gelesen/angeschaut habe«) und auf einer 6-Punkte-Skala erhoben. Geringe Werte weisen auf eine geringe Aufmerksamkeit hin und hohe Werte auf

3 Die Werte beziehen sich bereits auf den bereinigten Datensatz.

eine hohe Aufmerksamkeit ( $M = 4.94$ ,  $SD = 0.88$ )<sup>4</sup>. Um die Reliabilität der Aufmerksamkeit zwischen den Individuen und innerhalb der Individuen zu berechnen, nutzen wir einen Ansatz von Shrout und Lane (2012). Während die Reliabilität zwischen den Individuen hoch ist ( $R_{KRN} = .96$ ), ist die Reliabilität innerhalb der Individuen eher gering ( $R_{CN} = .59$ ). Dies deutet auf eine stabile Rangfolge zwischen den Individuen im Zeitverlauf hin und zeigt außerdem, dass die individuelle Aufmerksamkeit im Zeitverlauf variiert.

Als Kontrollvariablen wurden Geschlecht, Alter, Bildung und politisches Interesse in das Modell aufgenommen. *Das politische Interesse* wurde mit fünf Items (OTTO/BACHERLE 2011) auf einer 6-Punkte-Skala gemessen ( $M = 4.03$ ,  $SD = 1.04$ ). Da politisches Interesse nur zu einem Zeitpunkt gemessen wurde, dient Cronbach's Alpha als Reliabilitätsmaß ( $\alpha = .90$ ).

### 5.3 Analyse

Um kurzfristige dynamische Prozesse messen und abbilden zu können, bedarf es, wie oben beschrieben, nicht nur eines besonderen längsschnittlichen Designs. Auch die Auswertungsmethode sollte einerseits den Nachweis von selbstverstärkenden Spiralprozessen ermöglichen, andererseits jedoch auch kurzfristige dynamische Prozesse abbilden können. Für die Analyse der Daten schlagen wir deshalb multivariate Wachstumskurvenmodelle vor.

Während die Modellierung von Wachstum für eine abhängigen Variable (univariat) an anderer Stelle ausführlich beschrieben wurde (z. B. NEZLEK 2012; RAUDENBUSH/BRYK 2002), stellt die Modellierung von dynamischen, zusammenhängenden Wachstumsprozessen Forscher immer wieder vor Herausforderungen (CURRAN et al. 2014). Herausforderungen, die auch auf das RSM zutreffen: Um sich gegenseitig verstärkende Spiralprozesse zwischen Emotionen und Aufmerksamkeit darzustellen, müssen zwei parallele Wachstumskurven modelliert werden. Hierbei bietet die hierarchische Struktur von Experience-Sampling-Daten, in der die Zeit in Personen geschachtelt ist (jeder Person werden mehrere Messzeitpunkte zugeordnet), die Möglichkeit, Effekte zwischen den Teilnehmern (between-person effects) und Effekte in-

4 Mittelwerte und Standardabweichungen wurden über alle Messzeitpunkte und Individuen hinweg berechnet. Zusätzlich wurden individuelle Mittelwerte bzw. Gruppenmittelwerte und -standardabweichungen berechnet: Wut ( $M = 2.79$ ,  $SD = 1.02$ ), Freude ( $M = 2.05$ ,  $SD = 0.83$ ), Angst ( $M = 1.99$ ,  $SD = 0.84$ ), Aufmerksamkeit ( $M = 4.72$ ,  $SD = 0.72$ ).



nerhalb der Teilnehmer (within-person effects) zu modellieren. Somit können intraindividuelle Wachstumskurven für Emotionen und Aufmerksamkeit berechnet werden. Die Kovarianz beider Wachstumsparameter kann dann als interindividueller Effekt interpretiert werden.

Neben der Modellierung von parallelem Wachstum muss die Reziprozität zwischen beiden Konstrukten überprüft werden. Deshalb werden kreuzverzögerte Effekte und – in einem weiteren Schritt – autoregressive Effekte modelliert (s. hierzu auch das ALT-Modell, BOLLEN/CURRAN 2004). Werden die kreuzverzögerten Effekte nicht berücksichtigt, besteht die Möglichkeit einer Scheinkorrelation zwischen Emotionen und Aufmerksamkeit. In einem weiteren Schritt werden gleichzeitige Effekte in das Modell aufgenommen, um eine mögliche korrelative Verbindung beider Konstrukte zum gleichen Messzeitpunkt zu überprüfen.

In der nachfolgenden Gleichung ist zunächst ein univariates Wachstumskurvenmodell mit den Kontrollvariablen ›Geschlecht‹, ›Alter‹, ›Bildung‹ und ›politisches Interesse‹ abgebildet.

$$y_{ti} = \beta_{00} + \beta_{01}(sex_i) + \beta_{02}(age_i) + \beta_{03}(education_i) + \beta_{04}(pol. Int._i) + r_{0i} + \beta_{10}(time_t) + r_{1i} + \beta_{20}(aut. Effekt_t) + \beta_{30}(lag. Effekt_t) + \beta_{40}(con. Effekt_t) + e_{ti}$$

Wobei  $\beta_{00}$  den durchschnittlichen Intercept für die abhängige Variable  $y$  zum Zeitpunkt  $t$  für Individuum  $i$  und  $r_{0i}$  die korrespondierende Abweichung beschreibt.  $\beta_{01}$  bis  $\beta_{04}$  symbolisieren die Effekte der Kontrollvariablen. Außerdem beschreibt  $\beta_{10} + r_{1i}$  das durchschnittliche Wachstum für Variable  $y$  und die Abweichung von diesem für Individuum  $i$ . Autoregressive (›aut.«), kreuzverzögerte (›lag.«) und gleichzeitige Effekte (›con.«) werden durch  $\beta_{20}$  bis  $\beta_{40}$  beschrieben.

Da das RSM von mindestens zwei abhängigen Variablen ausgeht, deren Wachstumskurven miteinander korrelieren, ist die Modellierung von multivariaten Wachstumskurvenmodellen notwendig. Dies ist durch einen Vektor, der beide abhängigen Variablen in einer Variable zusammenfasst, und den Einsatz von Dummy-Variablen, die die jeweils zugehörige, abhängige Variable ›auslöst‹, möglich (MACCALLUM et al. 1997). Die folgende Gleichung beschreibt ein multivariates Wachstumskurvenmodell mit zwei Wachstumskurven.

$$y_{ti} = \delta_1\beta_{00} + \delta_1\beta_{01}(sex_i) + \delta_1\beta_{02}(age_i) + \delta_1\beta_{03}(education_i) + \delta_1\beta_{04}(pol. Int._i) + \delta_1r_{0i} + \delta_1\beta_{10}(time_t) + \delta_1r_{1i} + \delta_1\beta_{20}(aut. Effekt_t) + \delta_1\beta_{30}(lag. Effekt_t) + \delta_1\beta_{40}(con. Effekt_t) + \delta_1e_{ti} + \delta_2\beta_{00} + \delta_2\beta_{01}(sex_i) + \delta_2\beta_{02}(age_i) + \delta_2\beta_{03}(education_i) + \delta_2\beta_{04}(pol. Int._i) + \delta_2r_{0i} + \delta_2\beta_{10}(time_t) + \delta_2r_{1i} + \delta_2\beta_{20}(aut. Effekt_t) + \delta_2\beta_{30}(lag. Effekt_t) + \delta_2\beta_{40}(con. Effekt_t) + \delta_2e_{ti}$$

Wobei  $\delta_1$  (Dummy für Emotionen) und  $\delta_2$  (Dummy für Aufmerksamkeit) bestimmen, welche Variable in Vektor  $y_t$  als abhängige Variable erkannt wird. Somit können die Effekte für beide abhängigen Variablen geschätzt werden. Durch die resultierende Kovarianzmatrix des multivariaten Wachstumskurvenmodells, kann dann die Korrelation zwischen den Wachstumsparametern ermittelt werden (BALDWIN et al. 2014; MACDONALD-WALLIS et al. 2012). Das oben beschriebene Mehrebenenmodell eignet sich somit in besonderem Maße zur Modellierung des RSM, da es neben der Untersuchung von autoregressiven, kreuzverzögerten und gleichzeitigen Effekten die Modellierung zweier, paralleler Wachstumskurven ermöglicht.

Für die finalen Analysen wurden alle Variablen auf Level 1 (Zeit, Emotionen, Aufmerksamkeit) am Gruppenmittelwert zentriert. Alle Variablen auf Level 2 (Kontrollvariablen: Alter, Geschlecht, Bildung und politisches Interesse) wurden am Gesamtmittelwert zentriert (s. SLATER et al. 2003 bzgl. eines ähnlichen Vorgehens). Für jede Emotion wurde ein multivariates Mehrebenenmodell mit SPSS (v.24) berechnet. Unsere Analysestrategie zur Berechnung der multivariaten Modelle orientiert sich an der Vorgehensweise, die in Baldwin et al. (2014) beschrieben wird.

## 6. Ergebnisse

In den Tabellen 1, 2 und 3 sind die autoregressiven, kreuzverzögerten und gleichzeitigen Effekte für Wut und Aufmerksamkeit (Modell 1), Angst und Aufmerksamkeit (Modell 2) sowie Freude und Aufmerksamkeit (Modell 3) abgebildet. Um zu bestimmen, ob sich Emotionen und Aufmerksamkeit im Zeitverlauf gegenseitig verstärken, muss zum einen der Nachweis von Wachstum für beide Konstrukte erbracht werden. Zum anderen müssen reziproke Effekte (kreuzverzögerte Effekte) zwischen den Konstrukten nachgewiesen werden.

Demzufolge zeigt sich keine Reinforcing Spiral für den Zusammenhang zwischen Wut und Aufmerksamkeit (s. Tab. 1). Im Laufe der Zeit erhöht sich lediglich die Aufmerksamkeit der Probanden ( $\delta_2\beta_{10} = 0.03, S.E. = 0.01, p < .05$ ). Die schwache Effektstärke lässt jedoch den Schluss zu, dass das Konstrukt innerhalb des Untersuchungszeitraum nur ein geringes Wachstum aufweist. Im Gegensatz dazu zeigt die Emotion ›Wut‹ nur marginal signifikantes Wachstum ( $\delta_1\beta_{10} = 0.06, S.E. = 0.03, p < .10$ ). Auch die kreuzverzögerten Effekte zeigen keine signifikanten Auswirkungen. Somit kann nicht von ei-

nem reziproken Zusammenhang gesprochen werden. Betrachtet man die Kovarianzparameter der zufälligen Effekte ( $cov = -0.00$ ,  $S.E. = 0.00$ ,  $p = .88$ ), bestätigt sich, dass die beiden Wachstumskurven nicht zusammenhängen: Die Teilnehmer, deren Aufmerksamkeit über die Woche steigt, werden im Zeitverlauf nicht wütender.

Ein ähnliches Bild zeigt sich für die Emotion ›Angst‹ und Aufmerksamkeit (s. Tab. 2). Auch hier gibt es keine Hinweise auf ein Wachstum der Angst im Zeitverlauf ( $\delta_1\beta_{10} = 0.00$ ,  $S.E. = 0.02$ ,  $p = .85$ ). Die Angst der Probanden erweist sich demnach im Zeitverlauf als relativ stabil und verharrt auf geringem Niveau.

Der marginal signifikante, kreuzverzögerte Effekt auf die Emotion ›Angst‹ ( $\delta_1\beta_{30} = -0.09$ ,  $S.E. = 0.05$ ,  $p < .10$ ) könnte ein Anzeichen dafür sein, dass eine erhöhte Aufmerksamkeit zu einer Abnahme von Angst führt. Insgesamt deutet das Ausbleiben der kreuzverzögerten Effekte und ein nicht vorhandener Zusammenhang zwischen den beiden Wachstumskurven ( $cov = -0.00$ ,  $S.E. = 0.00$ ,  $p = .89$ ) jedoch auf keinen reziproken Wachstumsprozess hin.

TABELLE 1

**Multivariates Wachstumskurvenmodell für Wut und Aufmerksamkeit (feste Effekte, Modell 1)**

	Wut				Aufmerksamkeit			
	Koef.	Est.	S.E.	t	Koef.	Est.	S.E.	t
Intercept	$\bar{\delta}_1\beta_{00}$	2,69	0,11	25,57***	$\bar{\delta}_2\beta_{00}$	4,85	0,07	66,62***
Zeit	$\bar{\delta}_1\beta_{10}$	0,06	0,03	1,78†	$\bar{\delta}_2\beta_{10}$	0,03	0,01	2,05*
Aut. Effekt	$\bar{\delta}_1\beta_{20}$	0,00	0,04	-0,01	$\bar{\delta}_2\beta_{20}$	-0,08	0,04	-2,26*
Lag. Effekt	$\bar{\delta}_1\beta_{30}$	-0,04	0,07	-0,52	$\bar{\delta}_2\beta_{30}$	-0,01	0,02	-0,77
Con. Effekt	$\bar{\delta}_1\beta_{40}$	0,13	0,07	1,73†	$\bar{\delta}_2\beta_{40}$	0,03	0,02	1,80†

Anmerkungen: 775 Observationen, 76 Individuen. Aut. Effekt = autoregressiver Effekt, Lag. Effekt = kreuzverzögerter Effekt und Con. Effekt = gleichzeitiger Effekt. Kontrollvariablen wurden aufgenommen aber nicht abgebildet, \*\*\* $p < .001$ , \* $p < .05$ , † $p < .10$

Auch zwischen der positiven Emotion ›Freude‹ und Aufmerksamkeit konnte keine Reinforcing Spiral gefunden werden (s. Tab. 3). Die Freude der Teilnehmer verringert sich zwar im Zeitverlauf, jedoch ist dieses negative Wachstum nur marginal signifikant ( $\delta_1\beta_{10} = -0.04$ ,  $S.E. = 0.02$ ,  $p < .10$ ). Auch die kreuzverzögerten Effekte zeigen keine signifikanten Effekte auf Freude ( $\delta_1\beta_{30} = -0.05$ ,  $S.E. = 0.05$ ,  $p = .35$ ) und Aufmerksamkeit ( $\delta_2\beta_{30} = -0.01$ ,  $S.E. = 0.02$ ,  $p = .68$ ). Lediglich die gleichzeitigen Effekte deuten auf einen

korrelativen, negativen Zusammenhang zwischen der positiven Emotion ›Freude‹ und Aufmerksamkeit zum gleichen Messzeitpunkt hin. Sowohl für Freude ( $\delta_1\beta_{40} = -0.11$ ,  $S.E. = 0.06$ ,  $p < .10$ ) als auch für Aufmerksamkeit ( $\delta_2\beta_{40} = -0.05$ ,  $S.E. = 0.02$ ,  $p < .10$ ) sind beide Effekte nur marginal signifikant. Der Zusammenhang zwischen den Wachstumsgeraden ist sehr schwach und nicht signifikant ( $COV = -0.00$ ,  $S.E. = 0.00$ ,  $p = .17$ ).

TABELLE 2

**Multivariates Wachstumskurvenmodell für Angst und Aufmerksamkeit (feste Effekte, Modell 2)**

	Angst				Aufmerksamkeit			
	Koef.	Est.	S.E.	t	Koef.	Est.	S.E.	t
Intercept	$\delta_1\beta_{00}$	1,94	0,09	21,79***	$\delta_2\beta_{00}$	4,85	0,07	66,74***
Zeit	$\delta_1\beta_{10}$	0,00	0,02	-0,20	$\delta_2\beta_{10}$	0,03	0,01	2,14*
Aut. Effekt	$\delta_1\beta_{20}$	-0,11	0,04	-3,04**	$\delta_2\beta_{20}$	-0,09	0,04	-2,57*
Lag. Effekt	$\delta_1\beta_{30}$	-0,09	0,05	-1,93†	$\delta_2\beta_{30}$	-0,03	0,03	-1,16
Con. Effekt	$\delta_1\beta_{40}$	-0,07	0,05	-1,48	$\delta_2\beta_{40}$	-0,04	0,03	-1,52

Anmerkungen: 775 Observations, 76 Individuen. Aut. Effekt = autoregressiver Effekt, Lag. Effekt = kreuzverzögerter Effekt und Con. Effekt = gleichzeitiger Effekt. Kontrollvariablen wurden aufgenommen aber nicht abgebildet, \*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ , † $p < .10$

TABELLE 3

**Multivariates Wachstumskurvenmodell für Freude und Aufmerksamkeit (feste Effekte, Modell 3)**

	Freude				Aufmerksamkeit			
	Koef.	Est.	S.E.	t	Koef.	Est.	S.E.	t
Intercept	$\delta_1\beta_{00}$	2,02	0,09	23,24***	$\delta_2\beta_{00}$	4,85	0,07	66,75***
Zeit	$\delta_1\beta_{10}$	-0,04	0,02	-1,70†	$\delta_2\beta_{10}$	0,03	0,01	2,00†
Aut. Effekt	$\delta_1\beta_{20}$	0,00	0,04	0,10	$\delta_2\beta_{20}$	-0,09	0,04	-2,62**
Lag. Effekt	$\delta_1\beta_{30}$	-0,05	0,05	-0,94	$\delta_2\beta_{30}$	-0,01	0,02	-0,41
Con. Effekt	$\delta_1\beta_{40}$	-0,11	0,06	-1,86†	$\delta_2\beta_{40}$	-0,05	0,02	-1,89†

Anmerkungen: 775 Observations, 76 Individuen. Aut. Effekt = autoregressiver Effekt, Lag. Effekt = kreuzverzögerter Effekt und Con. Effekt = gleichzeitiger Effekt. Kontrollvariablen wurden aufgenommen aber nicht abgebildet, \*\*\* $p < .001$ , \*\* $p < .01$ , † $p < .10$

## 7. Diskussion

Während dynamische Prozesse in der Kommunikationswissenschaft allgegenwärtig zu sein scheinen, spielen kurzfristige Dynamiken bisher eher eine untergeordnete Rolle, wenn es um die Analyse von Rezeption, Verarbeitung und Auswahl von Medien geht (aber s. WANG 2014; WANG/TCHERNEV 2012; WANG/TCHERNEV/SOLLOWAY 2012). Dieser Forschungslücke wollen wir innerhalb dieses Kapitels begegnen und stellen daher mit der ESM erstens eine geeignete Methode zur Erfassung kurzfristiger Dynamiken vor und zweitens zeigen wir, wie intensive Längsschnittdaten und kurzfristige, selbstverstärkende Dynamiken mit multivariaten parallelen Wachstumsmodellen adäquat ausgewertet werden können. Die Methode berücksichtigt einerseits die hierarchische Struktur der Daten und ist andererseits in der Lage, die parallelen Wachstumskurven, wie sie innerhalb des RSM angedacht sind, zu modellieren.

Durch die Befragung unmittelbar nach der Rezeptionssituation kann die Mediennutzung und die Reaktion auf die Nutzung zum einen valider untersucht werden, da mit geringeren Verzerrungen durch Erinnerungsfehler zu rechnen ist. Zum anderen bietet die ESM durch die wiederholten Befragungen die Möglichkeit, kurzfristige Dynamiken zu erheben und zu analysieren. Als Anwendungsbeispiel aus der politischen Nachrichtenrezeptionsforschung dient der Zusammenhang zwischen emotionalen Reaktionen (Wut, Angst, Freude) auf politische Nachrichten und die Aufmerksamkeit gegenüber diesen Nachrichten. Diese Variablen scheinen sich besonders gut zur Untersuchung kurzfristiger Dynamiken zu eignen, da es sich bei Emotionen und Aufmerksamkeit eher um fluide, schnell veränderliche Variablen handelt. Außerdem wurde der vermutete selbstverstärkende Spiralprozess aus emotionalen Reaktionen und Aufmerksamkeit gegenüber politischen Informationen bereits nachgewiesen (SCHEMER 2012).

Während es innerhalb der Untersuchung Hinweise auf die korrelativen Zusammenhänge zwischen Wut, Angst und der selbstberichteten Aufmerksamkeit gab, konnten wir – im Unterschied zu Schemer (2012) – keine selbstverstärkenden Prozesse zwischen den affektiven Zuständen und der selbstberichteten Aufmerksamkeit nachweisen. Dies mag verschiedene Gründe haben: Einerseits könnte, wie bei anderen Untersuchungen, eine der Variablen zu stabil gewesen sein, um eine Veränderung oder Wachstum nachweisen zu können. Andererseits könnte man ebenfalls davon ausgehen, dass die Dynamik kleineren Abständen folgt und die Modellierung

des Wachstums über den gesamten Zeitraum hier eine Fehlannahme ist. Es könnte beispielsweise angenommen werden, dass der emotionale Zustand während der Medienrezeption die Aufmerksamkeit der nächsten Rezeption beeinflusst, wenn sie unmittelbar auf die vorherige Rezeptionssituation erfolgt, nicht jedoch, wenn die nächste Mediennutzung erst Stunden später stattfindet. Die Zeitabstände zwischen der Medienrezeption, die innerhalb einer *ESM* Studie, im Gegensatz zur klassischen Panelstudie, untersucht werden können, könnte eine entscheidende Variable zu Erklärung des Auftretens oder Ausbleibens von dynamischen Prozessen sein.

Zukünftige Studien sollten daher den dynamischen Prozess noch kleinteiliger untersuchen: So können Multilevel-Spline-Modelle, die es ermöglichen, dynamische Prozesse in kürzeren Sequenzen zu erfassen, paralleles Wachstum und Dynamiken noch detaillierter darstellen. Dieser sequenzielle Ansatz ermöglicht eine event-basierte Stückelung der Wachstumskurven (MACDONALD-WALLIS et al. 2012).

Das Ermitteln geeigneter Messzeitpunkte, die eine Erfassung von Dynamiken ermöglichen, erscheint als eine der größten Herausforderungen für die Erforschung dynamischer Prozesse. Hierzu bedarf es einerseits geeigneter Erhebungs- und Analysemethoden. Andererseits sollte die Wahl der Messzeitpunkte und des Designs (mehr als bisher) theoretisch begründbaren Entscheidungen folgen und nicht ausschließlich forschungsökonomischen Überlegungen (s. dazu auch CURRAN et al. 2014; EVELAND/MOREY 2011). Dies impliziert, dass vor der Durchführung einer längsschnittlichen Befragung fundierte Aussagen zur Stabilität und Veränderbarkeit der Variablen theoretisch hergeleitet und empirisch geprüft werden müssen (SCHNEIDER et al. 2014, 2017). Diesen Überlegungen folgend, wurde innerhalb dieses Kapitels die *ESM* als geeigneter Ansatz zur Messung von emotionalen Reaktionen und Aufmerksamkeitsprozessen genutzt. Wie jedoch bereits angeklungen, könnten die Messzeitpunkte weiterhin zu grobkörnig zur Erfassung der kurzfristigen Prozesse gewesen sein. Die stärkere Berücksichtigung der *time-lags* oder Instruktionen an die Probanden, die eine noch feinere Messung möglich machen, könnten hier Abhilfe schaffen.

Trotz der Vorzüge der *ESM*, hat das Design selbstverständlich auch Schwächen und ist daher für bestimmte Fragestellungen nicht geeignet. Erstens sind solche Designs mit einem erheblichen Aufwand für die Versuchspersonen verbunden. Fehlendes Engagement der Versuchspersonen kann zu hohen *Drop-out*-Quoten und geringerer Datenqualität führen (NAPA SCOLLON/PRIETO/DIENER 2009). Zweitens ist die Teilnahme an einer *Experience*

Sampling-Studie selbstverständlich mit einem gewissen Maß an Reaktivität verbunden. Das intensive Design kann zu Veränderungen des Erlebens und Verhaltens von Versuchspersonen führen. Innerhalb der vorgestellten Studie wäre beispielsweise anzunehmen, dass die Aufmerksamkeit durch das Hochladen der Medienitems und die Beantwortung von Fragen zum Medieninhalt bereits zu einer intensiveren Auseinandersetzung mit den Informationen führt als bei einer regulären Medienrezeption außerhalb der Studie. In ähnlicher Weise bleiben, drittens, die Schwächen und Verzerrungen von Selbstauskünften bestehen. So sind selbstverständlich auch In-situ-Befragungen von sozial erwünschtem Antwortverhalten, Reihenfolgeeffekten und anderen bekannten systematischen Verzerrungen betroffen. Dies gilt in besonderem Maße für die hier präsentierte Studie. Schließlich sind Selbsteinschätzungen von emotionalen Zuständen von Rationalisierungen betroffen, und die Messung unterschiedlicher, distinkter Emotionen ist häufig nicht reliabel und valide. Neuere Entwicklungen innerhalb der *ESM*-Forschung, die versuchen, Experience Sampling mit physiologischen Maßen von Smartphone, Smartwatches und anderen sogenannten ›Wearables‹ zu verknüpfen, sind eine sinnvolle Ergänzung zu Selbstauskünften über affektive Reaktionen. Viertens bieten *ESM*-Studien zwar, wie bereits erwähnt, über die intensive Längsschnittmessung den Vorteil von höherer Varianz innerhalb der Versuchsperson, allerdings erreichen *ESM*-Studien selbstverständlich aufgrund forschungsökonomischer Umstände nicht die Stichprobengrößen klassischer Panelstudien und büßen daher Between-Varianz ein.

Hohe Drop-out-Quoten, Reaktivität und verzerrtes Antwortverhalten sind selbstverständlich mögliche Erklärungen für das Ausbleiben von selbstverstärkenden dynamischen Prozessen innerhalb der vorgestellten Studie. Es sollte jedoch ebenfalls beachtet werden, dass selbstverstärkende Spiralprozesse nur eine, unter Umständen seltene, Beschreibung von dynamischen Prozessen in der Kommunikationswissenschaft sind. Die Voraussetzung von verzögerten Effekten und gleichzeitigem korrelierten Wachstum ist schwer nachweisbar. Möglicherweise sind die Informationen aus Experience-Sampling-Befragungen durch andere dynamische Modelle besser erklärbar als durch paralleles Wachstum innerhalb des *rSM*. Der Vorteil der *ESM* gerade innerhalb der Kommunikationswissenschaft liegt jedoch nicht nur in der Erhebung und Beschreibung von dynamischen Prozessen, vielmehr können Dynamiken – seien es nun Wachstumsprozesse oder temporäre ›Schocks‹ – über den Medieninhalt, der innerhalb von Experience Sampling einfach und sehr detailliert erfassbar ist, erklärt

werden. Anders ausgedrückt eignet sich die ESM nicht nur zur genaueren Erfassung von Mustern und Zeitverläufen, gleichzeitig kann auch der rezipierte Medieninhalt deutlich genauer erfasst und mit den Reaktionen des Rezipienten verknüpft werden als in traditionellen ›Linkage-Studien‹ (DE VREESE et al. 2017). Die ESM kann demnach nicht nur eine intensive Längsschnittbefragung, sondern auch eine längsschnittliche Erfassung von rezipierten Medieninhalten liefern. Somit können intensive längsschnittliche Linkage-Analysen (ILLA) durchgeführt werden, die dynamische Selektions- und Wirkungsprozesse in Kombination mit dem rezipierten Medieninhalt sehr genau abbilden können.

Auch wenn die Erfassung und Analyse von dynamischen Kausalzusammenhängen eine der schwierigsten Aufgaben der Sozialwissenschaften darstellt (CURRAN et al. 2014), bedeutet dies selbstverständlich nicht, dass die Anstrengungen, diese Forschungslücke zu bearbeiten und zu schließen, vernachlässigt werden sollte. Schließlich bilden diese (kurzfristigen) dynamischen Prozesse die ›Realität‹ von Medienrezeption deutlich detailgenauer ab, als klassische Paneldesigns und Linkage-Analysen in der Kommunikationswissenschaft. Somit stellt die ESM eine sinnvolle Ergänzung des Methodenrepertoires zur Analyse und Beschreibung kurzfristiger dynamischer Prozesse dar.

## Literatur

- ARNDT, C.; T. LISCHETZKE; C. CRAYEN; M. EID: The assessment of emotional clarity via response times to emotion items: shedding light on the response process and its relation to emotion regulation strategies. In: *Cognition and Emotion*, 2017, S. 1-19. doi: 10.1080/02699931.2017.1322039
- BALDWIN, S. A.; Z. E. IMEL; S. R. BRAITHWAITE; D. C. ATKINS: Analyzing multiple outcomes in clinical research using multivariate multilevel models. In: *Journal of Consulting and Clinical Psychology*, 82(5), 2014, S. 920-930. doi: 10.1037/a0035628
- BARRETT, L. F.; L. ROBIN; P. R. PIETROMONACO; K. M. EYSEL: Are Women the ›More Emotional‹ Sex? Evidence from Emotional Experiences in Social Context. In: *Cognition and Emotion*, 12(4), 1998, S. 555-578. doi: 10.1080/026999398379565



- BEAM, M. A.; M. J. HUTCHENS; J. D. HMIELOWSKI: Facebook news and (de) polarization: reinforcing spirals in the 2016 us election. In: *Information, Communication & Society*, 2018, S. 1-19. doi: 10.1080/13669118X.2018.1444783
- BOLLEN, K. A.; P. J. CURRAN: Autoregressive Latent Trajectory (ALT) Models. A Synthesis of Two Traditions. In: *Sociological Methods & Research*, 32(3), 2004, S. 336-383. doi: 10.1177/0049124103260222
- BRADER, T.; G. E. MARCUS; K. L. MILLER: Emotion and public opinion. In: EDWARDS, G. C.; L. R. JACOBS; R. Y. SHAPIRO (Hrsg.): *The Oxford Handbook of American Public Opinion and the Media*. Oxford, New York [Oxford University Press] 2011. doi: 10.1093/oxfordhb/9780199545636.003.0024
- CONNER, T. S.; H. TENNEN; W. FLEESON; L. F. BARRETT: Experience Sampling Methods: A Modern Idiographic Approach to Personality Research. In: *Social and Personality Psychology Compass*, 3(3), 2009, S. 292-313. doi: 10.1111/j.1751-9004.2009.00170.x
- CURRAN, P. J.: Have Multilevel Models Been Structural Equation Models All Along? In: *Multivariate Behavioral Research*, 38(4), 2003, S. 529-569. doi: 10.1207/s15327906mbr3804\_5
- CURRAN, P. J.; A. L. HOWARD; S. A. BAINTEK; S. T. LANE; J. S. MCGINLEY: The separation of between-person and within-person components of individual change over time: A latent curve model with structured residuals. In: *Journal of Consulting and Clinical Psychology*, 82(5), 2014, S. 879
- DE VREESE, C. H.; M. BOUKES; A. SCHUCK; R. VLIAGENTHART; L. BOS; Y. LELKES: Linking Survey and Media Content Data: Opportunities, Considerations, and Pitfalls. In: *Communication Methods and Measures*, 11(4), 2017, S. 221-244. doi: 10.1080/19312458.2017.1380175
- DE VREESE, C. H.; P. NEIJENS: Measuring Media Exposure in a Changing Communications Environment. In: *Communication Methods and Measures*, 10(2-3), 2016, S. 69-80. doi: 10.1080/19312458.2016.1150441
- DVIR GVIKSMAN, S.; L. R. HUESMANN; E. F. DUBOW; S. F. LANDAU; K. SHIKAKI; P. BOXER: The Effects of Mediated Exposure to Ethnic-Political Violence on Middle East Youth's Subsequent Post-Traumatic Stress Symptoms and Aggressive Behavior. In: *Communication Research*, 41(7), 2014, S. 961-990. doi: 10.1177/0093650213510941
- EVELAND, W. P.; A. C. MOREY: Challenges and opportunities of panel designs. In: BUCY, E. P.; R. L. HOLBERT (Hrsg.): *The sourcebook for political communication research: Methods, measures, and analytical techniques*. New York/NY [Routledge] 2011, S. 19-33

- FRALEY, R. C.; N. W. HUDSON: *Review of intensive longitudinal methods: An introduction to diary and experience sampling research*. [Taylor & Francis] 2014
- FRIJDA, N. H.: The psychologists' point of view. In: LEWIS, M.; J. M. HAVILAND-JONES; L. F. BARRETT (Hrsg.): *Handbook of emotions*. Vol. 2, New York/NY [Guilford Press] 2000, S. 59-74
- FRÜH, W.; K. SCHÖNBACH: Der dynamisch-transaktionale Ansatz. Ein neues Paradigma der Medienwirkungen. In: *Publizistik*, 27, 1982, S. 74-88
- FRÜH, W.; K. SCHÖNBACH: The dynamic-transactional approach III: An interim appraisal. In: *Publizistik*, 50(1), 2005, S. 4-20. doi: 10.1007/s11616-005-0115-7
- GREENBERG, B. S.; M. S. EASTIN; P. SKALSKI; L. COOPER; M. LEVY; K. LACHLAN: Comparing Survey and Diary Measures of Internet and Traditional Media Use. In: *Communication Reports*, 18(1-2), 2005, S. 1-8. doi: 10.1080/08934210500084164
- HART, W.; D. ALBARRACÍN; A. H. EAGLY; I. BRECHAN; M. J. LINDBERG; L. MERRILL: Feeling validated versus being correct: A meta-analysis of selective exposure to information. In: *Psychological Bulletin*, 135(4), 2009, S. 555-588. doi: 10.1037/a0015701
- HUTCHENS, M. J.; J. D. HMIELOWSKI; B. E. PINKLETON; M. A. BEAM: A Spiral of Skepticism? The Relationship Between Citizens' Involvement With Campaign Information to Their Skepticism and Political Knowledge. In: *Journalism & Mass Communication Quarterly*, 2016
- KARNOWSKI, V.: Befragung in situ: Die Mobile Experience Sampling Method (MESM). In: MÖHRING, W.; D. SCHLÜTZ (Hrsg.): *Handbuch standardisierte Erhebungsverfahren in der Kommunikationswissenschaft*. Wiesbaden [Springer Fachmedien Wiesbaden] 2013, S. 235-247. doi: 10.1007/978-3-531-18776-1\_13
- KARNOWSKI, V.; S. DOEDENS: Mobile Experience Sampling. Eine Methode zur Untersuchung mobilen Mediennutzungsverhaltens. In: JACKOB, N.; T. ZERBACK; O. JANDURA; M. MAURER (Hrsg.): *Methoden der Online-Forschung: Das Internet als Forschungsinstrument und -gegenstand der Kommunikationswissenschaft*. Köln [Herbert von Halem] 2010, S. 211-226
- KRUIKEMEIER, S.; A. SHEHATA: News Media Use and Political Engagement Among Adolescents: An Analysis of Virtuous Circles Using Panel Data. In: *Political Communication*, 34(2), 2017, S. 221-242. doi: 10.1080/10584609.2016.1174760

- KUBEY, R.; R. LARSON; M. CSIKSZENTMIHALYI: Experience Sampling Method Applications to Communication Research Questions. In: *Journal of Communication*, 46(2), 1996, S. 99-120. <http://dx.doi.org/10.1111/j.1460-2466.1996.tb01476.x>
- MACCALLUM, R. C.; C. KIM; W. B. MALARKEY; J. K. KIECOLT-GLASER: Studying Multivariate Change Using Multilevel Models and Latent Curve Models. In: *Multivariate Behavioral Research*, 32(3), 1997, S. 215-253. doi:10.1207/s15327906mbr3203\_1
- MACDONALD-WALLIS, C.; D. A. LAWLOR; T. PALMER; K. TILLING: Multivariate multilevel spline models for parallel growth processes: application to weight and mean arterial pressure in pregnancy. In: *Statistics in Medicine*, 31(26), 2012, S. 3147-3164. doi: 10.1002/sim.5385
- MARCUS, G. E.; M. MACKUEN; W. R. NEUMAN: Parsimony and Complexity: Developing and Testing Theories of Affective Intelligence. In: *Political Psychology*, 32(2), 2011, S. 323-336. doi: 10.1111/j.1467-9221.2010.00806.x
- MARCUS, G. E.; W. R. NEUMAN; M. MACKUEN: *Affective intelligence and political judgment*. Chicago [University of Chicago Press] 2000
- MOELLER, J.; C. DE VREESE: Spiral of Political Learning: The Reciprocal Relationship of News Media Use and Political Knowledge Among Adolescents. In: *Communication Research*, 2015. doi: 10.1177/009365021505148
- NABI, R. L.: A cognitive-functional model for the effects of discrete negative emotions on information processing, attitude change, and recall. In: *Communication Theory*, 9(3), 1999, S. 292-320
- NABI, R. L.: Exploring the Framing Effects of Emotion: Do Discrete Emotions Differentially Influence Information Accessibility, Information Seeking, and Policy Preference? In: *Communication Research*, 30(2), 2003, S. 224-247. doi: 10.1177/0093650202250881
- NABI, R. L.: The Case for Emphasizing Discrete Emotions in Communication Research. In: *Communication Monographs*, 77(2), 2010, S. 153-159. doi: 10.1080/03637751003790444
- NAPA SCOLLON, C.; C.-K. PRIETO; E. DIENER: Experience Sampling: Promises and Pitfalls, Strengths and Weaknesses. In: DIENER, E. (Hrsg.): *Assessing Well-Being: The Collected Works of Ed Diener*. Dordrecht [Springer Netherlands] 2009, S. 157-180. doi: 10.1007/978-90-481-2354-4\_8
- NEZLEK, J. B.: *Diary methods for social and personality psychology*. Thousand Oaks/CA [Sage Publications] 2012

- NOELLE-NEUMANN, E.: The spiral of silence a theory of public opinion. In: *Journal of Communication*, 24(2), 1974, S. 43-51
- OHME, J.; E. ALBAEK; C. H. DE VREESE: Exposure Research Going Mobile: A Smartphone-Based Measurement of Media Exposure to Political Information in a Convergent Media Environment. In: *Communication Methods and Measures*, 10(2-3), 2016, S. 135-148. doi: 10.1080/19312458.2016.1150972
- OTTO, L.; P. BACHERLE: Politisches Interesse Kurzskaala (PIKS): Entwicklung und Validierung. In: *Politische Psychologie*, 1(1), 2011, S. 19-35
- OTTO, L.; F. THOMAS; M. MAIER: Everyday dynamics of media skepticism and credibility: An ambulatory assessment study. In: OTTO, K.; A. KÖHLER (Hrsg.): *Trust in Media and Journalism. Empirical Perspectives on Ethics, Norms, Impacts and Populism in Europe*. Wiesbaden [Springer vs] 2018, S. 111-133. doi: 10.1007/978-3-658-20765-6\_7
- PETER, J.; P. M. VALKENBURG: Adolescents' Exposure to Sexually Explicit Internet Material and Notions of Women as Sex Objects: Assessing Causality and Underlying Processes. In: *Journal of Communication*, 59(3), 2009, S. 407-433. doi: 10.1111/j.1460-2466.2009.01422.x
- PRIOR, M.: Media and Political Polarization. In: *Annual Review of Political Science*, 16(1), 2013, S. 101-127. doi: 10.1146/annurev-polisci-100711-135242
- RAUDENBUSH, S. W.; A. S. BRYK: *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Thousand Oaks/CA [Sage] 2002
- SCHARKOW, M.; M. BACHL: Reinforcing spirals and the issue of stable attitudes and behavior. In: *International Communication Association*, 2017
- SCHEMER, C.: Reinforcing Spirals of Negative Affects and Selective Attention to Advertising in a Political Campaign. In: *Communication Research*, 39(3), 2012, S. 413-434. doi: 10.1177/0093650211427141
- SCHNEIDER, F. M.; L. OTTO; D. ALINGS; M. SCHMITT: Measuring Traits and States in Public Opinion Research: A Latent State-Trait Analysis of Political Efficacy. In: *International Journal of Public Opinion Research*, 26(2), 2014, S. 202-223. doi: 10.1093/ijpor/edu002
- SCHNEIDER, F. M.; L. OTTO; A. BARTSCH: Das ist doch kein Zustand! Zur Messung von States und Traits in der Kommunikationswissenschaft. In: *M&K Medien & Kommunikationswissenschaft*, 65(1), 2017, S. 83-100. doi: 10.5771/1615-634X-2017-1-83

- SHROUT, P. E.; S. P. LANE: Psychometrics. In: MEHL, R. S.; T. S. CONNER (Hrsg.): *Handbook of research methods for studying daily life*. New York/NY [Guilford Press] 2012, S. 302-320
- SLATER, M. D.: Reinforcing Spirals: The Mutual Influence of Media Selectivity and Media Effects and Their Impact on Individual Behavior and Social Identity. In: *Communication Theory*, 17(3), 2007, S. 281-303. doi: 10.1111/j.1468-2885.2007.00296.x
- SLATER, M. D.: Reinforcing Spirals Model: Conceptualizing the Relationship Between Media Content Exposure and the Development and Maintenance of Attitudes. In: *Media Psychology*, 18(3), 2015, S. 370-395. doi: 10.1080/15213269.2014.897236
- SLATER, M. D.; A. F. HAYES: The Influence of Youth Music Television Viewership on Changes in Cigarette Use and Association With Smoking Peers: A Social Identity, Reinforcing Spirals Perspective. In: *Communication Research*, 37(6), 2010, S. 751-773. doi: 10.1177/0093650210375953
- SLATER, M. D.; K. L. HENRY; R. C. SWAIM; L. L. ANDERSON: Violent Media Content and Aggressiveness in Adolescents. In: *Communication Research*, 30(6), 2003, S. 713-736. doi: 10.1177/0093650203258281
- STROUD, N. J.: Media Use and Political Predispositions: Revisiting the Concept of Selective Exposure. In: *Political Behavior*, 30(3), 2008, S. 341-366. doi:10.1007/s11109-007-9050-9
- STROUD, N. J.: Polarization and Partisan Selective Exposure. In: *Journal of Communication*, 60(3), 2010, S. 556-576. doi:10.1111/j.1460-2466.2010.01497.x
- WANG, Z.: Bridging Media Processing and Selective Exposure: A Dynamic Motivational Model of Media Choices and Choice Response Time. In: *Communication Research*, 41(8), 2014, S. 1064-1087. doi:10.1177/0093650214534963
- WANG, Z.; J. M. TCHERNEV: The »Myth« of Media Multitasking: Reciprocal Dynamics of Media Multitasking, Personal Needs, and Gratifications. In: *Journal of Communication*, 62(3), 2012, S. 493-513. doi: 10.1111/j.1460-2466.2012.01641.x
- WANG, Z.; J. M. TCHERNEV; T. SOLLOWAY: A dynamic longitudinal examination of social media use, needs, and gratifications among college students. In: *Computers in Human Behavior*, 28(5), 2012, S. 1829-1839. doi: <https://doi.org/10.1016/j.chb.2012.05.001>

## Autorinnen und Autoren

MARKO BACHL, Dr., wissenschaftlicher Mitarbeiter am Institut für Kommunikationswissenschaft der Universität Hohenheim. Dr. rer. soc. Universität Hohenheim, Institut für Kommunikationswissenschaft; Master of Arts, Medienmanagement, Hochschule für Musik, Theater und Medien Hannover, Institut für Journalistik und Kommunikationsforschung; Bachelor of Arts, Medien & Kommunikation, Universität Augsburg. Seine Forschungsschwerpunkte sind empirische Methoden in der Kommunikationsforschung, Politische Kommunikation und Gesundheitskommunikation.

JOHANNES DAXENBERGER, Dr., ist leitender wissenschaftlicher Mitarbeiter am Lehrstuhl für Ubiquitäre Wissensverarbeitung an der Technischen Universität Darmstadt, wo er 2015 im Bereich Textklassifikation promovierte. Seine Forschungsschwerpunkte sind die Anwendung von maschinellen Lernverfahren auf Problemstellungen der Geistes- und Sozialwissenschaften, die automatische Erkennung und Verarbeitung natürlichsprachlicher Argumentation sowie Digital Humanities. Er koordiniert das Projekt ›ArgumentText‹, das aktuelle Forschungsergebnisse aus den Bereichen Argument Mining und Textanalyse für industrielle Anwendungen validiert.

ANDREAS FAHR, Prof. Dr., Studium der Publizistik, Psychologie und Volkswirtschaft an der Johannes Gutenberg-Universität Mainz. 1995 bis 2000 wissenschaftlicher Mitarbeiter am Medien Institut Ludwigshafen. 2000 bis 2012 wissenschaftlicher Assistent, Akademischer Rat und Oberrat am Institut für Kommunikationswissenschaft und Medienforschung (IfKW) der LMU-München. 2001 Promotion zum Dr. rer. pol., 2010 Habilitation. 2012 bis 2013 Professor für Kommunikationswissenschaft mit dem Schwerpunkt Soziale Kommunikation an der Universität Erfurt. Seit Herbstse-

mester 2013 Professor für empirische Kommunikationsforschung an der Universität Fribourg (Schweiz). Forschungsschwerpunkte: Mediennutzung, Medienrezeption und Medienwirkung, Medienpsychologie, Empirische Methoden, Psychophysiologie.

STEFAN GEISS, Prof. Dr., ist Assoziierter Professor für politische Kommunikation am Institut für Soziologie und Politikwissenschaft der Technisch-Naturwissenschaftlichen Universität Norwegens (NTNU), Trondheim. Seine Forschungsschwerpunkte sind Agenda Setting und Building, Frame Setting und Building sowie Skandalforschung. Preisträger u.a. des Dissertationspreises der Johannes Gutenberg-Universität Mainz.

IRYNA GUREVYCH, Prof. Dr., ist Professorin für Ubiquitäre Wissensverarbeitung an der Technischen Universität Darmstadt. Ihre Gruppe ist international bekannt für ihre Forschung in den Bereichen Natural Language Processing, Informationsverarbeitung und Digital Humanities. Sie hat in weit über 100 Fachzeitschriften und Konferenzen veröffentlicht, wobei ihr Schwerpunkt auf der computergestützten Analyse natürlicher Sprache und der Extraktion von Bedeutung aus unstrukturiertem Text liegt. Sie leitet das interdisziplinäre Graduiertenkolleg AIPHES (Adaptive Aufbereitung von Informationen aus heterogenen Quellen) und ist Sprecherin des Zentrums für Digitale Forschung in den Geistes-, Sozial- und Bildungswissenschaften (CEDIFOR).

MARIO HAIM, Dr., ist Postdoctoral Fellow am Institut für Medien und Sozialwissenschaft der Universität Stavanger, Norwegen. Seine Forschungsschwerpunkte umfassen die Digitalisierung und Algorithmisierung, Journalismusforschung sowie die Methodenforschung.

PABLO JOST, M.A., ist Doktorand am Institut für Publizistik der Universität Mainz und promoviert über die Bedeutung digitaler Spurendaten in der Kommunikationsforschung. Zu seinen weiteren Forschungsinteressen zählen die journalistische Selektion und Verarbeitung gesellschaftlicher Streitfragen, die Kommunikation politischer Entscheidungsträger – insbesondere unter den Bedingungen des digitalen Medienwandels – sowie deren Wirkung auf Rezipienten.

PASCAL JÜRGENS, M.A., ist Doktorand am Institut für Publizistik der Universität Mainz. In seiner Forschung geht er dem Einfluss von Maschinen auf menschliches Verhalten nach – insbesondere in Bezug auf politische Informationsnutzung, Manipulation und die Struktur sozialer Netzwerke. Gegenwärtig promoviert er über die Fragmentierung der Gesellschaft durch Algorithmen.

SIMON KRUSCHINSKI, M.A., ist wissenschaftlicher Mitarbeiter am Institut für Publizistik der Johannes Gutenberg-Universität Mainz. Dort forscht und lehrt er im Bereich der Politischen Kommunikation zu technologiegestützter Wahlkampfkommunikation und Fragestellungen der Computational Communication Science. Er promoviert über die Organisation und Wirkung von professionalisierter interpersonaler Kommunikation am Beispiel des daten-gestützten Tür-zu-Tür-Wahlkampfes bei der Bundestagswahl 2017.

ANNA SOPHIE KÜMPEL, Dr., arbeitet seit Oktober 2013 als wissenschaftliche Mitarbeiterin am Institut für Kommunikationswissenschaft und Medienforschung der LMU München. Ihre Forschungsschwerpunkte liegen in der Rezeptions- und Wirkungsforschung und dabei vorrangig im Bereich der Wahrnehmung, Nutzung und Verbreitung von Nachrichten auf sozialen Netzwerkseiten (SNS).

DANIEL MAIER, Dr., Studium der Sozialwissenschaften an der Universität Passau, Kommunikationswissenschaft an der FU Berlin und Universität Zürich, sowie Studium der Epidemiologie an der Charité Berlin. Seit 2013 wissenschaftlicher Mitarbeiter an der FU Berlin. Forschungsinteressen: politische und soziale Netzwerke, Methoden der computerbasierten Textanalyse, Netzwerkanalyse sowie soziale Simulation.

FRANK MANGOLD, Dr., ist wissenschaftlicher Mitarbeiter an der Universität Hohenheim. Seine Forschungsgebiete umfassen Mediennutzungsforschung, Online-Kommunikation, interpersonale Kommunikation und Meinungsführung, Netzwerkanalysen sowie Methoden der Datenerhebung und quantitativen Datenanalyse.

PHILIPP K. MASUR, Dr., ist wissenschaftlicher Mitarbeiter am Institut für Publizistik der Johannes Gutenberg-Universität Mainz. Zuvor war er wissenschaftlicher Mitarbeiter am Lehrstuhl für Medienpsychologie an der Universität Hohenheim sowie als Gastwissenschaftler am Department of



Communication an der Cornell University (Ithaca, USA). Er studierte Publizistik, Wirtschaftswissenschaften und Philosophie an der Universität Mainz und der Macquarie University in Sydney. In seiner Forschung untersucht er verschiedene Formen der computer-vermittelten Kommunikation mit einem besonderen Fokus auf Privatheits- und Selbstoffenbarungsprozesse, Gruppen- und Normprozesse sowie Medien- und Privatheitskompetenz.

MARCUS MAURER, Prof. Dr., ist Professor für Politische Kommunikation am Institut für Publizistik der Johannes Gutenberg-Universität Mainz. Zuvor war er Professor für Empirische Methoden am Institut für Kommunikationswissenschaft der FSU Jena, sowie Vertretungs- und Gastprofessor an der FU Berlin, der LMU München und der Universität Zürich. 2006 bis 2010 Sprecher der Fachgruppe Methoden der DGPK. 2011 bis 2017 Projektleiter in der DFG-Forscherguppe ›Politische Kommunikation in der Online-Welt‹. Seine Forschungsschwerpunkte sind Politische Kommunikation, Empirische Methoden und Nonverbale Kommunikation.

PHILIPP MÜLLER, Dr., ist Akademischer Rat am Institut für Medien- und Kommunikationswissenschaft der Universität Mannheim. Zuvor war er wissenschaftlicher Mitarbeiter an der Johannes Gutenberg-Universität Mainz und der Ludwig-Maximilians-Universität München. Seine Arbeitsschwerpunkte sind die Rezeption und Wirkung journalistischer Medienberichterstattung, Medienwandel und Digitalisierung sowie politische Kommunikation.

TERESA K. NAAB, Dr., hat an der Hochschule für Musik, Theater und Medien Hannover als wissenschaftliche Mitarbeiterin gearbeitet und dort 2012 mit einer Arbeit zu Gewohnheiten und Ritualen der Fernsehnutzung promoviert. Seit 2013 arbeitet sie als Akademische Rätin a.Z. am Institut für Medien, Wissen und Kommunikation der Universität Augsburg. Ihre Forschungsschwerpunkte umfassen Mediennutzung, digitale Kommunikation, Meinungs- und Medienfreiheit und empirische Methoden der Sozialforschung.

MATTHIAS ORLIKOWSKI, M.Sc., ist wissenschaftlicher Mitarbeiter und Doktorand am Lehrstuhl für Digitale Kulturwissenschaften an der Universität Paderborn. Seine thematischen Schwerpunkte sind automatische textanalytische Methoden für die Kulturwissenschaften sowie Fairness,

Verantwortlichkeit und Transparenz sprachverarbeitender Systeme. Er studierte Soziologie, Informatik und Computerlinguistik an der Universität Bielefeld mit Aufenthalten an der University of Warwick, der Copenhagen Business School und der TU Darmstadt.

ALEXANDER ORT, M. Sc., ist Diplomassistent an der Université de Fribourg. Davor war er wissenschaftlicher Mitarbeiter am Wissenschaftscampus Tübingen. Er studierte Empirische Kommunikationswissenschaft an der Universität Hohenheim. Seine aktuellen Forschungsschwerpunkte umfassen kommunikationswissenschaftliche, empirische Medienrezeptions- und Medienwirkungsforschung sowie Medienpsychologie. In seiner Promotion beschäftigt er sich mit dem Einsatz, der Verarbeitung sowie der Wirkung von Emotionen in der persuasiven Gesundheitskommunikation.

LUKAS P. OTTO, Dr., ist Assistenzprofessor für politische Kommunikation und Journalismusforschung an der Amsterdam School of Communication Research (ASCoR) der Universität Amsterdam. Seine Forschungsschwerpunkte liegen in der politischen Medienwirkungsforschung, insbesondere Wirkung politischer Kommunikation auf politisches Vertrauen und Medienvertrauen, Emotionen in der politischen Kommunikation, sowie dynamische Prozesse in der Rezeptions- und Wirkungsforschung.

CHRISTINA PETER, Dr., 2004 bis 2009 Studium der Kommunikationswissenschaft und Politischen Wissenschaft an der Ludwig-Maximilians-Universität München. Seit Oktober 2009 wissenschaftliche Mitarbeiterin und seit 2015 akademische Rätin a. Z. am Lehrstuhl von Prof. Brosius am Institut für Kommunikationswissenschaft und Medienforschung der Ludwig-Maximilians-Universität München. Forschungsschwerpunkte: Rezeptions- und Wirkungsforschung, Politische Kommunikation, Medienpsychologie.

MICHAEL SCHARKOW, Prof. Dr., wissenschaftlicher Mitarbeiter am Institut für Kommunikationswissenschaft der Universität Hohenheim. Studium der Publizistik- und Kommunikationswissenschaft, Politikwissenschaft und Soziologie an der Freien Universität Berlin. Von 2006 bis 2009 wissenschaftlicher Mitarbeiter am Institut für Theorie und Praxis der Kommunikation an der Universität der Künste Berlin. Forschungsschwerpunkte: sozialwissenschaftliche Methoden (insb. Inhaltsanalyse) und Statistik, Mediennutzungs- und Online-Forschung.

CHRISTIAN SCHEMER, Prof. Dr., ist seit 2014 Professor für Allgemeine Kommunikationsforschung am Institut für Publizistik der Johannes Gutenberg-Universität Mainz. Seine Forschungsschwerpunkte sind Mediendarstellungen sozialer Gruppen und deren Wirkungen auf Vorurteile.

BIRGIT STARK, Prof. Dr., ist Professorin für Kommunikationswissenschaft mit dem inhaltlichen Schwerpunkt Medienkonvergenz am Institut für Publizistik der Johannes Gutenberg-Universität Mainz. Forschungsschwerpunkte sind Themen der Medienkonvergenz, Nutzungs- und Rezeptionsforschung, Methoden der Markt- und Kommunikationsforschung sowie vergleichende Medienforschung.

DANIELA STOLTENBERG, M.A., Studium der Publizistik- und Kommunikationswissenschaft sowie der Politikwissenschaft an der Freien Universität Berlin, Abschluss mit meiner methodischen Masterarbeit zum Thema ›Entdeckung kohäsiver Subgruppen in Hyperlink-Netzwerken‹. 2017 bis 2018 wissenschaftliche Mitarbeiterin an der Arbeitsstelle Kommunikationstheorie und Medienwirkungsforschung der FU Berlin. Seit 2018 wissenschaftliche Mitarbeiterin im Projekt ›Translokale Netzwerke: Öffentlichkeit im Social Web‹ am Arbeitsbereich Digitalisierte Öffentlichkeiten an der Westfälischen Wilhelms-Universität Münster (Teilprojekt des DFG-geförderten Sonderforschungsbereichs: ›Re-Figuration von Räumen‹).

FABIAN THOMAS, M.A., Universität Koblenz-Landau, ist wissenschaftlicher Mitarbeiter und Doktorand am Institut für Kommunikationspsychologie und Medienpädagogik (IKM) an der Universität Koblenz-Landau, Campus Landau. Er interessiert sich für dynamische Prozesse in der Kommunikationswissenschaft (z.B. die reziproke Beziehung zwischen Mediennutzung und Einstellungen oder Verhalten) und die statistische Modellierung solcher Prozesse.

MARTIN WETTSTEIN, Dr., Studium der Publizistikwissenschaft, Soziologie und Computerlinguistik an der Universität Zürich. Derzeit wissenschaftlicher Assistent in der Abteilung Medienpsychologie und -effekte am Institut für Kommunikationswissenschaft und Medienforschung (IKMZ) der Universität Zürich. Seit 2009 Projektassistent am NCCRdemocracy. Arbeitsgebiete: Empirische Methoden der Kommunikationswissenschaft, quantitative Inhaltsanalyse, Rezeptions- und Wirkungsforschung, Politische Kommunikation, Meinungsdynamiken und politische Partizipation über soziale Netzwerkdienste.

## Methoden / Grundlagen



CHRISTINA PETER / TERESA K. NAAB /  
RINALDO KÜHNE (Eds.)

### **Measuring Media Use and Exposure. Recent Developments and Challenges**

*Methoden und Forschungslogik der  
Kommunikationswissenschaft, 14*  
2019, 360 S., 213 x 142 mm, engl.  
ISBN 978-3-86962-287-3

The precise measurement of media use and exposure to media content posits currently one of the main methodological challenges in communication research. Against this background, new communication technologies have been gaining particular importance because they change existing patterns of media use and create new types of media use. At the same time, these technologies do not only present a challenge for communication research, but they also provide new opportunities for the assessment of media use.

The volume regards current developments and trends in the measurement of media use and exposure. It includes contributions on the refinement and advancement of classical measurement approaches, new methods for the assessment of media use and exposure, such as the use of online behavioral data as indicators of media exposure, and the combination of measurement approaches to increase accuracy and validity.



**HERBERT VON HALEM VERLAG**

Schanzenstr. 22 · 51063 Köln  
<http://www.halem-verlag.de>  
[info@halem-verlag.de](mailto:info@halem-verlag.de)

## Methoden / Grundlagen



ANDREAS NIEKLER

### **Automatisierte Verfahren für die Themenanalyse nachrichtenorientierter Textquellen**

*Methoden und Forschungslogik der  
Kommunikationswissenschaft, 13*

2018, 264 S., 41 Abb., 18 Tab., Broschur,  
213 x 142 mm, dt.

ISBN 978-3-86962-261-3

Im Bereich der medienwissenschaftlichen Inhaltsanalyse stellt die Themenanalyse einen wichtigen Bestandteil dar. Für die Analyse großer digitaler Textbestände hinsichtlich thematischer Strukturen ist es deshalb wichtig, das Potenzial automatisierter computergestützter Methoden zu untersuchen. Dabei müssen die methodischen und analytischen Anforderungen der Inhaltsanalyse beachtet werden und es muss abgebildet werden, welche davon auch für die Themenanalyse gelten.

In dieser Arbeit werden die Möglichkeiten der Automatisierung der Themenanalyse und deren Anwendungsperspektiven untersucht. Dabei wird auf theoretische und methodische Grundlagen der Inhaltsanalyse und auf linguistische Theorien zu Themenstrukturen zurückgegriffen. Den wesentlichen Beitrag stellt die Untersuchung der Potenziale und Werkzeuge aus den Bereichen des Data- und Text-Minings dar, die für die inhaltsanalytische Arbeit in Textdatenbanken hilfreich sind und gewinnbringend eingesetzt werden können.



**HERBERT VON HALEM VERLAG**

Schanzenstr. 22 · 51063 Köln  
<http://www.halem-verlag.de>  
[info@halem-verlag.de](mailto:info@halem-verlag.de)

Öffentliche Kommunikationsprozesse sind im Zeitalter der Digitalisierung von einer wachsenden Dynamik geprägt. Dies stellt die Kommunikationsforschung vor erhebliche methodische Herausforderungen. Die Methodenentwicklung steckt noch in den Kinderschuhen, wenn es darum geht, die eng getakteten und komplexen Interaktionsmuster menschlicher Akteure und technischer Strukturen der digitalen Öffentlichkeit adäquat abzubilden. Empirische Studien sind dazu gezwungen, die Komplexität der Dynamiken in der sozialen Realität zu reduzieren, um diese fassbar zu machen. Damit geht jedoch stets die Gefahr einher, entscheidende Aspekte zu übersehen.

Die im vorliegenden Band versammelten Beiträge widmen sich diesem Dilemma am Beispiel verschiedener Anwendungsfelder, von der Kommunikator- und Medieninhaltsforschung bis zur Rezeptions- und Wirkungsforschung. Die Beiträge liegen auf verschiedenen Stufen des Forschungsprozesses und befassen sich mit einer Vielzahl methodischer Ansätze wie der automatisierten Inhaltsanalyse, der Netzwerkanalyse oder der qualitativen Beobachtung. Sie eint die Suche nach innovativen Lösungen für ein gemeinsames Problem, nämlich die zunehmende Dynamik öffentlicher Kommunikationsprozesse adäquat abzubilden.

ISBN 978-3-86962-404-4



9 783869 624044 >

EUR [D] 29,50  
EUR [A] 30,40