Dissertation zur Erlangung des Grades eines Doktors der Naturwissenschaften
(Dr. rer. nat.) der Fakultät für Wirtschaftswissenschaften der
Universität Duisburg Essen.

ONLINE-PROFILING

ANALYSE UND INTERVENTION ZUM SCHUTZ VON PRIVATHEIT

Vorgelegt von

Martin Degeling

geboren in Bocholt.

Datum der mündlichen Prüfung: 29. April 2016

Erstgutachter: Prof. Dr.-Ing. Thomas Herrmann, Ruhr-Universität Bochum

Zweitgutachterin: Prof. Dr. Bettina Berendt, Katholieke Universiteit Leuven

1. ZUSAMMENFASSUNG / ABSTRACT

Online-Profiling wird insbesondere im Online-Marketing intensiv genutzt, um eigentlich anonyme Webseitenbesucher_innen zu kategorisieren und ihnen auf Basis ihres Surfverhaltens bestimmte Eigenschaften zuzuschreiben. Erweiterte Einsatzzwecke, wie etwa das Kreditscoring, zeigen allerdings, dass die vermeintlich anonymen Profile Auswirkungen auf die informationelle Selbstbestimmung und letztlich die Autonomie des_der Betroffenen haben. Neben einer Analyse der Hintergründe von Profiling und den Auswirkungen auf Privatheit, stellt diese Arbeit Möglichkeiten für Nutzer_innen vor, die Transparenz von und Intervenierbarkeit in Profile herstellen können. Dazu wurde ein neues Verfahren zur user-zentrierten Online-Tracking-Analyse entwickelt und der Nachweis erbracht, dass die vorgestellten Methoden zur *Obfuscation* (Verschleierung) einen Effekt auf die Interessenprofile eines Anbieters haben.

Online-Profiling is heavily used in online marketing to categorize website visitors that consider themselves to be anonymous. Profiles are assigned based on Online-Tracking techniques that try to reconstruct a user's web history. The use for these profiles is continuously expanded and currently also in use for credit scoring. This development highlights the influence profiling has on the informational self-determination and therefore the autonomy of those being profiled. Besides an analysis of Online-Profiling this dissertation contains new approaches to transparency and intervenability for web users. A new methodology is used for a user-centered Online-Tracking study and we provide evidence that the obfuscation method we developed is effective in influencing the interest profiles created by one large tracking provider.

2. INHALTSVERZEICHNIS

1. Einleitung 7
1.1 Online-Tracking und -Profiling 9
1.2 Forschungskontext 11
1.2.1 Privacy Enhancing Technologies 11
1.2.2 Privatheit und Datenschutz 12
1.2.3 Forschung zu den Folgen von Profiling 13
1.3 Leitfragen 15
1.4 Zusammenfassung der Ergebnisse 16
1.4.1 Grenzen der Autonomie gegenüber Profiling 16
1.4.2 Nutzer_innen-zentrierte Tracking Analyse 17
1.4.3 Transparenz über Profile 17
1.4.4 Intervention in Profile 17
1.5 Aufbau der Arbeit 17
1.5.1 Weitere Anmerkungen 18
2. Profiling und Privatheit 19
2.1 Begriffsbestimmung: Privatheit und informationelle Selbstbestimmung 20
2.1.1 Die Differenz öffentlich/privat 22
2.1.2 Kritik und Grenzen des Autonomiebegriffs 24
2.1.3 Der Wert des Privaten 25
2.2 Datenschutz: Schutz informationeller Privatheit 30
2.2.1 Recht auf informationelle Selbstbestimmung 30
2.2.2 Datenschutzrecht und Profiling 31
2.2.3 Grenzen des Rechts 33
2.3 Begriffsbestimmung Profiling 38
2.3.1 Profiling und Profile 38
2.3.2 Data-Mining und Profiling 39
2.3.3 Anwendungsbeispiele von Profiling 42
2.4 Eine Gesellschaft des Profilings 47
2.4.1 Profiling als Bevormundung 47
2.4.2 Profiling als Teil des Marktgeschehens 49
2.4.3 Profiling und Diskriminierung 51
2.4.4 Liquid Profiling und Identitätskonstruktion 52
2.4.5 Profiling und Kybernetik 56
2.5 Vorschläge zu Regulierung von Profiling 59
2.5.1 Konzeptionelle Ansätze 60

2.5.2 Rechtliche Ansätze....61
2.5.3 Technische Ansätze....63
2.6 Zusammenfassung....66
3. Onlinewerbung, User-Tracking und Privacy Enhancing Technologies....68
3.1 Grundlagen des Marketings....68
3.2 Funktionsweise von OnlineMarketing....70
3.3 Online-Tracking Techniken....74
3.3.1 Differenzierung von Profilen....77
3.4 Theoretische Ansätze zum Schutz von Privatheit gegen Tracking....79
3.4.1 Privacy Enhanced Tracking....79
3.4.2 Nutzer_innenkontrolle und Obfuscation....80
3.5 Analyse von Privacy/Transparency Enhancing Technologies....84
3.5.1 Beschreibung der Bewertungskriterien....84
3.5.2 Tracking-Dienste & Transparenz....87
3.5.3 Anbieterseitige Tools zur Nutzer_innen-kontrolle....95
3.5.4 Blocking....98
3.5.5 Obfuscation....104
3.5.6 Transparenz....107
3.6 Zusammenfassung....112
4. Empirische Analyse von Online-Profiling und Obfuscation....114
4.1 Übergeordnete Anforderungen....114
4.2 Methode zur Automatisierten Analyse von Profiling....116
4.2.1 Werkzeuge....116
4.2.2 Grenzen....117
4.3 Datenerhebung mit TrackTrack....119
4.3.1 Datenquelle Reddit....121
4.3.2 Datenquelle Google....124
4.3.3 Datenquelle Quantcast....127
4.3.4 Datenquelle Alexa....130
4.3.5 Datenquelle Compete....131
4.3.6 Datenquelle Open Directory Project....132
4.4 Analyse des Trackings....132
4.5.1 Durchschnitt der Nutzer_innen....132
4.5.2 Aufgerufene Webseiten....133
4.5.3 HTTP-Requests....134
4.5.4 Cookies....135
4.5.5 Umfang des Trackings....136

4.6 Interessenprofile....137
4.7 Soziodemografische Profile....139
4.8 Evaluation eigener Profiling und Obfuscation Verfahren....142
4.8.1 Berechnung eines Interessenprofils....142
4.8.2 Berechnung eines soziodemografischen Profils....146
4.8.3 Obfuscation-Optionen....147
4.9 Zusammenfassung....155
5. Entwicklung und Evaluation eines Transparenz- und Obfuscation-AddOns....156
5.1 Designziele....156
5.1.1 Transparenz....156
5.1.2 Privacy Literacy....158
5.1.3 Usability....159
5.2 TrickTrack: Das Browser Plugin....159
5.2.1 Basis Informationen....160
5.2.2 Interessenprofil....161
5.2.3 Soziodemografisches Profil....163
5.2.4 Weiterführende Informationen....164
5.2.5 Abwägungen und Zusammenfassung....165
5.3 Evaluation....167
5.3.1 Durchführung....167
5.3.2 Auswertung der Interviews....170
5.3.3 Auswertung Beobachtung....177
5.4 Ergebnis der Evaluation....183
5.5 Verbesserungspotentiale....186
5.6 Zusammenfassung....187
6. Fazit....189
6.1 Beantwortung der Leitfragen....189
6.2 Zentrale Forschungsbeiträge....192
6.3 Anknüpfungspunkte für weitere Arbeiten....193
7. Literaturverzeichnis....194
8. Danksagung....207
9. Eigenständigkeitserklärung....208
10. Anhang....209
10.1 Abbildungsverzeichnis....209
10.2 Tabellenverzeichnis....210
10.3 Quelltextverzeichnis....211
10.4 Formelverzeichnis....211

10.5 Verwendete Software und Bibliotheken....212
10.6 Vergleich Aller ObfuscationMethoden....213
10.7 Korrelationen von Interessen....214
10.8 Korrektheit der soziodemografischen Profile im Tests....215
10.9 Dokumente der Nutzer_innen Befragung....216
10.9.1 Einverständniserklärung....216
10.9.2 Einleitung in die Befragung....217
10.9.3 Vorabinterview....217
10.9.1 Nutzung....218
10.9.2 Nachbefragung oder Beobachtung....218
10.10 QUIS Fragebogen (Auswahl)....219
10.10.1 Gesamteindruck....219
10.10.2 Darstellung....220
10.10.3 Terminologie und System-Informationen....221
10.10.4 Lernfortschritt....223
10.10.5 Systemeigenschaften....224
10.10.6 Statistiken....225

1. EINLEITUNG

Profiling ist nicht erst seit dem Aufkommen von *Big Data* ein Teil vielfältiger Systeme und Dienstleistungen. Das automatische Einschätzen und Kategorisieren von Personen anhand des beobachteten Verhaltens oder gemessener Eigenschaften gehört zum Alltag der Informationsgesellschaft. Es wird im Versicherungs- oder Kreditwesen genauso eingesetzt wie bei der inneren und äußeren Sicherheit oder dem Internetmarketing. Dabei werden große Mengen von personenbezogenen und nicht-personenbezogenen Daten erhoben, verarbeitet, aggregiert, korreliert und genutzt, um Dienstleistungen und Produkte zu verändern. Am deutlichsten zeigt sich diese Entwicklung im Internet. Für Nutzer_innen[1] manifestieren sich automatisch erstellte Profile am häufigsten in personalisierter Werbung, die Effekt des O*nline Behavioural Advertising* und T*argeting* sind. Wie die Profile beziehungsweise die Verfahren, die diese errechnen, aussehen, ist in der Regel Geschäftsgeheimnis und für den_die Betroffene_n weder transparent noch beeinflussbar. Dabei hängt die Frage nach der Funktionsweise und dem Sinn dieser Technologie nicht nur mit dem Recht auf informationelle Selbstbestimmung und der Freiwilligkeit solcher Maßnahmen zusammen, sondern auch, allgemeiner gesprochen, mit dem Einfluss auf die Autonomie des oder der Einzelnen. Die Möglichkeiten, autonom zu handeln, sind eingeschränkt, wenn kybernetisch agierende Systeme sich auf Basis automatisierter Entscheidungen anpassen und versuchen, den oder die Nutzer_in zu beeinflussen. Insbesondere wenn unklar ist, aufgrund welchen Profils man in einer Situation wie angesprochen wird, ist dies eine Einschränkung von Autonomie, unabhängig davon, ob die Personalisierung im Sinn des_der Betroffenen ist oder nicht.

Das Ziel dieser Arbeit ist es daher, das Wissen über die Funktionsweise und Wirkungen von Profiling in einem davon besonders stark durchdrungenen Bereich – dem auf Online-Tracking basierenden Marketing – zu vertiefen.

Ein populäres Beispiel (nach Duhigg, 2012) verdeutlicht die Folgen von Profiling: Die US-amerikanische Drogerie und Einzelhandelskette Target ermittelt, automatisiert für alle Kundinnen, einen *pregnancy prediction score*. Anhand des Einkaufsverhaltens, das mit Bonus- und Kreditkarten online und offline getrackt wird, errechnet das Unternehmen die Wahrscheinlichkeit einer Schwangerschaft und auch gegebenenfalls des aktuellen Status und erstellt daraus ein Schwangerschaftsprofil. Nach dem Be-

1 In dieser Arbeit wird weitestgehend der Gender_Gap verwendet. Durch den Unterstrich soll Geschlechteridentitäten jenseits einer binären Ordnung symbolisch Platz eingeräumt werden. Die Irritation des Leseflusses, die möglicherweise einige Leser_innen auf den ersten Seiten erfahren, ist dabei durchaus intendiert (siehe HTTP://DE.WIKIPEDIA.ORG/WIKI/ GENDER_GAP_(LINGUISTIK), letzter Zugriff 26.09.2016).

richt wurde dieses Profil auch für eine 16-Jährige erstellt, die daraufhin postalisch Werbung für Schwangerschaftspflege und Babyprodukte zugesandt bekam. Durch diese offensichtliche Werbung wurden die Eltern misstrauisch, und die bis dahin geheim gehaltene Schwangerschaft wurde offenbart. Das Ziel von Target, durch den Algorithmus die Kundin davon zu überzeugen, mehr Produkte zu kaufen, hat hier die Handlungsmöglichkeiten der 16-Jährigen eingeschränkt und die Entscheidungsfreiheit darüber, die Schwangerschaft geheim zu halten, genommen. Selbst wenn die Vorhersage falsch gewesen wäre, hätte das Profil sicher unangenehme Folgen gehabt.

Im stationären Handel braucht es noch die Zuordnung von Adressen zum Profil, um Werbung zu versenden und so Einfluss nehmen zu können. Im Online-Handel wäre dasselbe Beispiel wesentlich einfacher zu realisieren und eine Mitwirkung der Betroffenen, etwa über Kunde_innenkarten, nicht notwendig. Bei jedem Seitenaufruf ist eine Vielzahl von Akteuren involviert, die Annahmen über Interessen, Gender, Alter, Einkommen oder Ethnizität treffen, um darauf aufbauend Werbekampagnen zu steuern und Versuche in die Wege zu leiten, das (Kauf-)Verhalten des_r Seitenaufrufenden zu beeinflussen. Steigt der Umfang dieser Beeinflussungsversuche und der Personalisierung immer weiter, führt dies zu einem Diskurs um die individuelle Autonomie in Informationsräumen. Häufig äußert sich dieser Diskurs in Datenschutzdebatten und rechtlichen Einschätzungen über die Grenzen solcher Verfahren.

Im Gegensatz zum obigen Beispiel ist es im Internet nicht nötig, den Namen oder die Adresse der Person zu kennen, über die ein Profil erstellt wird. Personenbezogene Daten im Sinn des Datenschutzes werden nur eingeschränkt verarbeitet. Oft wird mit abstrahierten und pseudonymisierten Profilen gearbeitet, die auf einer temporären Zuordnung von einem Profil zu einer Browsersession beruhen. Vom Datenschutzrecht wird dieses Vorgehen nur mangelhaft erfasst, weil – so eine zentrale These dieser Arbeit – die entindividualisierte Datenverarbeitung des Profiling auch konzeptionell ohne einen Personenbezug arbeitet und daher durch Datenschutz nicht regulierbar ist. Dennoch verletzt Profiling Dimensionen von Privatheit, die eigentlich im Schutzbereich des Datenschutzes liegen. Es mangelt vor allem an Transparenz und Interventionsmöglichkeiten für den_die Einzelne_n, aber auch für alle User_innen als Gruppe.

Zusammenfassung

In dieser Arbeit wird Profiling auf zwei Arten untersucht. Zuerst erfolgt eine theoretische Analyse des Einflusses von Profiling auf verschiedene Aspekte von Privatheit. Dabei werden die Grenzen der aktuellen technischen wie datenschutzrechtlichen Auseinandersetzung aufgezeigt, die vom Individuum und dessen Recht auf Selbstkontrollle und -bestimmtheit ausgehen. Methoden des Profiling arbeiten dagegen mit einer kybernetischen Hypothese, die auch ohne den Bezug auf ein Individuum auskommt. Aus dieser Diskussion werden dann Anforderungen an P*rivacy* und *Transparency Enhancing Technologies* entwickelt, die diese grundlegend andere, implizite Konzeption des-

sen, was eine Person ist, berücksichtigen und dennoch Möglichkeiten schaffen, Privatheit zu erhalten. Diese Anforderungen werden dann am Beispiel einer Browser-Erweiterung umgesetzt, deren Ziel es ist, Internetnutzer_innen Einsicht in Profile zu geben, die über sie erstellt werden können, und es ermöglicht, sie zu beeinflussen. Abschließend werden Nützlichkeit und Nutzbarkeit des entworfenen Werkzeugs evaluiert und die weiteren Entwicklungsmöglichkeiten aufgezeigt.

1.1 ONLINE-TRACKING UND -PROFILING

Profiling

Profiling ist in vielen unterschiedlichen Bereichen im Einsatz, im Bankgeschäft wie im Versicherungswesen, im Online-Marketing, aber auch in verschiedenen Bereichen der Sicherheitsdienste (siehe Kapitel 2). Auch wenn die Grundannahmen ähnlich sind, ist die konkrete Ausprägung dessen, was als „Profil“ definiert wird, kontextabhängig.[2] Die anwendungsbezogene Analyse dieser Arbeit fokussiert vor allem das Profiling von Internetnutzer_innen. Hier sind bereits einige informatische Vorarbeiten geleistet worden, und das Feld ist leichter zugänglich, als etwa Profiling im Bereich der Sicherheitsdienste.

Online-Tracking

Webseitenbetreibende und Werbeanbietende haben ein Interesse daran, etwas über die Besucher_innen einer Seite zu erfahren, um Werbung und Inhalte dem vermuteten Publikum anzupassen. Um das Verhalten der Nutzer_innen analysieren zu können, versuchen Tracking-Services über möglichst viele Webseitenaufrufe informiert zu werden (siehe Abbildung 1 und Kapitel 3). Neben den sichtbaren Inhalten der Webseite, die ein_e Anbieter_in bereit stellt, werden sichtbare und unsichtbare Zusatzfunktionen geladen, die auf unterschiedlichste Weise den_die Nutzer_in identifizieren und dessen_deren Interaktion mit der Webseite verfolgen. Es wird erfasst, welche Links angeklickt werden, wie lange an welcher Stelle innegehalten wird, welche Einstellungen (Sprache, Bildschirmauflösung u. a.) der Browser hat, mit dem die Seite aufgerufen wird, und aus welcher Region der_die User_in stammt. Der am weitesten verbreitete Tracking-Service zurzeit ist Google, dessen Skripte auf über 80 % der genutzten Webseiten verwendet werden (vgl. Kapitel 4). Ziel der Dienste ist es, einen möglichst großen Teil des Surfverhaltens einer_s Nutzer_in zu tracken. Das Funktionieren des Trackings in der aktuell verbreitetsten Form hängt von der Verfügbarkeit und dauerhaften Speicherung von Cookies im Browser des_der Nutzer_in ab. Außerdem können Profile so nur über die Nutzung auf einem Gerät erstellt werden.

Die Nutzungsdaten, die durch Tracking gewonnen werden, werden beim Profiling mit weiteren Datenquellen verknüpft, um über das Verhalten hinaus persönliche Eigenschaften des oder der getrackten Person zu errechnen. Dazu gehören Alter, Ge-

2 Vgl. zur Begriffsbestimmung des Profils (Weich 2016).

schlecht, Einkommensgruppen oder auch die politische Orientierung (vgl. Abschnitt 4.3), die zu einem Profil zusammengefügt werden. Dieses Profil wird dann genutzt, um Webseiten und Werbeanzeigen anzupassen und so die Interaktion der User_in mit den Seiten zu beeinflussen.

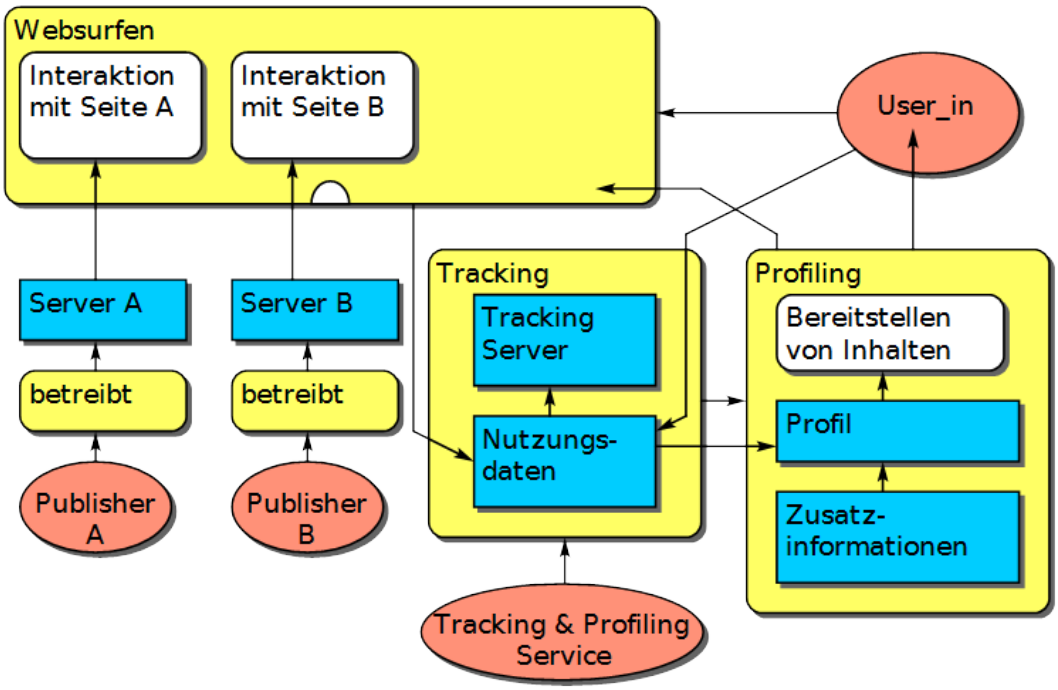


Abbildung 1: Schematische Darstellung[3] von Online-Tracking mit 2 Webseiten (A und B) und einem Trackingservice

Surfverhalten als sensible Daten

Im Gegensatz zur Allgegenwart des Profiling, das auf Seiten der Webseitenbetreiber_innen als notwendig und nützlich betrachtet wird, steht die Wahrnehmung der Nutzer_innen. Nach einer Umfrage des Pew Research Center (2014) ist die Liste der Webseiten, die ein_e Internetnutzer_in aufgerufen hat (Browserverlauf) für 60 % der Befragten eine sensible Information. Das Melden dieser Information an Google und andere Anbieter_innen wird daher häufig als Überwachung und Verletzung der Privatsphäre wahrgenommen. Darüber hinaus zeigt die Untersuchung von Turow, Hennessy und Draper (2015), dass die meisten Nutzer_innen diese Daten nur widerwillig preisgeben. Die rechtlich notwendige Einwilligung basiert in den meisten Fällen darauf, dass die Einwilligung nur implizit gegeben wird oder kein Widerspruch möglich ist. Ähnlich ist das Ergebnis einer auf Europa beschränkten Befragung von Symantec (2015), nach der Nutzer_innen die Diskrepanz zwischen der eigenen Ein-

3 Die schematische Darstellung in Abbildung 1 folgen den Regeln der Modellierungsnotation SeeMe (Herrmann 2006). Kreise stellen Rollen dar, Aktivitäten werden durch Rechtecke mit abgerundeten Ecken repräsentiert und die Rechtecke zeigen Entitäten. Die Richtung der Relationen/Pfeile ist ebenfalls von Bedeutung. Entitäten können von Aktivitäten genutzt (E zu A) oder verändert (A zu E) werden und Rollen beschreiben (E zu R). Der weiße Halbkreis zeigt die Unvollständigkeit des (Sub-)Modells an der jeweiligen Stelle an.

schätzung, wie privat etwas sein sollte und für wie privat die Befragten es in der Praxis halten, beim Browserverlauf am größten ist. Nicht zuletzt möchten viele Nutzer_innen das Surfen im Internet, d. h. ihre Bewegung im Informationsraum Internet, als Verhalten in einem privaten, von Dritten unbeobachtet und unbeeinflussten Raum geschützt wissen. In derselben Umfrage lehnten zudem 61 % der Befragten die Personalisierung von Internetdiensten auf Basis ihrer personenbezogenen Daten ab.

1.2 FORSCHUNGSKONTEXT

Unter anderem aufgrund dieses Widerspruchs zwischen technisch Möglichem und gesellschaftlich Erwünschten ist Online-Tracking regelmäßig Thema – nicht nur öffentlicher Berichterstattung, sondern auch der wissenschaftlichen Forschung. Neben Fragen von Privatheit/Privacy und Datenschutz, die in verschiedenen wissenschaftlichen Disziplinen diskutiert werden, ist Online-Tracking für die Informatik eines der wesentlichen Entwicklungsfelder für *Privacy* und *Transparency Enhancing Technologies (PET/TET)*.

1.2.1 Privacy Enhancing Technologies

Beitrag/ Forschungslücke

Die Gestaltung solcher Technologien wird in verschiedenen Feldern der Informatik, wie Mensch-Maschine-Interaktion (CHI) und Computerunterstützung (CSCW), in speziell darauf ausgerichteten Workshops und auf Konferenzen diskutiert (SOUPS und PETS, CPSP). Wissenschaftliche Veröffentlichungen untersuchten bisher vor allem die technischen Bedingungen von Tracking, Möglichkeiten für Transparenz und Kontrolle von Tracking sowie die Nutzbarkeit von PETs. Darüber hinaus werden zu dem Begriff der *Obfuscation* Verfahren entwickelt, die Nutzer_innen eine aktivere Rolle zuweisen.[4] Dabei handelt es sich um meist automatisierte Verfahren, Profilinformationen zu verschleiern, indem die Auswertung von Rohdaten erschwert wird. Allerdings kritisieren Brunton und Nissenbaum (2011) den Mangel an wissenschaftlichen Untersuchungen zu den Effekten von Verschleierung nicht nur beim Online-Tracking. Dieser Nachweis der Effektivität wird in dieser Arbeit anhand einer Praxisstudie erstmals erbracht.

Die Notwendigkeit der weiteren wissenschaftlichen Auseinandersetzung besteht auch aus der Perspektive einer anwendungsorientierten Wissenschaft, die den Nutzen technischer Verfahren auch immer von der Anwendbarkeit in einem sozialen Kontext abhängig macht. McDonald und Cranor (2010) stellen fest, dass, obwohl die technischen Möglichkeiten vorhanden sind, nur eine begrenzte Zahl von Nutzer_innen diese kennen und für einen selbstbestimmten Umgang – etwa mit Cookies oder Opt-out Me-

4 Zuletzt wurde nach der Veröffentlichung des Buchs „Obfuscation: a user's guide for privacy and protest“ (Brunton und Nissenbaum 2015) eine breitere Debatte zu dem Thema angestoßen. Ein ausführlicher Überblick über den Stand der Forschung in der Informatik findet sich in Kapitel 3.

chanismen – einsetzen können. Leon u. a. (2012) führten dies unter anderem auf die schlechte Gestaltung der Tools aus Sicht der Usability zurück.

1.2.2 Privatheit und Datenschutz

Neben der technisch geprägten Auseinandersetzung in der Informatik wird Profiling interdisziplinär in (medien-)philosophischen, juristischen und politik-/sozialwissenschaftlichen Fächern erforscht. Personalisierung im Zusammenspiel mit anderen Techniken der Überwachung wird in Graduiertenkollegs[5], Forschungsprojekten (z. B. „Forum Privatheit“ oder „Strukturwandel des Privaten“[6]) und in Forschungsbereichen wie den S*urveillance Studies* (Lyon 2002) untersucht.

Die interdisziplinäre Diskussion wird in dieser Arbeit aufgenommen und davon ausgehend ein eigener Beitrag zum Verständnis von Profiling im Verhältnis zu Privatheit vorgestellt. Aus der theoretischen Auseinandersetzung werden dann Anforderungen abgeleitet, die wiederum die technische Perspektive informieren. Um diese unterschiedlichen Perspektiven einnehmen zu können, wird nicht nur der technische Aspekt des Trackings, sondern der sozio-technische Profiling-Prozess als Ganzes betrachtet. Dieser umfasst neben der Erhebung der Tracking-Daten auch die Überführung in ein Profil, das Annahmen über den_die Nutzer_in trifft. Aus der Analyse dieses Profilings wird eine *Privacy Enhancing Technology* entwickelt, die nicht nur Tracking, sondern auch die Auswertung und Profilerstellung betrifft. Bisherige Arbeiten zu Online-Tracking haben durch die Betonung der technischen Seite häufig einen abstrakten Blick auf die Personalisierung geworfen. So existieren vor allem Studien zur Verbreitung bestimmter Tracking-Mechanismen oder zum Umfang von Tracking im Allgemeinen. Die in dieser Arbeit vorgestellte Analyse wiederum betrachtet Tracking nicht isoliert, sondern immer in Bezug auf die Nutzer_innen.

Informationelle Selbstbestimmung

Online-Tracking erfolgt meist ohne explizite und informierte Einwilligung der Nutzer_innen. Dies widerspricht einer Idee von informationeller Selbstbestimmung, wie sie im europäischen Raum gesetzlich verankert ist und in Deutschland aus dem Grundgesetz abgeleitet wird. Anstatt zu wissen, wer was wann über einen weiß, ist im

5 Beispiele sind die Graduiertenkollegs *Automatismen* (HTTP://WWW2.UNI-PADERBORN.DE/INSTITUTE-EINRICHTUNGEN/GK-AUTOMATISMEN) und *Privatheit* (HTTP://PRIVATHEIT.UNI-PASSAU.DE/). Bei letzterem ist der Autor kooptiertes Mitglied. (letzter Zugriff auf beide Seiten 26.09.2016)

6 Dies ist nur eine Auswahl aktueller, deutschsprachiger Projekte vgl. HTTPS://WWW.FORUM-PRIVATHEIT.DE/ und HTTP://WWW.STRUKTURWANDELDESPRIVATEN.DE/ (im europäischen Rahmen wurden bereits vor einiger Zeit Projekte wie FIDIS (HTTP://FIDIS-PROJECT.EU/) oder SPION (HTTP://WWW.SPION.ME/) gefördert, in den USA ist „Privacy“ im Fokus mehrerer Forschungseinrichtungen wie dem *Center for Internet and Society* (HTTP://CYBERLAW.STANFORD.EDU/FOCUS-AREAS/PRIVACY) oder dem *CyLab* (HTTPS://WWW.CYLAB.CMU.EDU/). (letzter Zugriff auf alle Seiten 26.09.2016)

Internet häufig eher das Gegenteil der Fall. Diejenigen Akteure, die Profile erstellen und nutzen, sind intransparent, da nicht ansprechbar und vielzählig. Die Art der erhobenen Informationen ist unbekannt und wie und zu welchem Zweck die Informationen genutzt werden, bleibt vollständig im Dunkeln.

Privatheit

Das in Deutschland aus dem Grundgesetz abgeleitete Recht auf informationelle Selbstbestimmung findet seine Begründung in der Idee der individuellen Autonomie, die alle liberalen Gesellschaftsordnungen propagieren. Der Wert von Privatheit liegt darin, dass er Ausbildung und Ausübung von Autonomie gewährleisten kann. Autonomie wird dabei allerdings heute nicht mehr als absolute, individuelle Eigenschaft gesehen, sondern im Rahmen von Beziehungen erlernt (vgl. Rössler 2001). Wie Privatheit funktioniert und geschützt werden kann, entzieht sich allerdings einer eindimensionalen Beschreibung. Neben informationeller Privatheit beschreibt Rössler dezisionale Privatheit als Raum zur autonomen Entscheidungsfindung. Auch dies ist eine Dimension von Privatheit, die durch Profiling beeinflusst wird, indem Entscheidungsspielräume – teilweise mit guten Absichten – einschränkt werden.

kybernetisch

Dennoch gelingt es den Gesetzgebern kaum, Tracking und Profiling, etwa durch Datenschutz, einzuschränken, da sie, anders als andere Überwachungstechnologien, die Einzelnen nur indirekt betreffen. Im Grunde beziehen sich Profiling-Systeme nicht auf ein Individuum, dessen Privatheit zu schützen ist. Beim Profiling geht es vielmehr um die Regulierung aller Internetsurfer_innen, die als ein System verstanden werden. Es geht daher nicht um personenbezogene, sondern um gruppenbezogene Daten, nicht um die genaue Beschreibung einer individuellen Person, sondern die Aggregation von *dividuellen* Repräsentationen durch Daten (Deleuze 1990; Galloway 2004). Deren Zweck ist nicht, absolute Aussagen zu treffen, sondern Wahrscheinlichkeiten von Merkmalen und Attributen zu ermitteln. Daraus können Schlüsse gezogen werden, wie das System zu beeinflussen ist. In diesem Sinn agieren Profiling und andere Big-Data-Verfahren nicht auf Basis einer liberalen, sondern einer kybernetischen Hypothese über die Menschen.

1.2.3 Forschung zu den Folgen von Profiling

Auch wenn der Fokus in weiten Teilen dieser Arbeit auf Online-Profiling liegt, bezieht sich die Motivation dieser Arbeit auf die vielfältigen Anwendungsbereiche von Profiling. Die negativen Folgen dieser und ähnlicher Technologien sind bereits Thema vielfältiger Veröffentlichungen gewesen.

Ausschluss

Im Jahr 2000 hat ein amerikanisches Unternehmen einen simplen Profiling-Mechanismus eingesetzt, um Nutzer_innen auf Basis ihres Postleitzahlcodes von seinen Diensten auszuschließen (Danna und Gandy 2002). Diese als *Redlining* bekannte Praxis ist

zwar auch in den USA verboten, aber komplexere Verfahren (z. B. Scoring) können heute dazu genutzt werden, Ausschlüsse zu produzieren, indem bestimmte Produkte den Ausgeschlossenen verborgen bleiben oder zu überhöhten Preisen angeboten werden.

Preisanpassung

Weniger binär ist die automatische Preisanpassung auf Webseiten anhand von Profilen, die nach einer Untersuchung des Wall Street Journals breit diskutiert wurde (Valentino-DeVries, Singer-Vine, und Soltani 2012). Ein Anbieter, z. B. ein Online-Shop, kann versuchen, über die Analyse eines Kund_innenprofils eine Vermutung über den maximalen Preis anzustellen, den ein_e Kund_in bereit ist zu zahlen. Auch wenn es Zweifel an der Verbreitung dieser Methode gibt (Vissers u. a. 2014), so stellt die Idee der perfekten Preisdifferenzierung *(price discrimination)* für Anbieter_innen doch immer noch ein zu erreichendes Ideal dar. Der Internethändler Amazon ändert einer Studie zu folge die Preise von bis zu 20 % seiner Produkte täglich (360pi 2014), und zudem ist die Preisdifferenzierung durch selektiv vergebene Coupons weit verbreitet, die weniger kritisch gesehen wird (Narayanan 2013).

Scoring

Ein möglichst umfangreiches Profil versuchen auch Kreditanbieter_innen über potentielle Kund_innen zu erstellen. Während in 2012 ein Kooperationsprojekt mit der SCHUFA zur Auswertung öffentlich verfügbarer Daten aus *social networks* für das Kreditscoring abgesagt wurde, (Hasso-Plattner-Institut 2012) basieren die Geschäftsmodelle im Ausland tätiger Unternehmen vollständig auf dem *„Big Data Scoring“* (Müller, Rosenbach, und Schulz 2013), bei dem Kleinkredite mit umgerechnet fast 300 % Zinsen unter anderem auf Basis von *Social Media*Daten vergeben werden.

Kriegsführung

Auch in der modernen Kriegsführung werden unter anderem Bewegungs- und Kontaktprofile errechnet. Diese dienen dazu, Personen zu ermitteln, die Kontakt zu Verdächtigen haben oder sich an denselben Orten wie diese aufhalten, und quantifizieren die Gefahr, die von einer Person vermutlich ausgeht. Wie durch die Veröffentlichungen von Edward Snowden seit 2013 bekannt wurde, ist in einigen Fällen die tatsächliche Identität der durch dieses Profiling identifizierten, und teilweise bei Drohnenangriffen getöteten, nicht bekannt (Scahill und Greenwald 2014).

Online-Tracking und personalisierte Werbung haben bei weitem nicht die Konsequenzen wie das letztgenannte Beispiel, dennoch basieren sie auf ähnlichen Prinzipien des Profiling, bei dem ein Datensatz durch Zusatzinformationen angereichert wird, um Entscheidungsprozesse zu automatisieren. In dieser Arbeit wird vornehmlich Profiling im Internet diskutiert, weil es eine Vorreiterrolle bei der Entwicklung dieser Technologien einnimmt und gleichzeitig mit den Methoden der Informatik gut zu untersuchen ist. Gerade das Online-Marketing ist ein rasant wachsender Markt mit hoher Innovationsgeschwindigkeit, von dem aus technische Entwicklungen im Bereich Profiling in andere Bereiche expandieren. In kaum zwanzig Jahren der Internetpopularisierung

wurde aus einem einfachen Werbebanner, das 1994 geschaltet wurde (Turow 2012), eine Echtzeitanzeigenbörse, die aus den immer gleichen Daten (Browser rufen Webseiten auf) immer umfangreichere Profile errechnet, die Persönlichkeiten und deren Verhaltensweisen vorherzusagen versucht.

1.3 LEITFRAGEN

Das übergeordnete Ziel dieser Arbeit ist es, das Wissen über und Analysemöglichkeiten von Profiling-Mechanismen am Beispiel von Online-Tracking in einem interdisziplinären Kontext zu ergänzen. Die Arbeit verwendet und entwickelt Methoden und Theorien aus der Informatik, um die technischen Zusammenhänge zu untersuchen und Möglichkeiten von *Privacy Enhancing Technologies* voranzubringen. Dabei wird die Arbeit inspiriert von einer kulturwissenschaftlich orientierten Auseinandersetzung mit Techniken des Profilings und leistet auch hier Beiträge zum Privatheitsdiskurs.

Geleitet wird die Analyse und Diskussion von vier Leitfragen:

1. *Inwiefern werden Privatheit und Autonomie individuell wie strukturell durch technologische Entwicklungen wie Profiling beeinflusst?*

In einem ersten Schritt steht die Frage im Zentrum, auf welcher Basis die Annahmen zum Profiling fußen und welche Kritik es daran gibt. Dabei zeichnet sich ab, dass die Theorien zur *kybernetischen Hypothese* eine gute Möglichkeit bieten, Profiling zu beschreiben und wirksame wie unwirksame Argumente der Kritik abzuleiten.

2. *Was ist Online-Tracking und Online-Profiling? Wie funktioniert es, wie und zu welchem Zweck wird es eingesetzt?*

Es geht hier einerseits darum, die technischen Grundlagen des Online-Trackings, auf denen Profiling beruht, sowie die Verfahren und die Möglichkeiten der Tracking-Services – Cookies, Browser, Fingerprinting u. a. – zu verstehen. Andererseits wird eine Untersuchung der Anwendung dieser Verfahren in der Praxis durchgeführt. Dazu wurde im Rahmen dieser Arbeit das *TrackTrack*-Framework zur automatisierten Datenerhebung entwickelt.

3. *Wie sehen Profile aus, die durch Online-Tracking generiert werden, und wie werden diese ermittelt?*

Auf Basis der ermittelten Daten wird analysiert, wie Profile aussehen (können) und welche Möglichkeiten es gibt, diese darzustellen. Hierzu wurden Metriken entwickelt, und mit *TrackBack* wurde ein Werkzeug zur Analyse und Visualisierung vorgestellt.

4. *Wie kann das gewonnene Wissen genutzt werden, um von Profiling Betroffenen zu helfen, Profiling zu verstehen und zu beeinflussen?*

Die Erkenntnisse werden genutzt, um das Browser-AddOn *TrickTrack* zu entwickeln. Ziel dieser Software ist es, Transparenz und Intervenierbarkeit in Bezug auf Profile zu schaffen und einen kritischen Umgang mit Profiling zu schulen.

Ziel der Arbeit ist also, eine strukturierte Analyse von Profiling – einerseits durch die Erarbeitung des theoretischen Hintergrunds und andererseits durch eine Erhebung des Umfangs von Online-Tracking in der Praxis, insbesondere der Möglichkeiten aus den zur Verfügung stehenden Daten, Profile zu ermitteln. Das Ergebnis dieser Analyse soll dann in zwei Bereichen nutzbar gemacht werden. Erstens sollen die Erkenntnisse dazu beitragen, ein Verfahren zu entwickeln, das die Profilbildung einzelner Nutzer_innen erschwert. Zweitens soll ein Hilfsmittel entstehen, das Internetnutzer_innen aufzeigen kann, wie und welche Profile über sie entstehen können und welche Folgen dies haben kann.

1.4 ZUSAMMENFASSUNG DER ERGEBNISSE

Insgesamt lassen sich aus der Arbeit an den Leitfragen vier zentrale Ergebnisse ziehen, die in verschiedenen Forschungsbereichen Beiträge leisten, um (Online-)Profiling besser zu verstehen und Schutz von Privatheit in Bezug auf Profiling zu ermöglichen.

1.4.1 Grenzen der Autonomie gegenüber Profiling

In Kapitel 2 wird gezeigt, inwiefern Profiling sich einbetten lässt in die Diskussion um Grundfragen des Datenschutzes. Diese sind bisher geprägt von einer individualisierten Sicht auf den_die Einzelne_n und „sein_ihre Daten“. Dieses an statischen Informationen angelehnte Denken ist kaum in der Lage, die flüchtigen und auf Wahrscheinlichkeiten basierenden Informationen zu berücksichtigen, die beim Profiling entstehen und in vielen Fällen keinen Personenbezug haben. Diese Arbeit soll zeigen, dass die Folgen von Profiling durchaus den Zielen von Datenschutz entgegenstehen, dies aber in vielen Fällen kaum rechtlich reguliert wird. Diese Schlussfolgerung schließt an die Argumentation an, nach der Profile als etwas verstanden werden können, dass Einfluss hat auf das, was in liberalen Gesellschaften als Privatheit verstanden wird, die wiederum begründet ist in dem Wunsch einer gewissen Autonomie. Profiling ist aus dieser Perspektive nicht kompatibel mit einer liberalen Hypothese, auf der demokratische Gesellschaften beruhen. Stattdessen folgt es einer kybernetischen Hypothese, die nur Systeme betrachtet, die sie zu regulieren versucht. Dabei geht es nicht darum, feste Profile zu ermitteln, deren Zuweisung zu einzelnen Personen eindeutig ist. Vielmehr lässt sich gerade im Internet, ein *Liquid Profiling* beobachten, bei dem Klassifizierungen und Profile ständig im Fluss und damit kaum mehr zu regulieren sind.

1.4.2 Nutzer_innen-zentrierte Tracking Analyse

In den Kapiteln 4 und 5 werden Analysen von Online-Tracking vorgestellt, die über bekannte Analysen des Forschungsbereichs hinausgehen. Statt den Umfang von Tracking abhängig von den Internetseiten zu betrachten, die sie einsetzen, wird in dieser Arbeit erstmals der Fokus auf die dabei entstehenden Profile gelegt. Dabei werden die jeweils einem *Surfprofil* zugeordneten Seiten als Ganzes betrachtet. So kann nachgewiesen werden, dass Google theoretisch in der Lage ist, bis zu 80 % der von einer_m Nutzer_in durchgeführten Webseitenaufrufe zu beobachten. Weitere Unternehmen wie Facebook und Twitter, aber auch spezielle Analysedienste, beobachten ebenfalls große Teile eines Surfprofils.

1.4.3 Transparenz über Profile

Die ausführliche Analyse der entstehenden Profile wird genutzt, um das Browser Add-On *TrickTrack* zu entwickeln (vgl. Kapitel 5). Es gibt Nutzer_innen erstmals die Möglichkeit, ausführliche Informationen zu den über sie erstellten (oder potentiell erstellbaren) Profilen zu erlangen. In einer qualitativen Studie mit zehn Nutzer_innen wird nachgewiesen, dass TrickTrack auch für Nicht-Experte_innen benutzbar ist und dabei hilft, die Art und den Umfang von Profilen zu verstehen. Das Plugin hilft dabei, Wissen über Profiling zu vermitteln und damit eine *privacy literacy* zu schulen.

1.4.4 Intervention in Profile

TrickTrack erlaubt außerdem, auf Basis der vorherigen Analyse, die Intervention in die Profile durch das gezielte Ansurfen von Seiten zu ermöglichen, um so ein Profil zu verschleiern (Obfuscation). Diese Form der *informierten Obfuscation* (vgl. 4.8) wurde im Rahmen der Arbeit entwickelt und evaluiert. Im Unterschied zu bisherigen Arbeiten ist die Verschleierung nachvollziehbar und nachweislich erfolgreich.

1.5 AUFBAU DER ARBEIT

Die Arbeit gliedert sich in sechs Kapitel wobei die Kapitel 2 bis 5 die Leitfragen behandeln. Im nächsten Kapitel folgt daher zuerst eine ausführliche Analyse von Profiling mit Bezug auf Privatheit. Im dritten Kapitel werden technische Grundlagen geklärt sowie auf Basis der Vorarbeit die Anforderungen an PET vorgestellt, anhand derer existierende Werkzeuge analysiert werden. In Kapitel 4 folgt die Beschreibung und Auswertung einer Analyse von Online-Profiling und die Konzeption des Verfahrens zu *Obfuscation*. In Kapitel 5 folgt die Beschreibung der Entwicklung und anschließende Evaluation des Browser-AddOns TrickTrack, das wesentliche Erkenntnisse der vorherigen Analyse für nicht-wissenschaftliche Zwecke nutzbar macht.

1.5.1 Weitere Anmerkungen

Beispiel für eine Marginalie

Teile dieser Arbeit sind bereits in Vorträgen und schriftlichen Ausarbeitungen einem Fachpublikum dargelegt worden. Eine kürzere Fassung der in Kapitel 2 angestellten Überlegungen wurde in einem Buchbeitrag veröffentlicht (Degeling 2014). Ein Teil der Analyse aus Kapitel 4 wurde ebenso auf wissenschaftlichen Konferenzen und Workshops präsentiert (Degeling 2015b, 2016) wie das in Kapitel 5 vorgestellte Browser-AddOn (Degeling 2015a, 2015c). Die Vorgehensweise zur Entwicklung datenschutzfreundlicher Technologien ist ein Forschungsfeld, in welchem der Autor mit weiteren Kolleg_innen publiziert hat und deren Ergebnisse an vielen Stellen, teilweise implizit, genutzt werden (Degeling und Nierhoff 2013; Loser und Degeling 2014; Loser, Degeling, und Herrmann 2012).

`Beispiel/für/einen/Verweis`

In der Randspalte werden in Marginalien Stichworte, die den nebenstehenden Abschnitt zusammenfassen, dargestellt. Außerdem finden sich, insbesondere in den späteren Kapiteln, Verweise auf Programme und Skripte, die im Rahmen der Arbeit entwickelt wurden. Diese sind auf der Seite *tricktracking.com* abrufbar.

2. PROFILING UND PRIVATHEIT

In diesem Kapitel wird das Verhältnis von Profiling und Privatheit beschrieben. Es fasst eine, mehrere Disziplinen berührende, Diskussion zusammen und benennt die Besonderheit von Techniken des Profilings in juristischen, medienwissenschaftlichen und sozial-philosophischen Debatten. Am Ende steht die Erkenntnis, dass Profiling einen (negativen) Einfluss auf Privatheit und Autonomie sowie demokratische Grundprinzipien hat, sich aber trotzdem der Regulierung durch Datenschutz entzieht. Daraus wird die Notwendigkeit abgeleitet, eine weitergehende Auseinandersetzung mit konkreten Phänomenen des Profiling zu unternehmen, es werden Rahmenbedingungen für weitere Analysen vorgestellt sowie Möglichkeiten Profiling transparent zu machen und zu beeinflussen.

Motivation

Dieses Kapitel stellt darüber hinaus die Motivation der übrigen Arbeit dar. Andere Arbeiten zu Privatheit, die in einem technischen Fachgebiet entstehen, argumentieren meist juristisch-normativ oder technisch-kompetitiv. Juristisch-normativ ist der Bezug auf gesetzliche Rahmenbedingungen, wie das allgemeine Recht auf informationelle Selbstbestimmung oder konkrete Anwendungsgesetze auf Bundes- und Länderebene in Deutschland, welche die Entwicklung von Schutztechniken begründen. Andere sehen ihre Arbeiten eher auf einer technischen Ebene kompetitiv als *privacy enhancing*, also als Gegenmaßnahme zu der als *privacy disruptive* wahrgenommenen, allgemeinen Entwicklung. In diese Arbeit werden beiden Linien Eingang finden, aber um eine sozial-philosophische Perspektive ergänzt.

Begründung der Herangehensweise

Diese Herangehensweise lässt sich auf drei Arten begründen. Erstens funktionieren die folgenden Argumente auch jenseits einer konkreten rechtlichen Regelung, die darzustellen bei der Betrachtung der Praxis global agierender Konzerne eine eigene Arbeit rechtfertigen würde. Zweitens scheint eine alternative Auseinandersetzung auf Grund der Tatsache lohnenswert, dass die juristische Argumentation über Datenschutz und informationelle Selbstbestimmung in vielen Fällen des Profilings nicht wirksam ist, weil behauptet wird, dass nur anonymisierte Daten verarbeitet werden würden. Und drittens kann so dem oft schwarz-weiß geführten Diskurs um *privacy enhancing* und *privacy disruptive* Technologien eine Facette hinzugefügt werden, die grundsätzlicher danach fragt, wie Privatheit und Autonomie durch technologische Entwicklungen wie Profiling beeinflusst werden.

Zusammenfassung

Im Folgenden werden einige Begriffe erläutert, die helfen, Probleme von Profiling und Privatheit zu analysieren. Das in Deutschland verfassungsrechtlich begründete Grundrecht auf *informationellen Selbstbestimmung* ist Ausdruck eines Verständnisses von liberal-demokratischen Gesellschaften, in denen *Privatheit* der Herstellung und dem

Schutz von Autonomie dient und damit auch Grundbedingung einer demokratischen Gesellschaft ist. Im Rahmen einer vertiefenden Diskussion sollen die bereits im ersten Kapitel erwähnten Beispiel ausgeführt und in die Diskussion um Big Data eingebettet werden. Bei der Gegenüberstellung von technischen Verfahren und liberal-demokratischen Idealen wird klar, dass sie auf Basis unterschiedlicher Prämissen argumentieren. Die These dieser Arbeit lautet, dass im Fall von Profiling die *kybernetische Hypothese* die liberale abgelöst hat und in einem de-individualisierten Verständnis von Identität mündet. Diese Analysen werden zusammengeführt mit unterschiedlichen kulturwissenschaftlichen wie ökonomischen Perspektiven.

Struktur des Kapitels

Das vorliegende Kapitel ist wie folgt organisiert: Zu Beginn erfolgt ein kurzer Abriss der Begriffsgeschichte von Privatheit inklusive einer Differenzierung in unterschiedliche Dimensionen des zeitgenössischen Verständnisses von Privatheit als ein Element von Autonomie. Darauf folgt eine Auseinandersetzung mit den juristischen Begriffen *informationelle Selbstbestimmung* und *Datenschutz* (2.2). An deren Ende werden die grundsätzlichen Grenzen dieser Konzeption in Bezug auf Profiling deutlich, die trotz der Möglichkeit mittels Datenschutz auf die Gestaltung von IT-Systemen Einfluss zu nehmen, bisher keinen wirksamen Schutz von Privatheit im Rahmen von Profiling erreicht hat. Im Abschnitt 2.3 wird dann näher auf Profiling und die technischen Aspekte des Data-Mining eingegangen. Darauf folgt eine genauere Beschreibung des negativen Einflusses von Profiling auf Privatheit und Autonomie (2.4). In den letzten beiden Abschnitten dieses Kapitels werden unterschiedliche Herangehensweisen zur Regulierung von Profiling vorgestellt (2.5) sowie abschließend die Vorteile einer aktiven Strategie von *Privacy Enhancing Technologies* am Beispiel der auch im weiteren Verlauf der Arbeit verfolgten *Obfuscation* aufgezeigt (2.6).

2.1 BEGRIFFSBESTIMMUNG: PRIVATHEIT UND INFORMATIONELLE SELBSTBESTIMMUNG

Seit einigen Jahren ist der Verlust von Privatheit in Verbindung mit informationstechnischen Entwicklungen kontinuierlich Thema des öffentlichen wie wissenschaftlichen Diskurses.[7] Die Verbreitung des Internet – insbesondere die Nutzung sozialer Online-Netzwerke –, staatliche Überwachung, Identitätsdiebstahl und sogar das Abfotografieren von Häuserwänden im Auftrag eines Internetunternehmens – dies alles hat einen vermeintlich negativen Effekt auf das, was wir als Privatsphäre zu kennen glauben. Regelmäßig ist sie „jetzt“ und „endgültig“ verloren. Dabei wird das Ende des Privaten

7 Wie Kammerer (2014) zeigt, ist der Diskurs um das Ende des Privaten noch viel älter und eng mit den Entwicklungen jeweils „neuer“ Medien verbunden. Im Folgenden soll aber vor allem die letzte Episode dieses Ende-Diskurses besprochen werden.

genauso häufig ausgerufen wie seine Notwendigkeit hervorgehoben.[8] Immer wieder betont werden dabei die besondere Stellung des Internet, die Entwicklung von Umgangsweisen mit dem Internet sowie deren Einfluss auf die Konzeption des Privaten. Gleichzeitig wird die Notwendigkeit einer Rückbesinnung auf Privatheit als Element moderner Gesellschaften hervorgehoben. Ein Verlust von Privatheit wird von den Autor_innen meist gleichgesetzt mit einem Verlust von Freiheit der Einzelnen, und damit der Grundlage westlicher Demokratien, in deren Folge das Bestehen liberaler Gesellschaftsordnungen grundsätzlich in Frage stünde. Die Aufgabe, der „wir" uns stellen müssten, beschreibt etwa Peter Schaar in seiner Funktion als Beauftragter des Bundes für Datenschutz und Informationsfreiheit wie folgt:

> Im Ergebnis geht es um nicht weniger als die Entwicklung einer globalen Ethik des Informationszeitalters, in deren Mittelpunkt die Bewahrung und Entwicklung der individuellen Selbstbestimmung steht: Verantwortung statt Kontrolle! (Schaar 2009:215)

Eine ähnlich umfangreiche Zeitenwende beschreiben die Autor_innen Trojanow und Zeh:

> Heute, am Anfang des 21. Jahrhunderts, zeigt sich der Schutz von Privatsphäre und persönlicher Kommunikation in völlig neuer, zentraler Bedeutung für das Fortleben der demokratischen Gesellschaft. (Trojanow und Zeh 2010)

Es ginge also um das Verhältnis Demokratie, Privatheit und Selbstbestimmung in diesem, einem neuen Zeitalter. Dabei existiert bei weitem kein einheitliches Verständnis von dem, was als privat oder was als Privatheit gilt. Im Gegenteil setzt sich in den Diskursen eher ein Verständnis durch, das eine vielschichtige, kontext- und rollenabhängige Beschreibung von Privatheit notwendig macht. Im Folgenden soll in einem kurzen Abriss diese Wandlung von Privatheit als einem emphatischen liberal-demokratischen Anspruch bis hin zur juristischen Umsetzung informationeller Selbstbestimmung nachgezeichnet werden. Dabei stützt sich die Beschreibung insbesondere auf die Arbeit von (Rössler 2001).

Privatheit und Politische Philosophie

Die Arbeit von Rössler vermittelt die vielschichtigen und interdisziplinären Diskurse zur Privatheit auf sozio-technischer wie sozial-philosophischer Ebene. Die folgenden Überlegungen sind aber meist belegt mit Rückgriff auf Autor_innen der politischen Philosophie, der Rössler entstammt. In Anbetracht der Breite der Diskussion um Privatheit in unterschiedlichen Disziplinen könnte man meinen, dass ebenso gut eine soziologische oder - mit Blick auf die folgenden technischen Beschreibungen - medienwissenschaftliche Analyse angebracht wäre. Allerdings besteht der Zweck dieser

8 Dieses Motiv findet sich zum Beispiel bei Kurz und Rieger (2011), Schaar (2009), Sofsky (2009) und Whitaker (1999). Zuletzt erschien im Januar 2015 eine Ausgabe des *Science Magazin* mit dem Titel „The End of Privacy" (Ausgabe 347, Band 6221).

Überlegung darin, im Weiteren etwas zur Diskussion um die Rolle und Ausgestaltung von informatischen Systemen beizutragen. Die dazu häufig genutzten Bezüge zum Datenschutz begründen Überlegungen zur Art und Weise der Ausgestaltung liberaldemokratischer Prozesse, die sich in juristischen Instrumenten zur Regulierung von Datenverarbeitung manifestieren. Der Rückbezug auf die Grundsätze der politischen Philosophie dient nicht nur der Rückversicherung über die im Datenschutz umgesetzten Annahmen. Er bietet darüber hinaus die Möglichkeit einer Neubestimmung der notwendigen Maßnahmen zum Schutz von Privatheit in einer Welt, in der informationelle Selbstbestimmung in zunehmenden Maße durch Profiling eingeschränkt wird.

2.1.1 Die Differenz öffentlich/privat

Um die Differenzierung zwischen den unterschiedlichen Dimensionen von Privatheit besser zu verstehen, ist es hilfreich, die Unterscheidung zwischen dem Öffentlichen und dem Privaten nachzuvollziehen, die dem heutigen Verständnis historisch vorausgeht.

Die Trennung der Polis und des Hauses

Die Trennung von öffentlichem und privatem Leben, wie sie heute vielfach noch als Gesellschaftsordnung angenommen oder zumindest als notwendige soziale Sphärentrennung anerkannt ist, ist nach Hannah Arendt (1960:31) aus der griechischen Antike überliefert.[9] Arendt beschreibt, angelehnt an die Schriften Aristoteles', die griechische Ordnung als eine, bei der jeder Bürger zwei Seinsordnungen angehörte; einerseits der privaten Ordnung des Hauses, als der des „naturhaften Zusammenlebens", sowie andererseits der öffentlichen Ordnung, der „Polis", als Form politischen Zusammenlebens. Dabei ist die *Polis* eine der Naturhaftigkeit entwachsene Form, die nicht mehr nur das Überleben im Verbund von Stämmen und Familien sichern soll. Stattdessen treten die Fähigkeiten des Handelns und Redens, die der Organisation und Strukturierung des gemeinsamen Lebens in der Öffentlichkeit dienen, in den Vordergrund. Das Private ist demgegenüber vor allem auf den Erhalt der_s Einzelnen sowie der Gattung „Mensch" als Ganzes ausgerichtet und gilt somit als durch eine natürliche Ordnung bestimmt. Zu dieser „natürlichen Ordnung" gehörte der Besitz von Sklaven_innen genauso wie die Geschlechtertrennung, bei der die Frauen durch ihre Arbeit die Existenz der Polis erst möglich machen (Benhabib 1991), während der öffentliche Raum das „Reich der Freiheiten" (der freien Männer) darstellt. Im aristotelischen Konzept sind alle Freiheiten rein öffentliche Angelegenheiten.

Übernahme in die Moderne

Nach Arendt unterscheidet sich der neuzeitliche Individualismus wesentlich von der aristotelischen Konzeption, denn die klassische politische Philosophie der Griechen kennt ein solches Individuum im neuzeitlichen Sinn nicht und schützt insofern auch

9 Für eine ausführliche Auseinandersetzung mit dem Privatheitsbegriff bei Arendt siehe Mönig (2015).

keine individuellen Rechte oder Besitztümer. All dies liegt zwar im Bereich des Privaten, meint aber eher etwas eine Gruppe Betreffendes, denn der Bezug auf ein abgrenzbares Individuum mit jeweils eigener Privatsphäre existierte nicht. Letztlich wurde, so Arendt, das Verständnis vom „Privaten“ in den liberalen Theorien der Neuzeit ergänzt um ein positives Verständnis des Privaten als Zufluchtsort, dem ebenfalls Freiheiten zugesprochen werden. Dazu gehört die Freiheit vom Staat, die bei Aristoteles nur dem Öffentlichen zugesprochen werden konnte (Rössler 2001:47 in Bezug auf Arendt), da in der antiken Ordnung die Einflussmöglichkeit des Staates grundsätzlich an den Mauern der Privathäuser endete. In aktuellen Diskussionen wird die Differenz zwischen dem Privaten und dem Politischen/Öffentlichen immer wieder aufgegriffen, beispielsweise um einen Eingriff des, als Teil des Öffentlichen wahrgenommen, Staates in den privaten Bereich zurückzuweisen[10]. Viele Klassiker der liberalen Ideengeschichte[11], die maßgeblich zu den Grundsätzen heutiger Demokratieverständnisse beigetragen haben, adaptieren die Differenzierung zwischen öffentlichem und privatem Leben. Dabei werden beide Bereiche aber um Konnotationen ergänzt, die mit einem antiken Verständnis teilweise in Widerspruch stehen und so zu den vielfältigen Konflikten um die Definition von Privatheit und Privatheitsverletzung beitragen.

Gesellschaft und das Private

Für Arendt ist die liberal-demokratische Gesellschaftsordnung, die sich auch um den Erhalt und das Wohlergehen ihrer Mitglieder in einem Sozialstaat kümmert, allerdings keine politische im antiken Sinn. Sie sieht darin eher das Aufkommen eines *dritten Raumes*, dem der *Gesellschaft*, „[deren] politische Organisationsform die Nation bildet“ (Arendt 1960:32).

> [Die Gesellschaft existiert], seitdem der private Haushalt und das in ihm erforderliche Wirtschaften eine Sache der Öffentlichkeit geworden sind […] (Arendt 1960:46)

Dieser zusätzliche Raum deckt sich in Teilen mit den ursprünglich zwei Sphären, wodurch klar wird, warum sich der Privatheitsbegriff in der Moderne in einem Spannungsverhältnis befindet und einer klaren Definition versperrt. Eine überlappungsfreie Trennung zwischen Öffentlichem und Privatem ist nicht möglich, wenn Elemente des Privaten als gesellschaftliche und, im neuzeitlichen Sinne, damit als öffentliche begriffen werden. Das trifft insbesondere auf ökonomisches Handeln zu, das in der griechischen Antike als „privates“ (heute „privatwirtschaftliches“) galt, obwohl es in der Öffentlichkeit stattfand. Während manch Radikalliberale_r diese Zuordnung fortschreiben will – vor allem, um jeglicher staatlichen Regulierung zu entgehen – wird

10 Üblicherweise wird hier Westin (1967) herangezogen, der für die Privatheit in den eigenen Räumen argumentiert, aber auch das relativ junge „Grundrecht auf Integrität informationstechnischer Systeme“ (Petri 2008) geht in diese Richtung.

11 Rössler untersucht die amerikanischen Klassiker John Locke (17. Jhd.) und John Stuart Mill (19. Jhd.). Geuss (2001) benennt als Autoren ähnlicher Theorien zentral-europäische Denker wie Wilhelm von Humboldt und Benjamin Constant (beide 19. Jhd.).

die Unterscheidung schwierig, wenn sich z. B. Gesellschaften entscheiden, die Nutzung privater Daten zum Schutz vor privatwirtschaftlicher Ausbeutung zu regulieren.

Dennoch verliert Privatheit damit nicht ihre Existenzberechtigung – ganz im Gegenteil macht die Aufweichung der Grenzen eher eine Unterscheidung zwischen verschiedenen Dimensionen von Privatheit notwendig, wie sie Rössler vornimmt. Sie beschreibt keine voneinander abgeschlossenen Räume und Sphären, sondern bezieht die neue Figur des *Individuums* in die Beschreibung mit ein und fragt, wie konkret der Schutz von welcher Privatheitsdimension aussehen kann.

2.1.2 Kritik und Grenzen des Autonomiebegriffs

Unterdrückung im Privaten

Wie schon kurz angedeutet, bezog sich die Trennung von öffentlichem und privatem Leben in der Antike wie auch in modernen, liberalen Modellen vor allem auf männliche, weiße Personen. Trotz der Erklärung individueller Freiheiten für „den Menschen“, blieben Sklaven_innen, Frauen und Kinder Teil der privaten Sphäre. Das liberale Freiheits- und Gleichheitsansprüche nur zwischen (erwachsenen) Männern ausgehandelt wurden, ist nach Rössler „der Effekt dieser falsch interpretierten Trennung zwischen den beiden Sphären“ (Rössler 2001:48), die aus der Antike übernommen wurde. Dieser Ausschluss bestimmter Gruppen von Freiheitsrechten war zwar als Konflikt von Anfang an wahrgenommen worden, führte aber eben auch dazu, dass jene Freiheiten über die Zeit eingefordert wurden. Eine Auseinandersetzung mit den unterdrückenden Aspekten von Privatheit findet vor allem seit den 1970er Jahren statt. Verschiedene Feminist_innen zeigen auf, wie die Gleichberechtigung im Konflikt steht mit einer Unterscheidung von Öffentlichem und Privaten[12].

Relationale Privatheit

Einem Menschenbild, das von autonom und selbstbestimmt lebenden Männern ausging, deren Privatheit von allen zu respektieren sei, wurden, im Zuge der feministischen Debatten, Konzepte entgegengestellt, die eine *relationale Perspektive*[13] einnahmen. Auch wenn individuelle, genderunabhängige Selbstbestimmung als Ziel von Gleichberechtigung formuliert wird, ist klar, dass sie nicht außerhalb sozialer Kontexte existiert, sondern sich viel mehr innerhalb dieser und damit in sich wechselnden Verhältnissen zu anderen konstruiert. Da solche wechselnden Positionierungen, in deren Rahmen einige Beziehungen als öffentlich, andere als privat behandelt werden, eher mehr Konflikte zwischen den Sphären produzieren, formuliert Rössler eine *Dimension des Privaten*, deren Grenzen verhandelbar sind, deren Anerkennung (und Schutz) aber gleichzeitig notwendig ist, um an einer Gesellschaft zu partizipieren. Damit ist aber auch kein absolutes Verständnis von Autonomie mehr möglich. Sie gilt

12 Die Notwendigkeit Öffentlich und Privat zusammen zu denken beschreibt auch Ritter (2008).

13 Siehe dazu auch Rössler und Mokrosinska (2013).

vielmehr als ein Konstrukt, das sich bildet und auch in Abhängigkeit zu anderen Personen eingefordert wird. Damit kann und muss auch die Frage, inwieweit etwas privat ist oder nicht, ständig neu diskutiert werden.

Autonomie und Fortschritt

Gleichzeitig ändert sich mit dem technologischen Fortschritt auch grundsätzlich die Art, wie Beziehungen geführt werden (können) und unter welchen Bedingungen Autonomie hergestellt werden kann. Rosa (2014) argumentiert, dass der Wunsch nach Autonomie eng verknüpft ist mit dem Fortschrittsversprechen der Moderne. So sei ein Mehr an Autonomie immer erst mit dem (technischen) Fortschritt möglich geworden, da bestimmte Aufgaben automatisiert und rationalisiert werden konnten. In den letzten Jahren, so Rosa, habe sich der Fortschritt aber so weit beschleunigt, dass die Ausübung von Autonomie wieder erschwert werde. Ähnlich argumentiert auch Stalder (2010) der feststellt, dass Beziehungen, die helfen, Autonomie zu konstruieren, vermehrt in einem mediatisierten Raum stattfinden würden, welcher (wie soziale Netzwerke) als „öffentlich" beschrieben werden kann. Stalder folgert daraus, dass ein Verständnis von Autonomie entwickelt werden muss, das ohne Privatheit auskommt. Ein so grundsätzlicher Kurswechsel ist allerdings nicht nötig. Die von Rössler und anderen beschriebenen Dimensionen des Privaten lassen sich durchaus auch in eine vernetzte Welt übertragen, selbst wenn ihre Funktionen durch Mechanismen wie Profiling gestört sind.

2.1.3 Der Wert des Privaten

Definition des Privaten

Rössler argumentiert, dass das Private keine vom Öffentlichen getrennte „Sphäre" ist, sondern es sich stattdessen lohnt, bestimmte Funktionen von Privatheit zu identifizieren und unabhängig von der konkreten Sphäre zu denken. Sie definiert: „[A]ls Privat gilt etwas dann, wenn man selbst den Zugang zu diesem »etwas« kontrollieren kann - umgekehrt bedeutet der Schutz von Privatheit dann einen Schutz vor unerwünschtem Zutritt anderer" (Rössler 2001:23). Dieses »etwas« und das Verständnis von Zutritt ist abhängig von der Dimension der Privatheit, die man betrachtet. Rössler[14] nennt drei dieser Dimensionen: lokale, dezisionale und informationelle Privatheit, die sie nicht nur auf ein einzelnes Individuum bezieht, sondern auch auf soziale Kontexte und Räume. Die lokale Dimension bezieht sich auf eine räumliche Beschreibung, wie etwa das private Zimmer, ohne aber dem Raum an sich Privatheit zuzuschreiben. Vielmehr macht die Art der Nutzung einen Raum zum Privatraum. Mit dezisionaler Privatheit bezieht sich Rössler auf eine sozialere Dimension, in der Privatheit Möglichkeiten schafft, Entscheidungen unabhängig von nicht-privater Einflussnahme zu treffen. Und zuletzt beschreibt sie die Dimension der informationellen Privatheit als eine, die vor

14 Pohle (2016) weist darauf hin, dass diese Differenzierungen bereits bei Kang (1997) als drei *Cluster* des Privaten vorgestellt werden.

allem die Selbstbeschreibung einer Person betrifft und deren Kontrolle es zu bewahren gilt.

Drei Dimensionen von Privatheit

Die verschiedenen Ebenen lassen sich anhand der Diskussionen um internationale staatliche Überwachung beschreiben. Lokale Privatheit ist dann verletzt, wenn sich, ganz wörtlich, Zugang verschafft wird, etwa in eine Wohnung, um dort Überwachungsutensilien anzubringen. Aber auch die Durchsuchung eines Koffers kann als Verletzung lokaler Privatheit gesehen werden, bei dem die Kontrolle über den Zugang, etwa durch ein Schloss, umgangen wird. Die informationelle Privatheit eines_r Einzelnen ist betroffen von den breit angelegten Überwachungsmaßnahmen, etwa im Internet. Die Betroffenen, die bisher davon ausgingen, im Rahmen ihrer Möglichkeiten kontrollieren zu können, wer Zugriff auf Informationen über sie hat, sehen sich durch die anlasslose Überwachung des Datenverkehrs dieser Kontrollmöglichkeit beraubt. Zuletzt ist dezisionale Privatheit eingeschränkt, wenn auf Grund von Sicherheitsmaßnahmen z. B. Reisebeschränkungen erlassen werden oder Personen in ihren privaten Entscheidungen, ein Land zu bereisen, eingeschränkt werden.

Privatheit und Autonomie

Rössler stellt die Diskussion um Privatheit und den Wert des Privaten direkt am Anfang in den Kontext des politischen Liberalismus. Wie auch in Abschnitt 2.1.1 beschrieben, sieht sie die Trennung zwischen einem öffentlichen und einem privaten Bereich als ein historisches Element liberaler Gesellschaftsformen:

> […][D]ie Unterscheidung zwischen einem öffentlichen und einem privaten Bereich [ist] für den Liberalismus nicht irgendeine Unterscheidung, sondern konstitutiv: denn in dieser Trennung bringt sich der für den Liberalismus grundlegende Gedanke des Schutzes der individuellen Freiheit und Autonomie von Personen vor unzulässigen Eingriffen oder Bestimmungen des Staates zum Ausdruck. (Rössler 2001:27)

Im Anschluss an die oben beschriebene Entwicklung dieser Differenz sieht sie aber den Wert des Privaten nicht in der Aufrechterhaltung von voneinander abgeschlossen Bereichen, sondern beschreibt Privatheit vielmehr als Voraussetzung für Autonomie:

> Ein autonom gelebtes Leben ist also zunächst einmal eines, für das wir uns selbst (gute) Gründe geben können; man würde eine Person, die einfachhin ohne Nachdenken auch wichtige Entscheidungen trifft, ohne sich dabei die Frage zu stellen, wie sie »wirklich« leben möchte, nicht unfrei nennen, wohl aber nicht autonom. (Rössler 2001:96)

Und später heißt es:

> Autonom ist eine Person, wenn sie sich mit ihren handlungsleitenden Wünschen, mit ihren Zielen und Projekten authentisch identifizieren, wenn sie diese Ziele auch verfolgen kann; wenn sie im Prinzip darauf reflektiert, wie sie leben will, welche Person sie sein will, und dann auch so lebt und leben kann. Für diese Autonomie einer Person ist, in unterschiedlichen Hinsichten und Dimensionen, der Schutz des Priva-

> ten notwendig, um Bedingungen zu gewährleisten, unter denen sie allererst Autonomie entwickeln, lernen und ausüben kann. (Rössler 2001:331)

Diese Beschreibung von Autonomie ist es, um die es in dieser Arbeit bei der Beschreibung von Profiling und den daraus entstehenden Folgen geht. Inwiefern beeinflusst Profiling die Art und Weise, wie Personen ihre Autonomie ausüben? Wie können sie diese überhaupt entwickeln, wenn in der Interaktion mit technischen Systemen immer bereits ein Bild von ihnen vorhanden ist, das die Interaktion beeinflusst und Handlungsmöglichkeiten vorgibt?

Im Folgenden soll kurz auf die einzelnen Aspekte von Privatheit, die Rössler mit Rückgriff auf liberale Theorien in ihrem Buch beschreibt, eingegangen und diese sollen dann mit den beschriebenen Phänomenen zusammengebracht werden.

Lokale Privatheit

Ein traditionelles Verständnis von Privatheit bezieht sich auf eine Trennung zwischen öffentlichem und privatem Raum, wie in 2.1.1 beschrieben wurde. In Anerkennung der Kritik, die an der Form des privaten = natürlichen = weiblichen Raums geübt wurde, bezieht Rössler lokale Privatheit nicht auf einen Machtraum, in dem der „Herr im Haus“ über ein Weisungs- und Gewaltmonopol verfügt. Stattdessen wird privaten Räumen eine Funktion in dem oben beschriebenen Prozess der Konstruktion von Autonomie zugeschrieben.

> Die ist offenbar ein fundamentaler Aspekt der normativen Begründung des Schutzes privater Räume: das Sich-einrichten-Können in einer bedeutungsvollen Umgebung. […] er garantiert Bedingungen dafür, Weisen des Sich-zu-Verhaltens auszuprobieren, die verstanden werden können als Versuch der Selbst-Definition. (Rössler 2001:261)

Verhältnis zu sich selbst

Notwendige Element lokaler Privatheit sind dabei nicht die Eigenschaften des Raumes, die Anzahl der Personen darin oder Besitzrechte. Vielmehr geht es um die Abwesenheit von (ungewollten) Störungen und Eingriffen in diesen Raum. „Ruhe für das Verhältnis zu sich selbst“ (Rössler 2001:262) kann man nicht nur in physikalischen, sondern auch in virtuellen Räumen suchen, etwa beim Surfen im Internet. Gerade dann, wenn man nicht partizipiert oder mit Fremden kommuniziert, sondern konsumierend die Nachrichten anderer liest, Videos schaut oder spielt, geht man davon aus, allein zu sein oder zumindest nicht den Blicken derjenigen ausgesetzt, die einen wertend beobachten. Gerade im Internet ist man in solchen Fällen zwar mit keiner weiteren Person in Kontakt und in dem Sinn unbeobachtet, ist aber dennoch, häufig unbemerkt, den Einflussnahmen von automatischer Personalisierung ausgesetzt, die, etwa durch das Vorfiltern von Nachrichten oder Suchergebnissen, die genannte Ruhe stören.

Lokale Privatheit in virtuellen Räumen

Boyd und Marwick (2011) haben diese Funktion privater Räume auch bei der Nutzung sozialer Online-Netzwerke durch Teenager beobachtet. Während sich Jugendliche in ihrem meist als „privat“ definierten, eigenen Zimmer oft nicht privat fühlen, da sie der Kontrolle durch die Erwachsenen ausgeliefert sind, werden soziale Online-Netzwerke und Chaträume, in denen sie sich mit Freund_innen austauschen, als private Räume wahrgenommen, auch wenn sie es technisch gesehen nicht sind. Hier zeigt sich, dass die Nutzung von Webdiensten für eben jenen Zweck, das „Sich-zu-Verhalten“ auszuprobieren, genutzt werden, etwa in anonymen Chats oder Weblogs. Diese „Ruhe für sich“ ist für Rössler auch notwendige Bedingungen zur Erprobung von *Selbstdarstellungen* sowie für *Selbsterfindung* und unerlässlich für die Entwicklung der eigenen Persönlichkeit, um Rollen zu erproben, die im Auftreten außerhalb privater Räume nützlich sein können.[15]

Dezisionale Privatheit

> […] mit der dezisionalen Privatheit und ihrem Schutz wird ein sozialer Handlungs- und Spielraum geschaffen, der notwendig ist für individuelle Autonomie. (Rössler 2001:145)

Negative Freiheitsrechte

Als (negatives) Freiheitsrecht erlaubt dezisionale Privatheit es Einzelnen, Entscheidungen zu treffen, ohne sich für diese rechtfertigen zu müssen und ohne dass dies öffentlich als unrecht oder sonst wie besprochen wird (so lange die Entscheidung nicht überwiegend andere beeinflusst). Gerade letzteres kann man als eine Frage der Privatheit beschreiben, da es um „die Möglichkeit [geht] mich unbehelligt so zu verhalten, so zu leben, wie ich möchte […]“ (Rössler 2001:151). Es geht also um die Autonomie der Lebensführung und gerade in sozialen Kontexten auch darum, dass „das je eigene Leben nicht von solchen anderen Personen kommentiert oder interpretiert, oder auch stärker: beeinflusst wird, denen sie gerade keine solche Interpretationshoheit über ihre Leben zubilligen will.“ (Rössler 2001:153) Diese grundsätzliche Aussage muss relativiert werden vor dem Hintergrund der unterschiedlichen sozialen Kontexte, in denen man sich bewegt und in denen auch Konventionen gelten, die Unterschiedliches zur Privatsache machen oder eben nicht.

Positive Freiheitsrechte

Auf die Problemstellung bezogen wird aber auch schnell deutlich, dass gerade die – ungewollte oder unbewusste – Beeinflussung der Handlungsräume durch Profiling-Algorithmen eine Verletzung der dezisionalen Privatheit der_s Einzelnen ist und damit der (positiven) Freiheit, selbstbestimmt leben zu können. Es lässt sich aber auch argumentieren, dass durch Profiling eine Inanspruchnahme des negativen Freiheitsrechts, sich nicht rechtfertigen zu müssen, erschwert wird. Wenn Faktoren, die (statistisch) maßgeblich, zum Beispiel für eine Kaufentscheidung, waren, berechnet werden, wird

15 Es ist wichtig zu bemerken, dass dieses Verständnis von Rollen nicht impliziert, dass ein wahres oder „echtes“, nicht durch Rollen vermitteltes Selbst existiere.

eine (nachträgliche) Transparenz über den Entscheidungsprozess hergestellt, ohne dass der_die Betroffene es verhindern kann, davon weiß oder auch nur dazu Stellung beziehen kann.

Informationelle Privatheit

Die dritte Dimension von Privatheit, die Rössler beschreibt, ist wesentlich für die Diskussion von Profiling. Zentrales Element informationeller Privatheit ist die *Information* über eine Person. Eine personenbezogene Information meint dabei eine breite Palette von Beschreibungen eines Menschen: von Eigenschaften (wie Name, Alter oder Lieblingssportverein) über Verhaltensweisen („isst dienstags immer auswärts", „fährt öfter zu schnell"), aber auch weiteren Attributen wie zum Beispiel Beziehungsverhältnisse („kennt X", „war mit Y zusammen"). Den Zusammenhang zwischen den eine Person beschreibenden Informationen und der Privatheit beschreibt Rössler wie folgt:

> […] der Schutz informationeller Privatheit ist […] so wichtig für Personen, weil es für ihr Selbstverständnis als autonome Personen konstitutiv ist, (in ihnen bekannten Grenzen) Kontrolle über ihre Selbstdarstellung zu haben, also Kontrolle darüber, wie sie sich wem gegenüber in welchen Kontexten präsentieren, inszenieren, geben wollen, als welche sie sich in welchen Kontexten verstehen und wie sie verstanden werden wollen. (Rössler 2001:209)

Kontrolle über die Selbstdarstellung

Der Verlust der Kontrolle des Informationsflusses[16] einer Person über die sie beschreibende Daten bedeutet in diesem Sinn das Ende informationeller Privatheit. Zum Erhalt von Autonomie ist aber auch der Erhalt der „Kontrolle über ihre Selbstdarstellung" wichtig, denn:

> eine Person reguliert mit den Informationen, die sie anderen über sich mitteilt oder die andere über sie, wie sie weiß, immer schon haben, zugleich die ganz unterschiedlichen sozialen Beziehungen, in denen sie lebt. Ohne diese Form der selbstbestimmten Kontrolle darüber, wen man was über sich wissen lassen möchte […], wäre die selbstgewählte Unterschiedlichkeit von Beziehungen nicht möglich, damit auch nicht das selbstbestimmte kontextuell je unterschiedliche authentische Verhalten anderen gegenüber und damit nicht die unterschiedlichen, selbstgewählten Formen der Auseinandersetzung mit anderen, […] und damit auch nicht: das *authentische* Finden einer Antwort auf die Frage, wie man leben will. (Rössler 2001:209)

Natürlich kann Autonomie nur im Kontext ausgeübt werden und individuelle Freiheit ist immer abzuwägen gegen die Freiheiten der Anderen. Diese Regel gilt auch in Bezug auf Informationen: Informationelle Privatheit muss immer abgewogen werden gegen das Recht, etwas über das Gegenüber zu erfahren, oder gegen das Recht auf freie Meinungsäußerung. Informationelle Privatheit meint daher auch, dass eine Person in gewissem Maß eigenständig darüber entscheiden können muss, inwiefern sie Wissen über sich preisgibt. Diese Entscheidungen müssen abhängig vom Kontext getroffen

16 Auch hier ist Rössler nicht alleinige Urheberin des Gedankens. Die Konzeption von „Privacy as control" wird Westin (1967) zugeschrieben.

werden, in dem man sich bewegt, um so auch sich selbst gegenüber authentisch sein zu können.

Keine absolute Kontrolle

Dieses relationale Verständnis von Privatheit geht in aktuellen Diskussionen zur „Verteidigung der Privatsphäre“ häufig unter. Hier liegt der Fokus vor allem auf der unfreiwilligen Preisgaben von Informationen entweder gegenüber dem Staat oder übermächtig erscheinenden, datenverarbeitenden Unternehmen.[17] Diese starke Polarisierung führt bei der Entwicklung von *Privacy Enhancing Technologies* häufig dazu, dass die Verhinderung von Preisgabe von Informationen einseitig überbetont wird. Dieses Phänomen wird auch als *privacy as confidentiality-Paradigma* (Gürses, Preneel und Berendt 2009) beschrieben (siehe dazu 2.5.3).

Zusammenfassung

Bis hierhin wurde eine Differenzierung von Privatheit in drei Dimensionen nachvollzogen, die die häufig zitierte Differenz von *öffentlich* und *privat* hinter sich lässt, die Notwendigkeit relationaler Autonomie betont und so über eine rein individualistische Sicht auf Privatsphäre hinausgeht. Auf ähnlichen Annahmen beruht die im Folgenden beschriebene, juristische Debatte um informationelle Selbstbestimmung und Datenschutz, deren Ziel die normative Regulierung von Datenverarbeitung ist. Obwohl der im Datenschutz vorhandene systemische Ansatz auch eine weitgehende Kritik von Profiling erlaubt, hat er keinen wirksamen Schutz vor diesen Techniken entwickeln können.

2.2 DATENSCHUTZ: SCHUTZ INFORMATIONELLER PRIVATHEIT

Der informationellen Selbstbestimmung und dem daran anschließende Datenschutzrecht liegt das, als wegweisend geltende, Urteil des Bundesverfassungsgerichts (BVerfG) zur Volkszählung von 1983 (vgl. BVerfG 1983) zu Grunde. Die Richter_innen stellten darin die Notwendigkeit des Individual- wie Systemdatenschutzes heraus. Sie begründen dies mit dem Schutz individueller Autonomie durch Privatheit und der Notwendigkeit eben jener Autonomie für das Funktionieren demokratischer Gesellschaften.

2.2.1 Recht auf informationelle Selbstbestimmung

Definition informationelle Selbstbestimmung

Im Folgenden soll kurz auf die Datenschutzgesetzgebung in der Bundesrepublik Deutschland und Europa eingegangen werden, die internationale Gesetzgebung muss außen vor bleiben, da sie den Rahmen der Arbeit sprengen würde, ist aber häufig an

17 Siehe dazu wiederum Trojanow und Zeh (2010) und Kurz und Rieger (2011).

ähnlichen Grundsätzen orientiert. Das *Recht auf informationelle Selbstbestimmung*[18] (BVerfG 1983) ist wesentlicher Ankerpunkt für die Datenschutzgesetzgebung in Deutschland. Die Richter_innen leiteten in ihrem Urteil dieses Recht aus dem Artikel 1 Absatz 1 des Grundgesetzes, der Unantastbarkeit der Menschenwürde, sowie Artikel 2 Absatz 1, dem Recht auf freie Entfaltung der Persönlichkeit, ab. Neben dieser persönlichkeitsrechtlichen Komponente (Individualdatenschutz), ist Datenschutz zu Beginn auch systemisch gedacht worden.

> Mit dem Recht auf informationelle Selbstbestimmung wären eine Gesellschaftsordnung und eine diese ermöglichende Rechtsordnung nicht vereinbar, in der Bürger nicht mehr wissen können, wer was wann und bei welcher Gelegenheit über sie weiß. Wer unsicher ist, ob abweichende Verhaltensweisen jederzeit notiert und als Information dauerhaft gespeichert, verwendet oder weitergegeben werden, wird versuchen, nicht durch solche Verhaltensweisen aufzufallen. (BVerfG 1983)

Das Urteil benennt auch Folgen für die Gesellschaftsordnung beziehungsweise die (informationellen) Machtverhältnisse zwischen Bürger_innen und Staat, auch weil zur Zeit des Urteils eine umfassende Erfassung vor allem von staatlichen Institutionen durchgeführt werden konnte. Diese Perspektive wird von Vertreter_innen des Systemdatenschutzes auch heute noch betont (Rost 2013; Simitis 1987; Simitis, Dammann, und Arendt 2011). Allerdings ist mit der Verbreitung der Informations- und Kommunikationstechnologie der wesentlich schwächer regulierte „nicht-öffentliche Bereich“ stark gewachsen. Obwohl die ökonomische Bedeutung der Verarbeitung personenbezogener Daten heute einen so starken Einfluss auf die informationelle Selbstbestimmung hat, dass von einer systematischen Beeinträchtigung gesprochen werden kann, ist eine rechtliche Regulierung nicht in Sicht.

2.2.2 Datenschutzrecht und Profiling

Personenbezogene Daten und bestimmbare Personen

Das Recht auf informationelle Selbstbestimmung ist seit 1983 Grundlage des deutschen Datenschutzes und schlägt sich in einer Vielzahl von bereichsspezifischen Regelungen nieder[19]. Im Folgenden werden einige zentrale Datenschutzbegriffe kurz vorgestellt und am Beispiel von Online-Tracking und Profiling erläutert.[20]

Personenbezogene Daten im Sinne des Gesetzes sind solche Informationen, die direkt Eigenschaften einer *bestimmten* oder *bestimmbaren Person* beschreiben. „Joseph

18 Zur Entstehungsgeschichte des Begriffs siehe Steinmüller und Podlech (2007).

19 Es existieren z. B. Regelungen für den Datenschutz im Telekommunikations- und Telemediengesetz, in den Sozialgesetzbüchern genauso wie in den Melde- und Polizeigesetzen, um nur eine kleine Auswahl zu nennen.

20 Wesentliche Quellen hierfür sind Bull (2009), Gola u. a. (2012) sowie Simitis, Dammann, und Arendt (2011).

Weizenbaum wurde am 08. Januar 1923 in Berlin geboren“ enthält dementsprechend das personenbezogene Datum des Geburtstags von Herrn Weizenbaum[21]. Ein Satz wie „Der Autor des Buches 'Die Macht der Computer und die Ohnmacht der Vernunft' war Jude“ enthält dagegen personenbezogene Daten einer bestimmbaren Person. Bestimmbar in dem Sinn, dass das genannte Buch nur einen Autor hat und damit die Information über einen kurzen Wikipedia-Umweg einer Person zugeordnet werden kann. Beim Online-Tracking basieren Profile häufig auf Pseudonymen, die zwar nicht direkt einer Person zugeordnet werden können, aber, etwa im Fall von Browser-Fingerprinting (vgl. Abschnitt 3.3), auf eine bestimmbare Person verweisen.

Besondere Arten personenbezogener Daten

Darüber hinaus zählt die Information über die Religionszugehörigkeit, genauso wie die über „rassische und ethnische Herkunft, politische Meinungen, […] philosophische Überzeugungen, Gewerkschaftszugehörigkeit, Gesundheit oder Sexualleben“, zu den *besonderen Arten personenbezogener Daten*. Beim Online-Profiling werden unter anderem Angaben zur politischen Überzeugung generiert (vgl. Abschnitt 4.3.3).

Verantwortliche Stelle

Betroffene_r einer Datenverarbeitung ist diejenige Person, deren personenbezogene Daten verarbeitet werden. Die Institution, welche die Datenverarbeitung in Auftrag gibt und mit dem_r Betroffenen einen Vertrag eingeht, ist die *verantwortliche Stelle*. Wichtig ist die Unterscheidung zur datenverarbeitenden Stelle, die nicht identisch mit der verantwortlichen Stelle sein muss, z. B. dann, wenn die Datenverarbeitung outgesourct ist (wie im Fall von Online-Werbung).

Grundregeln des Datenschutz

Aus Datenschutzsicht unterscheidet man bei der Datenverarbeitung die Prozesse der Erhebung, Speicherung und Nutzung von personenbezogenen Daten. Bizer (2007) hat aus dem BDSG sieben Aspekte abgeleitet, die bei der Bewertung der Zulässigkeit einer Datenverarbeitung zu berücksichtigen sind. Eine solche Einschätzung ist immer in Bezug auf eine konkrete Maßnahme zu treffen. Im Fall von Online-Tracking auf einer gewerblich betriebenen Webseite ergibt sich die *Rechtmäßigkeit*[22] aus § 15 TMG (2007), der zu Nutzungsdaten festlegt:

> Der Diensteanbieter darf personenbezogene Daten eines Nutzers nur erheben und verwenden, soweit dies erforderlich ist, um die Inanspruchnahme von Telemedien zu ermöglichen und abzurechnen (Nutzungsdaten).

Darüber hinaus geht man davon aus, dass die Nutzer_innen in die Datenspeicherung zum Profiling durch den Besuch der Webseite einwilligen. Details zur Art und Weise

21 Ich möchte Herrn Weizenbaum als Beispiel anführen, auch wenn er, da er bereits verstorben ist, nicht mehr als Person im Sinne des Datenschutzes zählt.

22 Die Argumentation geht hier davon aus, dass bei einer solchen Maßnahme überhaupt personenbezogene Daten verarbeitet werden und Datenschutzrecht anzuwenden ist (Steidle und Pordesch 2008). Der Bundesverband der Digitalen Wirtschaft ist allerdings schon an diesem Punkt anderer Meinung (BVDW e. V. 2014).

der *Einwilligung* ist im § 13 TMG geregelt und sieht vor, dass Nutzer_innen ausreichend informiert und über ihre Rechte aufgeklärt werden. In der Praxis muss die Wirksamkeit einer solchen impliziten Einwilligung allerdings in Frage gestellt werden, da die Betroffenen häufig nicht informiert und sich der möglichen Folgen nicht bewusst sind (vgl. Beisenherz und Tinnefeld 2011; Kamp und Rost 2013). Grundsätzlich dürfen Daten nur für den Zweck verarbeitet werden, für den Sie erhoben wurden (*Zweckbindung*). Das Telemediengesetz erlaubt in § 15 Abs. 3 die Erstellung von Profilen zum Zweck „der Werbung, der Marktforschung oder zur bedarfsgerechten Gestaltung" der Dienste, so lange diese nur mit einem Pseudonym versehen sind. Daraus ergibt sich auch die Erforderlichkeit der Datenerhebung, die mit einer *Datenminimierung* einhergeht. Allerdings ist nicht abschließend geregelt, was die minimal notwendigen Daten sind, die zur Erstellung eines „Nutzungsprofils" gespeichert werden dürfen. Die Definitionsmacht liegt hier meist bei den Dienstanbieter_innen. Nach Bizer ist es zudem notwendig, Maßnahmen zu ergreifen, die die *Transparenz* gegenüber den Betroffenen herstellen, etwa indem eine Auskunft über die erstellten Profile gegeben wird. Hier gehen die Anbieter_innen unterschiedliche Wege. So denn überhaupt eine Zuordnung zu einer Person möglich ist und die Daten als ein Profil erhoben werden, statt erst im Bedarfsfall ein Profil zu generieren, können die Daten bei den Anbieter_innen angefragt werden. *Datensicherheit* zu gewährleisten ist meist auch im Sinne des_r Betreibers_in einer Infrastruktur, da die gespeicherten Daten als Geschäftsgrundlage zu schützen sind. Den gesetzlich vorgeschriebenen Mechanismen zur *Kontrolle* der Datenschutzprinzipien kommt in Deutschland nach, wer eine_n Datenschutzbeauftragten bestellt. Seltener werden durch Audits zusätzliche Zertifizierungen erworben.

2.2.3 Grenzen des Rechts

Wie bereits erwähnt, sind nur solche Profilingverfahren aus Datenschutzsicht relevant, bei denen personenbezogene Daten verarbeitet werden. Häufig argumentieren Dienstanbieter_innen, dass sie nur anonymisierte Daten verarbeiten. Neben den Interessenverbänden ist dies insbesondere im US-amerikanischen[23] Raum der Fall, wo die Regulierung schwächer ist (vgl. Schwartz und Solove 2011, S. 1855). Das BDSG definiert

Anonymisierung

> [A]nonymisieren [als] das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können. (BDSG 1990 § 3 Abs. 6)

23 Nicht weiter behandelt wird hier das Problem, dass global agierende Konzerne, die die Anwendbarkeit nationaler Rechtsprechung insgesamt in Frage stellen.

Diese Definition[24] ist eine Abschwächung gegenüber einer informationstechnischen Anonymisierung, wie sie etwa Pfitzmann und Hansen (2008) beschreiben, bei der die Zuordnung von einem Datum zu einer natürlichen Person nachweislich nicht möglich ist. Dazu ist es notwendig, dass eine ausreichend große Gruppe (*anonymity set*) dieselben Attribute hat.[25] Profile sind daher aus informationstechnischer Sicht meist pseudonymisiert, weil sie zwar keiner Person direkt zugeordnet werden können, aber einer vorgelagerten „pseudonymen" Entität. Dieses Pseudonym kann im Internet etwa ein Nutzer_innenname sein, in einer Datenbank aber auch eine Zahlenkombination (ID). Anonym sind hingegen häufig die zur Erstellung von Profilen verwendeten Zusatzinformationen, wie Zuordnungen von Attributen zueinander, die jenseits der konkreten Profile vorgehalten werden und nicht unter das Datenschutzrecht fallen.

Verbot automatisierter Einzelentscheidung

Relevanz erhalten diese Daten erst, wenn es zur Anwendung eines Profils kommt, zum Beispiel um auf Basis von Erfahrungswerten zu entscheiden, welche Werbung in einem Browser angezeigt wird. Hier findet dann eine automatisierte Einzelentscheidung statt, die im BDSG in § 6a („Verbot automatisierter Einzelentscheidungen") geregelt ist. Dieses Verbot bezieht sich auf (abschließende) Entscheidungen, die nicht von Menschen, sondern von Automaten getroffen werden und sich auf Persönlichkeits*merkmale* beziehen. Merkmale meint, im Gegensatz zu Eigenschaften, nicht etwa Name und Adresse, sondern Informationen, die „die Persönlichkeit des Betroffenen unter bestimmten 'einzelnen Aspekten' beschreiben" (Gola u. a. 2012 § 6a Rn 7-9). Die EU-Datenschutzrichtlinie nennt hier als Beispiele „berufliche Leistungsfähigkeit, Kreditwürdigkeit, Zuverlässigkeit einer Person" (Richtlinie 95/46/EG, Art 15.). Allerdings ist die Anwendbarkeit dieses Paragraphen stark eingeschränkt. Er betrifft nur automatisierte Einzelentscheidungen, die folgenden Kriterien erfüllen:

1. Es muss sich um eine *abschließende, negative Entscheidung* handeln. Nicht reguliert sind Hinweise für eine Entscheidung oder solche mit positivem Ausgang.
2. Die Entscheidung darf *nicht trivial* sein (zum Beispiel, ob etwas zu groß oder zu klein ist);
3. Eine *rechtliche Folge* oder eine *erhebliche Beeinträchtigung* muss gegeben sein.
4. Der Paragraph greift auch nicht, wenn *informiert*, *erklärt* und *Widerspruch ermöglicht* wurde.

---

24 Siehe dazu auch Simitis u. a. (2011 § 3 Rn 196ff).

25 Zur Verifizierung von Anonymität siehe (Article 29 Data Protection Working Party 2014; Brunst 2009; Machanavajjhala u. a. 2007; Sweeney 2002).

5. Man muss der Entscheidung „unterworfen“ sein, das heißt die Bedingungen werden einseitig von der verarbeitenden Stelle festgelegt.

Die auf Basis von Profiling gemachten Entscheidungen sind also nur dann datenschutzrechtlich relevant, wenn sie eine gewisse Wichtigkeitsschwelle überwinden können. In der Praxis wurde daher dieser Absatz fast ausnahmslos auf Kreditscoring und Personalentscheidungen angewendet (Gola u. a. 2012 § 6a Rn. 3-6).

Systemischer Datenschutz und das Privacy Paradox

Privacy Paradox

Ein Problem in der Anwendung von Datenschutzrecht auf Profiling liegt in der Ausgestaltung der informationellen Selbstbestimmung als Individualrecht. Obwohl das Bundesverfassungsgericht und die meisten Kommentare in der systemischen Perspektive eine weitere Säule des Datenschutzes sehen, setzt das Datenschutzrecht voraus, dass die Rechte einer natürlichen Person betroffen sein müssen[26]. Mit der Definition personenbezogener Daten wird eine Dichotomie fortgeführt, die der antiken Unterscheidung zwischen dem Öffentlichen und dem Privaten angelehnt ist. Personenbezogene als private Daten sind gleichzeitig Besitz und geschützter heimischer Bereich, über den der_die Bürger_in selbst verfügen kann – im Prinzip. Demgegenüber sind alle Daten, die als öffentlich deklariert werden, Teil des öffentlichen Diskurses und ohne Einschränkungen von allen zu nutzen. So bleiben die Konfliktlinien erhalten, die auch Arendt mit dem Entstehen der „Gesellschaft“ verbindet. Der Erlaubnisvorbehalt, der Teil der Datenschutzgesetzgebung ist, sieht vor, das (mit gesetzlich geregelten Zwecken) in das Private eingedrungen und personenbezogene Daten für einen gemeinschaftlichen Zweck genutzt werden können. Ähnliches passiert im ökonomischen Bereich, wenn Nutzer_innen „ihre“ Daten „verkaufen“, etwa im Tausch gegen Rabatte oder, um kostenlos Dienste im Internet nutzen zu können, die sich durch Werbung finanzieren. Diese nicht-lösbaren Konflikte werden häufig als *privacy paradox* bezeichnet (siehe Norberg, Horne und Horne 2007).

Gruppenbezogene Daten

Die Fokussierung auf einzelne Personen zeigt sich in der gesetzlichen Unterscheidung zwischen personenbezogenen und nicht-personenbezogenen Daten. Erstere sind nach § 2 Abs. (1) „[…] Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person (Betroffener).“ Hiervon sind statistische Informationen, die sich damit nicht auf einzelne Personen, sondern auf Gruppen beziehen, genauso ausgenommen wie „anonymisierte Daten“, so lange sie nicht wieder auf eine konkrete Person „durchschlagen“ (Gola et al. 2012 § 3 RN 3). Das Datenschutzgesetz findet, etwa im Fall von Scoring erst dann Anwendung, wenn die Daten,

26 Deutlich wird der Charakter des Individualrechts im Verweis auf das Persönlichkeitsrecht, aus dem es abgeleitet ist. Siehe dazu Bull (2009).

die zu einer Gruppe gespeichert sind (Gola u. a. 2012 § 3 RN 3a), direkt auf eine Person bezogen werden.

Systemischer Datenschutz

Diese individualrechtliche Konzeption entspricht allerdings nicht der ursprünglichen Konzeption von Datenschutzrecht, bei der dem Systemdatenschutz eine ebenso wichtige Funktion zugewiesen wurde. Die zentrale juristische Stellungnahme zur Datenschutzfrage aus dem Jahr 1971[27], die einige wesentliche Grundlagen für die späteren Gesetze geliefert hat, hatte noch eine etwas breitere Bestimmung der Daten, die in den Anwendungsbereich des Datenschutzes fallen sollten:

> Gegenstand des Dsch [Datenschutz] sind demnach grundsätzlich
> 1. alle Personeninformationen,
> 2. alle personenbezogenen Informationen,
> 3. alle individualisierbaren (statistischen Personen- oder personenbezogenen) Informationen,
> 4. alle Gruppen(-bezogenen) Informationen;
>
> kurz: die Individual- und Gruppeninformationen.
> (Steinmüller u. a. 1971:41)

Die Punkte 3 und 4 beziehen also explizit gruppenbezogene Daten in das Datenschutzrecht mit ein, die notwendig sind, um aus dem Profiling entstehende Probleme als Datenschutzprobleme zu betrachten. Hierfür ist eine systemische Lesart von Datenverarbeitung notwendig, die sich eher an den Effekten von Datenverarbeitung orientiert und das Recht auf informationelle Selbstbestimmung zuerst als Schutzrecht gegen diese Effekte betrachtet. Während verschiedene Autoren (Pohle 2014; Rost 2013) die systemische Betrachtung[28] nicht zufällig mit dem Aufkommen der Diskussionen um *Big Data* wieder betonen, hat sich in der Praxis, wie beim Profiling, die Annahme etabliert, dass anonyme Daten datenschutzrechtlich nicht relevant seien.

---

27 Siehe Steinmüller u. a. (1971); zum Einfluss des Papiers auf die spätere Gesetzgebung vgl. Steinmüller und Podlech (2007). Die Ebene des Systemdatenschutzes wird auch von Simitis (u. a. 2011 RN 111 ff.) wiederholt betont.

28 In eine ähnliche Kerbe schlägt auch die Kritik von Vertreter_innen der *Surveillance Studies* am Datenschutz. Zurawski (2014) argumentiert mit Coll (2014), dass eine Fokussierung der Diskussion auf Privatheit und die individuellen Gefahren eine größere Perspektive aus dem Blick geraten lässt.

Grenzen der (informationellen) Selbstbestimmung

Kontrollverlust

Neben der Überbetonung der individuellen Rechte im Datenschutz, gibt es auch Einwände gegen ein Verständnis von Privatheit und Autonomie, das diese einseitig als Möglichkeit der Kontrolle von (persönlichen) Informationen auf den Kontext der Entwicklung neuer Informations- und Kommunikationstechnik versteht.[29] Häufig wird Privatheit, etwa in der computergestützten Kommunikation, gleichgesetzt mit der Notwendigkeit, Informationsflüsse kontrollieren und beschränken zu können. Verschiedene Autor_innen argumentieren, dass diese Form der Kontrolle nicht mehr umsetzbar sei (Heller 2011; Stalder 2010), stattdessen müssten Formen von Privatheit entwickelt werden, die den *Kontrollverlust* (Seemann 2014) akzeptieren und berücksichtigen.

Ontological friction

Die Idee des Kontrollverlusts geht von der Beobachtung aus, dass informationelle Selbstbestimmung, in dem Sinn dass eine Person wissen kann, wer, was, wann über sie weiß, in einer Informationsgesellschaft ein zunehmend komplexe Aufgabe ist, die nicht mehr zu bewältigen sei. Floridi (2006) beschreibt die Entwicklung von Informations- und Kommunikationstechnologien (IuK) als eine, bei der Informationsflüsse – und personenbezogene Daten sind nur eine von vielen Informationsarten – immer mehr beschleunigt werden. Er prägt dafür den Begriff der *ontological friction* und beschreibt mit der Metapher der „Reibung im Informationsraum“ die Geschwindigkeit, mit der sich Informationen ausbreiten können. IuK-Technologien verringern dabei nach Floridi nicht nur diese Reibung und sorgen dafür, dass Informationen in kürzerer Zeit immer mehr Menschen erreichen, sondern erhöhen gleichzeitig auch die Menge der verbreiteten Informationen. Aus dem dadurch wahrgenommen Kontrollverlust ziehen nicht wenige die Konsequenz, Privatheit als Konzept sei aufzugeben und die Gesellschaften müssten zu einer *Post Privacy* (sieh dazu auch 2.5.1) übergehen. Dabei übersehen die Befürworter_innen des Kontrollverlusts, dass die Verringerung der *ontological friction* vor allem von denen vorangetrieben wird, die die Technologie entwickeln. Die Asymmetrie in den Machtverhältnissen (Rost 2013) geht einher mit der *kognitiven* und *voluntativen Asymmetrie* (Rössler 2003). Häufig ist denen, die ihre informationelle Selbstbestimmung aufgeben, entweder nicht bewusst, dass dies passiert (kognitiv) oder sie nehmen die Entwicklung hin, ohne dass dies freiwillig (voluntativ) passiert.

Zusammenfassung

Aktuell sind die Möglichkeiten zur Regulierung von Profiling auf juristischer Ebene begrenzt. Mit der Fokussierung auf die individuelle Perspektive von informationeller Selbstbestimmung wurde den Entwickler_innen eine Möglichkeit gegeben, auf Basis „anonymisierter“ Daten zu agieren und sich so einer Regulierung zu entziehen. Diese

29 Die Beschreibung der „informationellen Privatheit“ geht mit der Entstehung von Informations- und Kommunikationstechnologien einher. Schon der einflussreiche Artikel „The right to privacy“ von 1890 (Warren und Brandeis 1890) bezog sich auf den Verlust von „privacy“ durch die Fotografie.

Schwäche schlägt sich auch in der Kritik am Datenschutz nieder, die, ebenso wie die Post-Privacy-Bewegung, dessen Abschaffung fordert. Die mittelbaren Effekte von Profiling, eine Schwächung von Autonomie und Privatheit, sind allerdings nicht immer individuell und werden nur mit einem systemischen Blick auf Datenverarbeitung nachvollziehbar. Durch die Betonung der individuellen Kontrolle sind Prinzipien wie Datenminimierung und Zweckbindung, die auf den systemischen Datenschutz der frühen Debatten zurückgehen, in den Hintergrund geraten und haben, etwa im Zuge der Entwicklung von *Big Data*, an Bedeutung stark verloren.

2.3 BEGRIFFSBESTIMMUNG PROFILING

Die Entwicklung von Techniken des Profilings hat sicherlich mit dazu beigetragen, dass eine Verringerung der *ontological friction* und ein Kontrollverlust wahrgenommen werden. Gleichzeitig zeigt sich an ihm, wie schon beschrieben, auch die Grenze einer individuell konzeptionierten Privatheit. Im Folgenden soll „Profiling“ noch einmal genauer aus zwei unterschiedlichen Blinkwinkeln beschrieben werden. Zuerst werden Begriffe und technische Grundlagen von Profiling erläutert, im Anschluss dann konkrete Anwendungsfelder von Profilen vorgestellt: Personalisierung, Risikoeinschätzung und Methoden der datenbasierten Vorhersagen. Im zweiten Teil dieses Abschnitts werden Beiträge nicht-technischer Wissenschaften vorgestellt, die die Anwendung dieser Technologien in ihre jeweiligen Disziplinen einordnen.

2.3.1 Profiling und Profile

Definition Profiling

Profiling bezeichnet nach Gutwirth und Hildebrandt (2010) den Vorgang des Zusammenstellens von Informationen über Einzelpersonen sowie über eine oder mehrere Gruppen von Personen. Diese Zusammenstellung kann einerseits darin bestehen, bekannte Informationen in strukturierter Form zu ordnen und andererseits neue Informationen aus unterschiedlichen Zusammenstellungen zu generieren. Ersteres ist vergleichbar mit einer Personalakte, die verschiedene Informationen (Name, Ausbildung, Abteilung) über eine_n Angestellte_n in strukturierter Form sammelt und so mit anderen Personalakten vergleichbar macht. Im zweiten Schritt würde diese Akte angereichert werden mit Informationen, die aus der Zusammenführung weiterer Daten ermittelt werden können. So ließe sich, etwa aus den durchschnittlichen Krankheitstagen einer Abteilung, eine Vermutung darüber anstellen, wie viele Tage im Jahr eine weitere Person dieser Abteilung krank sein wird. Bei diesem Beispiel ist die Information über die durchschnittliche (oder auch absolute Zahl der Krankmeldungen pro Jahr) Teil des Profils der Abteilung als Gruppe. Vergleichbares passiert beim Profiling im Internet, wo Informationen, wie etwa die Information über aufgerufene Webseiten, in einem Profil zusammengestellt werden. Dieses Profil wird mit dem abstrahierten Wis-

sen über das Verhalten einer bestimmten Gruppe anderer Internetnutzer_innen in ein Verhältnis gesetzt, wodurch zusätzlich Informationen gewonnen werden können.

Personenprofile

Im Unterschied zu den Informationen, die in einer Akte einer Person über den Namen direkt zuordenbar sind, ist die Erstellung von Profilen über Dritte im Internet schwieriger. Die Art der Profile, die durch Google oder andere Tracking-Provider erstellt werden, sind sowohl in dem, was sie beschreiben, als auch in der Art, wie sie es beschreiben, weniger absolut. Die Entität, über die ein Profil erstellt wird, ist meist keine natürliche Person, sondern das Gerät oder der Browser der zum Internetsurfen benutzt wird. Profile auf Basis von Online-Tracking können zudem unterschiedliche Attribute umfassen, die wiederum verschiedene Abstraktionsniveaus (von der groben Genderkategorisierung bis zum Psychogramm, vgl. dazu 2.3.3) enthalten. Darüber hinaus werden häufig keine absoluten Zuweisungen getätigt, sondern für jedes Attribut nur eine Wahrscheinlichkeit berechnet, mit der die profilierte Entität in eine der Kategorien fällt. So kann sich ein durch Online-Tracking erstelltes Profil also auf mehre Personen beziehen, die dasselbe Gerät (z. B. ein Tablet) zum Surfen verwenden. Jeder Webseitenaufruf verändert dabei das Profil, auch in Abhängigkeit zum Verhalten anderer Internetnutzer_innen. Während etwa der Aufruf einer Sportseite an einem regulären Samstagnachmittag, bei dem für das Tablet erstellten Profil, die Wahrscheinlichkeit steigen lässt, dass es sich bei dem_der Nutzer_in um eine als männlich wahrgenommene Person handelt, kann dieselbe Aktion am Tag eines großen Fußballturniers ohne Effekt auf die Berechnung des Genders bleiben.

2.3.2 Data-Mining und Profiling

Das Erweitern eines Profils durch die Produktion von weiteren Informationen durch Data-Mining oder auch *knowledge discovery in databases* (KDD) ist wichtiger Bestandteil des Online-Profilings. Gutwirth und Hildebrandt (2010) beschreiben aktuelle Formen des Profilings als eine Bewegung hin zur Ermittlung von Zusammenhängen (Korrelationen), die sich vom klassischen *Messen* hin zum *Entdecken* verschieben.[30] Um auf ein oben genanntes Beispiel zurückzukommen, ist die Zahl der Krankmeldungen pro Jahr ein klar messbares Attribut einer Person oder einer Abteilung. Mittels Data-Mining kann versucht werden weitere Zusammenhänge zwischen unterschiedlichen Attributen herzustellen, zum Beispiel zwischen der Anzahl der Krankmeldungen und der Anzahl der (nicht) genommenen Urlaubstage. Dieser Zusammenhang könnte in den Daten „entdeckt“ und den einzelnen Profilen jeweils als Attribut hinzugefügt werden.

30 Diese Form der Datenanalyse ist auch wesentlicher Bestandteil dessen, was als *Big Data* diskutiert wird, vgl. etwa Geiselberger (2013) oder Mayer-Schönberger und Cukier (2013).

Construction und Application von Profilen

Durch die Zunahme dieser Form des Profilings und der Zunahme von Informationen über Zusammenhänge, die nicht direkt auf die Attribute des_r der von Profiling betroffenen Person oder Gruppe hervorgehen, verschiebt sich der Punkt, zu dem ein Profil erstellt (*construction*) wird oder werden kann, vor das Erheben von Informationen des klassischen Profis. Eine Aussage über die wahrscheinliche Anzahl der Krankheitstage einer Person, die neu in einer Abteilung ist, kann bereits gemacht werden, bevor sie den ersten Tag Urlaub genommen hat oder das erste Mal krank gewesen ist. Die Folgerungen, die aus einem Profil gezogen werden, können zu Aktionen führen (*application*) bevor das Profil ein Merkmal tatsächlich aufweist.

Distributive und nicht-distributive Attribute

Bei dieser Folgerung wird außer Acht gelassen, dass nicht alle Attribute für alle Entitäten einer Gruppe berechnet werden können oder sollten. Hildebrandt (2006b, 2008) differenziert innerhalb von Gruppenprofilen zwischen *distributiven* und *nicht-distributive* Attributen. Distributive Attribute sind solche, die auf alle Mitglieder der Gruppe gleichermaßen zutreffen. Das Gruppenattribut „Abteilung A ist auf Etage 3“ lässt sich auf alle Personen der Abteilung übertragen (Person B der Abteilung A arbeitet auf Etage 3). Anders verhält es sich mit nicht-distributiven Attributen, die nicht auf alle Personen einer Gruppe übertragen werden können. Dies lässt sich anhand des oben genannten Beispiels und des Zusammenhangs zwischen Urlaubs- und Krankheitstagen erläutern: Wenn 75 % der Personen, die ihre Urlaubstage nicht ausschöpfen, mehr Krankheitstage haben als der Durchschnitt, lässt sich dieser Wert nicht automatisch generalisieren. Nicht nur weil eine Wahrscheinlichkeit von 75 % weniger als 100 % ist, sondern auch weil weitere Analysen nötig sind, um auszuschließen, dass nicht andere Faktoren für diesen Zusammenhang wichtiger sind, wie zum Beispiel dass 100 % der Angestellten, die ihre Urlaubstage nicht ausschöpfen, keinen Balkon haben.

Die Berechnungen funktionieren in der Regel auf Basis von Methoden des Data-Minings[31]. Wie auch in den genannten Beispielen werden Data-Mining-Prozesse oft als Blackbox dargestellt. Ein Algorithmus wird mit einem Datenset als Input gestartet und auf dem Bildschirm erscheint eine Zahl, die alle Informationen aggregiert und auf den Punkt bringt. In der Praxis und vor allem der Entwicklung von Data-Mining- Methoden ist die Sache komplizierter. Dabei treten regelmäßig Probleme auf (Domingos 2012), die im Ergebnis zu Ungenauigkeiten und einer Einschränkung der Aussagekraft von Profiling führen. Diese sind einer kurzen Betrachtung wert.

Datenerhebung

Zuallererst müssen Daten erhoben werden, deren Verarbeitung und Analyse sich anschließt. Merkmale über Personen, Dinge und Verhalten müssen als Information ab-

31 Vor allen Dingen in der Literatur um 2000 wird auf den gesamten Prozess häufig als *Knowledge Discovery in Databases* verwiesen, von dem Data-Mining nur ein Schritt ist. In den letzten Jahren hat sich der Begriff Data-Mining aber weitestgehend durchgesetzt und meint in der Regel heute alle Elemente.

strahiert und digitalisiert werden. Danach muss eine geeignete Form der Daten-Repräsentation gefunden werden. Das meint auf der einen Seite rein technisch die Frage, ob Daten klar strukturiert in Tabellen vorliegen, die schon pro Spalte eine Kategorie vorsehen und in einer relationalen Datenbank gespeichert werden, oder sich, wie etwa bei einer Warenkorbanalyse[32], die zu ermittelnden Zusammenhänge auf die Beziehungen zwischen den Artikeln beziehen. Klar ist, dass bereits das Aufstellen von Kategorien eine Filterung mit sich bringt. Dinge, die nicht in einer Tabelle erfasst werden, können auch nicht mit anderen zusammen analysiert werden. Auch beim Erstellen einer Datenrepräsentation werden bereits erste Kategorisierungen vorgenommen. Nur wenige Datenbanken erlauben zum Beispiel eine Differenzierung zwischen Geschlecht und Gender, geschweige denn eine andere Klassifizierung als die binäre in solchen Datenfeldern.

Unstrukturierte Daten

Technisch komplexer ist die Analyse unstrukturierter Daten wie von Fließtexten oder Videos, bei denen der erste Schritt darin besteht, analysierbare Einheiten zu konstruieren (wie etwa das Erkennen von Augen in einem Gesicht). Weniger komplex ist dagegen die Datenerhebung bei Interaktionen, die bereits vermittelt stattfinden. Der Mausklick auf einen Link wird, etwa beim Online-Tracking, als digitale Repräsentation der Entscheidung eines_r Nutzer_in zur weiteren Recherche über den Begriff erfasst, der in dem Link dargestellt wurde. Die Bewegung der Hand mit der Maus, die Bewegung des Mauszeigers auf dem Computerbildschirm, der vom Betriebssystem gesteuert wird, und der Klick im Browserfenster auf einen bestimmten Punkt kumulieren in einem „Event“, das der Browser für die Skripte messbar macht.

Skalierbarkeit und domänenspezifisches Wissen

Erwähnenswert ist, dass existierende Data-Mining-Algorithmen nicht beliebig skalieren. Das Hinzufügen weiterer Datenfelder bedeutet in der Regel eine nicht-lineare Komplexitätssteigerung. Je mehr „Dimensionen“ miteinander in Verbindung gesetzt werden, desto länger braucht ein Computer zur Berechnung der Zusammenhänge. Und auch wenn viele dieses Problem als vorübergehend ansehen, weil die Rechenleistung stetig steigt, so erhöht eine größere Menge an Daten auch das Risiko, dass sich keine Zusammenhänge mehr errechnen lassen, weil am Ende `alles mit allem` zusammenhängt. Am Beispiel der Drogeriemarktkette aus der Einleitung lässt sich dieses Phänomen gut veranschaulichen. Obwohl der *Target* Konzern sicherlich die technische Ausstattung besitzt, um eine Berechnung aller Produkte durchzuführen, wird nur eine Liste von 25 Artikeln betrachtet, aus denen der *pregnancy score* ermittelt wird. Data-Mining setzt domänenspezifisches Wissen voraus. Wer die Produktpalette kennt, kann eine Menge Artikel schon aus der Erfahrung heraus ausschließen. An dieser

32 Das Beispiel wird meist in Verbindung gebracht mit Amazons Kaufempfehlungen. Die Analyse von Konsument_innendaten ist allerdings schon lange ein wichtiges Anwendungsfeld für Data-Mining (siehe etwa Berry und Linoff 1997).

Stelle lassen sich viele Data-Mining Algorithmen kritisieren, da durch diese Vorauswahl häufig auch Vorurteile in die vermeintlich objektiven Berechnungen einfließen.

Auswertung; Trainings- und Testdaten

Die Auswertung so erhobener Daten erfolgt zwar in der Regel mit bekannten statischen Methoden, deren Ausführung allerdings von Fall zu Fall unterschiedlich ist. Bei der Entwicklung von Algorithmen wird häufig mit Trainings- und Test-Daten gearbeitet. Die Trainingsdaten enthalten Informationen zu allen Dimensionen und auch zu dem gewünschten Ergebniswert (ob schwanger oder nicht). Aus dem Testdaten-Set wird diese Information für einen Test entfernt. Nachdem mittels der Trainingsdaten Gewichtungen für die einzelnen Dimensionen errechnet worden sind, wird mit einer Evaluierungsfunktion getestet, ob die Werte für diesen anderen Datensatz ebenfalls plausibel sind. So soll ein *Overfitting* verhindert werden. Also dass der Algorithmus für einen konkreten Trainingsdatensatz eine sehr hohe Genauigkeit hat, weil etwa die *pregnancy prediction* immer 0 oder 1 ist, aber auf anderen Datensätzen nicht funktioniert, weil die Zusammenhänge `zu genau` gelernt wurden. Auf der anderen Seite muss auch eine zu große Generalisierung verhindert werden. Wenn es nicht um eine binäre Entscheidung geht, sondern Datensätze zu Clustern zusammengefügt, also gruppiert werden sollen, besteht die Gefahr, dass nur eine Kategorie entsteht, die alle Datensätze enthält und keine Differenzierung möglich ist.

Korrelation ist nicht Kausalität

Bei der Frage der Bewertung der Ergebnisse ist wiederum eine menschliche Einschätzung notwendig. Beliebt in Statistikeinführungskursen sind Beispiele für ermittelte Korrelationen zwischen Datenpaaren, die mehrere Erklärungen zulassen. So gibt es zum Beispiel einen statistischen Zusammenhang zwischen Einkommen und der Schuhgröße, weil sowohl die *Gender Pay Gap* in der Bezahlung als auch ein gewisses Mindestalter nicht mitbetrachtet wurden. Auch in der Detailbewertung ist ein Fokus auf die Zahlen nur bedingt aussagekräftig. In Brasilien wurde bei einer Marktanalyse festgestellt, dass es in den Armenvierteln eine steigende Nachfrage nach Fernsehern gegeben hat. Bei Feldstudien wurde dann aber klar, dass sich es nicht etwa um eine Frage des Fernsehers als Statussymbols handelte, sondern das Gerät als einfache Möglichkeit gesehen wurde, Kinder dazu zu bringen, das Haus nicht zu verlassen, um sie vor den Folgen der Bandenkriege zu schützen (Lee & Sobol 2012).

Während sich bei der einmaligen Analyse eines Datensatzes viele der beschriebenen Probleme umgehen lassen, können sie bei dynamischen, selbst-lernenden Analysen kaum ausgeschlossen werden.

2.3.3 Anwendungsbeispiele von Profiling

Personalisierung vs. Profiling

Bevor verschiedene Formen des Profilings vorgestellt werden, ist es notwendig Profiling abzugrenzen von einfacher Personalisierung, insbesondere von Webanwendun-

gen. Personalisierung meint unterschiedliche Formen der Anpassung von, insbesondere elektronischen, Dienstleistungen. Profiling ist nur eine (technische) Art, Personalisierung zu realisieren. Dabei reicht Personalisierung von der Bereitstellung verschiedener Sprachversionen, die Nutzer_innen auswählen können oder die auf Basis globaler Einstellungen im Betriebssystem automatisch vorgenommen werden, bis zu *collaborative filtering* (Linden, Smith und York 2003), bei dem Nutzer_innen einer Plattform gemeinsame Taxonomien erstellen, um Inhalte zu filtern. Das von Amazon eingesetzte Empfehlungssystem, bei dem Bücher mit ähnlichen Titeln oder von den gleichen Autor_innen vorgeschlagen werden, kann ebenfalls noch als Personalisierung verstanden werden. Erst das Vorschlagen von Buchtiteln auf Basis der aggregierten Profile („andere, die dieses Produkt gekauft haben, haben auch jenes gekauft"), geht über einfache Personalisierung hinaus.

Profiling zur Risikobewertung

Ein wichtiges Anwendungsfeld für Profiling ist die Risikobewertung. Von einer Person wird eine Liste von Eigenschaften zu einem Profil zusammengefasst und dieses wird ins Verhältnis zu anderen Profilen gesetzt. Dabei soll ein bestimmtes Merkmal verglichen und die Abweichung von einer gewünschten Norm berechnet werden.

Scoring

Ein Bereich, in dem diese Form des Profilings schon seit längerer Zeit angewendet und diskutiert wird, ist das Kreditscoring. Bei der SCHUFA[33] werden aus einer Menge an Einzelinformationen über Personen sogenannte Scores für die Kreditwürdigkeit errechnet. Die Einzelinformationen stammen dabei in der Regel von Banken und Vertriebspartnern der SCHUFA und umfassen neben Namen und Geburtsdatum aktuelle und vorherige Wohnadressen, eine Liste aller Kreditverträge, Kontoeröffnungen, oft auch Telefonverträge, vor allem aber auch Informationen über offene Forderungen und Einträge über „abweichendes Zahlungsverhalten". Anhand dieser Informationen werden mittels statistischer Verfahren *Scores* berechnet, die eine Aussage darüber treffen sollen, wie wahrscheinlich es ist, dass eine Person eine Kreditrate, eine Wohnungsmiete oder eine Telefonrechnung nicht bezahlen wird. Der Score wird vierteljährlich berechnet und anfragenden Unternehmen mitgeteilt, die dann damit unterschiedlich weiter verfahren. Je nach Geschäftskonzept entscheidet sich die Bank oder der Telefonanbieter einen Vertrag einzugehen oder nicht. In der Regel fällt die endgültige Entscheidung durch eine_n Mitarbeiter_in, um nicht eine verbotene automatisierte Einzelentscheidung (siehe 2.2.2) durchzuführen.[34] Verbreitet sind dabei einfache Ampelsysteme. Ist ein Score im grünen Bereich – wie genau die Bereiche aufge-

33 Die SCHUFA, gegründet als Schutzvereinigung der Banken in Deutschland, ist das Unternehmen, das über den umfangreichsten Scoring-Datensatz verfügt.

34 Eine ausführliche Diskussion der Praxis des Scorings bei der SCHUFA findet sich unter anderem bei Kamp und Weichert (2005) und Weichert (2006).

teilt sind, variiert – wird der Vertrag meist ohne weitere Prüfung eingegangen. Im gelben Bereich liegt die Entscheidung in der Hand der Mitarbeiterin oder des Mitarbeiters – oft ist dann noch die Zustimmung einer anderen Person, etwa des oder der Vorgesetzten notwendig. Wer einen Score im roten Bereich hat, hat meist keine Chance.

Ziel des Scorings ist es, eine vermeintlich objektive Messgröße für die Vertrauenswürdigkeit zu haben, Vertrauen zu schaffen, wo es um hohe Risiken geht oder wo es nicht effizient wäre sich ein persönliches Bild von dem_der Geschäftspartner_in zu machen. Der Score trifft auf Basis von Informationen, die über vergangenes Verhalten, das in die Datenbank eingetragen wurden, eine Aussage über die Wahrscheinlichkeit von zukünftigem Verhalten. Möglichkeiten, diese personenbezogene Daten zu beeinflussen gibt es nur indirekt, indem man sich etwa möglichst konform verhält und dafür sorgt, dass der Score der eigenen Wahrnehmung der Vertrauenswürdigkeit entspricht. Die Frage, was genau `konform` ist, lässt sich hier nicht beantworten, weil die genaue Berechnungsweise der Scores nicht öffentlich ist.

Fehler in den Scores

Die Diskussion um die Genauigkeit der Daten und die Fairness des Scorings ist gleichzeitig auch regelmäßig Argumentationsgrundlage, um Profiling-Maßnahmen einzufordern, die auf vermeintlich genaueren Daten beruhen. Tatsächlich ist eine gewisse Fehlerquote nicht zu vermeiden. Bei der SCHUFA ermittelte die Stiftung Warentest vor einigen Jahren, dass 1 % der Daten falsch, 8 % veraltet und 28 % unvollständig sind (Stiftung Warentest 2010).

Big Data Scoring

Die eingangs beschriebene Erweiterung des SCHUFA-Profils um beliebige öffentlich abrufbare Informationen, zum Beispiel aus sozialen Netzwerken, hat die SCHUFA zwar bisher nicht umgesetzt, allerdings wurde dieser Markt mittlerweile von anderen Anbieter_innen besetzt. Die in Hamburg ansässige, aber in Deutschland nicht tätige Firma Kreditech[35] vergibt Kleinkredite über das Internet. Die Entscheidung über die maximale Kredithöhe und Zinsen trifft dabei ein Algorithmus automatisch anhand der Analyse von mehreren tausend Datenpunkten, die unter anderem aus sozialen Netzwerken aber auch der Benutzung der Webseite gewonnen werden.[36]

Risikobewertung

Ein weiteres wichtiges Anwendungsfeld für *Risikobewertungen* findet sich im Bereich der staatlichen Sicherheitsorgane. Insbesondere bei der Flugsicherung in den USA wurde nach den Anschlägen vom 11. September 2001, aber auch schon davor, auf Profiling zurück gegriffen. Einer der Versuche, Profiling in der Breite zu nutzen, war das CAPPS II-Projekt *(Computer Assisted Passenger Prescreening System II)* (vgl. Bar-

35 Siehe HTTP://KREDITECH.COM (letzter Zugriff 26.09.2016). Vgl. auch (Müller, Rosenbach, und Schulz 2013).

36 Eine Brücke zum im weiteren Verlauf diskutierten Profiling für Werbezwecke schlägt ein Patent von Apple, das den Kontostand bei Werbeanzeigen berücksichtigt (Aaltonen und Ahopelto 2015).

nett 2004; Caulkins 2004). Dabei sollten Informationen aus den Datenbanken der Fluggesellschaften in Kombination mit diversen Datenbanken der Sicherheitsbehörden gekoppelt werden, um potentiellen Terrorist_innen vorab identifizieren, beziehungsweise die Wahrscheinlichkeit, dass eine Person terroristische Absichten haben könnte, vorab berechnen und diese dann am Flughafen genauer kontrollieren zu können. Das Projekt stieß nach Bekanntwerden, unter anderem bei Bürgerrechtsinitiativen, auf große Ablehnung und wurde 2006 eingestellt, nachdem öffentlich geworden war, dass die Datenbank, auf deren Basis das Profiling durchgeführt wurde, zu über 38 % fehlerhafte Datensätze enthielt. Nichtsdestotrotz sind ein Großteil der Datenbanken, die zur Analyse des `Terroristen Scores` herangezogen werden sollten, weiterhin in Benutzung und dienen als Input für das heute eingesetzte *Secondary Security Screening Selection*-Verfahren, bei dem einzelne Fluggäste, deren Namen über Profiling ermittelt wurden, einer gesonderten Überprüfung und Befragung am Flughafen unterzogen werden.

Verhaltensvorhersage und Werbung

Anders wird Profiling und Data-Mining im Werbesektor genutzt. Hier wird mit großen Datenmengen versucht, Gruppen von Personen zu identifizieren, denen anschließend spezifische (englisch: *targeted*) Produkte angeboten werden. Solche Datenlisten existierten auch schon vor den Möglichkeiten der digitalen Verarbeitung (Gandy 1993). Das im Adresshandel tätige Unternehmen Schober gibt an, bis zu dreihundert Merkmale über jeden Haushalt – nicht die Einzelpersonen – bereitzuhalten und daraus Ziel- und Interessengruppen gebildet zu haben, die für Werbetreibende interessant seien. Die Daten werden durch die Bewertung von Häusern und Wohngegenden vorgenommen.

> Die Schober Einzelhausbewertung umfasst über 19 Mio. Gebäude in Deutschland, die persönlich vor Ort nach neun Kriterien bewertet werden. Über die Wohnverhältnisse werden signifikante Informationen zu Kaufkraft und Konsumverhalten gewonnen.[37]

Online Behavioural Targeting

Ähnlich operieren Tracking-Provider und Werbetreibende im Internet. Erstere sammeln Informationen und aggregieren sie zu Profilen anhand derer Werbung geschaltet wird. Die Funktionsweise des Trackings erlaubt aber auch ausgefallenere Werbetechniken. Acxiom – eine im Online-Tracking tätige Firma – beschreibt in einer Produktpräsentation folgendes Szenario:

> Mr Higgs sieht bei Facebook, dass eine Freundin den Onlineshop Bryce geliked hat, was ihn dazu verleitet sich die Webseite des Ladens an-

37 Aus einer Werbebroschüre der Schober Group HTTP://WWW.SCHOBER-DIRECT-MEDIA.DE/CMS/WP-CONTENT/UPLOADS/2014/10/SCHOBERDM_FACT-SHEET_MARKETBASE-LIVING.PDF (letzter Zugriff 26.09.2016).

> zuschauen und nach Druckern zu suchen, weil er vor hat sich bald einen neuen zu besorgen. Da der Onlineshop mit Facebook verknüpft ist, wird das Facebook Profil mit seinem Verhalten bei Bryce verknüpft. Er registriert sich im Weiteren auf der Seite des Shops, kauft aber keinen Drucker. (eigene Übersetzung nach Singer (2012))

Als er am nächsten Tag auf einer Sportnachrichtenseite surft, wird sein Profil wiedererkannt und ihm wird Werbung für jene Drucker angezeigt, die er am Vortag nicht gekauft hat. Als er daraufhin wieder die Seite des Shops besucht, wird ihm vom System ein Rabatt angeboten, wenn er jetzt kaufen würde. Der Rabatt erscheint nicht zufällig; durch die Verknüpfung mit Facebook und das Wissen um die Sportseite, die Mr. Higgs angesurft hat, konnte Acxiom ihn in die Kategorie „cleverer Single" einsortieren – eines von siebzig Clustern, denen unterschiedliche Profile zugeordnet werden. Das System weist ihm deswegen die Eigenschaften „mobil, obere Mittelklasse, online Banker, Profisport-Fan und preissensibel" zu. Für dieses Cluster empfiehlt Acxiom eben einen Rabatt. Im Szenario ist Mr Higgs begeistert von dem Angebot und kauft den Drucker.

Conversion Rate

Beim sogenannten *Online Behavioural Targeting* wird versucht, die *Conversion Rate* - d. h. die Anzahl der Verkäufe, Anmeldungen oder einfach Klicks – pro geschalteter Bannerwerbung zu erhöhen. Die Werbenden versuchen, ihre Werbung auf bestimmte Personengruppen zu lenken und diesen im Internet zu folgen und etwas über ihr Verhalten zu erfahren, um dann eine Anzeige möglichst häufig anzuzeigen und, etwa durch spezielle Preisgestaltung, Kund_innen zu gewinnen. Die Kategorisierung von Alter, Geschlecht, Ausbildung und Einkommen gehört zum Standardrepertoire der Tracking-Provider[38]. Um aber noch genauer Zielgruppen und Persönlichkeiten spezifizieren zu können, wie im Beispiel mit Mr Higgs, werden über bekannte Zusammenhänge Zusatzinformationen hinzugefügt.

Studien die Daten korrelieren

Chittaranjan, Blom und Gatica-Perez (2011) korrelieren die Art der Smartphone-Nutzung, die sich mit Apps tracken lässt, mit Persönlichkeitsprofilen.[39] Die Zuweisung gelingt ihnen mit einer Genauigkeit von 75 % auf Basis von Anruf- und SMS Listen, Bluetooth-Verbindungen und Nutzungsstatistiken von Apps. In einer anderen Studie (Epp, Lippold und Mandryk 2011) ermitteln die Autor_innen einen Zusammenhang zwischen der Art und Weise des Tippens auf einem Smartphone und einem emotionalen Zustand mit einer Genauigkeit von 84 %. Backstrom und Kleinberg (2014) wiederum nutzen die Analyse von Freundschaftsbeziehungen auf Facebook, um die Entwicklung

38 Untersuchungen dazu finden sich bei De Bock und Van den Poel (2010) oder Murray und Durrell (2000), ähnlich funktioniert auch das in Fußnote 36 erwähnte Verfahren von Apple.

39 Sie verwenden dazu die „Big Five" genannten Hauptdimensionen der Persönlichkeitspsychologie: Neurotizismus, Extraversion, Offenheit für Erfahrungen, Gewissenhaftigkeit und Verträglichkeit.

von Beziehungsstati vorherzusagen. Sie berechnen z. B. für zwei Personen, die viel über Facebook kommunizieren, die Wahrscheinlichkeit, dass sie in naher Zukunft ihren Status auf „verheiratet“ ändern werden. Eine weitere Studie ermittelt mit einer Genauigkeit von 80-90% sexuelle Orientierung, ethnische Kategorisierung und politische Ausrichtung anhand von Facebook Likes (Kosinski, Stillwell und Graepel 2013). Youyou, Kosinski und Stillwell (2015) konnten, ebenfalls anhand von Facebook Likes, Persönlichkeitsmerkmale konstruieren. Exemplarisch sind in dieser Studie auch Bewertungen für Genauigkeit angegeben, wie sie auch im Online-Profiling verwendet werden. Klassifizierungen mit einer Korrektheit über 85 % gelten als „Goldstandard“ (vgl. Epp, Lippold und Mandryk 2011) und 20 % falsche Zuordnungen (false positives) sind absolut akzeptabel.

Zusammenfassung

Im Unterschied zu den vorher beschriebenen Scoring- und Profiling-Mechanismen ist das Online-Profiling wesentlich dynamischer. Anbieter_innen versuchen, kontinuierlich mehr Informationen über Nutzer_innen in Erfahrung zu bringen, um die Kategorisierung zu verfeinern. Jeder Kauf bzw. auch Nicht-Kauf fließt dabei zurück in die Algorithmen, die das Feedback nutzen, um sie zu optimieren. Das Ziel ist dabei die Beeinflussung der Nutzerinnen und Nutzer in ihren Kaufentscheidungen und ihrem Verhalten (ausführlich dazu Abschnitt 3.1). Der Unterschied zur Risikobewertung liegt in der tendenziellen Ignoranz für die tatsächliche Identität, da die Aussage darüber, in welche Kategorie man einsortiert wird, wichtiger ist als der Name oder die Adresse.

2.4 EINE GESELLSCHAFT DES PROFILINGS

Profiling und dessen Folgen sind aus unterschiedlichen wissenschaftlichen Perspektiven beschrieben worden. Im Folgenden werden einige dieser Beschreibungen aufgegriffen, um das im Weiteren fokussierte Online-Profiling zu kontextualisieren. Auf die Beschreibung der konkreten individuellen wie gesellschaftlichen Folgen von Profiling und personalisierten Informationsströmen folgt eine Diskussion von Profiling als Technik der *liquide* gewordenen Überwachung.

2.4.1 Profiling als Bevormundung

Serendipity und Filterblasen

Einige Autor_innen kritisieren Profiling, beziehungsweise darauf aufbauende Personalisierung, weil es Nutzer_innen Informationen vorenthält. Die vermeintlich gute Absicht des automatischen Vorfilterns könne die negativen Effekte einer Filterblase (*filter bubble*) (Pariser 2012) oder einen Verlust von Zufallsfunden (*serendipity*) (Meckel 2012) nicht aufwiegen. Pariser und Meckel sehen eine Gefahr für das Zusammenleben darin, dass einer Person bestimmte Inhalte, vor allem im Internet, vorenthalten werden. Dies führe dazu, dass konträre Meinungen nicht mehr wahrgenommen werden oder über Ereignisse nicht mehr ausgewogen berichtet würde. Letztlich sehen sie dar-

in eine Gefahr für pluralistische Demokratien, die von kritischen und widersprüchlichen Meinungen profitieren würden. Auf Profilen basierte Vorfilterung, deren vordergründiges Ziel die automatische Aufwertung „relevanter" Inhalte zur Reduktion einer Informationsflut ist, könnte gleichzeitig zur (unerwünschten) Ausblendung alternativer Meinungen führen. Beispiele für solche Filtermechanismen sind zum Beispiel der automatisch angepasste Newsfeed im sozialen Netzwerk Facebook (Zuckerberg u. a. 2010) sowie die Filterung von Suchergebnissen bei Google (Pariser 2012).

Einschränkung dezisionaler Privatheit

In der Diskussion um Filterblasen wird weniger die Verwendung personenbezogener Daten kritisiert als die Art und Weise, wie sie ausgewertet und die Auswertungen angewendet werden. Das Fehlen von Kontingenz oder *Serendipity* wird damit weniger die informationelle als die dezisionale Privatheit beeinträchtigen. Die Befürworter argumentieren allerdings, dass solche Filterungen die Autonomie respektieren und getroffene Entscheidungen bestätigt werden, indem Beiträge von Personen, mit denen man viel interagiert, und zu Themen, die als relevant identifiziert wurden, häufiger angezeigt werden. Allerdings können autonome Entscheidungen gleichzeitig auch nur getroffen werden, wenn sie in einem Umfeld stattfinden, das eine Vielfalt von Entscheidungsmöglichkeiten bereitstellt und gegebenenfalls auch eine Rücknahme erlaubt. Wenn allerdings vor allen Dingen Meinungen und Informationen angezeigt werden, die der eigenen Position entsprechen, ist es schwieriger, davon abweichende Entscheidungen zu treffen und zu rechtfertigen. Genauso besteht die Gefahr, dass frühere Entscheidungen sich langfristig auf die Filterung auswirken, weil sie stabile Persönlichkeitsentwicklungen annehmen.

Begrenzte Beeinflussung

So verändert Personalisierung auch die Art, wie mit Informationen umgegangen wird. Informationen werden nicht gesucht, sondern, so Stelter (2008), es wird angenommen, dass man darüber „stolpern" wird, wenn etwas denn nur relevant genug ist. Hier wird also Autonomie an einen Informationsassistenten abgegeben, ohne dessen Funktionsweise zu kennen. Welchen Einfluss diese Informationsassistenten auch auf den emotionalen Status einer Person haben, hat eine viel kritisierte[40] Studie von Kramer, Guillory und Hancock (2014) beschrieben. Die Autor_innen nahmen minimale Veränderungen am Newsfeed-Algorithmus von Facebook vor und konnten so nachweisen, dass eine Person, die mehr schlechte Nachrichten liest, sich auch in den eigenen Beiträgen negativer äußert.

Studien zur Filterblase

Die Gefahr der Filterblase und mangelnder *Serendipity* für dezisionale Privatheit besteht vor allem dann, wenn die darunterliegenden Algorithmen von linearen Persönlichkeiten ausgehen, deren Positionen bestätigt werden sollen, und Nutzer_innen ungewollt und unbemerkt personalisierte Informationen angeboten bekommen. Wenn

40 Die Kritik bezog sich vor allem auf das Studiendesign, da die Testpersonen nicht über das Experiment informiert wurden.

nicht regelmäßig neue Profile mit „neuen Einstellungen“ angelegt werden, besteht die Gefahr, dass Änderungen im Verhalten durch die eigene Historie überschrieben werden. Tatsächlich zeigen erste Studien, dass sich der Effekt bei Filmempfehlungen mittelfristig umkehrt (Nguyen u. a. 2014), sich auch Neutralität technisch herstellen lässt (Kamishima u. a. 2012) und Werkzeuge existieren, die den negativen Effekten entgegen wirken (Bozdag und van den Hoven 2015). Darüber hinaus sind der informationellen Bevormundung in westlichen Demokratien aktuell auch noch real-weltliche Grenzen gesetzt. Einerseits sind Menschen in der Regel noch in weitere soziale Kontexte eingebettet und nutzen auch andere Informationsquellen, in denen andere Meinungen nicht vorgefiltert werden, und auch online bleiben nicht-personalisierte Nachrichtenseiten existent. Nichtsdestotrotz besteht die Gefahr, dass die negativen Effekte zukünftig realer werden, wenn Datensammlungen und Profiling zunehmen.

2.4.2 Profiling als Teil des Marktgeschehens

Ein weiterer, viel diskutierter Anwendungsbereich von Profiling ist Preisdifferenzierung *(price discrimination).* Es beschreibt das ökonomische Prinzip, den „optimalen“ Preis zwischen Anbieter_in und Nachfrager_in zu erreichen. Während das Prinzip sich real-weltlich, etwa in höheren Spritpreisen an Autobahnen im Vergleich zu anderen Tankstellen, manifestiert, ist es im Bereich des Online-Shoppings bereits wesentlich weiter ausdifferenziert (Calo 2013; Danna und Gandy 2002; Odlyzko 2003). Mittels Profiling wird versucht, einen angemessenen Preis, nicht etwa pro Produkt in einem bestimmten Kontext, zu ermitteln, sondern ihn pro Kund_in und pro Einkaufsvorgang aus Sicht der Anbieter_innen zu optimieren. Es soll der Preis ermittelt werden, den die_der Kund_in maximal bereit ist zu zahlen.

Price discrimination in der Praxis

Das Prinzip der *perfekten Preisdifferenzierung* ist umstritten. Während einige Dienste nach Experimenten und Protesten die Anwendung eingestellt haben (American City Business Journals 2000), gibt es immer wieder Studien, die in einzelnen Bereichen automatische Preisdifferenzierung und -schwankungen nachweisen. Zuletzt berichteten Hannak u. a. (2014) von massiven Preisveränderungen in Abhängigkeit vom Login Status eines_r Nutzer_in, aber auch mit Bezug zu dem verwendeten Betriebssystem und Gerät. Valentino-DeVries u. a. (2012) konnte eine Abhängigkeit vom vermeintlichen Wohnort des_der Kund_in nachweisen. Gleichzeitig konnten Studien, wie die von Vissers u. a. (2014), keine generelle Preisdifferenzierung auf Preisvergleichsportalen entdecken. Aktuell gehen Beobachter_innen (vgl. Broderick 2015) aber davon aus, dass Preisdifferenzierung weiter zunehmen wird.[41]

41 Eine nicht repräsentative Untersuchung zeigte zuletzt 2014 (360pi 2014), dass bei 15-20% der Produkte bei großen Online-Händlern die Preise täglich neu berechnet werden.

Neben der umstrittenen aktiven Preisdifferenzierung durch den_die Anbieter_in funktionieren personalisierte Bonuskarten und Gutscheine nach dem gleichen Prinzip, führen aber zu weit weniger Widerspruch (Narayanan 2013). Den Kund_innen werden Rabattgutscheine oder andere indirekte Preissenkungen angeboten, wie etwa der Verzicht auf Versandkosten (vgl. das Acxiom Beispiel oben). Diese Angebote haben dabei dieselbe Wirkung wie eine aktive Preisveränderung, nämlich den Preis, abhängig von der_m einzelnen Kund_in, zu ändern, allerdings meist nach unten. Zusätzliche Ziele sind die langfristige Kundenbindung und der Verkauf weiterer Produkte.

Politische Kampagnen

Nach, im Sinne des Marketings, erfolgreichem Einsatz dieser Strategien in der Produktbewerbung werden ähnliche Techniken auch in anderen Bereichen eingesetzt. Scherer (2012) berichtet z. B. über die Werbekampagne während des Präsidentschaftswahlkampfes in den USA, bei der Personen segmentiert und im Anschluss mit speziellen Argumenten umworben wurden, für die ermittelt wurde, dass sie für die konkrete Gruppe besonders überzeugend seien.

Panoptic Sort

Für den Prozess der Segmentierung von potentiellen Kund_innen oder Wähler_innen in Zielgruppen hat Gandy (1993) den Begriff des *panoptic sort* geprägt. Er beschreibt damit den Markt der Adresshändler_innen[42] und die Entwicklung hin zu immer ausgefeilteren Segmentierungen, die mit Beginn des 20. Jahrhunderts in Form von Umfragen und Zielgruppenanalysen ihren Anfang nahmen. Dazu werde, so Gandy weiter, eine immer größer werdende Zahl von Merkmalen herangezogen und Listen von Personen für alle möglichen Produkte und Dienstleistungen entworfen. Gandy teilt die Funktionsweise des *panoptic sort* in drei Schritte.

Identifikation, Klassifikation und Assessment

Identifikation ist im ökonomischen Geschäft einerseits notwendig, um Vertrauen bilden zu können, auf dessen Basis Transaktionen abgeschlossen werden. Andererseits erlaubt die Identifikation einzelner Konsument_innen im *panoptic sort* auch eine bessere Ausgestaltung der *Klassifikation*. Hier bezieht sich Gandy auf Foucault, der in seiner Arbeit beschrieb, wie Klassifikation in Disziplinargesellschaften[43] genutzt wird, um die Effizienz und Genauigkeit von Disziplinarmaßnahmen (die sowohl Belohnung als auch Bestrafung sein können) zu steigern. Nach der Identifizierung und Klassifizierung dient das *Assessment* der Komplexitätsreduktion. Dabei können die unterschiedlichen Klassifizierungen als Hinweise für eine Auswertung des entstandenen Profils genutzt werden, um eine Ausgangsfrage – etwa, ob ein_e Konsument_in es wert ist, Werbung gesendet zu bekommen oder nicht – zu entscheiden.

---

42 Nach einem aktuellen Bericht existieren allein in den USA 4000 Data-Broker (Dixon 2013: Senate Committee on Commerce, Science, and Transportation), deren Geschäftsmodel Adresshandel und Segmentierung zur Marketingzwecken ist.

43 Foucault verwendet als Beispiele häufig die Systeme von Schulen, Militär oder Krankenhäusern, in denen jeweils spezifische Regeln existieren und disziplinarisch durchgesetzt werden.

2.4.3 Profiling und Diskriminierung

Die Segmentierung von (potentiellen) Kund_innen in lohnenswerte und weniger lohnenswerte Ziele hat nicht nur ökonomische, sondern auch weitergehende soziale Aspekte. Wie fließend die Übergänge von Segmentierung zu Diskriminierung sind, zeigen die folgenden Beispiele.

Redlining

Danna und Gandy (2002) beschreiben in ihrem Aufsatz die Strategie eines Online-Versandhändlers, der im Jahr 2000 einen US-amerikanischen Lebensmittellieferdienst aufbauen wollte. In erster Linie, um Zahlungsausfälle zu verhindern, bot der Dienst Lieferungen nur in bestimmte Stadtteile an, die anhand ihrer Postleitzahl unterschieden wurden. Diese Praxis wird als *redlining* bezeichnet, da sich in vielen Städten der USA die ethnische Segregation mit der Einteilung in Stadtbezirke deckt. So lässt sich über die vermeintlich unbedenklichen Postleitzahlen eine (vereinfachte) ethnische Zuordnung und Segmentierung organisieren. Der Ausschluss bestimmter Stadtbezirke von der Verfügbarkeit einer Dienstleistung ist also gleichzeitig ein rassistischer Ausschluss.

Segregation in Werbeanzeigen

Ein zweites Beispiel betrifft die Wortwahl bei Online-Werbeanzeigen. Sweeney (2013) zeigte, dass sich auch hier die Segmentierung zu Marketingzwecken mit rassistischen Vorurteilen deckte. Eine Personensuchmaschine, die Werbung für ihren Service bei Google schaltet, adaptierte den Text in Anzeigen abhängig von der angenommenen Hautfarbe der Person, nach der gesucht wurde. Suchte man nach „Latisha Jefferson“ - ein Vorname, der eher Afroamerikanerinnen zugeordnet wird - warb die Seite mit dem Satz „Latisha.. was arrested?“. Bei „Emily Jefferson“ - ein `weiß` konnotierter Name - nur mit „We found Emily“. Die Beschriftung der Anzeigen stand jeweils in keinem Zusammenhang mit den Daten, die dem Dienst tatsächlich über die Person vorlagen. Die Werbeanzeigen wurden, abhängig von der Sucheingabe, durch den_die Werbeanbieter_in (hier Google, definiert durch die Regeln des Kunden) generiert. Sie hatten keinen Bezug zu den Daten in der Datenbank des Kundenunternehmens selbst. Das heißt, die Anzeige lässt keinen Schluss auf die beim beworbenen Unternehmen vorliegenden Daten zu. Dieser Umstand ist aber nicht zwangsläufig allen Internetnutzer_innen bekannt. Stattdessen suggeriert die Anzeige das Vorhandensein von Informationen und damit einen Zusammenhang zwischen der Person, nach der gesucht wird, und deren Vorstrafenregister, ausschließlich abhängig von dem Namen der Person. Hier werden also gleich zwei Annahmen getroffen, die ausschließlich auf statistischen Daten beruhen, aber auf eine einzelne Person angewendet werden: einerseits die Verteilung von bestimmten Vornamen und der wahrgenommenen Ethnie und andererseits die Verteilung von Ethnien in Bezug auf den Anteil derjenigen, die bereits inhaftiert waren und über die eine öffentliche Dokumentation vorliegt.

Sexistische Diskriminierung

Ähnliche Diskriminierungen konnten Datta, Tschantz und Datta (2015) in ihrer Untersuchung nachweisen. Sie fanden auf einer indischen Webseite mehr Anzeigen zu gut bezahlten Stellenangeboten, wenn sich der_die User_in zuvor in den werbebezogenen Einstellungen von Google als „männlich“ definiert hatte.

2.4.4 Liquid Profiling und Identitätskonstruktion

Segmentierung zur Disziplinierung

In Anlehnung an Foucaults Analyse des Panoptikums (Foucault 1977) beschreibt Gandy die Folgen des *panoptic sort* als eine Entwicklung, die durch die Digitalisierung und Informationssammlung aller Lebensbereiche vorangetrieben wird. Anders als im Panoptikum, wo die Gefangenen nicht wissen können, ob sie beobachtet werden, sei es nun die ständige und häufig nicht geheime Anwesenheit von Klassifizierungsmechanismen, die darüber entscheiden, welche Leistungen einer Person zugebilligt werden. Damit würden bestehende Strukturen verfestigt, aus denen nur ausbrechen kann, wer seine oder ihre Vergangenheit löscht und damit unauswertbar macht.

Auch anhand der für papierbasiertes Marketing entworfenen Listen stellt Gandy bereits fest, dass es bei der Verwendung soziodemografischer Daten, in Abhängigkeit von der Sortierung nach Postleitzahlen, schwierig ist, von *personenbezogenen Daten* zu sprechen, obwohl der Schutzbereich einer Person tangiert wird. Er beschreibt, dass die Nutzung der anonymisierten Daten erst bei der Anwendung auf Dritte einen Effekt auf eine einzelne Person entfaltet, die aber nicht die Person ist, auf die sich die Daten ursprünglich bezogen haben.

> [This is an] *inferential* difference machine. Its predictions are based on information gathered from samples of persons and samples of their behaviors. Information gathered from particular individuals is frequently most useful in developing approaches to *other* individuals who may remain unknown to the organization until they respond to a promotional appeal. (Gandy Jr 1996:142)

Social Sorting

Gandy thematisiert damit schon früh, was in Form von Online-Profiling aktuell perfektioniert wird. Die Klassifizierung und Segmentierung von Gruppen anhand verschiedener Eigenschaften, die im jeweiligen Kontext relevant sind, existierte auch schon vor der automatisierten Datenverarbeitung. Wie in Abschnitt 2.2 für den europäischen Raum diskutiert, kritisiert er auch für die amerikanische Jurisdiktion, dass die individualisierte Konzeption von *Privacy* nicht in der Lage ist, die Effekte dieser Machttechnik zu begrenzen.

Aufbauend auf den Arbeiten von Gandy formuliert Lyon (2002) die These von der *surveillance as social sorting*, nach der das Ziel staatlicher wie nicht-staatlicher Überwachung sich verschoben habe. Von der Überwachung konkreter „Gefährder_innen“ zu einer umfassenden, gleichzeitig aber verteilten Überwachung aller, deren Ergebnis

die ständige Klassifizierung und damit einhergehenden Diskriminierung jedes_r Einzelnen ist.

> Abstract data, now including video, biometric, and genetic as well as computerized administrative files, are manipulated to produce profiles and risk categories in a liquid, networked system. The point is to plan, predict, and prevent by classifying and assessing those profiles and risks. (Lyon 2002:13)

Dividuum

Dabei benennt Lyon an andere Stelle die Besonderheiten der Überwachung im Internet (Lyon 1998). Während Gandy ausgehend von Foucaults Arbeiten zur Biopolitik das *social sorting* als Spielart der Disziplinargesellschaft zur Klassifizierung und Regulierung der Bevölkerung beschreibt, geht Lyon mit Deleuzes Gedanken zur Kontrollgesellschaft (Deleuze 1990) davon aus, dass die neuen Strukturen sich anders verhalten als klassische Einschließungs-Milieus. Deleuze beschreibt den Übergang von der Disziplinar- in die Kontrollgesellschaft im Wesentlichen an der Verschiebung des Fokus von der Disziplinierung des einzelnen Körpers hin zur Kontrolle abstrakter *Dividuen*, deren Eigenschaften gemessen und in unterschiedlichen Situationen unterschiedlich bewertet werden, aber im Unterschied zum Individuum nicht als einzelne Einheit, sondern als Ansammlung von Merkmalen aufgefasst werden. Die Disziplinierung vollzieht sich auch nicht mehr in abgeschlossenen Einrichtungen, sondern kontinuierlich und abhängig von der jeweiligen Situation. Die Vielfältigkeit der Erscheinung des Dividuums, eben abhängig vom Kontext, steht im Widerspruch zur angenommen unteilbaren Einheitlichkeit des Individuums, das unabhängig vom Kontext mit sich selbst identisch ist.

Surveillance assemblage

Haggerty und Ericson (2000) bringen diese Entwicklung in Zusammenhang mit technischen Einrichtungen zur Überwachung, die sie, auf Grund ihrer Vielfältigkeit, als *surveillance assemblage* bezeichnen. Dieses Überwachungsgefüge umfasst verschiedene Überwachungs- und Kontrollmechanismen von staatlichen wie nicht-staatlichen Akteuren und einzelne Maßnahmen, wie Überwachungskameras genauso wie komplexe – gleichzeitig aber in ihren Zielen sehr beschränkte – Kontrolltechniken, wie die bereits erwähnte Fluggastkontrolle. Auch sie stellen fest, dass sich die Techniken nicht mehr auf den biologischen Körper des_der Einzelnen ausrichten. Anders als die vorherigen Autoren sehen sie die als Merkmale digitalisierten Eigenschaften des Körpers aber nicht vollständig losgelöst von seinem Ursprung.

> The observed body is of a distinctively hybrid composition. First it is broken down by being abstracted from its territorial setting. It is then reassembled in different settings through a series of data rows. The result is a decorporealized body, a ‘data double’ of pure virtuality. (ebd.)

data doubles

Social Sorting ist eine Praxis dieser Kontrollgesellschaft, die mit einem *data double* arbeiten. Sie bezieht sich weniger auf den Körper als auf seine Beschreibung und wichtiger noch: Ziel ist nicht die Normierung der Einzelnen und die Identifizierung von Ausreißer_innen abseits eines definierten Normalbereichs. Beim Profiling passieren vielfältige Klassifizierungen gleichzeitig. Auch wenn Ausschlüsse durchaus vorkommen, ist das eigentliche Ziel aber die Individualisierung und der Umgang mit vielfältigen Profilen.[44]

Online-Profiling lässt sich in seiner Funktionsweise als Element dieses Überwachungsgefüges beschreiben. Allerdings ist der Effekt der Deterritorialisierung nochmal verstärkt. Internetnutzer_innen sind gegenüber Online-Trackern nicht identifizierbar, die Person ist mit ihrem Namen und Körper nicht bekannt.

> [the] aggregation processes must omit much detail and only include what is deemed necessary for the pre-determined goal (the serving of relevant advertising). (McStay 2011:131)

Vielmehr werden aus zu einem bestimmten Zweck gesammelten Informationen (vgl. McStay) auf Basis der Interaktion eines_r Nutzer_in mit verschiedenen Webseiten *data doubles* entwickelt. Darüber lassen sich Profile generieren[45], die Daten enthalten, die gemeinhin als „personenbezogen“ beschrieben werden (Alter, Einkommen, Geschlecht). Dabei zeigen sich diese Profile in der Verarbeitung aber als die eigentlichen Subjekte, mit denen die Systeme interagieren, oder, wie Galloway schreibt: „The clustering of descriptive information around a specific user becomes sufficient to explain the identity of [a] user.“ (Galloway 2004:96)

Liquid profiling

Diese Merkmal-Cluster wiederum werden allerdings nicht mehr durch absolute Werte gemessen, sondern sind in Wahrscheinlichkeiten[46] repräsentiert und ändern sich potentiell mit jeder Interaktion des_der Nutzer_in:

> `You`, or any given subject are not seen as a stable entity but rather as an assemblage (a collection or set of relationships) who can be aggregated and disaggregated on the basis of advertising opportunities and the cluster of key user groups. (McStay 2011:137)

Die Profile können aus Data*streams*, also dem kontinuierlichen Fluss neuer Daten, gewonnen werden und sind selbst auch dynamisch. Dabei spielen die Interaktionen der Nutzer_innen mit einer Webseite, Informationen, die aus zusätzlichen Datenquellen zusammengetragen werden, sowie die Anforderungen an die Konstruktion der Ziel-

44 Diese Unterscheidung ist genau die, die bei Foucault die Disziplin von der Gouvernementalität unterscheidet. (Foucault 1982:787ff.)

45 Siehe dazu Abschnitt 4.7.

46 Analog zu der Beschreibung von Merkmal-Clustern (*data doubles)* als einer Identitätsverdoppelung beschreibt Esposito (2007), wie Wahrscheinlichkeitstheorie und Statistik zu einer Realitätsverdoppelung führen.

gruppen eine Rolle bei der Bestimmung des Profils. In Anlehnung an Lyon und Bauman (2013) kann man diese Form des Profilings daher auch als *liquid profiling* bezeichnen.

Identitäten unter Profiling

Die Beschreibung von Profiling als liquide, schwer zu überblickende Machttechnik der Überwachung und Klassifizierung lässt sich anschließen an die vorherige Diskussion um den Wert des Privaten für die Entwicklung von (weitestgehend) autonomen Subjekten. Infolge der Unmöglichkeit zu wissen, welche Profile über einen selbst vorhanden sind oder gebildet werden, sind auch die Möglichkeiten, über sie zu bestimmen und an der Konstruktion der eigenen Wahrnehmung (von sich selbst wie auch der Welt) entscheidend mitzuwirken, eingeschränkt.

Dass dies insbesondere auf Profile zutrifft, die im Bereiche des Online-Marketing erstellt werden, beschreiben Zwick und Dholakia.

> […] the ability of consumers to constitute their own identity in the digital marketspace is severely limited if not completely eliminated. (Zwick und Dholakia 2004:3)

Die Autor_innen führen die begrenzten Möglichkeiten der Selbstbestimmung in ihrer Arbeit einerseits darauf zurück, dass die Konsument_innen keinen Zugriff auf die Datenbanken selbst haben, und andererseits darauf, dass durch die Struktur der Werbeindustrie und das Gefüge der großen Zahl von Anbieter_innen eben auch vielfältige „Identitäten“ von einer Person existieren.

Nun fragen einige Autor_innen danach, wie die negativen Folgen abgewendet werden können, die diese Form der Identitätskonstruktion auf diejenigen hat, denen durch Profiling Identitäten zugewiesen werden. Hildebrandt (2006a) stellt die These auf, dass eine Kernidentität existiere, die nicht-vorhersehbar und damit profilierbar sei. Gleichzeitig fordert sie, eine Form von Privatsphäre zu bestimmen, die vor den Folgen des Profiling geschützt ist.

> Whatever our profile predicts about our future, a radical unpredictability remains that constitutes the core of our identity. […] we need a certain kind of personal space that is tuned to the multiplicity of the virtual. We need a design of public and private life that does not constrain us on the basis of past habits and inclinations. (Hildebrandt 2006a)

Selbstbestimmung und Informationsmodulation

Andere wiederum gehen davon aus, dass die technische Entwicklung nicht umkehrbar und diese Form der Identitätskonstruktion nicht mehr aufrechtzuerhalten sei. Stalder (2010) argumentiert, dass Identitäten sich nicht aus einem inneren Kern heraus bilden würden, sondern vor allem durch Interaktion. Diese seien heute geprägt von den technischen Bedingungen und Autonomie könne nur noch in Relation zu diesen Bedingungen gebildet werden. Auch Rössler hatte bereits festgestellt, dass sich autonome

Subjekte nicht ausschließlich entwickeln, indem jede_r für sich vereinzelt „sich selbst“ bestimmt, sondern relational in Interaktion mit anderen (siehe. 2.1.2). Allerdings argumentiert sie dennoch, ähnlich wie Hildebrandt, für die Notwendigkeit eines privaten Raums, in dem eine Person bestimmen können muss, mit wem und zu welchem Grad diese Interaktion stattfindet. Statt diese grundsätzliche Entscheidung über die Kontrolle von Informationsflüssen zu treffen, fordert Stalder, hier Möglichkeiten zu schaffen, die *Modulation* der Information zu kontrollieren. Allen Autor_innen ist gemein, dass sie an der Idee der autonomen Subjekte festhalten, die auf die eine oder andere Weise Akteure ihrer Identitätskonstruktion bleiben sollten. Stalders Argumentation ist allerdings insofern interessant, als er von der Idee, die auch häufig im Datenschutz vertreten wird, Abstand nimmt, dass die Technik immer durch die Einzelnen kontrollierbar sein muss. Er argumentiert auf der Ebene von Informationsnetzwerken, in denen der_die Einzelne agieren können müsse. Dieser Gedanke der Regulierung von Informationsflüssen ist im Prinzip ein kybernetischer, der davon ausgeht, dass die Blackboxen der Profiling-Systeme nicht als Ganzes zu kontrollieren sind, sondern nur der Fluss der Informationen in die Blackboxen regulierbar ist.

2.4.5 Profiling und Kybernetik

Das Verhältnis zwischen liberal-demokratischer Individualtheorie, die an die aktuelle Praxis des Datenschutzes angelehnt ist, und der Praxis automatisierten Profilings, das kaum noch Individuen kennt, lässt sich auch als ein Konflikt zwischen der liberalen und *kybernetische Hypothese* beschreiben. Zu letzterer lieferte das Autor_innenkollektiv Tiqqun (2007) eine Analyse, die die kybernetische Hypothese als Gesellschaftstheorie beschreibt, die angetreten sei, die liberale Hypothese abzulösen. Also jenes Verständnisses von liberalen Subjekten, das die Entwicklung des Autonomie- und auch Privatheitsbegriffs in den letzten Jahrhunderten geprägt hat.

Kybernetische Hypothese

Tiqqun benennen die kybernetische Hypothese als das wenig wahrgenommene Erbe der kybernetischen Forschung, die Mitte des 20. Jahrhundert begründet wurde. Als Gründungsmythos fungieren Norbert Wieners Arbeiten der 40er Jahre, in denen er Wahrscheinlichkeitsrechnung auf Verhalten anwendete (Scherffig 2009). Während sein Ziel anfänglich die Vorhersage der Bewegung von Betrunkenen und Bienen war, entstand in den Jahren des Zweiten Weltkriegs die Idee einer automatisierten Flugabwehr, die die Flugbahnen von Flugzeugen berechnen und vorhersehen sollte. Um Flugzeuge möglichst automatisch abschießen zu können, muss die Position eines Flugzeugs in der Zukunft vermutet werden, damit sich die Flugbahnen von Geschoss und Flugzeug, unter Berücksichtigung der Flugzeit der Rakete, kreuzen. Wieners Konzept nach sollte die Flugbahn aus der vergangenen Flugbahn ermittelbar sein, der

Algorithmus sich aber ständig in einer *Feedbackschleife*, anhand der Messungen der tatsächlichen Flugbahn, korrigieren.

Entstehung der Kybernetik

Basierend auf diesen frühen Arbeiten, entwickelte sich daraus in den 50er Jahren eine kybernetische Gesellschaftsidee, die von anderen Wissenschaften übernommen wurde. Zentraler Ort des Austauschs waren die Macy-Konferenzen, auf denen die kybernetischen Annahmen, Organismen seien durch kontinuierliche Messung und Feedback steuerbar, auf soziale Systeme übertragen wurden (Pias 2003). Mit Heinz von Förster begann in den 70er Jahren die Übertragung der kybernetischen Denkweise in Management-Strategien. Die Kybernetiker_innen behaupteten, dass „die Kontrolle über ein System durch einen optimalen Grad der Kommunikation zwischen seinen Teilen erreicht“ werden könne (Tiqqun 2007:23) und wendeten diese Denkweise unter anderem in Marketing-Instrumenten wie dem *behavioural advertising* an. Werbung wird dann nicht mehr als Kommunikation vom Sender (Werber) zum Empfänger (Konsument_in) gedacht, sondern als Feedbacksystem, in dem Kommunikation zirkuliert.

> [This] opens up the possibility of a blurring between transmission and reception, to the extent they may become indistinct. (McStay 2011:106)

Wieder ist in der Online-Werbung dieses Format optimiert. Werbesysteme und Webnutzer_innen stehen in einem kontinuierlichen „Austausch“, Anzeigen werden eingeblendet, das Verhalten beobachtet, dementsprechend die Anzeigen angepasst, was wiederum das Verhalten beeinflussen soll.

Regulierung in Feedback

Obwohl der Kybernetikbegriff heute nicht mehr so geläufig ist, entfaltet die Denkweise weiterhin ihre Wirksamkeit, zum Beispiel im Profiling. Eine Kommunikation, die eine Regulierung über Feedback erlaubt und im Anschluss auf den Gegenstand zurückwirkt, ist erst dann möglich, wenn eine Information von dem Gegenstand selbst abgetrennt verarbeitet werden kann. Wenn also Klickverhalten auf Webseiten abgefangen wird, um daraus, in Kombination mit von individuellen Bezügen befreiten statistischen Daten, Informationen über Verhalten und Persönlichkeitsprofile zu gewinnen, kann diese Information genutzt werden, um (meist mittels Werbung als Feedback) Verhalten zu regulieren. Die Änderung im Verhalten wird dann wiederum als Klickverhalten gemessen und die Feedbackschleife beginnt von vorne. Tiqqun stellen dazu fest, „[d]as Internet ermöglicht gleichzeitig, die Präferenzen des Konsumenten zu erkennen und sie durch die Werbung zu steuern.“ (Tiqqun 2007:37). Dabei „orientierte die Soziokybernetik sich in erster Linie als Untersuchung *des Individuums als Ort von Feedbacks*, als »selbstdisziplinierte Persönlichkeit«“ (ebd., S. 32). Die im Silicon Valley erdachte Denkschule[47] habe versucht, das Subjekt zu „einer fleischlosen

47 Zuerst hatten Barbrook und Cameron (1996) die Region als Entstehungsort der *kalifornischen Ideologie* ausgemacht.

Hülle [zu machen], zum bestmöglichen Leiter der gesellschaftlichen Kommunikation, zum Ort einer unendlichen Rückkopplung, die reibungslos vonstatten [gehe]“ (ebd.). Dies stehe im krassen Gegensatz zum liberalen Subjektbegriff, der alle Menschen mit dem „Mythos der Innerlichkeit“ aufgeladen habe. Dieses liberale Subjekt wird an vielen Stellen, insbesondere bei der Diskussion um Privatheit, adressiert, bei Persönlichkeitsrechten, wie dem der informationellen Selbstbestimmung genauso, wie bei der Beschreibung einer Kernidentität, wie sie auch Hildebrandt betont.

Das viele der Beispiele für Profiling aus den Bereichen (Un-)Sicherheitsmanagement und Marketing stammen hat viel mit der Anwendbarkeit der kybernetischen Ideen dort zu tun. Nach Morris (2012) sind sie nicht nur exemplarische Anwendungsfelder, sondern von ihrer Struktur und Funktionsweise bereits kybernetisch gedacht. Regulierung und Abweichungserkennung auf der einen sowie Segmentierung und Management der Aufmerksamkeitsökonomie auf der anderen Seite. Die *data doubles* oder *Dividuen,* die im Feedback analysiert und geformt werden, sind für beide notwendige Bedingungen und Voraussetzung, um zu funktionieren. Dabei sei noch einmal betont, dass Dividuen keine „vollständigen Abbildungen“ der Menschen in Profilen sind, sondern diese eine Auswahl aller erfassbaren Merkmale darstellen. Bei ihrer Erstellung nehmen viele Faktoren, technische wie organisatorische, Einfluss auf die Auswahl der „relevanten“ Merkmale, die zu – sich möglicherweise auch widersprechenden – Profilen führen.

Während die in 2.4.3 beschriebene Perspektive aus den *surveillance studies* das Profiling in eine gesamtgesellschaftliche Entwicklung einbettet und kommentiert, grenzt die Beschreibung über die kybernetische Hypothese diese klar von bestehenden Gesellschaftsformen ab. Das eröffnet die Möglichkeit, eine Außenposition[48] einzunehmen und sich zu fragen, welche Handlungsoptionen, jenseits von denen, die auf gesamtgesellschaftliche Entwicklungen zielen, bestehen.

Scrambling the informatics network

Morris folgert genauso wie Tiqqun, dass nur Anonymität die Möglichkeit schafft, sich der Feedbackschleife zu entziehen. Anonymität nicht im Sinne einer Maske, die die Identifizierung verhindert – auf die Identität ist Profiling nicht angewiesen. Tiqqun verwendet die Metapher des *Nebels*, der erst hergestellt werden muss, um innerhalb einer undefinierbaren Menge nicht mehr adressierbar zu sein. Anonymität muss dafür eher kollaborativ hergestellt werden, so dass die auf einzeln regulierbare Systeme (von Profilen oder data doubles) fokussierten Feedbackschleifen aufgebrochen werden. Als konkrete, wenn auch immer noch allgemein gehaltene, Maßnahme schlägt Morris das Durcheinanderbringen der Informationsnetzwerke („scrambling the informatics networks“) vor und verweist als Theorie auf das Modell des „black boxing the

48 Mit dieser Außenposition ist nicht der kybernetische Beobachter zweiter Ordnung gemeint, sondern tatsächlich eine Position außerhalb der Systemtheorie.

self“ (Galloway 2010). Eine Möglichkeit zur Umsetzung dieser Idee sind Methoden der Verschleierung (*Obfuscation*) gegenüber Datensammlern, auf die später noch ausführlich eingegangen wird (vgl. 3.4.2).

Zusammenfassung

Mit dem Verständnis von Profiling als Mechanismus, der auf einer kybernetischen Hypothese über die Funktionsweise von Gesellschaften basiert, kann der Konflikt mit den klassischen Datenschutztheorien deutlicher werden. Dazu wurden verschiedene Beispiele für die negativen Folgen von Profiling (Bevormundung, Beeinflussung und Diskriminierung) diskutiert und als Elemente einer liquide gewordenen Überwachung gekennzeichnet. Den Beispielen ist gemeinsam, dass sie nur noch bedingt mit autonomen Subjekten im liberalen Sinne interagieren, welche dadurch gekennzeichnet sind, dass selbst unter der Bedingung, das sich Personen in Rollen unterschiedlich verhalten und sich kontinuierlich in Abhängigkeiten befinden, eine Identität existiert, die sich durch personenbezogene Daten beschreiben lässt. Diese Verbindung zwischen einem mit sich identischen Subjekt und den es beschreibenden Daten ist unter der kybernetischen Hypothese (nahezu) aufgelöst. Beim Profiling werden Merkmale immer in Bezug auf Andere und in Abhängigkeit vom jeweiligen Kontext errechnet. Zielgruppen werden nach ihren (wahrscheinlichen) Persönlichkeitsmerkmalen und ihrer angenommenen Lebensweise ausgewählt, obwohl die Tracking-Provider eigentlich nur die besuchten Webseiten und die Version des verwendeten Browsers kennen. Gegebenenfalls werden aus Teilen der jeweiligen Profile in anderen Zusammenhängen weitere Wahrscheinlichkeiten für Eigenschaften generiert, die Herkunft dieser Information, genauso wie sie selbst, nach dem Ausliefern der Werbung vergessen oder – nach Evaluation der Reaktion – geändert. Die Entkoppelung eines Profils von einem Individuum ist nahezu absolut und vor allen Dingen jeder direkten Kontrolle – oder informationellen Selbstbestimmung – entzogen, da diese selbst im Profiling keine Relevanz hat.

2.5 VORSCHLÄGE ZU REGULIERUNG VON PROFILING

Nachdem im vorherigen Abschnitt eine *Analyse* von Profiling und dessen Verhältnis zu Gesellschaftsmodellen und -entwicklungen vorgenommen wurde, sollen im nächsten Abschnitt aktuelle Ansätze zum *Umgang* mit Profiling vorgestellt und diskutiert werden. Bis zu diesem Punkt wurden vor allem Kritik und Umschreibungen des (kybernetischen) Profilings besprochen. An einigen Stellen wurden aber bereits Vorschläge für den Umgang mit Profiling kurz angerissen, wie etwa die aktive Modulation der Informationsströme 2.4.4 oder zuletzt die Konstruktion von Anonymität durch Verschleierung 2.4.5. Bevor in den nächsten Kapiteln eigene Vorschläge zur Konkretisierung dieser Ideen gemacht werden, sollen zunächst bestehende Vorschläge vorgestellt werden. Diese sind unterteilt in normative, rechtliche, technische sowie marktorientierte

Ansätze[49], wobei die letzten beiden in Kapitel 3 noch ausführlicher besprochen werden.

2.5.1 Konzeptionelle Ansätze

Post-Privacy Utopien

Neben denjenigen, die einen Erhalt von Privatheit in der Form der ebenfalls normativen liberalen Hypothese fordern, sind insbesondere Verfechter_innen einer Welt der *Post-Privacy* (Brin 1999; Heller 2011) in den Diskursen präsent. Ausgehend von der Beschreibung des Kontrollverlusts (siehe 2.2.3) und der Beobachtung, dass scheinbar immer weniger Menschen sich um den Schutz ihrer Privatsphäre kümmern (können oder wollen), begegnen sie der Technologie affirmativ. Sie postulieren, dass früher oder später – teilweise ist der *point of no return* auch schon erreicht – die (Wieder)-Erlangung der Kontrolle über Informationsflüsse, zumindest für den_die Einzelne, nicht mehr möglich ist. Auf dieser Erkenntnis basiert die Forderung an die Welt, das normative Konstrukt Privatheit durch vollständige Transparenz aller Vorgänge zu ersetzen. Die Praxis der sozialen Netzwerke zeige, so Heller (2011), dass insbesondere bei jungen Menschen die Norm Privatheit keine Rolle mehr spiele und zu den Akten gelegt werden sollte. Dass eben jene allerdings nicht unbedingt aus einer affirmativen Haltung heraus so agieren, zeigen Studien wie die von Boyd und Marwick (2011). Sie weisen die These, dass junge Menschen weder Interesse an noch Verständnis für Privatheit hätten, zurück und beschreiben in ihren Arbeiten anhand vielfältige Beispiele, wie Jugendliche versuchen, Privatheit unter den zeitgenössischen medialen Bedingungen herzustellen. Gerade Minderjährige befänden sich aber in der prekären Situation, daran gehindert zu werden, die Kontrolle über Situationen, Orte oder Daten selbst herzustellen und *agency* zu übernehmen. Einerseits würden Eltern und andere Erwachsene die Normen definieren, etwa wenn es darum geht, Zugang zu privaten Räumen zu bekommen, andererseits würden technische Bedingungen, die von den Entwickler_innen der Plattformen aufgestellt werden, verhindern, dass bestimmte Informationen so „privat“ ausgetauscht werden können, wie es dem Bedürfnis der Jugendlichen entspricht. So heißt es an anderer Stelle: „While it's easy to be private in public offline, doing so online can be quite difficult and frustrating.” (Boyd 2010). Verantwortlich dafür seien insbesondere die Betreibenden von Online-Netzwerken, die bestimmte Funktionen, etwa zur Einschränkung der Sichtbarkeit von Beiträgen, nicht oder nur unzureichend implementieren.

Transparenz und Machtverhältnisse

Während Post-Privacy eine Verringerung der *ontological friction* evolutionär-innovativ als quasi naturgegeben annehmen, zeigen nicht zuletzt die Enthüllungen von Edward Snowden, dass nicht eine quasi natürliche Informationsverbreitung stattfindet, son-

49 Diese Struktur ist angelehnt an Lessig (2006), der in „Code“ ausführt, wie diese vier regulatorischen Formen zusammenspielen.

dern viel mehr eine auf klassischen Machtasymmetrien basierende Hierarchie Informationsflüsse steuert und auch Transparenz nur in eine Richtung schafft. Dementsprechend wendete sich die Diskussion um Post-Privacy nach einer kurzen Hochphase mit starken Utopien hin zu der Frage, welche Vorbedingungen existieren müssen, um die positiven Effekte von Transparenz und freien Informationsflüssen zum Vorteil aller nutzbar machen zu können.[50]

2.5.2 Rechtliche Ansätze

Möglichkeiten der Regulierung in Deutschland

Bereits in 2.2 wurden die aktuell gültigen gesetzlichen Regulierungen durch Datenschutzgesetze diskutiert sowie die Problematik beschrieben, die sich daraus ergibt, dass beim Profiling verarbeitete Daten in der Regel nicht-personenbezogen sind und damit aus dem Regelungsrahmen fallen. Nichtsdestotrotz werden aus der systemischen Datenschutzperspektive heraus Forderungen an eine Regulierung von Profiling aufgestellt. Pohle (2014) hat dazu neun Anforderungen formuliert, die eine Kontrolle solcher Systeme ermöglichen sollen. Grundannahme dafür ist allerdings, dass die eingesetzten Systeme voll kontrollierbar sind bzw. voll kontrolliert werden, was, wie Pohle selbst schreibt, der aktuellen Praxis widerspricht.

Das eine Regulierung in der bestehenden rechtlichen Situation möglich ist, beschreibt auch Schuler-Harms (2005). Sie sieht zwar keine Möglichkeit der Anwendung des Rechts auf informationelle Selbstbestimmung unter subjektiv-rechtlichen Aspekten, wie sie die Datenschutzgesetze vorsehen, meint aber, objektiv-rechtliche Aspekte, nach denen jede_r wissen können müsse, „wer, was, wann über einen weiß“, seien durchaus betroffen. Bekannte Urteile, die Profiling auf Basis dieser Rechtsauslegung untersagen, existieren allerdings nach Kenntnis des Autors aktuell nicht.

Schutz vor unberechtigter Anwendung von Profilen

Auf europäischer Ebene wurden schon vor einigen Jahren Forderungen laut, den Anwendungsbereich von Datenschutz zu erweitern, beziehungsweise weitere gesetzliche Regelungen zum Schutz von Privatheit aufzunehmen. Gutwirth und Hert (2008) argumentieren, mit Bezug auf Profiling, für eine Trennung von Privatheit und Datenschutz. Ersteres sei als Grundrecht aufzufassen und zweites zur Regulierung der Verarbeitung personenbezogener Daten einzusetzen. Auch wenn verschiedene Autor_innen das Urteil des Bundesverfassungsgerichts ähnlich interpretieren, hat sich, wie oben beschrieben, ein Verständnis von Profiling als Grundrechtsverletzung in der Rechtsprechung bisher nicht durchgesetzt. An anderer Stelle (Gutwirth 2008; Hildebrandt 2006a) schlagen die Autor_innen vor, neben dem Schutz personenbezogener Daten auch die Anwendung von Profilen in den Schutzbereich des Datenschutzrechts aufzu-

50 Beispielhaft ist hier Seemann (2014) mit der Forderung nach Plattformneutralität.

nehmen.[51] Die Erweiterung von Datenschutz auf jegliche Datenverarbeitung, so argumentieren sie, sei keine Revolution, da bereits die Direktive 2008/58 den Schutz von (nicht zwangsläufig personenbezogenen) Daten aus der Verkehrserfassung wegen der Privatheitsverletzung untersagt habe.

Profiling im europäischen Recht

Ebenfalls mit dem aktuell gültigen europäischen Recht beschäftigen sich Forscher_innen im Projekt „Profiling“. Analog zur möglichen Auslegung des deutschen Rechts auf informationelle Selbstbestimmung konstatieren sie, dass die aktuelle Praxis des Profilings einen Grundrechtseingriff darstellt (Ferraris, Bosco, Cafiero, u. a. 2013; Ferraris, Bosco und D’Angelo 2013). Allerdings können sie, nach einer Befragung mehrerer europäischer Datenschutzbehörden, feststellen, dass weiterer und genauerer Regelungsbedarf besteht. Gleichzeitig ist eine Empfehlung des Ministerrats der Europäischen Union zu dem Thema nur in wenigen Mitgliedsstaaten umgesetzt (Bosco, D’Angelo und Vermeersch 2014).

EU Datenschutzgrundverordnung

Bewegung in die juristische Regulierung von Profiling kam 2012 durch den Vorschlag einer gemeinsamen Datenschutzgrundverordnung für die Europäische Union. Der erste Entwurf der EU Kommission enthielt einen Artikel, der Profiling regulieren und die Rechte der Betroffenen stärken sollte. Dabei ging der Entwurf weiter als der in der Intention ähnliche § 6a BDSG. *Erwägungsgrund 58* sah einen Erlaubnisvorbehalt für Profiling vor (Hildebrandt 2012). Allerdings enthält bereits der vom Parlament überarbeitete Vorschlag (siehe Albrecht 2013) Abschwächungen, die die Reichweite des Erlaubnisvorbehalts auf das Niveau des § 6a BDSG begrenzen, dafür aber ein Diskriminierungsverbot. Das Diskriminierungsverbot ist auch in der vom Parlament erlassenen Fassung enthalten, sieht aber, analog zum BDSG, nur vor, dass eine erhebliche rechtliche Wirkung nachweisbar sein muss, damit eine explizite Erlaubnis notwendig wird.

> Die betroffene Person sollte das Recht haben, keiner Entscheidung – was eine Maßnahme einschließen kann – zur Bewertung von sie betreffenden persönlichen Aspekten unterworfen zu werden, die ausschließlich auf einer automatisierten Verarbeitung beruht und die rechtliche Wirkung für die betroffene Person entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt, wie die automatische Ablehnung eines Online-Kreditantrags oder Online-Einstellungsverfahren ohne jegliches menschliches Eingreifen. (Erwägungsgrund 70, (O A 2016)

Informierte Einwilligung

Wie auch bei den meisten anderen Datenverarbeitungen ist aber vor allem der Einwilligungsvorbehalt ein Kritikpunkt. In der Regel werden solche Einwilligungen leichtfer-

51 Ähnlich schlussfolgert auch Schallaböck (2014) in einem Bericht für eine Bundestagsfraktion.

tig gegeben. Eine informierte Einwilligung sähe zudem vor, dass der_die Nutzer_in über mögliche Folgen eines Profilings vorab informiert würde (vgl. 2.2.2).

In der langen öffentlichen Diskussion um die Datenschutzgrundverordnung waren allerdings auch weitergehende Vorschläge gemacht worden. So forderte die Datenschutzgruppe des EU-Parlaments eine Art Whitelist für Bereiche, in denen Profiling gestattet seien solle, statt die zu bestimmen, in denen es verboten sei.

> In the Working Party's view, an approach should be taken that clearly defines the purposes for which profiles may be created and used, including specific obligations on controllers to inform the data subject, in particular on his or her right to object to the creation and the use of profiles.
> (Article 29 Data Protection Working Party 2013, 14)

Widerspruch und Transparenz

Darüber hinaus schlug Hildebrandt (2012) vor, einen Artikel zu Profiling so zu gestalten, dass dieser auch die Gruppenperspektive berücksichtigen solle. Sie forderte: „protection against the undesired application of profiles and the creation of transparency rights regarding group profiles" (ebd.:2). Dieser Schutz müsse erstens eine Art Widerspruchsrecht gegen die Anwendung von Profiling und zweitens weitergehende Rechte auf Transparenz in Bezug auf Gruppenprofile umfassen. Der Unterschied zur klassischen Datenschutzgesetzgebung ist im ersten Fall eklatant. Statt nur der Erhebung und Verarbeitung von personenbezogenen Daten widersprechen zu können, möchte Hildebrandt das Widerspruchsrecht ausweiten auf die Anwendung von Datenverarbeitungsverfahren. Diese Forderung nach einer Intervenierbarkeit in Datenverarbeitungsprozesse setzt allerdings voraus, dass Systeme derart gestaltet sein müssten, dass ihre Funktionsweise nachvollziehbar und transparent gestaltet ist.

Beobachter_innen gehen allerdings momentan nicht davon aus, dass diese Punkte ihren Eingang in die EU Datenschutzgrundverordnung finden werden. Stattdessen wird erwartet, dass es im Rahmen der weiteren Verhandlungen zu einer weiteren Schwächung der Regulierung von Profiling kommt oder dies nicht mehr als eigener Regelungsbereich gefasst wird.

2.5.3 Technische Ansätze

Um den negativen Effekten von Profiling auf der technischen Ebene zu begegnen, gibt es zwei unterschiedliche Ansätze. Zum einen solche, die auf der Seite der Datenverarbeitung eingreifen, und zum anderen solche, die die Betroffenen des Profilings zu unterstützen suchen.

Privacy preserving Data-Mining

In die erste Gruppe fällt Forschung im Bereich der Statistik, die sich seit Anfang des Jahrtausends unter anderem mit Verfahren des *privacy preserving* Data-Mining be-

schäftigt.[52] Dabei geht es darum, wie Datenanalysen und insbesondere Datenbankabfragen stattfinden können, ohne dass personenbezogene Daten dabei einsichtig sind. Beim *discrimination-aware* Data-Mining (Pedreshi, Ruggieri und Turini 2008) liegt der Fokus auf der Optimierung von Klassifizierungs-Algorithmen. Die Autor_innen versuchen, Diskriminierung in Klassifizierung messbar zu machen, indem sie solche Datenkategorien identifizieren, die einen besonders großen Einfluss auf das Ergebnis haben. Ziel solcher Anpassung der statistischen Methoden ist es, diese mit juristischen Anforderungen in Einklang zu bringen. An der grundsätzlichen Problematik der beschriebenen Einschränkungen von Privatheit ändern sie nichts.

Recht auf Vergessen

Anders verhält es sich mit solchen Algorithmen, deren Ziel die Datenzerstörung ist und die das *Recht auf Vergessen* umsetzen, wie es auch im Zuge der Diskussion um die EU-Datenschutzgrundverordnung diskutiert wurde. Aus technischer Sicht ist das eine schwierige Aufgabe, da sie den grundsätzlichen Prinzipien von Computersicherheit, bei der auf Redundanz und möglichst dauerhafte Speicherung gesetzt wird, entgegensteht. Tools wie Vanish (Geambasu u. a. 2009) oder Ephemerizer (Tang 2010) versuchen, ein Verfallsdatum für digitale Informationen durchzusetzen und ermöglichen es so, zumindest einen Teil der digital erfassten Vergangenheit zu verschleiern. Ein erfolgreicher *digitaler Radiergummi* würde Profiling zumindest insoweit erschweren, als dass weniger vergangenes Handeln auswertbar ist. Allerdings erweisen sich diese Strategien als wenig nützlich unter den beschriebenen Bedingungen der Anreicherung von Profilen durch zusätzliche Daten.

Privacy Enhancing Technologies

Weitere *Privacy Enhancing Technologies* konzentrieren sich auf die Anonymisierung der Daten der Nutzer_innen im Internet. Sei es, dass sie jedwede Internetverbindungen anonymisieren[53] oder zumindest die Inhalte der Kommunikation zwischen Nutzer_innen oder Nutzer_innen und Servern verschlüsseln.[54] Allerdings verhindert die Verschlüsselung von Inhaltsdaten nicht die Auswertung der sogenannten Metadaten, also solcher Informationen über Kommunikationsnetzwerke, die ebenfalls zum Profiling genutzt werden.

Eine wesentlich größere Anzahl technischer Entwicklungen fällt in den Bereich solcher *Transparency* und *Privacy Enhancing Technologies*, die direkt Tracking und Profiling zum Thema haben und den Fokus auf die Anwendung durch die_den Nutzer_in

52 Siehe dazu Dwork und Smith (2009); Grosskreutz, Lemmen und Rüping (2010); Matwin (2013).

53 Am weitesten verbreitet ist das TOR Netzwerk (HTTP://TORPROJECT.ORG; letzter Zugriff 26.09.2016).

54 Darunter fallen Protokolle wie OTR, GPG, S/MIME, HTTPS, PSYCH aber auch Anwendungen, die Daten innerhalb existierender Anwendungen verschlüsseln und verschleiern (*Cloaking*) vgl. D'Angelo, Vitali und Zacchiroli (2010) und Wanying Luo, Qi Xie und Urs Hengartner (2009).

legen. Es existiert eine Vielzahl von technischen Werkzeugen, die von Nutzerinnen und Nutzern eingesetzt werden können, um Profiling zu blockieren oder unbrauchbar zu machen. Neben der Möglichkeit, Werbung auszublenden oder mittels AdBlockern zu verhindern, dass Werbeserver mit sogenannten Tracking-Cookies das Surfverhalten protokollieren, gibt es mehrere Browsererweiterungen, deren Ziel es ist, die Profile unbrauchbar zu machen. Diese Anwendungen werden in Kapitel 3 ausführlich besprochen.

privacy as confidentiality

Technische Lösungen, die dem Schutz unterschiedlicher Dimensionen von Privatheit zuträglich sind, stehen in der Kritik, einseitig die Folgen des Kontrollverlusts durch Verschlüsselung rückgängig machen zu wollen. Dies trifft insbesondere auf Anonymisierungs- und Verschlüsselungswerkzeuge zu. Dieses *privacy as confidentiality* Paradigma (Gürses u. a. 2009) wird häufig als technikzentriert kritisiert, wobei die Bedürfnisse von Nutzer_innen ignoriert werden.[55] Der Fokus auf Vertraulichkeit impliziere, dass Daten, die einmal Dritten zugänglich gemacht worden sind, unwiederbringlich öffentlich seien. Dadurch werden all jene von der Nutzung der Tools ausgeschlossen, die sich mit einer, auch nur teilweisen, Preisgabe ihrer Daten arrangiert haben. Darüber hinaus kann man ergänzen, dass sich auch weitere, speziell mit Profiling zusammenhängende Probleme mit dem Vertraulichkeitsparadigma nicht fassen lassen. So kann auch die Nichtveröffentlichung von Daten schon zur Zuweisung von Merkmalen führen (zum Beispiel des Profils der_s „Datenschutzbewussten“). Diesem verkürzten Blick auf Privatheit stellen die Autor_innen unter anderem performative Strategien entgegen, deren Ziel nicht die (Rückgewinnung der) Kontrolle, sondern die Erprobung vom Umgang mit dem Kontrollverlust sind. Als eine dieser performativen Strategien wird auch hier *Obfuscation* (siehe 2.4.5) genannt, da hierbei nicht das Verhindern von Datenpreisgaben Anonymität schaffen soll, sondern die gezielte Steuerung von Informationen durch den_die Nutzer_in vorgesehen ist.

Privacy literacy und critical data literacy

Weitere Beispiele für die konstruktive Auseinandersetzung mit den Folgen von Data-Mining beschreibt Berendt (2012). Sie formuliert *critical data literacy* und darauf aufbauend eine *privacy literacy* als notwendige Kompetenz für diejenigen, die in einer Informationsgesellschaft partizipieren wollen und sich den Folgen, zu denen Profiling gehört, ausgesetzt sehen. Sie führt aus, dass sich diese insbesondere durch die (kritische) Anwendung von Data-Mining schulen lasse, wie sie durch viele der *Privacy Enhancing Technologies* angewendet werden. Werkzeuge, die *critical data literacy* schulen, kategorisiert Berendt anhand der Anwendung einer (Analyse-)Methode auf ein (Daten-)Objekt oder Thema. Diese Kategorisierung lässt sich auch auf Online-Pro-

55 Ähnlich, wenn auch populärwissenschaftlicher, argumentiert auch Morozov (2013), der auf generellerer Ebene Technologieentwicklung (und die Unternehmen und Personen, die diese vorantreiben) getrieben sieht von einem *Solutionism*, der an den tatsächlichen Problemen vorbeigehe und nicht-technische Lösungen weitgehend ignoriere.

filing übertragen und gibt Ideen für Werkzeuge, die eine *privacy literacy* in diesem Bereich fördern könnte:

- *Datenanalyse als Methode, Beliebiges als Objekt*: Nutzer_innen ohne spezielle Kenntnisse könnten durch die eigenhändige Anwendung von Data-Mining die Grenzen der Methoden kennenlernen und so einen kritischen Umgang mit Profiling und Statistik üben.
- *Datenanalyse als Methode, Datenanalyse als Objekt*: Durch die Untersuchung der Datenanalyse selbst, etwa durch die Darstellung der Grenzen der Auswertung, kann eine kritische Auseinandersetzung mit Profiling und Tracking stattfinden.
- *Datenanalyse als Methode, Verhalten in Data Spaces als Objekt*: Durch Datenanalyse soll ein kritischer Umgang mit (dem eigenen) Verhalten in Informationsräumen wie dem Internet eingeübt werden.

Eine Auseinandersetzung mit den Funktionsweisen von Profiling und den zu Grunde liegenden Daten soll eine kritische Auseinandersetzung mit dem Thema bewirken und Profiling so nachvollziehbar und die Folgen abschätzbar machen. Auf diese Weise kann zumindest der *kognitiven Asymmetrie* (vgl. 2.2.3) der Einzelnen im Umgang mit datenverarbeitenden Systemen entgegengewirkt werden. Diese Aspekte sowie die Idee der Obfuscation sind Anforderungen, die bei der Entwicklung, die im Weiteren vorgestellt wird, berücksichtigt und im nächsten Kapitel noch weiter diskutiert werden.

2.6 ZUSAMMENFASSUNG

In diesem Kapitel wurde die erste der in der Einleitung gestellten Leitfragen beantwortet (*Inwiefern werden Privatheit und Autonomie individuell wie strukturell durch technologische Entwicklungen wie Profiling beeinflusst?).* Zu Beginn wurde eine grundsätzliche Beschreibung und Herleitung von Privatheit vorgenommen und dabei die Dimensionen lokale, informationelle und dezisionale Privatheit vorgestellt (vgl. 2.1). Es wurde erläutert, wie diese Dimensionen von Privatheit in die Jurisdiktion liberaler Gesellschaften übernommen wurden, die den Fokus auf ein selbstbestimmtes Subjekt legt, dem ein Recht auf informationelle Selbstbestimmung zugesprochen wird (vgl. 2.2). Dieses Konzept wurde kontrastiert mit Beschreibungen von Profiling am Beispiel von Scoring und Vorhersagen von Verhalten und Persönlichkeitsmerkmalen (vgl. 2.3). Aufgrund verschiedener Merkmale wurde Profiling als Element einer *surveillance assemblage* identifiziert, die, auf viele verschiedene Systeme verteilt, Personen überprüft und im Rahmen eines *panoptic sort* Optionen einschränkt, nach denen Entscheidungen getroffen werden können. Dabei wird eine Person repräsentiert

durch eine Liste von Attributen, die als „beschreibend“ angenommen wird. Diese selektiven und beeinflussbaren *data doubles* wurden als widersprüchlich zur liberalen Hypothese befunden und die kybernetische Hypothese als Denkmuster benannt, die diesen Formen des Verständnisses der Menschen zu Grunde liegt (vgl. 2.4). Zuletzt sind dann Umgangsweisen beschrieben worden, die Einfluss nehmen wollen auf die Arbeitsweise der kybernetischen Profiling-Mechanismen. Dabei wurde festgestellt, dass entweder nicht zu erwarten ist, dass eine rechtliche Regulierung in naher Zukunft bevorsteht, oder dass sie, wie viele technische Entwicklungen, auf die (Wieder-)Herstellung eines autonomen Individuums fokussiert sind. Mehrfach sind Strategien der Verschleierung genannt worden, um Kontrolle zu verbinden mit der Möglichkeit, Autonomie über die Auswertung zu gewinnen. Alle Verfahren setzen aber einen bewussten Einsatz durch den_die Nutzer_in voraus, der_die um seine_ihre Privatheit und die Folgen von Profiling wissen muss (vgl. 2.5).

Die zweite Leitfrage richtet sich auf die Möglichkeiten einer genaueren Analyse von Online-Profiling. Ausgehend von der vorliegenden Analyse lassen sich erste abstrakte Anforderungen an ein technisches Verfahren zur Untersuchung von Online-Tracking und Interventionsmöglichkeiten beschreiben.

Weiterführende Erkenntnisse

Die Methode zur Untersuchung von Online-Tracking sollte diese als **Blackboxen** annehmen und die Variabilität der Daten berücksichtigen. Anbieter_innen von Online-Tracking haben zudem kein Verständnis von (Kern-)Identitäten, sondern arbeiten mit dem wahrscheinlichsten data double, das Identitäten der Nutzer_innen in Form von Profilen abbildet, für die abhängig vom Kontext bestimmte Attribute bestimmt werden. PETs sollten daher die Profile nicht als Abbilder der_s Nutzer_in verstehen. Stattdessen sollten sie einen performativen Umgang mit **(multiplen) data doubles unterstützen**. Dabei wird die Existenz von Profiling nicht zum Anlass genommen, es grundsätzlich zu unterbinden, sondern stattdessen nach Möglichkeiten gesucht, Nutzer_innen eine Auseinandersetzung mit Profiling zu ermöglichen (um *privacy literacy* zu schulen) und einen Aktionsraum zu schaffen, der über das „Erlaubnis/Widerspruch“-Prinzip hinausgeht. **Anonymität** kann dabei hergestellt werden durch Verschleierung und die Modulation von Datenflüssen in einer Art, dass Einzelne nicht mehr zu erkennen sind.

In den nächsten Kapiteln sollen diese Ziele in einem konkreten Projekt, der Untersuchung von Online-Profiling, angewendet werden. Dazu wird im nächsten Kapitel zuerst den Stand der Forschung mit Fokus auf informatische Anwendungen betrachten, um Best Practices zu identifizieren und Lücken zwischen den Zielen und aktuellen *Privacy* und *Transparency Enhancing Technologies* aufzeigen.

3. ONLINEWERBUNG, USER-TRACKING UND PRIVACY ENHANCING TECHNOLOGIES

Im vorangegangenen Kapitel wurde ausführlich auf die Beeinträchtigung von Privatheit durch Profiling und die Ansätze verschiedener Disziplinen zum Umgang damit eingegangen. Im Folgenden soll ein konkreter Anwendungsbereich, das Online-Profiling, in den Fokus genommen werden. Dazu erfolgen zuerst die Beschreibungen der technischen und ökonomischen Rahmenbedingungen anhand derer wichtiger Grundbegriffe sowie der Stand der Forschung im Bereich *Online-Profiling* und *Privacy Enhancing Technologies* erläutert werden. Nach einem kurzen Exkurs in den Bereich des Online-Marketings (3.1) werden die technischen Grundlagen von Online-Tracking erläutert, die das Profiling ermöglichen (3.3). Aufbauend auf der Erkenntnis des zweiten Kapitels, dass auch aus informatischer Sicht einige Möglichkeiten bestehen, mit Profiling aktiver umzugehen, folgt in 3.4 eine ausführliche Diskussion bereits existierender *Privacy and Transparency Enhancing Technologies*, die Transparenz und Interventionsmöglichkeiten anbieten. Das Kapitel schließt mit der Bewertung des Status Quo, einer Auflistung guter Beispiele von bereits vorhandenen Tools und weiteren Anforderungen an Software zur Beeinflussung von und Transparenz im Online-Profiling.

3.1 GRUNDLAGEN DES MARKETINGS

Werbung als antirationales Element

Vorweg einige kurze Bemerkung, die sich auf Werbung allgemein beziehen und im Anschluss an das vorangegangene Kapitel klar machen, wo der Zusammenhang zwischen Werbung und der Frage von Autonomie besteht. Im Marketing selbst wird Werbung nicht ausschließlich als Mittel zur Konsument_innenbeeinflussung gedacht. In der ökonomischen Literatur sei, so McStay, Werbung wie folgt definiert: Durch Werbung, die auf eine Vielzahl von Angeboten hinweist, würden Konsument_innen erst in die Lage versetzt, eine Entscheidung für eine_n bestimmten Anbieter_in und so eine optimale Entscheidung am Markt treffen zu können (McStay 2011). Werbung sei nicht dazu in der Lage, Konsument_innen dazu zu zwingen, Produkte zu kaufen. Stattdessen würde sie den Wettbewerb fördern und Informationen vermitteln. In diesem Sinne sei Werbung eine Maßnahme, von der alle Seiten profitieren würden: die Konsument_innen durch Informationen über das Angebot, die Anbieter_innen durch die Möglichkeit, Produkte zu verkaufen. Zusätzlich profitieren, insbesondere bei der Online-Werbung, die unterschiedlichen Marketingunternehmen, die zwischen beiden vermitteln. Diese Position wird in den Analysen von McStay als klassisch marktliberal

identifiziert. Die Grundannahme sei, dass vereinzelte Konsument_innen nur das Ziel haben würden, egoistisch die optimale Entscheidung am Markt zu treffen. Diese Haltung übergehe, so McStay weiter, die sozialen Normen und Werte, die ebenfalls einen Einfluss auf jede Entscheidung haben (z. B. das Produkt als Statussymbole) zu Gunsten einer Theorie, die in der Maximierung des Nutzens der_des Einzelnen auch eine Maximierung des Gemeinwohls sieht. Aber auch innerhalb der Ökonomie werde Werbung als gegensätzlich zum Marktinteresse kritisiert. Werbung lasse die Unterscheidung zwischen notwendigen Bedürfnissen und Wünschen verschwimmen und fördere anti-rationale Charaktere. Die Widersprüchlichkeit wurde in der Privatheitsforschung als „privacy paradox“ bezeichnet (2.2.3) und in Studien nachgewiesen. Acquisti und Grossklags (2005) haben nachgewiesen, wie sich die rationale Selbsteinschätzung in Bezug auf Datenpreisgaben von der tatsächlichen, werbe-geleiteten Praxis unterscheidet. Selbst wenn man also den Informationscharakter von Werbung in den Vordergrund stellt, lässt sich nicht abstreiten, dass eine gezielte Manipulation und Konstruktion von Wünschen Teil der Information ist. Dieser Manipulation ist es dann geschuldet, dass Autonomie eingeschränkt wird. Während diese Argumentation für einfache Anzeigen in einer Zeitschrift auch etwas zugespitzt klingt, ist sie in Bezug auf Online-Profiling leichter nachvollziehbar. Hier wird informationelle und dezisionale Privatheit (vgl. 2.1.3) durch das Erstellen und die Weitergabe von Profilen, die eben den Zweck haben, die Manipulation zu perfektionieren, beeinträchtigt.

Geschichte der Online Werbung

Im Folgenden nun ein Blick auf die Geschichte und Strukturen der Online-Werbung. Die Werbeindustrie hat das Internet früh für sich entdeckt und so wurde bereits ein Jahr nach der Einführung des grafischen Webbrowsers 1993 das erste Werbebanner geschaltet (Turow 2012). Von Beginn an waren Werbetreibende daran interessiert, die neuen Möglichkeiten der Messbarkeit von Anzeigen sowie deren Sichtbarkeit und Reichweite zu testen. Messbar war nicht mehr nur die Auflage (wie im Fall einer Zeitung), sondern jede einzelne *Impression* (=Aufruf) einer Seite, die gleichzeitig Details über den_die Aufrufende_n enthielt. „There’s no question that the Web is the most measurable of all media by far“, schrieb Arial Poler – Vorstandsmitglied einer der ersten Internetwerbeagenturen – in der Zeitschrift AdWeek bereits 1995.[56] Seither hat sich der Markt stark weiterentwickelt und kontinuierlich einen größeren Teil der Werbebudgets auf sich gezogen (Advertising Research Foundation 2007). Mit der Einführung von Cookies und JavaScript wurde die Messbarkeit der *clickstreams*, also des Verhaltens der Nutzer_innen im Internet, immer weiter ausgebaut. Mit dem steigenden Datenvolumen, das pro Webseite, pro Nutzer_in und pro Anzeige verarbeitet werden musste, stieg gleichzeitig die Anzahl und die Komplexität der Dienste, die die Daten automatisiert auswerteten. An Online-Werbung verdienen heute nicht nur diejenigen, die Werbeplätze anbieten (Publisher) und diejenigen, die Anzeigen schalten wol-

56 Zitiert nach (Turow 2012:45).

len (Marketing), sondern auch viele Zwischenhändler, die die Werbeflächen personalisiert vermitteln oder für diese Vermittlung Zusatzinformationen bereitstellen (Busch 2014). Der Markt der Vermittlungsdienste und die Möglichkeiten der Steuerung von Werbekampagnen haben sich stark ausdifferenziert und firmiert aktuell unter Begriffen wie *Programmatic Advertising* und *Marketing Clouds* (Born 2015). Im ständigen Wettbewerb überbieten sich die Anbieter_innen in den Versprechen, die sie in der Werbung für sich selbst machen; etwa in Bezug auf die Genauigkeit, mit der sie vermeintlich Anzeigen ganz bestimmten Zielgruppen (Segmenten) anzeigen können. Zudem werden regelmäßig neue Metriken entwickelt, die den Erfolg der Werbung messbar machen sollen.

3.2 FUNKTIONSWEISE VON ONLINEMARKETING

Abbildung 2 zeigt in einer Übersicht die Ausdifferenzierung der Rollen und Aktivitäten im Online-Anzeigen-Business. Die beiden Hauptakteure sind diejenigen, die Produkte oder Dienstleistungen vermarkten wollen, (Marketer) sowie diejenigen, die Anzeigeflächen anbieten (Publisher). In einer Broschüre des *Bundesverbands Digitale Wirtschaft (Fokusgruppe Targeting im BVDW 2014)* werden die beiden Akteure wie folgt beschrieben:

Publisher

Publisher sind Webseitenbetreibende, die auf ihren Seiten Werbeplätze anbieten. Im Prinzip kann auf jeder Webseite Werbung eingebettet werden, auf privaten Homepages genauso wie beispielsweise auf stark frequentierten Nachrichtenportalen. Webseitenbetreiber_innen können ihre Anzeigefläche entweder an eine_n Einzelne_n oder auch an mehrere verschiedene Zwischenhändler vermieten. Insbesondere vielbesuchte Webseiten können Exklusivverträge mit einzelnen Anzeigenvermarktern abschließen oder ihre Anzeigenplätze über eine *Supply Side Platform* (SSP) vermarkten.

Abrechnungsmetriken

Die Publisher verdienen dabei, je nach Geschäftsmodell des Zwischenhändlers, entweder dann, wenn eine Anzeige dargestellt wird (*Impression*), oder aber nur, wenn ein Link angeklickt wird (*per click*).

Demand Side und Data Management Platformen

Auf der anderen Seite stehen Marketingabteilungen von Unternehmen, die Werbung für ein Produkt oder eine Dienstleitung an eine möglichst genau definierte Zielgruppe bringen wollen. Dabei werden sie von Anbieter_innen so genannter *Demand Side Platforms* (DSP) unterstützt. Diese Plattformen erlauben eine genaue Steuerung und Kontrolle der Werbekampagnen in Echtzeit. Das Verfahren unterscheidet sich dabei nicht nur technisch von klassischen Methoden der Werbeplatzierung. Während in Printmedien Artikel oder Themenseiten ausgewählt werden, in deren Kontext eine Anzeige dargestellt wird (z. B. eine Autoanzeige auf der Seite eines Automagazins), kann die Platzierung im Internet über die Auswahl von Zielgruppensegmente oder auch Stich-

worten gesteuert werden (s. u. „Strategien“). In den meisten Fällen ist die Webseite, auf der die Werbung geschaltet wird, den Werbeschaltenden unbekannt. Die Anzeige zu einem neuen Auto kann also genauso gut auf der Seite eines IT Magazins erscheinen, dass über Fahrassistenzsysteme berichtet, oder auf einer Musikseite, bei der für eine_n konkrete_n Besucher_in gespeichert wurde, dass sie_er vor einigen Tagen den Namen eines Autoherstellers in eine Suchmaschine eingegeben hat oder auch generell Teil einer Zielgruppe ist, die sich tendenziell für Autos interessiert.

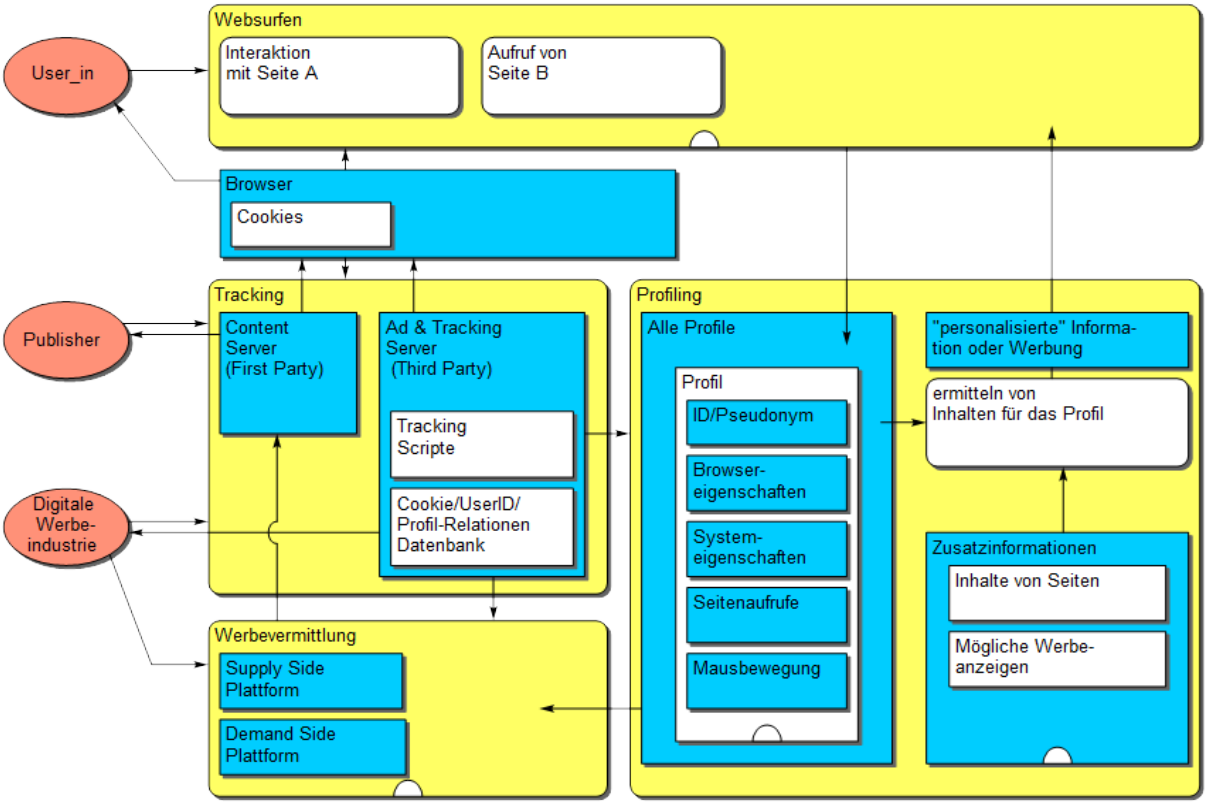


Abbildung 2: Funktionsweise von Online-Tracking und -Profiling inklusive weiterer Akteure.

Zwischen diesen Hauptakteuren, den *Publishern* und *Marketern*, sind verschiedene weitere Unternehmen tätig, die Teile der Werbevermittlung automatisieren oder Prozessteile zusammenführen, die unter Begriffen wie *Data Management Plattformen (DMP)* oder *Marketing Clouds* gefasst werden. Deren Ziele sind es, die unterschiedlichen Daten zu aggregieren und den Marketingabteilungen einen kontinuierlichen Überblick über laufende Kampagnen und Erfolge der geschalteten Anzeigen bereitzustellen.

Verarbeitete Daten

Um Werbung aber unabhängig von einer Webseite personalisiert für den_die aktuelle_n Benutzer_in anzuzeigen, müssen über diese eine Reihe von Daten gespeichert und ausgetauscht werden. Dabei werden in der Regel drei Arten von verarbeite-

ten Daten unterschieden (BVDW e. V. 2014). Unter *1st-Party*-Daten[57] versteht man solche, die bei dem_r Kunden_in direkt erhoben werden, wie E-Mailadressen oder sonstige Kontaktdaten, aber auch Daten über das Nutzungsverhalten eines_r Kunden_in auf der Webseite des Unternehmens selbst. *2nd-Party*-Daten sind solche, die das Verhalten eines_r potentiellen Kund_in auf externen Webseiten betreffen. Darunter fällt zum Beispiel, welche Suchbegriffe benutzt werden oder ob und wie mit einer Werbeanzeige interagiert wird. In den Bereich der *3rd-Party*-Daten fallen solche, die nicht direkt erhoben, sondern aus weiteren Quellen ermittelt werden. Damit gemeint sind „sämtliche soziodemografischen Daten wie Alter, Geschlecht oder Wohnort, psychografische Daten zu Einstellung, Verhalten und Meinung des Nutzers." (Brosche 2014:190).

Audience Measurement

Diese Informationen sind nicht nur für das Marketing interessant, sondern, im Sinne eines ökonomischen Managements von Online-Angeboten, auch für die Publisher. Aus Sicht der Publisher, die *Content Marketing* betreiben, ist *es* zum Beispiel von Interesse zu wissen, welche Altersstruktur Leser_innen haben, um Angebote darauf abzustimmen, gegebenenfalls anzupassen oder auszudifferenzieren. Gerade Informationswebseiten versuchen, die Zeit, die Nutzer_innen auf den Seiten verbringen, zu erhöhen und somit mittelfristig die Werbeeinnahmen zu steigern. Dem Marketing wiederum nutzt eine große Anzahl bekannter Eigenschaften dabei, Zielgruppen für Werbeanzeigen genauer zu spezifizieren und Kampagnen auf diese auszurichten. Auch im *e-commerce* zielt die Beobachtung der Besucher_innen, etwa eines Online-Shops, auf die Umsatzsteigerung und Kund_innenbindung. Als Messwert dienen hier die *Conversions*, das Verhältnis zwischen denjenigen, die eine Online-Shop nur besuchen, um sich Produkte anzusehen, und solchen, die auch tatsächlich kaufen.

A/B Testing

Ein weiterer Einsatzbereich der Nutzerverfolgung dient der Messung von Effekten von Änderungen an einer Webseite (ob in der Struktur, dem Inhalt oder dem Design). Beim *A/B-Testing* (Ash, Ginty und Page 2012:214 ff.) werden unterschiedlichen Nutzer_innen unterschiedliche Varianten (etwa einer Produktbeschreibung) angezeigt und beobachtet, welche Variante zu einer höheren *conversion rate* führt. So kann getestet werden, welche Formulierungen besonders zum Kauf animieren. Jede_r einzelnen Nutzer_in nimmt mit ihrer_seiner Interaktion an einem Versuch teil und gibt (meist ungewollt) Feedback an die Betreiber_innen.

---

57 Ähnlich der Diskussion in Abschnitt 2.2.2 erklärt der BVDW, dass aus juristischer Perspektive nur die *1st-Party*-Daten als personenbezogene Daten gelten, so lange alle weiteren Daten pseudonymisiert oder anonymisiert verarbeitet werden. Außerdem erklärt der BVDW Leitfaden mit Bezug auf einen Kommentar der Bundesregierung und der EU-Kommission, dass eine Aufklärung über Tracking in den Nutzungsbedienungen sowie die Möglichkeit eines Opt-Out als ausreichend rechtssicher anzusehen sei (BVDW e.V. 2014:26). Dieser Auslegung wurde jedoch zuletzt in einer Entschließung der Konferenz der Datenschutzbeauftragten des Bundes und der Länder (2015) widersprochen.

Real Time Bidding

Real Time Bidding (RTB) ist das Verfahren, mittels dessen ausgehandelt wird, welche Werbung für eine_n Nutzer_in auf einer Webseite angezeigt wird (Yuan, Wang, und Zhao 2013). Ruft ein_e Surfer_in eine Webseite auf, werden Informationen über die Seite (beispielsweise Daten über den Inhalt der Webseite) sowie über den_die Nutzer_in, der_die die Seite aufruft, zu einer RTB Plattform gesendet. Hier sind über *Demand Side Plattformen* Gebote für bestimmte Kombinationen, von User-Segmenten aber auch Webseitenkontext, hinterlegt. Der Anzeigenplatz geht dann an den_die meistbietende_n Werbekunden_in. Das Verfahren ist vergleichbar mit einer Hochgeschwindigkeitsbörse[58], auf der Werbeplätze angeboten werden, deren Wert, abhängig von verschiedenen Variablen, ermittelt wird. Die Marketingseite gibt dazu in der Regel den Maximalpreis an, den sie bereit ist, pro Anzeige zu zahlen. Neben hochpreisigen Werbeplätzen am oberen Ende von Webseiten, die auch ohne Scrollen sichtbar und häufig personalisiert sind, entstehen so auch Restposten, die häufig nicht oder nur auf wenige Attribute hin personalisiert sind.

Targeted/Behavioural Advertising

Eine der Strategien hinter der Personalisierung wird als *Targeting* bezeichnet (Advertising Research Foundation 2007:7–32; Hass und Willbrandt 2011). Von den bekannten Methoden setzen vor allem das *Re-Targeting* sowie das *Online Behavioural Targeting* auf Profiling. Beim *Re-Targeting* wird der Inhalt von Werbeanzeigen für solche Internetnutzer_innen angepasst, die die zu bewerbende Webseite (in der Regel einen Online-Shop) bereits besucht haben. Werbeanzeigen enthalten dann häufig eine Liste von Produkten, die der_die Nutzer_in sich bereits angeschaut, aber nicht gekauft hat. Ziel ist, wie im eingangs beschriebenen Beispiel von Acxiom, durch zusätzliche Preisanpassungen zu versuchen, die_den potentielle_n Kund_inn_en zu einem Kaufabschluss zu bewegen. Beim *Online Behavioural Targeting* oder *Advertising* (OBA) wird Werbung entsprechend angenommener Interessen ausgeliefert, die durch eine Analyse des Surfverhaltens einer_s Nutzer_in ermittelt werden. So können, unabhängig von der aktuell besuchten Seite, Werbeanzeigen zu Produkten geschaltet werden, nach denen vor einiger Zeit gesucht wurde und an denen die Person interessiert zu sein scheint. Dass die Personalisierung tatsächlich einen Effekt auf das Kaufverhalten hat, wurde in verschiedenen Studien nachgewiesen (Chen und Stallaert 2010; Farahat und Bailey 2012; Lambrecht und Tucker 2011).

Aus Sicht des Online-Marketings ist es also nachweislich sinnvoll *Online Behavioral Advertising* zu betreiben. Die Manipulation der Entscheidungsfindung und damit letztlich die Unterminierung der Autonomie der Betroffenen erfolgt auf Basis von Profilen, die zu den Nutzer_innen erstellt werden. Der nächste Abschnitt stellt konkrete Techniken vor, mit denen das Profiling durchgeführt wird.

---

58 In den Selbstbeschreibungen der Börsen wird angegeben, dass eine Transaktion nicht mehr als 50 ms dauert (vgl. Busch 2014).

3.3 ONLINE-TRACKING TECHNIKEN

Um Profile erstellen zu können, ist es nötig, die Bewegung einer_s Nutzer_in im Internet über mehrere Webseiten hinweg zu verfolgen (engl.: *tracking*). Die Liste der besuchten Webseiten wird häufig als Browserverlauf oder -chronik (engl.: *Browser History)* bezeichnet. Beim Tracking werden verschiedene Browser- und Web-Technologien verwendet und kombiniert, die im Folgenden vorgestellt und deren Verwendung im weiteren Verlauf der Arbeit analysiert werden. Diese Aufbereitung und die sich daran anschließende Änderung von dargestellten Anzeigen oder auch Webseiteninhalten ist *Profiling*.

Cookies

Am weitesten verbreitet ist Online-Tracking durch Cookies. Cookies sind textbasierte Elemente von Webbrowsern und Teil des HTT-Protokolls. Sie bestehen aus einem Key-Value-Paar (zum Beispiel. „Name = Martin“) und einer definierten Speicherdauer, die von dem_r Webseitenbetreiber_in festgelegt werden kann und nach deren Ablauf der Cookie automatisch gelöscht wird. Während dieser Speicherdauer wird ein Cookie bei jeder weiteren HTTP-Anfrage an diesen Server mitübertragen und ermöglicht so die Re-Identifizierung des Clients über den Browser. Über diese Funktion von HTTP lassen sich Webanwendungen gestalten, die einen Session-Status (zum Beispiel: „User_in ist eingeloggt“) voraussetzen, um eine Personalisierung zu ermöglichen. Da Cookies an den Browser gebunden sind, haben sie Vorteile gegenüber anderen Session-Varianten, wie die Bindung an die IP-Adresse der_s Nutzer_in. Allerdings birgt ihre Verwendung auch Risiken. Sie lassen sich bei unverschlüsseltem HTTP oder direkt auf dem Rechner auslesen und an andere Rechner übertragen.[59]

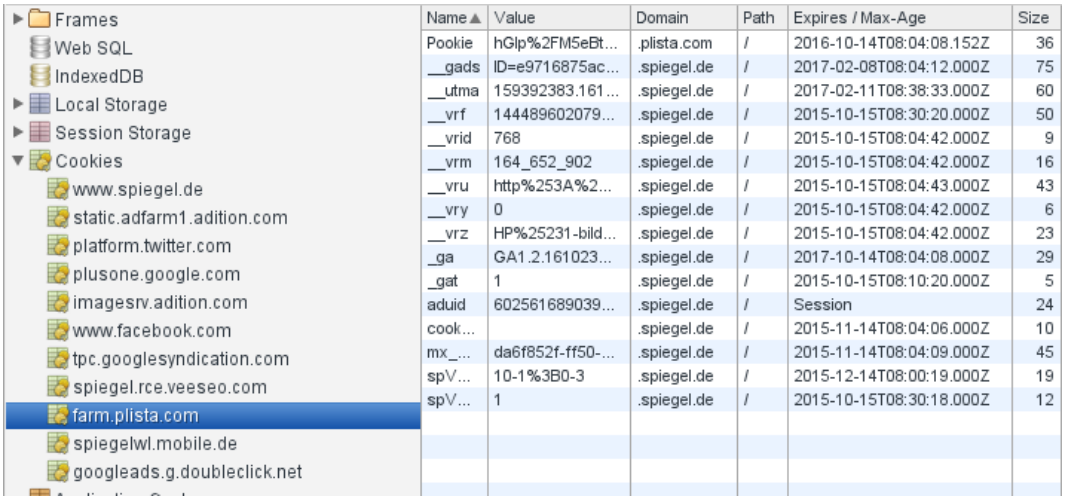


Abbildung 3: Screenshot des Browsers „Chrome“, der die Cookies darstellt, die beim Aufruf der Webseite http://spiegel.de auf dem Rechner gespeichert werden.

59 Mit *Firesheep* (Butler 2010) wurde eine Browsererweiterung vorgestellt, die Session-Diebstahl durch das Auslesen von Cookies in offenen WLAN vereinfacht.

First- und Third-Party Cookies

Abbildung 3 zeigt beispielhaft die Cookies, die beim Aufruf eines Artikels auf spiegel.de gesetzt werden. In vielen Browsern wird zwischen *1st-Party* und *3rd-Party* Cookies unterschieden. Erstere werden vom Server gesetzt, den der_die User_in zuerst angefragt hat, im Beispiel spiegel.de. Beim Aufruf der Seite könnten dann Ressourcen von weiteren Servern, z. B. Anzeigen von plista.com, nachgeladen werden. Bei diesen Aufrufen werden *3rdParty Cookies* von plista.com gespeichert, die *Plista* auch beim Laden von Werbung auf anderen Portalen nutzen kann, um den_die Nutzer_in zu reidentifizieren. Die Nutzung von Cookies zum Online-Tracking ist seit vielen Jahren der Standard und wird häufig in Kombination mit Tracking-Skripten oder Zähl-Pixeln verwendet.

Evercookies und Respawning

Flash-Cookies (auch *Zombie-Cookies*) sind eine Technik, die Cookie-ähnliche Prinzipien anwendet, aber für den_die Nutzer_in schwieriger zu löschen ist. Adobe Flash spielt dabei als am weitesten verbreitetes Browser-Plugin eine besondere Rolle. Mit Flash können LocalStorageObjects (LSO) auf dem Rechner des_der Nutzers_in abgelegt werden. Ähnlich wie bei Cookies ist ihr eigentlicher Zweck, Einstellungen, wie beispielsweise die bevorzugte Sprache oder Zeitzone, über die Dauer der Nutzung einer Flash-Animation hinaus zu speichern, sie werden aber eben auch zum Tracking verwendet. Während gängige Browser allerdings seit längerer Zeit Funktionen implementiert haben, die ein einfaches Entfernen von Cookies ermöglichen – über das Löschen der Browser-Historie – ist dies bei Flash nicht der Fall. Wie Soltani u. a. (2009) zeigen konnten, wird diese persistente Form von Cookies unter anderem auch benutzt, um den Inhalt von HTTP-Cookies nach deren Löschung wieder herzustellen (*respawn*). In einer weiteren Studie (Ayenson u. a. 2011) wurde gezeigt, dass sich auch andere Browserfunktionen wie *LocalStorage* und *eTags* dazu eignen, den Inhalt eines Cookies nach dessen Löschung wiederherzustellen. Nachdem sich ein Anbieter dieses Trackings vor Gericht auf einen Vergleich einließ (Singel 2012), zeigte sich in einer aktuellen Studie (G. Acar et al. 2014), dass Respawning in der beschriebenen Form kaum noch eingesetzt wird und sich auf chinesische und russische Webseiten beschränkt.

Browser- / Device-Fingerprinting

Darüber hinaus werden regelmäßig neue Formen der User-Identifizierung entwickelt, deren Funktionieren unabhängig vom Verhalten (und der Einwilligung) der Nutzer_innen ist. Weite Verbreitung hat *Browser-* oder *Device-Fingerprinting* (Boda u. a. 2012) gefunden. Dabei ermitteln JavaScripte und nicht sichtbare Flash-Animationen Eigenschaften[60] des Gerätes und des benutzten Browsers, die in vielen Fällen eine eindeutige Merkmalskombination ergeben, die den_die Nutzer_in auf verschiedenen Webseiten identifiziert. Acar et al. (2013) konnten einige Anbieter_innen identifizieren, die

60 Eine Liste von Merkmalen, die sich über einen Browser und das Betriebssystem, auf dem der Browser läuft, ermitteln lassen, führt die Seite HTTP://BROWSERSPY.DK/ (letzter Zugriff 26.09.2016).

diese Technik anwenden und so ermitteln, welche Schriftarten auf dem Rechner des_der Nutzer_in installiert sind. Zum Zeitpunkt der Studie wurde Browser-Fingerprinting auf 51 der 100.000 meistbesuchten Webseiten eingesetzt. Einige der Skripte testeten die Verfügbarkeit von 500 Schriftarten, um deren Kombination zur Identifikation der Browser/Betriebssystem-Kombination zu nutzen.

Canvas-Fingerprinting

Zuletzt konnte dieselbe Forschergruppe (Acar u. a. 2014) nachweisen, dass auch eine andere Variante des Fingerprintings, die erst 2012 (Mowery und Shacham 2012) als theoretische Möglichkeit vorgestellt wurde, in der Praxis verbreitet ist. *Canvas Fingerprinting* nutzt das Canvas-Element von HTML5, das Entwickler_innen in neueren Browsern zur Verfügung steht, um die Unterschiede im Rendering von Schriftarten zu vergleichen und so unterschiedliche Geräte und Browser zu differenzieren. Ziel der Unterscheidung einzelner Browser anhand ihrer Merkmale ist die Wiederzuweisung von Profilen unabhängig von der Existenz eines Identifikationsmerkmales, wie einem Cookie, das durch den_die Nutzer_in kontrolliert und gelöscht werden kann.

Die steigende Verbreitung von Browser-Fingerprinting führen Beobachter_innen auf eine stärkere Beschränkung der Standard-Browser-Einstellungen für 3rd-Party-Cookies sowie die Nutzung von Werbeblockern zurück. Zuletzt hat sich allerdings die *Technical Architecture Group* der Web-Standardisierungsorganisation W3C gegen die Nutzung von Browser-Fingerprinting ausgesprochen (W3C Technical Architecture Group 2015).

Cookie Syncing

Wesentliche Beschränkung beim Erstellen eines Profils ist die in allen Browsern implementierte *Same Origin Policy*. Dadurch, dass für jeden Cookie hinterlegt ist, welcher Server dessen Speicherung veranlasst hat, wird der Inhalt eines Cookies bei jedem weiteren Request nur an eben diesen Server übertragen und kann nicht von anderen abgerufen werden. Diese Technik, die den Diebstahl von Zugangsdaten innerhalb des Browsers erschwert, verhindert gleichzeitig das Tracken von Nutzer_innen über mehrere Webseiten hinweg, wenn diese nicht denselben Tracking-Dienst verwenden. Vergibt Tracker A auf einer Webseite für den User U die ID 123 und Tracker B für U auf einer anderen Seite die ID 234, lassen sich die an die jeweiligen IDs gebundenen Profile nicht zusammenführen, da Tracker A keinen Zugriff auf die von Tracker B vergebene ID hat. Nach Olejnik, Minh-Dung und Castelluccia (2013) ist es das Ziel von *Cookie Syncing* (manchmal auch *Cookie Matching*) eben jene Profilzusammenführung zu ermöglichen. Tracker A und B tauschen dabei ihre IDs untereinander aus, um U auch auf solchen Webseiten verfolgen zu können, auf der nur ein Tracking-Element von A oder B vorhanden ist. So lassen sich Profile erstellen, die einen größeren Teil des Browserverlaufs berücksichtigen können. Nach Acar et al. (2014) nutzen einige große Online-Werbevermarkter *cookie syncing* in Kombination mit anderen oben beschriebenen Techniken.

Tracking-Header

Ebenfalls zur Vermarktung von Werbezwecken wird das Tracking beim amerikanischen Mobilfunkanbieter und *Internet Service Provider* (ISP) Verizon betrieben (Mayer 2014, 2015). Dieser fügt in unverschlüsselte HTTP-Anfragen einen zusätzlichen Tracking-Header ein, um so User (bzw. genauer gesagt deren Geräte mittels ihrer IP-Adresse, die damit personalisiert sind) auch auf Seiten tracken zu können, die keine Werbung enthalten. Verizon bietet Dritten seine Daten an, um eine Verbindung zwischen der geräteeindeutigen Verzion-ID und den, möglicherweise multiplen, Tracking-IDs der Werbenetzwerke herzustellen.

Cross-Device-Tracking

Dadurch, dass Internetzugriff nicht mehr nur über PCs und Laptops, sondern vermehrt auch über Smartphones, Tablets oder Fernseher erfolgt, teilt sich auch die Menge der besuchten Webseiten auf mehrere Browser auf. Die Profile sind bisher, bei der Verwendung von Cookies, an einen Browser gebunden. Tracking-Dienste versuchen diese Lücke mit Cross-Device-Tracking (Ante 2013) zu schließen. Ziel ist es mehrere Geräte, die einer_m Nutzer_in zugeordnet werden können, zu markieren und so ein geräteunabhängiges Tracking zu ermöglichen. Solche Gerätegruppen lassen sich unter anderem dadurch identifizieren, dass sie gelegentlich im selben Netzwerk (z. B. Haus-WLAN) sind und über dieselbe externe IP-Adresse verfügen.[61] Mit den technischen Anforderungen an moderne Webbrowser steigt gleichzeitig auch die Anzahl der Möglichkeiten, sie eindeutig zu identifizieren. So war es kurzzeitig möglich, Nutzer_innen durch Ausnutzung eines Sicherheitsfeatures für verschlüsselte Verbindungen zu tracken, auch wenn diese den „privaten Modus“ aktiviert hatten, der eigentlich kein Tracking zulassen soll (Greenhalgh 2015). Tracking ist aber nicht nur über den Browser möglich. Banse, Herrmann, und Federrath (2012) konnten über die aufgezeichneten Anfragen eines DNS-Servers die Webnutzung von Nutzer_innen verfolgen. Hierbei ist für die_den DNS-Betreiber_in allerdings nur der erstmalige Aufruf einer Domain (und nicht die konkret angesehene Seite) nachvollziehbar.

3.3.1 Differenzierung von Profilen

Je nach genutzter Technik ist es den Trackern möglich, in unterschiedlichem Maße Teile des Verlaufs zu beobachten, um daraus Profile zu generieren. Analog zur Differenzierung von Pseudonymen nach Pfitzmann und Hansen (2008) wird die folgende Unterscheidung von Profilen vorschlagen.

Transaktionsprofile

Sehr einfache Profile können bereits auf der Ebene von einzelnen Transaktionen erstellt werden. Bei jedem HTTP-Verbindungsaufbau eines Clients (Browser) mit einem Webserver sendet der Client einfache Identifizierungsmerkmale mit (Version des

61 Darüber hinaus wird an weiteren Techniken gearbeitet, wie der Identifizierung von Geräten in einem Raum über AudioBeacons (vgl. HTTP://WWW.SILVERPUSH.CO/#!/AUDIO letzter Zugriff 26.09.2016).

Browser und des Betriebssystems, vgl. Text 1 Zeile 1). Darüber hinaus kann die IP-Adresse als Zuordnungsmerkmal mehrerer Anfragen zu einem Profil genutzt werden. Zusätzlich wird bei vielen Requests auch der *Referer* (vgl. Text 1 Zeile 2) mit übertragen. Damit wird die Adresse der Seite bezeichnet, über die ein_e Besucher_in auf eine Webseite gelangt ist. Der *Referer* wird aber auch mit übertragen, wenn z. B. ein Bild oder ein Tracking-Skript von einem Drittanbieter geladen wird. So kann ein Transaktionsprofil erstellt werden, das pro User und Server die Nutzungsdauer und die besuchten Webseiten enthält. Die Grenze dieser Art des Trackings ist allerdings erreicht, wenn sich eines der Merkmale ändert, zum Beispiel durch eine Änderung der IP-Adresse beim Wechsel zwischen zwei WLANs oder einem Update des Browsers.

1. user-agent:Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/45.0.2454.85 Safari/537.36

2. referer:https://incentiveswidget.appspot.com/w/2Go6i/?autofocus=0

Text 1: Ausschnitt aus einem HTTP-Request Header

Dienstprofile

Ein größerer Teil des Browserverlaufs lässt sich mit den oben erläuterten Techniken (Cookies oder Browser-Fingerprinting) beobachten und, analog zu Rollen- oder Beziehungspseudonymen, zu einem Dienstprofil zusammenfügen. Der Umfang der hierbei entstehenden Profile hängt stark davon ab, in welcher Beziehung ein_e Nutzer_in zu einem_r Dienstleister_in steht. Wer kontinuierlich auf allen Internetzugangsgeräten (Laptop, Smartphone, Tablet) im Browser in den Account eines bestimmten Dienstes eingeloggt ist (z. B. Google, Apple oder Facebook) erlaubt diesen Diensten einen größeren Teil des Browserverlaufs zu beobachten. Andere Dienstleister_innen mit weniger exklusivem Zugang zu den Geräten einer_s Nutzers_in, wie die oben besprochenen Werbenetzwerke, haben weniger privilegierten Zugriff und nutzen Techniken wie *cookie syncing,* um ihre Reichweite zu erhöhen. Diese Form des Profiling nutzt weniger die auf den Servern anfallenden Daten als Eigenschaften des Browsers zur Reidentifizierung. Durch das Einbetten von Facebooks „Like“-Button oder Googles Werbenetzwerk können auch Besuche auf Webseiten registriert werden, die Nutzer_innen nicht über den Umweg der Suchmaschine oder des Newsfeed erreicht haben.

Personenprofile

Letztendlich ist es das Interesse der Tracker, den kompletten Browserverlauf zu kennen, um ein vollständiges Personenprofil erstellen zu können. Während das *Cross-Device-Tracking* der registrierten Nutzer_innen durch Google in diese Richtung geht, sind den meisten kleineren Anbieter_innen hier Grenzen gesetzt. Nur große, monopolartige Dienstleister_innen (Facebook), Betriebssysteme (Google, Microsoft), Geräte (Apple, Microsoft), oder solche Akteure, die wie Verizon (s. o.) auf der Netzwerkebene Zugriff auf den Internetverkehr haben, sind annähernd in der Lage, Personenprofile zu erstellen.

3.4 THEORETISCHE ANSÄTZE ZUM SCHUTZ VON PRIVATHEIT GEGEN TRACKING

Mit der Ausweitung und Verbesserung der Trackingverfahren entstand das Bedürfnis, die Anonymität der Nutzer_innen im Internet zu schützen und verschiedene Aspekte des Datenschutzes zu stärken. Im Rahmen der Entwicklung von *Privacy Enhancing Technologies* wurden sowohl Verfahren entwickelt, die auf Seiten der Betreibenden datenschutzfreundliches Tracking erlauben, als auch solche, die Einflussmöglichkeiten der Nutzer_innen stärken (Werbeblocker etc.) können.

3.4.1 Privacy Enhanced Tracking

Nicht durchgesetzt haben sich in der Praxis Verfahren, die Systeme für Werbenetzwerke nach den Prinzipien der datenschutzfreundlichen Technikgestaltung (*privacy-by-design*) entworfen haben. Dazu gehören *Adnostic* (Toubiana u. a. 2010), *Privad* (Guha, Cheng, und Francis 2011) und *Repriv* (Fredrikson und Livshits 2011). Sie alle eint die Idee, Profiling auf Seite der Nutzer_innen im Browser stattfinden zu lassen. Anstatt die Analyse des Verhaltens und die Berechnung von Interessen durch die Anbieter_innen vorzunehmen, sollen Browsererweiterungen das Surfverhalten beobachten und so Profile anlegen. Beim weitestgehenden Modell Adnostic sollte dieses Profil den Computer des_r Nutzer_in nicht verlassen. Stattdessen sollen mehrere Werbeanzeigen vom Werbenetzwerk übertragen werden, von denen der Browser dann selbstständig die am besten auf das Profil passende auswählt. Privad und Repriv erlauben, dass zumindest ein Teil (Privad) des Profils oder nur mit expliziter Zustimmung (Repriv) des_der Nutzer_in an ein Werbenetzwerk oder eine_n Webseitenbetreiber_in gesendet wird, um das Angebot entsprechend anzupassen.

Kaum Anwendung in der Praxis

Obwohl insbesondere Adnostic und Privad die Businessmodelle der Werbenetzbetreibenden bereits mitgedacht haben und Vorschläge für datenschutzfreundliche Werbeauktionierung machen, konnte sich keines der Modelle durchsetzen. Die Gründe für den Misserfolg sind dabei sicherlich vielfältig. Die Systeme erfordern, zumindest zu Beginn, die Interaktion des_der Nutzer_in oder gar der Browserhersteller, um eine kritische Masse an Nutzer_innen zu erreichen. Darüber hinaus gibt und gab es allerdings auch für die Betreiber_innen kaum Gründe, diese Form des Profiling zu übernehmen. Im Gegenteil würde durch die dezentrale Speicherung der Profile eine weitere Verwertungsmöglichkeit verloren gehen. Die Aggregation von Profilen zur Bestimmung von Trends und Einflussfaktoren ist nur möglich, wenn diese zentral gesammelt werden. Eine Dezentralisierung ist daher nicht im Sinne der Tracker, da weitere, vielleicht sogar gewinnbringendere, Datenverarbeitungen und -aggregationen so nicht mehr möglich sind.

3.4.2 Nutzer_innenkontrolle und Obfuscation

Privacy Enhancing Technologies für Nutzer_innen lassen sich anhand ihrer Zwecke grob einteilen: zum einen existieren solche, deren Ziel das Verhindern von Tracking ist, zum anderen solche, die das Verschleiern der Spuren technisch unterstützen.

Blocking

Systeme der ersten Kategorie arbeiten in der Regel daran, bestimmte Tracking-Skripte zu blocken, oder sie verhindern, dass Tracking durch Cookies ausgeführt werden kann.[62] Ihr Ziel ist es, das Tracking zu unterbinden und so den_die jeweilige Nutzer_in gegenüber den Trackern nicht zählbar zu machen. Dieses Verfahren setzt voraus, dass die Tracking-Methode und deren Anbieter_innen bekannt sind. Tracking-Blocker haben zudem damit zu kämpfen, dass sie keinen Einfluss auf die Datenverarbeitung auf Seiten der Server haben. Tracking, das auf Serverseiten stattfindet, wie die von Verizon genutzte Header-Ergänzung, lässt sich kaum unterbinden, da die Nutzer_innen in der Regel keinen Einfluss auf die Infrastruktur haben.

Obfuscation

Eine andere Strategie verfolgt die Verschleierung (engl. *Obfuscation*). Hierbei soll nicht das Tracking verhindert, sondern die Profile, also die Auswertung des Trackings, unbrauchbar gemacht werden (Brunton und Nissenbaum 2011). Brunton und Nissenbaum definieren *Obfuscation* als „producing misleading, false, or ambiguous data to make data gathering less reliable and therefore less valuable“. Das Ziel ist nicht, sich aus der Menge der Getrackten auszunehmen und sich aus dem System herauszunehmen oder es zu zerstören, sondern das System derart zu verwirren, dass es seine Funktion nicht korrekt ausführen kann – eine Taktik, die in 2.4.5 als „scrambling the informatics network” beschrieben wurde. Dabei müssen die störenden Informationen aber gleichzeitig so glaubhaft sein, dass sie nicht direkt als solche zu erkennen sind. Im Rahmen des Austauschs mit anderen Systemen handelt es sich daher um *konstruktive Desinformation* (Alexander und Smith 2010). In Bezug auf Tracking müssen zur Obfuscation also zusätzliche Daten erzeugt werden, die scheinen als wären es Nutzer_inneninteraktionen und so erschweren, dass das „tatsächliche“ Profil eines_r einer Nutzer_in ermittelt werden kann.

Brunton und Nissenbaum unterscheiden zwischen vier Arten von Obfuscation:

1. T*ime-based obfuscation* ist mit dem Prinzip eines Störsenders vergleichbar. Für eine kurze Zeit werden so viele Daten erzeugt, dass die tatsächlichen Informationen vom Rauschen nicht mehr unterscheidbar sind.

2. C*ooperative obfuscation* funktioniert, in dem sich mehrere Nutzer_innen zusammenschließen und sich voneinander nicht unterscheidbar machen, so dass Einzelne

62 Siehe 3.5.4 für Beispielanwendungen dieser Kategorie.

nicht identifizierbar sind. Die Autor_innen nenne hier als Beispiele Programme zum Bonuskartentausch oder das Anonymisierungsnetzwerk TOR.

3. *Selective Obfuscation* hat nicht zum Ziel, eine_n Nutzer_in vollständig zu verbergen, sondern innerhalb eines bestehenden Systems einen Teil – wie bestimmte Nachrichten – zu verbergen.[63]

4. *Ambiguating obfuscation* hat zum Prinzip, Zweifel zu streuen und Kontingenz zu erzeugen, indem ähnlich dem Störsender-Prinzip zusätzliche Daten erzeugt werden, die aber nicht die Auswertung unmöglich machen sollen, sondern bei der Profilerstellung für Ungenauigkeiten sorgen.

Die Autor_innen weisen darauf hin, dass Obfuscation grundsätzlich auch in die andere Richtung stattfinden. Statt zur Verschleierung der Identität oder des Profils eines_r Einzelnen vor einem mächtigeren Gegenüber, kann sich dieses Gegenüber auch selbst verschleiern und so die Obfuscation erschweren. Sie nennen als Beispiel (staatliche) Bürokratien, die sich der Verantwortung für den Einzelfall entziehen, indem sie eine undurchsichtige Gesetzeslage vorschieben oder die Verantwortung innerhalb der Bürokratie selbst verschleiern, indem Zuständigkeiten nicht offengelegt werden. Wie anhand der Praxisbeispiele (3.5) ersichtlich wird, verschleiern auch Tracking-Services ihre Methoden, indem Profile gar nicht oder nur teilweise transparent gemacht sowie die Einflussmöglichkeiten der Nutzer_innen hinter komplizierten Seiten versteckt werden.

Moralische Bedenken gegen Obfuscation

Darüber hinaus kritisieren Brunton und Nissenbaum, dass Obfuscation in einer einseitigen Nutzung eine Menge an Datenmüll produziert und im Extremfall – ähnlich einer *Denial Of Service (DoS)*-Attacke – zum Ausfall eines Systems führen kann, vor dem man sich zwar unkenntlich machen, das man aber eventuell trotzdem nutzen will. Diese, im Bereich Tracking eher unrealistischen Szenarien setzen voraus, dass eine große Zahl an Einzelpersonen Obfuscation einsetzt und in der Gesamtzahl der Nutzer_innen die Mehrheit bildet. Die Praxis lässt eine solche Entwicklung allerdings als nicht realistisch erscheinen, gerade weil es sich nicht um eine strategische Attacke, sondern um eine nutzer_innenbezogene Privacy Enhancing Technology handelt. Die moralische Bewertung hängt zudem sicherlich vom Einzelfall, also dem durch Obfuscation beeinflussten System ab; genauso wie die von Brunton und Nissenbaum aufgeworfene Frage, inwiefern Obfuscation nur so lange effektiv ist, wie es von einer Min-

63 Brunton und Nissenbaum nennen hier als Beispiel FaceCloak (Wanying Luo u. a. 2009), das Facebook Nutzer_innen die Wahl gelassen hat, ob sie tatsächlich Daten preisgeben oder FaceCloak gefälschte Daten einträgt. Durch einen separaten Dienst sollten weitere FaceCloak Nutzer_innen die tatsächlichen Daten einsehen können, während Facebook selbst nur die gefälschten Daten speichert und auswertet.

derheit betrieben wird, da andernfalls mit Gegenmaßnahmen derjenigen zu rechnen ist, deren Systeme beeinträchtigt werden.[64]

Obfuscation Modell

Balsa, Troncoso und Diaz (2012) haben eine Taxonomie für die Verschleierung von Suchanfragen entworfen, die nach dem Prinzip der *ambiguating obfuscation* funktioniert.

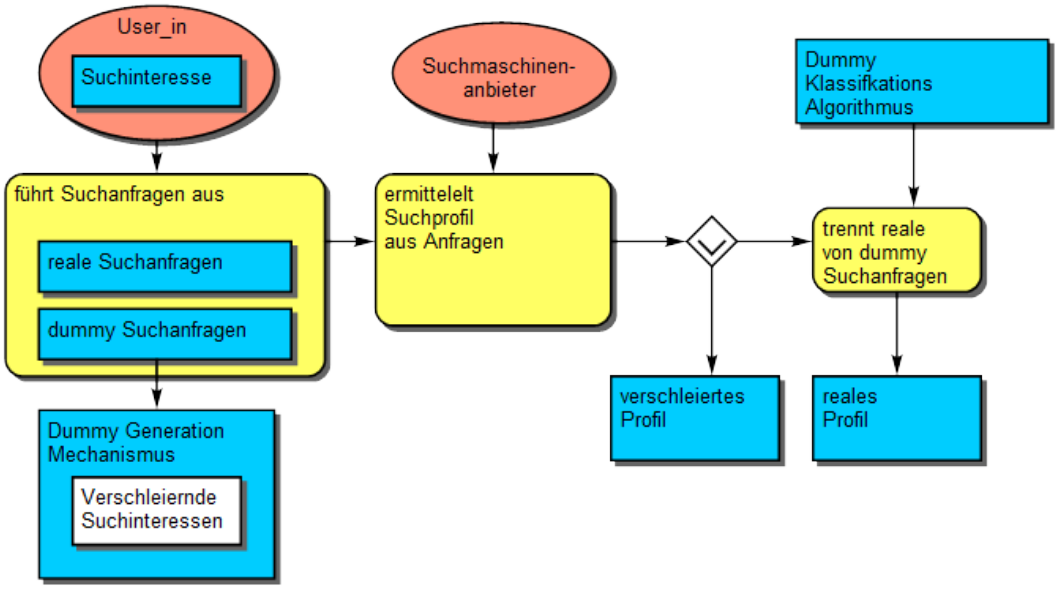


Abbildung 4: Funktionsweise von Tracking Obfuscation nach Balsa, Troncoso und Diaz (2012). Eigene Darstellung.

Abbildung 4 zeigt die Funktionsweise von Obfuscation bei Suchanfragen in einer Übersicht. Auf der Seite des_der Nutzer_in werden automatisiert Weblinks generiert (*Dummy Generator*), gegebenenfalls mit dem Wissen um das tatsächliche Profil ergänzt, die der Liste der besuchten Seiten hinzugefügt werden. Wenn der Tracking-Provider nun alle Informationen über alle besuchten Webseiten nutzt, um ein Profil zu generieren, schließt das die automatisch hinzugefügten Seiten mit ein. Das entstehende Profil basiert also nicht mehr nur auf den Seiten, die dem_r Nutzer_in direkt zugeordnet werden können, sondern zusätzlich auf den automatisch eingefügten Seiten. Hierdurch wird das ursprüngliche Profil verschleiert.

Dummy Classification Algorithm

Neben der Beschreibung notwendiger Komponenten eines Obfuscation-Systems beschreiben die Autor_innen auch Ansätze für Tracking-Provider Obfuscation auszuhebeln, am Beispiel einer Suchmaschine[65]. Dafür benötigt er einen *Dummy-Classifica-*

64 Die Nutzung von AdBlockern kann hier als Beispiel dienen. Trotz der immer noch geringen Verbreitung scheint seit einiger Zeit eine Schwelle überschritten zu sein, so dass Internetwerbeunternehmen sich genötigt sehen, Ad-Blocker-Blocker zu entwickeln (Kleinz 2015).

65 Das Beispiel ergab sich aus dem 2009 vorgestellten Obfuscation Werkzeug TrackMeNot (Howe und Nissenbaum 2009), dass Anfragen gegenüber Suchmaschinen verschleiern sollte.

tion-Algorithmus (DCA), der automatisch entscheidet, ob eine Suchanfrage oder ein Webseitenbesuch von einen Dummy-Generator oder von der_dem Nutzer_in selbst ausgelöst wurde. Auf diese Weise soll das „reale“ Profil vom Verschleierten getrennt werden können. Die vorgestellten Methoden sind:

1. Die *profile based analysis* basiert auf der Idee, dass ein Profilingservice bereits einige Profile und damit Muster und Durchschnittswerte kennt. So können unwahrscheinliche Profilelemente klassifiziert und das ursprüngliche Profil bis zu einem gewissen Grad rekonstruiert werden. Ein theoretisches Modell, um dieser Analyse beim Verschleiern von Webseitenaufrufen zu entgehen, wurde bereits vorgestellt (Dankar und El Emam 2013).

2. Bei der *query based analysis* sind die Rahmendaten jeder Anfrage dahingehend zu überprüfen und zu klassifizieren, ob die Quelle (Nutzer_inneninteraktion oder Dummy-Generator) erkennbar ist. So ließen sich z. B. für TrackMeNot (Howe und Nissenbaum 2009) die Suchanfragen aufgrund des technischen Aufbaus (der HTTP-Header-Informationen) als Dummy Traffic identifizieren. Im Fall der Obfuscation beim Webtracking wäre es zudem theoretisch möglich, die Interaktion mit der Webseite selbst zu untersuchen (Scroll- und Klickverhalten, Länge des Besuchs), um Dummy-Traffic zu identifizieren.

Messbarkeit der Verschleierung

Die Diskussion um Konzepte und Gegenkonzepte für Obfuscation impliziert die Messbarkeit der Effekte von beiden. Allerdings ist, wie sich auch im Weiteren noch zeigen wird, hierzu bisher in der Praxis kein Verfahren bekannt, da nur wenige Verschleierungsstrategien tatsächlich umgesetzt sind. Brunton und Nissenbaum stellen daher auch die Frage nach einer *science of obfuscation.*

> If there is to be a science of obfuscation it will need to identify key variables and create a systematic way of looking at the relationships between them. The set of variables will undoubtedly be hybrids of the social and the mathematical, including — goals (*i.e.*, time–based, ambiguating, selective), method (*i.e.*, whether group or individual, whether plausible data or obvious noise, whether hiding or protest), adversarial intent and resources (*i.e.*, time, opportunity cost), ratios (*i.e.*, of noise to signal), cost (*i.e.*, to obfuscator, to target), and more.

Jenseits der Modelle ist es darüber hinaus, notwendig für jede Obfuscations-Strategie Ziele und Methoden in Abhängigkeit vom Gegenüber zu bestimmen und insbesondere Metriken zu entwickeln, mit denen die Effektivität einer Strategie getestet werden kann.

3.5 ANALYSE VON PRIVACY/TRANSPARENCY ENHANCING TECHNOLOGIES

Wie bereits angedeutet, existieren einige *Privacy* und *Transparency Enhancing Technologies* im Bereich Online-Tracking und -Profiling. Im Folgenden sollen diese kurz vorgestellt und anhand einer Reihe von Kriterien untersucht werden. Das Ziel dieser Analyse ist, sowohl funktionierende Konzepte als auch Schwachstellen zu identifizieren, um daraus Anforderungen für die Entwicklung von TrickTrack abzuleiten. Untersucht wird die Umsetzung von funktionalen und nicht-funktionalen Anforderungen. Während sich die funktionalen Anforderungen aus Datenschutzschutzzielen ableiten lassen, betreffen die nicht-funktionalen einerseits die Usability, in Anlehnung an die Diskussion aus Kapitel 2 aber auch die Förderung eines kritischen Umgangs mit datenbasierten Diensten. Nach der Beschreibung der Anforderungen werden diese zum Vergleich auf verschiedene Dienstleister_innen aus dem Bereich Tracking und deren Praxis des Profilings angewendet sowie anschließend gängige Privacy Enhancing Technologies für Endnutzer_innen daran gemessen.

3.5.1 Beschreibung der Bewertungskriterien

Im vorherigen Kapitel wurde als Ergebnis festgehalten, dass die verschiedenen Regulierungen der Profiling-Praxis aktuell Mängel aufweist und es Nutzer_innen an Möglichkeiten fehlt, sich aktiv mit Profiling auseinanderzusetzen. Diese Anforderung lässt sich unter anderem mit den Datenschutzschutzzielen (Rost 2012) operationalisieren und überprüfen. Für die vorliegende Arbeit ist dabei insbesondere relevant, inwieweit entstehende Profile eingesehen (*Transparenz*) und beeinflusst (*Intervenierbarkeit*) werden können. Bei der Untersuchung von Werkzeugen, die hierzu technische Unterstützung bieten (Abschnitt 3.5.3), soll außerdem die Gebrauchstauglichkeit (*Usability*) und die Förderung einer kritischen Auseinandersetzung mit Profiling (*critical literacy*) eine Rolle spielen.

Datenschutzschutzziele

Anhand der Datenschutzschutzziele (auch Gewährleistungsziele, siehe Abb. 5) lassen sich Systeme, die personenbezogene Daten verarbeiten, auf die Einhaltung datenschutzrechtlicher Grundsätze überprüfen.[66] Die Liste der Schutzziele beinhaltet neben den klassischen Schutzzielen der Datensicherheit, die aus Datenschutzsicht interpretiert werden (Rost und Pfitzmann 2009) – Integrität, Verfügbarkeit und Vertraulichkeit –, zusätzlich die Ziele Transparenz, Intervenierbarkeit und Nichtverkettbarkeit.

66 Die Datenschutzziele sind seit kurzem Teil des Standard-Datenschutzmodells, das zur Verfahrensprüfung im Datenschutz empfohlen wird, vgl. 90. Konferenz der unabhängigen und Datenschutzbehörden des Bundes und der Länder (2015).

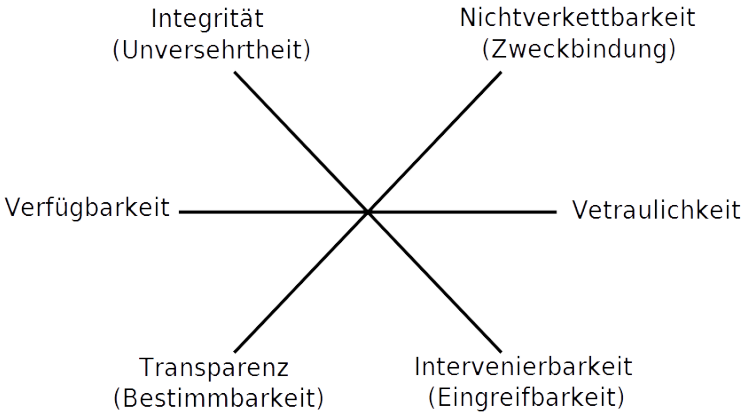


Abbildung 5: Datenschutzschutzziele; Darstellung nach Rost (2012).

Transparenz und Intervenierbarkeit

Die Schutzziele können in Form technischer oder organisatorischer Maßnahmen umgesetzt werden. Die Integrität von Daten kann so zum Beispiel durch Transportverschlüsselung realisiert werden, die vor unautorisierten Änderungen schützt. Die Vertraulichkeit von Daten kann, neben Verschlüsselung, auch durch organisatorische Maßnahmen, wie Zugangsprotokolle für Serverräume, gewährleistet werden. Verfügbarkeit bedeutet aus Sicht des Datenschutzes auch, die Verbindlichkeit einer Datenverarbeitung durch redundant arbeitende Systeme sicherzustellen. Die drei weiteren Schutzziele sind Transparenz, Intervenierbarkeit und Nichtverkettbarkeit. Sie beziehen sich stärker auf die Art und Weise der Datenverarbeitung und die Kontroll- und Einflussmöglichkeiten der Betroffenen und eignen sich daher besser für die Bewertungen von Werkzeugen für Endnutzer_innen. Für die folgende Untersuchung von besonderem Interesse sind die Ziele *Transparenz* und *Intervenierbarkeit*. Sie bilden die Schnittstelle zwischen den geschlossenen datenverarbeitenden Systemen und denen, deren Daten verarbeitet werden, um nach der Datenerhebung die im Datenschutz garantieren Rechte der Betroffenen (Bizer 2007:353 f.) abzusichern. Diese Ziele sind es auch, die solche PETs umzusetzen versuchen, die auf die geschlossenen Blackboxen der Datenverarbeitung Einfluss nehmen.

Usability-Kriterien

Neben der funktionalen Perspektive ist es notwendig zu betrachten, inwiefern das jeweilige System dazu geeignet ist, komplexe Funktionen für den_die User_in nutzbar zu machen. In der Regel geschieht dies durch die Bezugnahme auf *Usability*-Kriterien. Etabliert sind hier als Richtlinien nach ISO 9241-110: Aufgabenangemessenheit, Selbstbeschreibungsfähigkeit, Lernförderlichkeit, Steuerbarkeit, Erwartungskonformität, Individualisierbarkeit und Fehlertoleranz (Sarodnick und Brau 2011).

In einer Studie verschiedener Opt-Out- und Tracking-Blocker wurde insbesondere kritisiert, dass gängige Werkzeuge in den Punkten *Lernförderlichkeit* und *Erwartungskonformität* Mängel aufweisen, die eine Nutzung für ungeschulte Nutzer_innen erschweren (vgl. Cranor 2012; Leon u. a. 2012). Leon u. a. evaluierten mehrere Werk-

zeuge mit 45 Teilnehmer_innen und stellten fest, dass viele Tools weder dabei halfen zu verstehen, was Tracker sind, noch konnten sie vermitteln, wie genau das Werkzeug darauf Einfluss nimmt („Need for Feedback“, „Communication Problems“; ebd.). Zudem waren Standardeinstellungen teilweise so gewählt, dass erwartete Effekte (das Blocken von Werbung) nicht eintraten und die Funktionen zum Ändern dieser Einstellungen nicht ohne Weiteres erreichbar waren („Inappropriate Defaults“, „Confusing Interfaces“ ebd.).

Transparenz und ciritical literacy

Darüber hinaus stellt sich die Frage, wie Transparenz geschaffen werden kann, nicht nur im Sinne faktischer Sichtbarkeit (z. B. der erhobenen Daten, die eigentlich in einer Blackbox „liegen“), sondern auch derart, dass die Funktionsweise nachvollziehbar ist.

> Transparency of (personal) data flows contributes to privacy awareness of users. Technological means to provide transparency - soplummetedcalled transparency tools - can give information on intended collection and storage of personal data. (Pekárek und Pötzsch 2009)

Die Herausforderungen solcher *Transparency Enhancing Technologies* haben Camenish u. a. (2009) und Hildebrandt (2012, 53 f.) formuliert. Sie bilden eine Brücke zwischen Usability-Aspekten und einer Funktionalität, die eine *critical privacy literacy* (vgl. 2.6) schulen können.

- **Nachvollziehbarkeit**: Anwendungen, die ansonsten unsichtbare Aktionen wie Tracking transparent machen, stehen vor der Aufgabe, komplexe Informationen darstellen zu müssen. Zusätzlich ist es notwendig, nicht nur die aufgedeckten Mechanismen darzustellen, sondern gegebenenfalls auch die zur Aufdeckung genutzten, eigenen Funktionsweisen zu erläutern.
- **Folgenabschätzung**: TET sollte nicht nur die gegebenen Daten berücksichtigen, sondern es Nutzer_innen auch erlauben unterschiedliche Eingaben auszuprobieren, um Folgen erfahrbar zu machen.

Eine besondere Herausforderung sieht Hildebrandt beim Umgang mit Systemen, deren Datenverarbeitung selbst nicht transparent ist. Da deren Verhalten nicht vorhersehbar ist, müssen sie als Blackboxen behandelt und dabei berücksichtigt werden, dass das Herstellen von Transparenz selbst wieder als Feedback in das System zurückfließen kann.

Zusammenfassung der Kriterien

Die folgenden Kriterien sollen also für die Bewertung existierender Werkzeuge herangezogen werden.

1. Die **Funktionalität** der Werkzeuge ist interessant im Hinblick auf ihre Fähigkeiten Transparenz und Intervenierbarkeit bei Tracking und Profiling herzustellen und/oder zu unterstützen. Dabei können beide Punkte jeweils auf das zu Untersuchende (Tracking oder Profiling) sowie auf das Werkzeug selbst bezogen werden.

2. Bei der Untersuchung der **Usability** ist insbesondere auf Umsetzung von Maßnahmen zur Lernförderlichkeit und Selbstbeschreibungsfähigkeit zu achten.

3. Eine kritische Auseinandersetzung mit den zugrundeliegenden Prinzipien der Datenverarbeitung, also die Förderung einer **critical literacy,** in Bezug auf Tracking und Profiling setzt voraus, dass Nachvollziehbarkeit und Folgenabschätzung ermöglicht werden.

3.5.2 Tracking-Dienste & Transparenz

Zuerst werden Unternehmen präsentiert, die im Bereich Online-Tracking und Werbung tätig sind. Sie wurden ausgewählt, weil sie teilweise Funktionen zur Transparenz und Intervenierbarkeit umsetzen. Außerdem gehören sie zu den größten Anbieter_innen, generieren mit jeweils unterschiedlichen Mitteln Daten, sind prototypisch für viele weitere Unternehmen und werden außerdem im weiteren Verlauf der Arbeit als Datenquellen genutzt.

Google

Wie bereits frühere Studien gezeigt haben (Gomez, Pinnick und Soltani 2009) kann Google auf das umfangreichste Netzwerk an Diensten im Bereich Online-Tracking und Werbung zurückgreifen und so Nutzungsdaten aus unterschiedlichen Quellen zusammenführen. Dazu gehören:

Durch Google erhobene Daten

- **Suchanfragen**: Über Cookies können mehrere Suchanfragen, die ein_e Nutzer_in macht, einem Profil zugeordnet werden. Hieraus lassen sich Themenzusammenhänge ermitteln. Google betreibt darüber hinaus spezielle Suchmaschinen, wie für Konsumgüter oder Nachrichten, über die sich Interessen innerhalb bestimmter Kategorien in Verbindung bringen lassen.

- **Dienste**: Google bietet weitere Dienste an wie GMail, Youtube oder Google+, für die sich Nutzer_innen registrieren müssen. Durch diese Dienste können konkretere, accountbezogene Profile erstellt werden. Während in Google+ personenbezogene Daten wie Alter oder Geschlecht von den Nutzer_innen hinterlegt werden, die als Trainingsdaten dienen können, lassen sich in Ver-

bindung mit der Inhalts- wie Metadaten-Analyse des Mailverkehrs in GMail Abhängigkeiten zwischen Themen, Personen und Netzwerken herstellen.

- **Tracking:** Mit *AdWords* und *Analytics* betreibt Google zwei Dienste, die über Skripte in Seiten Dritter eingebunden werden. Das Werbenetzwerk AdWords spricht Webseitenbetreibende so doppelt an, sowohl als Kund_innen, die Werbung schalten, als auch als Werbeplatzanbieter_innen. Nutzer_innen von Google Analytics können Profilinformationen der Nutzer_innen ihrer Webseite aggregiert einsehen. Durch das Einbinden von AdWords oder Google-Analytics wird Google über alle Aufrufe einer Webseite informiert.

Google bietet Kund_innen, die Werbeanzeigen schalten wollen, zwei Verfahren an. Zum einen kann Werbung in der Suchmaschine geschaltet werden. Werbeanzeigen lassen sich so Anfragen an die Suchmaschine zuordnen. Sucht eine Person z. B. nach „Gebrauchtwagen“, können hierzu als passend eingestufte Anzeigen geschaltet werden. Zum anderen kann Werbung nicht kontextbezogen, sondern zielgruppenspezifisch gebucht werden, wobei die Anzeigen dann über das Werbenetzwerk auf andere Seiten verteilt anzeigt werden.

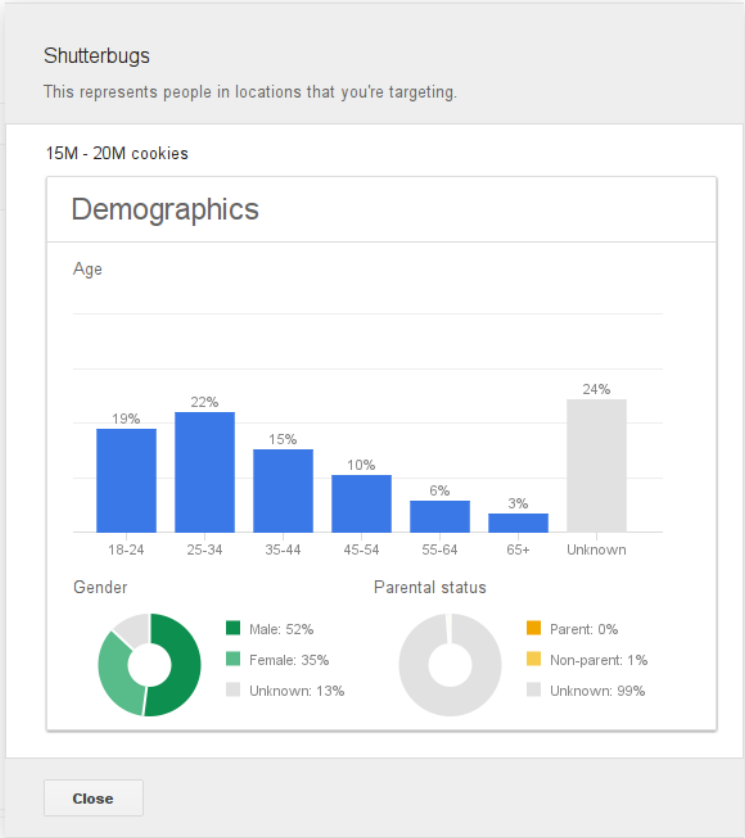


Abbildung 6: Beispiel für eine Zielgruppe für welche Werbung geschaltet werden kann. Quelle Google AdWords (letzter Zugriff 07.11.2014). „Shutterbug“ bezeichnet nach Wikipedia die Gruppe der „enthusiastischen Amateurfotografen“.

Verkettung mit Daten Dritter

Die Daten, die Google seinen Kund_innen aus dem Marketing zur Verfügung stellt, sind Alter, Geschlecht, Elternstatus[67] sowie eine Liste von Interessen, welche im späteren Verlauf dieser Arbeit noch ausführlich dargestellt werden. Diese Eigenschaften werden dann zu Zielgruppen zusammengefügt und, wie in Abbildung 6 dargestellt, Anzeigenkunden zum Zielgruppenmanagement zur Verfügung gestellt.

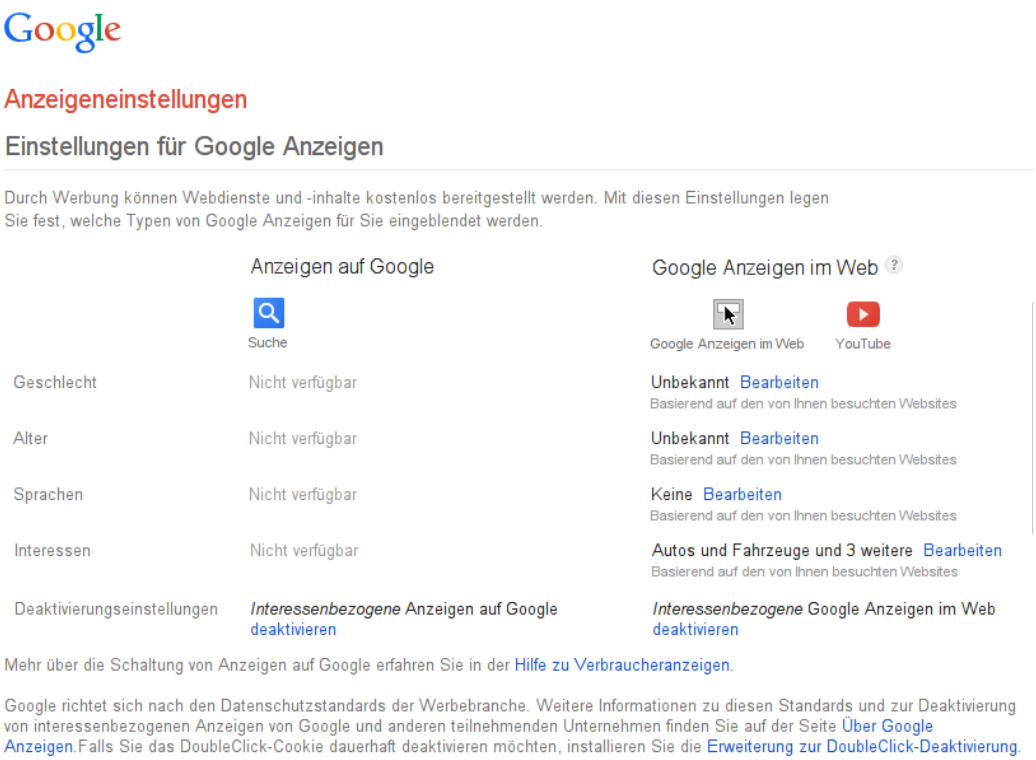


Abbildung 7: Anzeigeneinstellungsseite von Google. (Screenshot von HTTPS://WWW.GOOGLE.COM/SETTINGS/ADS vom 08.Juli 2015; die Seite wurde im Spätsommer 2015 neugestaltet)

Um diese soziodemografischen Eigenschaften von Nutzer_innen ermitteln zu können, greift Google auch auf die Daten Dritter zurück, wie es im Support-Forum heißt.

> Google kann auf Informationen zurückgreifen, die Nutzer auf diesen Partner-Websites in Bezug auf ihr Geschlecht, ihr Alter, ihre Interessen und auf sonstige demografische Merkmale angegeben haben. Außerdem lassen sich einige Informationen aus Website-Besuchen der Nutzer und aus Daten von Drittanbietern ableiten. Ein Beispiel sind Websites mit überwiegend weiblichen Nutzern. Anhand von Umfragedaten, die in der Vergangenheit von den Websitebesuchenden erfasst wurden, sind die statistischen Merkmale der Besucher bekannt. Daher kann das Cookie des Nutzers einer solchen Website dem demografischen Merkmal "weiblich" zugeordnet werden. (Google Support Webseite)[68]

67 Diese Eigenschaft lässt sich im Gegensatz zu den meisten anderen (siehe nächste Seite) in den Einstellungen, die Google Endnutzer_innen zur Verfügung stellt, nicht ändern.

Nutzer_innenkontrolle bei Google

Transparenz und **Intervenierbarkeit** versucht Google durch die Teilnahme am AdChoices-Programm zu ermöglichen (vgl. S. 95), die neben den meisten Werbeanzeigen angezeigt wird. Ein Link führt zu ausführlichen Informationsseiten über die Funktionsweise von personalisierter Werbung und der Möglichkeit zum Opt-Out. Wie am Beispiel des obigen Zitats zu sehen, versucht Google vor allem auf textueller Ebene die Funktionsweise des Profilings zu erläutern.

Googles Ad-Settings Seite

Etwas weiter geht eine personalisierte Informationsseite für Nutzer_innen, die Google bereitstellt. Unter „Anzeigeneinstellungen“ werden in mehreren Kategorien die Informationen gelistet, welche die Google-Algorithmen benutzen, um sie Werbeanzeigen zuzuweisen. Dabei unterscheidet die Seite zwischen solchen Anzeigen, die auf den Google-Suchseiten anzeigt werden, und solchen, die im restlichen Web (über Google-AdSense oder Doubleclick) und Youtube angezeigt werden. Während sich die Anzeigen in der Google-Suche auch aus Informationen speisen können, die ein_e Nutzer_in gegebenenfalls in ihrem Google-Profil preisgegeben hat (im Screenshot daher „nicht verfügbar“), basieren die Informationen für die übrigen Anzeigen auf „den von Ihnen besuchten Webseiten“ (siehe 7). Neben Geschlecht, Alter und Sprachen, die der_die Nutzer_in vermeintlich versteht, werden auch vermutete Interessen gelistet, die Google anhand der besuchten Webseiten schätzt. Die Möglichkeiten der Einflussnahme für den_die Nutzer_in bestehen auf dieser Seite darin, einerseits die Informationen zu korrigieren, beziehungsweise überhaupt erst preiszugeben, andererseits die interessenbezogene Werbung zu deaktivieren. Allerdings bestehen Zweifel an der Zuverlässigkeit der Kontrollmöglichkeiten für die Nutzer_innen über diese Seite. Datta, Tschantz, und Datta (2014) konnten zeigen, dass Werbeanzeigen von Google weiterhin personalisiert waren, obwohl die entsprechende Information aus den Anzeigeneinstellungen entfernt worden waren. Zudem hat Google die Möglichkeiten der Einflussnahme in 2015 eingeschränkt. Während es vorher auch Nutzer_innen ohne Google Account möglich war, die Informationen einzusehen, ist diese Funktion seit einer Umstellung auf registrierte Nutzer_innen beschränkt.

Darüber hinaus bietet Google nur wenige Informationen, die zur Nachvollziehbarkeit und Folgenabschätzung der Daten beitragen. Die erläuternden Texte ermutigen, weitere Informationen beizutragen bzw. die vorliegenden zu korrigieren, wobei die Vorteile einer besseren Personalisierung in den Vordergrund gestellt werden.

68 Siehe HTTPS://SUPPORT.GOOGLE.COM/ADWORDS/ANSWER/2497941?HL=DE&UTM_ID=AD Stand 07.11.2014.

Quantcast

> We know before they do. We know before you do. We can tell you not only where your customers are going, but how they're going to get there, so we can actually influence their paths.[69]

Quantcast ist, ähnlich Google, gleichzeitig Tracking- wie Werbenetzwerkanbieter und analysiert nach eigenen Angaben die Besucher_innen von 100 Millionen Webseiten. Durchschnittlich trackt Quantcast nach eigenen Angaben jede_n Amerikaner_in (der_die das Internet benutzt) im Durchschnitt 600-mal pro Monat (Quantcast o. J.). Darüber hinaus bietet Quantcast auch Informationen über Seiten, die nicht am Quantcast Netzwerk teilnehmen. Deren Besuchszahlen werden auf Basis der auf anderen Seiten erhobenen Daten approximiert, wobei es allerdings erheblich Zweifel an der Genauigkeit dieser Daten gibt (Kamerer 2013).

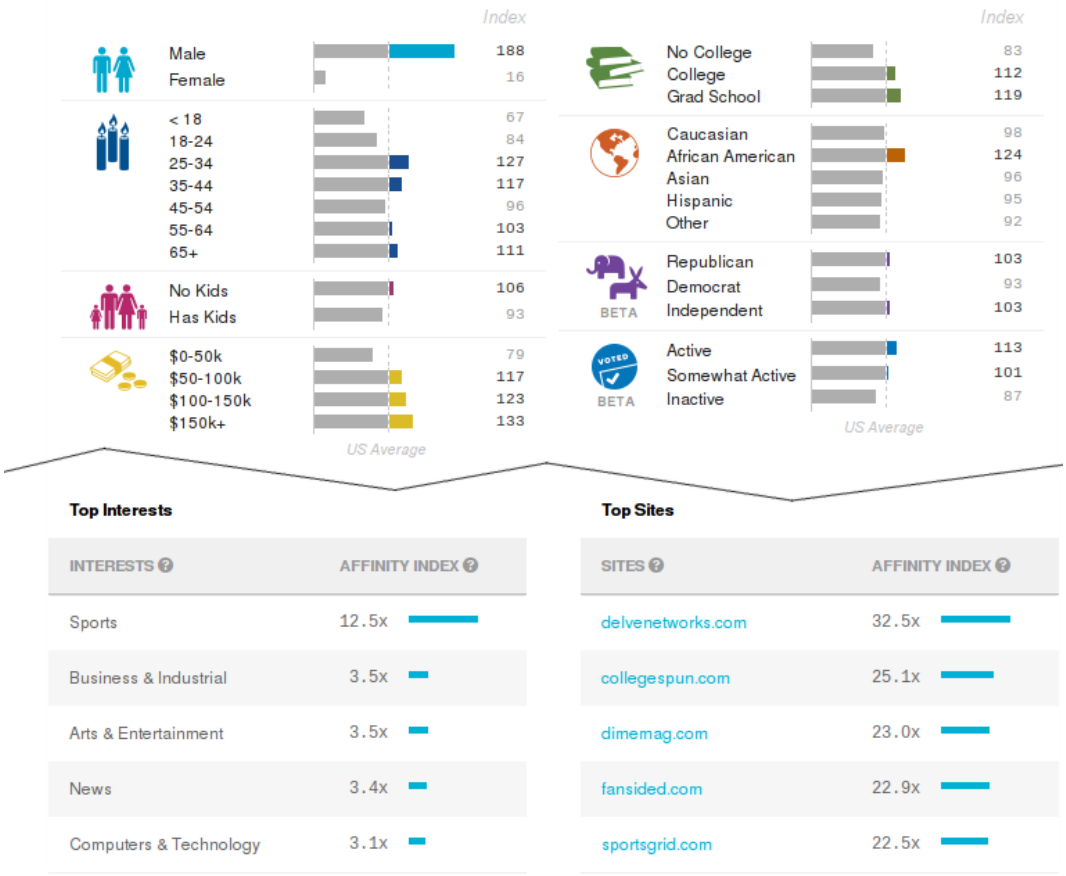


Abbildung 8: Beispiel der Informationen, die Quantcast über Besucher_innen einer Webseite bereitstellt (Stand 16.07.2015).

Von Quantcast erhobene Daten

Die Datenerhebung durch Quantcast erfolgt vor allen Dingen durch klassische Tracking-Skripte und Cookies, die die Nutzer_innen von Quantcast in ihre Seiten inte-

69 Werbeslogan Quantcast vgl. HTTPS://WWW.QUANTCAST.COM/ADVERTISE (Stadn Juli 2015).

grieren. Dabei ermittelt das Skript die Browsereigenschaften, die Sprachen sowie die besuchten Seiten und gegebenenfalls den *Referer*, über den die Seiten aufgerufen wurden. Zusätzlich verwendet Quantcast ebenfalls Daten aus Umfragen unbekannten Umfangs, in denen Nutzer_innen ihre „tatsächlichen" Attribute benennen. Daraus folgert Quantcast dann die durchschnittlichen Eigenschaften von Besucher_innen. Abbildung 8 zeigt die Daten, die Quantcast Nutzer_innen zur Verfügung stellt.

Keine Transparenz

Quantcast bietet den Betroffenen keinerlei Informationen zu den über sie konkret gesammelten Informationen an. Dabei bezieht das Unternehmen den Standpunkt keinerlei personenbezogene Informationen zu verarbeiten, sondern nur anonymisierte Daten zu erheben und diese aggregiert zu veröffentlichen.

> We do not intentionally use Personally Identifiable Information, or PII ever. In fact, our business is specifically designed to never need PII, and our Terms of Service explicitly forbid any partner to send us PII.[70]

Selbst, wenn die Basisinformationen über Umfragen anonymisiert erhoben werden, lässt sich, aus Sicht des deutschen Datenschutzes, argumentieren, dass die beim Tracking erhobenen IP-Adressen allerdings als personenbezogene Daten zu werten sind. Allerdings ist nicht bekannt, inwiefern die Daten zu Personalisierung von Werbung zu Profilen verarbeitet werden. Die Tatsache, dass Quantcast Mitglied des AdChoices Netzwerk ist, deutet aber darauf hin. Das Setzen eines Opt-Out-Cookies über AdChoices ist die einzige Interverventionsmöglichkeit in Bezug auf Quantcast.

Alexa

Von Alexa bereit gestellte Daten

Alexa ist ein Tochterunternehmen von Amazon und bietet im Rahmen sogenannter *Competitive Intelligence* Informationen zu Webseiten und der Anzahl ihrer Besucher_innen sowie deren soziodemografischen Profilen inklusive vermuteter Herkunftsländer. Darüber hinaus katalogisiert der Dienst jede Seite und listet auch die Suchbegriffe, unter denen eine Seite gefunden werden kann (vgl. Abb. 9). Anders als Quantcast und Google ist Alexa selbst kein Werbevermarktungsunternehmen, sondern finanziert sich durch ein Abo-Modell, das zahlenden Kund_innen Zugriff auf zusätzliche Informationen gewährt. Die Daten von Alexa werden häufig zur Suchmaschinenoptimierung und Marktbeobachtung verwendet, indem sie im Vergleich zu Konkurrenzunternehmen ausgewertet werden. Nach Kamerer (2013) und Eigenauskunft des Unternehmens[71] beruhen die Daten in großen Teilen auf den Analysen von Nutzer_innen, die ihre Internetnutzung durch die Alexa-Toolbar beobachten lassen (siehe unten), ein geringerer Teil entfällt auf das *direct measurement* durch TrackingPixel, die

70 Vgl. HTTPS://WWW.QUANTCAST.COM/COMPANY/OPT-OUT (letzter Zugriff 26.09.2016).

71 Siehe auch HTTPS://SUPPORT.ALEXA.COM/HC/EN-US/ARTICLES/200449744-HOW-ARE-ALEXA-S-TRAFFIC-RANKINGS-DETERMINED- (letzter Zugriff am 26.09.2016).

Webseitenbetreiber_innen einbetten können. Alexa veröffentlicht fortlaufend Listen, die die 1.000.000 meistbesuchten Webseiten der Welt aufzählen. Diese Toplisten werden in vielen Studien als Baseline genutzt, um zum Beispiel die Verbreitung von bestimmten Tracking-Verfahren zu ermitteln.

Alexa-Toolbar

Alexa erhebt bei den Nutzer_innen ihrer Toolbar die soziodemografischen Daten, auf deren Aggregation die seitenspezifischen Profile beruhen.[72] Diese werden nach der Installation aufgefordert, freiwillig Angaben über ihr Gender, Alter und Einkommen zu machen. Eine Verifikation dieser Daten erfolgt durch Alexa nicht. Im Anschluss beobachtet die Toolbar die Webseitenaufrufe der Nutzer_innen. Die Daten der einzelnen Nutzer_innengruppen können so, für die von den User_innen besuchten Seiten, aggregiert werden. Im Gegenzug zeigt die Toolbar Informationen über die aktuell besuchte Webseite wie die Platzierung auf der Topliste und Links zu ähnlichen Seiten. Die Nutzer_innen der Toolbar können per Opt-Out der Datenerhebung für Alexa widersprechen.

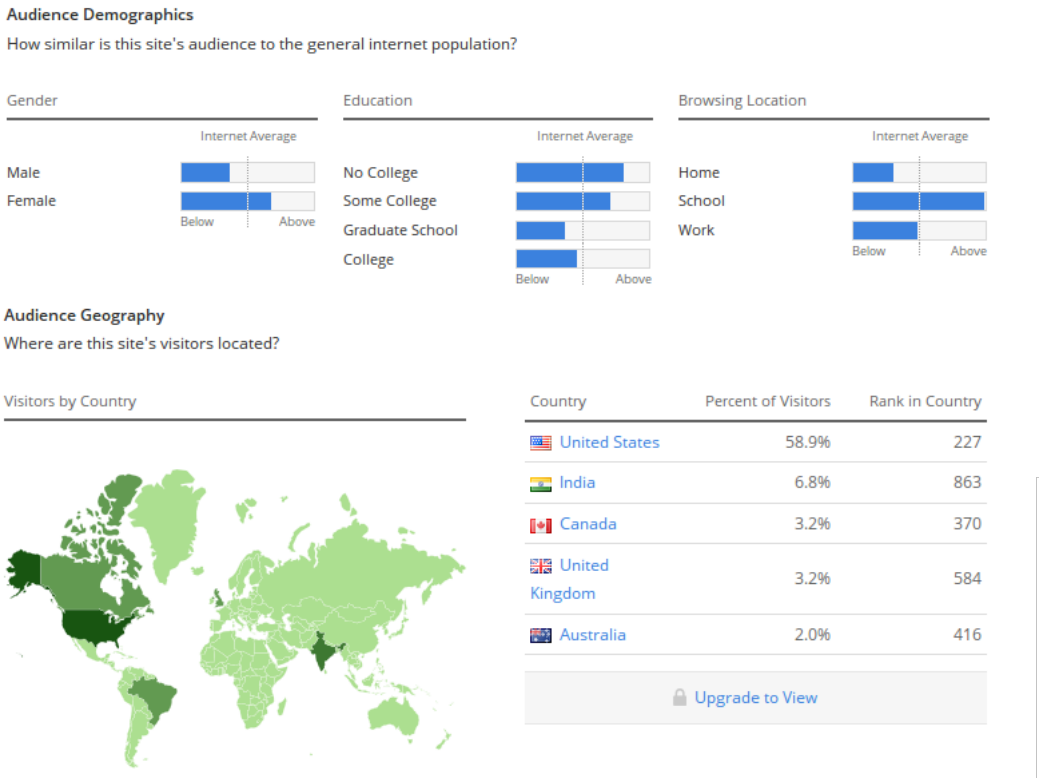


Abbildung 9: Screenshot der Alexa-Informationsseite zu wired.com (Stand 14.07.2015).

72 Siehe HTTP://WWW.ALEXA.COM/TOOLBAR (letzter Zugriff 06.10.2015).

Alexa Trackingpixel

Für das Opt-Out der Datenerhebung des Direct Measurements durch Tracking-Pixel, in dessen Nutzung User_innen nicht einwilligen, empfiehlt Alexa das Sperren von Hosts über die Konfiguration des Rechners, an dem man surft.[73] Auch wenn ausführliche Anleitungen für einige Betriebssysteme bereitgestellt werden, ist dies wohl die aufwendigste Form des Opt-Out, insbesondere im Vergleich zu dem im Folgenden beschriebenen Cookie-basierten Opt-out Verfahren. In der *Privacy Policy* macht Alexa keine Angaben zu den möglichen Folgen einer Datensammlung. Auch ist es nicht das Ziel der Toolbar, Transparenz bei den Nutzer_innen über die Datenerhebung zu schaffen, sondern sekundäre Informationen bereitzustellen.

	Funktionalität		**Usability**		**Critical Literacy**	
	Transparenz	Intervenierbarkeit	Erwartungs-konformität	Lernförder-lichkeit	Nachvoll-ziehbarkeit	Folgenab-schätzung
Google	erstellte Profile nur für registrierte Nutzer_innen einsehbar; genaue Ursache für ein Interesse nicht transparent	Profile können geändert, Tracking deaktiviert werden	die Seite ist übersichtlich gestaltet	nur im Sinne der Verbesserung der Genauigkeit des Profils	unklar, was genau zu welchem Interesse führt	positive Folgen werden hervorgehoben
Alexa	bei der Installation wird in die Datensammlung eingewilligt; keine weiteren Möglichkeiten der Einflussnahme.	keine Möglichkeit zur Intervention in die Profile keine, Tracking kann nur aufwendig deaktiviert werden	nicht anwendbar, da keine Einsicht in Profile	nicht vorhanden	nicht vorhanden	nicht vorhanden
Quantcast	keine_r der Anbieter_innen gibt an, keine personenbezogenen Daten zu verarbeiten	keine	-	-	-	-

Tabelle 1: Vergleich der Tracking-Anbieter_innen in Bezug auf die Anforderungen.

Tabelle 1 fasst die Eigenschaften der Informationsseiten der vorgestellten Anbieter_innen zusammen. Von den großen Tracking-Diensten bietet nur Google eine – eingeschränkte – Transparenz und Kontrollmöglichkeit. Alexa und Quantcast argumentieren, dass sie nur seitenbezogene und anonymisierte Daten erheben und keine personenbezogenen Daten verarbeiten. Darüber hinaus sind die konkreten Verfahren, die bei der Zusammenführung der Daten genutzt werden, nicht öffentlich, wodurch den Vorwürfen der Ungenauigkeit (Kamerer 2013) nicht nachgegangen werden kann.

73 Siehe HTTPS://SUPPORT.ALEXA.COM/HC/EN-US/ARTICLES/200685410-OPTING-OUT-OF-ALEXA-DIRECT-MEASUREMENT (letzter Zugriff 06.10.2015).

Neben Google stellt aktuell nur ein weiterer Anbieter, Bluekai[74], eine Informationsseite über angelegte Profile bereit.

3.5.3 Anbieterseitige Tools zur Nutzer_innen-kontrolle

Viele Tracking- und Profiling-Anbieter_innen stellen keine eigenen Möglichkeiten für Nutzer_innen bereit, mit dem Dienst zu interagieren, und bieten ihren Service nur der Publisher- und Marketingseite an. Gesetzlichen Verpflichtungen, z. B. Einsicht zu ermöglichen, kommen sie nicht nach. Der von der Industrie seblst ausgearbeiteten Selbstverpflichtung wird stattdessen nur in Form unübersichtlicher Opt-Out Portale nachgekommen.

AdChoices/Opt-Out

Opt-Out bei Anbietern

Als Möglichkeit des Widerspruchs gegen Online Tracking hat sich *Opt-Out* etabliert. Die meisten Anbieter_innen von Online-Werbung bieten Nutzer_innen die Möglichkeit, durch Opt-Out vom Tracking ausgenommen zu werden. Diese müssen dazu allerdings selbst aktiv werden und ein Opt-Out-Cookie setzen, dass dem jeweiligen Tracker signalisiert, dass ein Tracking nicht erwünscht ist. Ein großer Teil dieser Unternehmen hat sich in zwei Netzwerken (YourOnlineChoices[75] und AdChoices[76], vgl. Abb. 10) zusammengeschlossen, um das Prozedere zu vereinfachen.

Abbildung 10: Logo der Kampagne „AdChoices“. Quelle: http://www.youradchoices.com/ (letzter Zugriff 05.02.2015).

Nachteil des Opt-Out

Nachteil dieses Cookie-basierten Opt-Out ist, dass es an den Browser und das Vorhandensein eines Cookies gebunden ist. Löscht ein_e Nutzer_in ihre_seine Cookies, zum Beispiel, um andere (Tracking) Cookies zu entfernen, löscht sie_er auch gleichzeitig die Opt-Out-Cookies, die weiteres Tracking verhindern sollen. Acar u. a. (2014) wiesen zudem darauf hin, dass Opt-Out Cookies nur einen geringen Effekt auf verschiedene Tracking-Verfahren haben. Darüber hinaus bietet AdChoices keine Kontrolle des Trackings, sondern nur der Präsentation von verhaltensbasierter Werbung.

---

74 Vgl. HTTP://BLUEKAI.COM/REGISTRY/ (letzter Zugriff 26.09.2016). Bluekai gehört zum IT-Unternehmen Oracle, dass durch Zukäufe von kleineren Unternehmen am Aufbau einer „Marketing Cloud“ arbeitet.

75 Siehe HTTP://WWW.YOURONLINECHOICES.COM/UK/YOUR-AD-CHOICES (letzter Zugriff 26.09.2016).

76 HTTP://WWW.ABOUTADS.INFO/CHOICES/ (letzter Zugriff 05.02.2015). Das AdChoices Icon wird dabei interessanterweise von der Betreiberfirma von Ghostery (siehe S. 99) zur Verfügung gestellt.

> After you opt out, participating companies and the Web sites you visit may continue to collect and use information for purposes other than online behavioral advertising.[77]

Die Tracking-Anbieter_innen behalten sich also vor, das Surfverhalten der Nutzer_innen weiterhin zu verfolgen und nur die Auswirkungen, in Form angepasster Werbung, zu mindern.

Wenig überraschend ist auch, dass die Funktionen zum Opt-Out auf der Seite der Kampagne selbst nicht besonders benutzer_innenfreundlich gestaltet sind. Während die Startseite von AdChoices einige einfach gehaltene Videos bereitstellt, wie personalisierte Werbung funktioniert, führt der Link zum Opt-Out zu einer optisch weniger ansprechende Seite. Dort muss sich der_die Nutzer_in einige Sekunden gedulden, bis ermittelt wurde, welche_r Anbieter_in bereits Daten über den_die Nutzer_in bereithält. Danach ist es möglich, für alle Anbieter_innen gleichzeitig einen Opt-Out Cookie zu setzen.[78]

Do Not Track

Anti-Tracking Header

Einer der wenigen Ansätze der (Selbst-)Regulierung von Online-Tracking auf übergeordneter Ebene ist der *DoNotTrack* (DNT) Standard. DoNotTrack geht auf eine Initiative der amerikanischen *Federal Trade Commission* (FTC) zurück, deren Umsetzung in der *Tracking Protection Working Group*[79] des *WorldWideWeb Consortiums* (W3C) diskutiert wurde. Das W3C ist für die Standardisierung von Webtechnologien wie HTML und JavaScript zuständig. In einer Arbeitsgruppe wird seit 2011 der DNT-Standard entwickelt. Ergebnis ist die Standardisierung eines HTTP-Headers, der relativ zügig in den gängigen Browsern implementiert wurde. User_innen können ihren Browser konfigurieren, einen binär kodierten DNT-Header mitzusenden. Wert 1 teilte den Empfänger_innen des HTTP-Requests mit, dass der_die Sender_in nicht getrackt werden möchte, Wert 0 gestattet ein Tracking. Die Idee ist, dass Tracker diesen Header auslesen und dementsprechend Daten speichern sollten oder eben nicht. Dienstspezifische Opt-Out-Cookies wären damit unnötig. Das DNT-Verfahren wurde bereits zu Beginn kritisiert, unter anderem aus Gründen der Usability (Leon u. a. 2012) und des mangelnden Feedbacks über das Funktionieren bzw. die Akzeptanz durch die Anbieter_innen. Zudem decken sich die Erwartungen der Nutzer_innen nicht mit den Funktionen, die DNT tatsächlich bietet McDonald und Peha (2011).

77 Vgl. HTTP://WWW.ABOUTADS.INFO/HOW-INTEREST-BASED-ADS-WORK (letzter Zugriff 26.09.2016).

78 Die Zeitdauer ist deswegen bemerkenswert, weil dieselben Betreiber_innen beim Real-Time-Bidding innerhalb weniger Millisekunden ein Profil analysieren, eine Auktion starten, abschließen und die Anzeige bereit stellen können.

79 Siehe HTTP://WWW.W3.ORG/2011/TRACKING-PROTECTION/ (zuletzt abgerufen am 26.09.2016).

Trotz breiter Diskussionen im Vorfeld der Standardisierung und Einbindung der Industrie galt DNT bereits vor der abschließenden Verabschiedung als gescheitert (Campbell 2014). Zuletzt wiesen Acar u. a. (2014) nach, dass nur wenige Tracker die DNT Einstellungen berücksichtigen. Die Gründe für das Scheitern sind allerdings nicht nur auf der Seite der Tracking-Dienste zu suchen. Einerseits sorgt sicherlich der starke Konkurrenzdruck unter den Tracking-Diensten, die ihre Geschäftsmodelle durch DNT gefährdet sehen, für eine geringe Akzeptanz und Umsetzung der Selbstverpflichtung. Andererseits wurden die Befürchtungen durch die Ankündigung von Microsoft[80], den DNT Header standardmäßig zu aktivieren, bestätigt. Ein solches Vorgehen entspräche einem Opt-In in Online-Tracking – Nutzer_innen müssten sich aktiv dafür entscheiden, sich tracken zu lassen. Tracking-Dienste fürchten (vermutlich zu Recht), dass nur wenige Nutzer_innen DNT deaktivieren würden.

	Funktionalität		**Usability**		**Privacy Literacy**	
	Transparenz	Intervenierbarkeit	Erwartungskonformität	Lernförderlichkeit	Nachvollziehbarkeit	Folgenabschätzung
AdChoices	jede Werbung ist mit dem Icon markiert; Profile nicht einsehbar	Opt-Out-Cookies	die Seite ist ungewöhnlich träge und unübersichtlich	Funktionen der Seite werden ausführlich erläutert	die Seite verweist auf die Datenschutzbestimmungen der einzelnen Partner	ausschließlich positive Effekte der Personalisierung werden besprochen
Do Not Track	bei regulärer Benutzung des Browsers ist nicht klar, ob DNT aktiviert ist	im Browser einstellbar	browserabhängig; in den gängigen Browsern ist die Einstellung erst nach einigen Klicks zu erreichen	browserabhängig; die gängigen Browser verweisen auf weiterführende Hilfeseiten	weder ist ein Profil einsehbar, noch können Nutzer_innen mit Gewissheit sagen, ob eins erstellt wird	browserabhängig; in den gängigen Browsern nicht in Bezug auf Profiling

Tabelle 2: Vergleich der Anforderungen in Bezug auf Formen der Selbstregulierung.

Die wenigen Möglichkeiten, die von den Anbieter_innen selbst bereitgestellt werden, beschränken sich auf die Intervenierbarkeit durch Opt-Out. Sowohl AdChoices als auch DoNotTrack stellen diese Funktion allerdings nur unzureichend erklärt bereit und geben keinerlei Hinweise, die eine Nachvollziehbarkeit oder Folgenabschätzung ermöglichen. Im Gegenteil suggeriert das AdChoices Programm, dass man Tracking widersprechen kann, während tatsächlich nur die Folge, nämlich die Personalisierung von Werbeanzeigen, deaktiviert wird. Tabelle 2 fasst die verschiedenen Formen der Selbstregulierung in Bezug auf die Anforderungen zusammen.

80 Siehe HTTP://HEISE.DE/-1588863 (letzter Zugriff 26.09.2016).

3.5.4 Blocking

Neben diesen Opt-Out-Möglichkeiten sind Blocking-Tools weit verbreitet, die einen ähnlichen Effekt haben, aber nicht auf die Kooperation der Anbieter_innen angewiesen sind. Dabei handelt es sich um clientseitige Software, die meist als Browser-Erweiterung installiert wird und verhindert, dass der Browser Kontakt mit einem der Tracking-Dienste aufnimmt. Sie arbeiten in der Regel mit einer Blacklist von URLs, also einer manuell gepflegten Liste von Webseiten, die Werbe-/Tracking-Diensten zugeordnet werden. Dabei ist Blocking eine durchaus erfolgreiche Methode, um 3rd-Party-Requests zu unterbinden. Nach J. R. Mayer und Mitchell (2012) blocken gängige Plugins bis zu 80 % des Traffic an potentielle Tracking-Services. Allerdings tragen unter anderem erheblich Usability-Probleme (Cranor 2012; Leon u. a. 2012) dazu bei, dass in Europa länderabhängig nur zwischen 8 und 38 % der Nutzer_innen Blocking-Plugins benutzen, wobei technikaffine Nutzer_innen die Mehrheit bilden (Pagefair 2015).

AdBlockPlus/Edge

AdBlock Plus (ABP) ist mit ca. 50 Millionen Nutzer_innen der weltweit am weitesten verbreitete Werbe-Blocker[81]. ABP funktioniert als klassischer Werbeblocker, der das Anzeigen von Werbung unterbindet. Dazu gleicht das AddOn die Skripte und Bilder, die in eine Webseite eingebunden werden, mit einer Liste bekannter Werbenetzwerke und -anbieter_innen ab und blockiert den Request im Fall eines Treffers. ABP ist nach der Installation ohne weitere Eingaben betriebsbereit, wodurch sich das Tool positiv von vergleichbaren Lösungen abhebt (Leon et al. 2012). Nutzer_innen können zudem eigene Regeln definieren, nach denen Werbung blockiert werden soll. Diese Funktion ist allerdings eher rudimentär umgesetzt und erschloss sich in der genannten Studie den Teilnehmer_innen nicht.

AdBlockEdge

In den Standardeinstellungen blockiert ABP allerdings nur sichtbare Werbeanzeigen und nur einige, nicht sichtbare Tracking-Skripte. Zusätzliche Blacklisten, etwa gegen Tracker, müssen separat aktiviert werden. Stark in die Kritik geraten ist die Herstellerfirma von ABP Eyeo für ihr Geschäftsmodell[82], bei dem sich Werbetreibende gegen eine Gebühr in die Liste von *acceptable ads* aufnehmen lassen können. Die Werbung dieser Anbieter_innen wird anschließend in den Standardeinstellungen von ABP nicht

81 Siehe HTTPS://ADBLOCKPLUS.ORG/ (letzter Zugriff 26.9.2016). Die Nutzungszahlen basieren auf Aussagen des Unternehmens. Nachvollziehbar ist die Installationsstatistik für Firefox nach der etwa 21 Millionen Firefox Nutzer_innen AdBlock Plus installiert haben (vgl. HTTPS://ADDONS.MOZILLA.ORG/DE/FIREFOX/ADDON/ADBLOCK-PLUS/STATISTICS/?LAST=365 letzter Zugriff 26.09.2016).

82 Vgl. heise online, „Schwere Vorwürfe gegen Werbeblocker AdBlock Plus“ vom 26.06.2013; HTTP://HEISE.DE/-1897152 zuletzt abgerufen am 26.09.2016.

mehr ausgeblendet. Zusätzliche Einstellungen sind nötig, um alle Werbeanzeigen zu blockieren. Von dieser Möglichkeit haben bereits einige große Werbenetzwerke wie die von Google, Amazon und Microsoft[83] Gebrauch gemacht. AdBlock Edge[84] ist eine Variante von AdBlock Plus bei der dieselbe Codebasis verwendet wird, aber auch *acceptable ads* nicht angezeigt werden.

Abbildung 11: Logo (oben) und Informationsoverlay (unten) von hostery.

Über die Möglichkeiten des AdBlocking hinaus bieten beide Varianten den Nutzer_innen kaum Optionen. Die Standardeinstellungen sind so gewählt, dass keine weitere Interaktion notwendig ist, gleichzeitig aber auch keine Informationen über blockierte Inhalte angezeigt werden. Auch über die Funktionsweise von Online-Werbung und Tracking sowie die Grenzen der Tools werden Nutzer_innen nicht aufgeklärt.

Ghostery

Ghostery hat nach eigenen Angaben 20 Millionen Nutzer_innen und blockierte zuletzt[85] 2126 Tracker und Werbeanbieter_innen über eine Blacklist. Einer der Schwerpunkte des Plugins ist dabei die Steigerung der Transparenz durch die Visualisierung der blockierten Tracker auf jeder aufgerufenen Webseite. Darüber hinaus können einzelne Tracker und Dienste manuell (de)aktiviert werden. Leon u. a. (2012) beschreiben allerdings in ihrer vergleichenden UsabilityStudie, dass die Standardeinstellungen von Ghostery die Nutzung erschweren. Nach der Installationen ist eine zusätzliche Aktivierung des Blocking-Mechanimus notwendig, so dass einige Testnutzer_innen fälschlicherweise annahmen, den Tracking-Schutz bereits aktiviert zu haben, obwohl dem nicht so war.

Kritik an Ghostery

Auch bei Ghostery gibt es Kritik am Geschäftsmodell, dass auf der freiwilligen Teilnahme (Opt-In) von Ghostery Nutzer_innen am „Ghost Rank“ beruht. Hierbei zählt

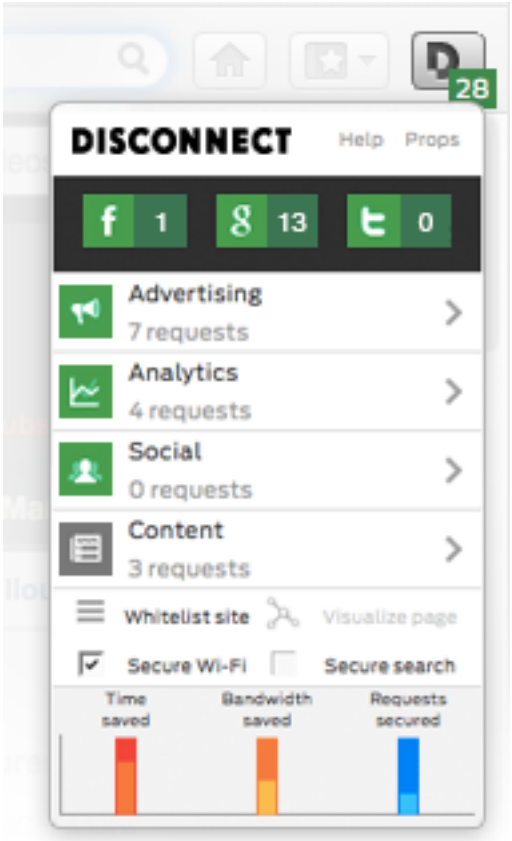
83 Vgl. Financial Times, „Google, Microsoft and Amazon pay to get around ad blocking tool“ vom 01.02.2015; HTTP://WWW.FT.COM/CMS/S/0/80A8CE54-A61D-11E4-9BD3-00144FEAB7DE.HTML?SITEEDITION=INTL#AXZZ3QBJ7OEJB zuletzt abgerufen am 04.02.2015.

84 Siehe HTTPS://ADDONS.MOZILLA.ORG/EN-US/FIREFOX/ADDON/ADBLOCK-EDGE/ zuletzt abgerufen am 26.09.2016.

85 Stand 16.11.2015.

Ghostery die geblockten Tracking-Scripte, um diese als Statistik über die Verbreitung wieder an die Tracking-Services verkaufen (Bilton 2012).

Durch die Visualisierung der gefunden Tracker (vgl. Abbildung 11, unten) sorgt Ghostery für Awareness über den Umfang des Tracking. Obwohl die Nutzer_innen angehalten werden, am *Ghost Rank* teilzunehmen, existiert keine Möglichkeit eben diese Daten selbst zu nutzen und so den Umfang von Tracking einschätzbar zu machen. Zudem sind Zusatzinformationen zu einzelnen Trackern, die Ghostery durchaus anbietet[86], nur schwer zu erreichen.



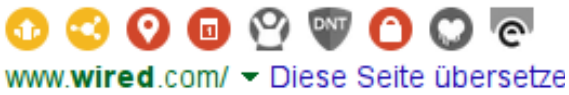


Abbildung 12: Disconnect.me Overlay im Browser (oben), Privacy Icons (unten).

Disconnect.me

Disconnect.me ist ebenfalls ein Browser-Plugin mit - nach eigenen Angaben – einer Million Nutzer_innen. Das auf einer Blacklist basierende Plugin blockiert Anfragen an ca. 2000 Tracking Seiten. Ein zweites Plugin desselben Herstellers ermöglicht es außerdem, Zusatzinformationen zu Webseiten in Form von *Privacy Icons* anzuzeigen. Dabei repräsentiert jedes Symbol einen Themenbereich wie den Umgang des_r Betreiber_in mit den Daten, mögliche Datenweitergaben, die Speicherdauer von Daten und einige mehr. Durch Einfärbung der Icons nach Ampelfarben (grau signalisiert, das keine Informationen vorliegen) soll die Einschätzung der Datenschutzbedingungen vereinfacht werden (siehe Abb. 12, unten).

Differenzierung von Tracking

Disconnect teilt Tracking-Dienste in verschiedene Kategorien ein und listet sie entweder als Werbeanbieter_innen (*Advertising*), Analysedienst oder Social Plugins (Abb. 12, oben). Auf diese Art werden den Nutzer_innen Arten und Umfang von Tracking dargestellt. Zusätzlich wird im unteren Bereich der Einblendung berechnet, wie viel Bandbreite durch das Blockieren von Skripten eingespart wurde. Die Privacy Icons wiederum geben Auskunft über verschiedene Arten der Datenerhebung, allerdings nicht über den Zweck der Verwendung.

86 Siehe HTTPS://APPS.GHOSTERY.COM/EN/APPS/ (letzter Zugriff 08.10.2015).

Die Kategorien der *Privacy Icons* sind ohne Zusatzwissen nur schwer einzuschätzen und beziehen sich vor allen Dingen auf solche Informationen, die automatisiert erhoben werden. Ob eine Seite über den Browser die genaue Position des Computers erfragt, wird auf derselben Ebene dargestellt wie die Information über die Speicherdauer der Daten beim jeweiligen Dienst. Daneben wiederum ist die Information gelistet ist, ob die Seite von einem externen Service ein Zertifikat erhalten hat. Die Einschätzung einer Webseite auf Basis der Privacy Icons gestaltet sich daher selbst für Expert_innen schwierig. Ähnlich verhält es sich mit den berechneten Einsparungen durch das Blocking. Hierbei handelt es sich um eine rein technische Information, die keinen Aufschluss über den Umfang eines Tracking-Netzwerks oder der erstellten Profile gibt, sondern nur auf der Ebene einzelner Seiten funktioniert.

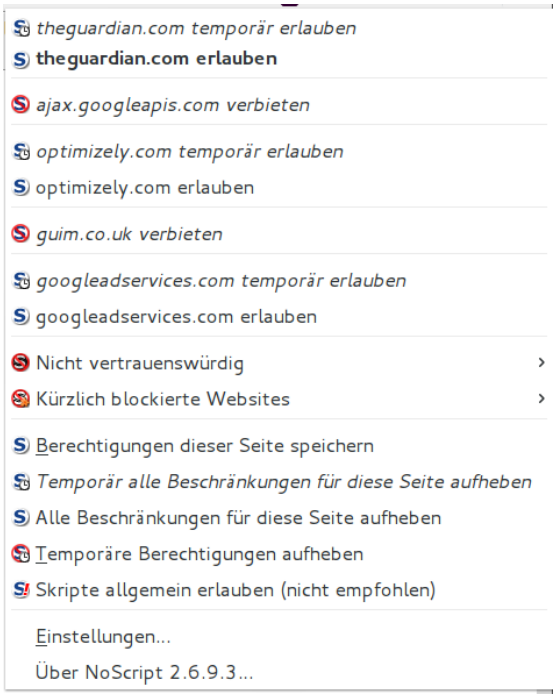


Abbildung 13: Optionsmenü zum Blockieren und Erlauben von Skripten auf theguardian.-com.

NoScript

NoScript ist ein Skriptblocker, der das Ausführen von jeglichem JavaScript unterbinden kann. NoScript unterscheidet dabei nicht zwischen Herkunft der Skripte (ob First oder Third-Party) und bringt nur eine kurze, vorkonfigurierten Whitelist[87] mit. Alle anderen Skripte werden blockiert, und der_die User_in muss einzeln oder pro Seite entscheiden, welche Skripte ausgeführt werden (vgl. Abb. 13). Da viele Tracking-Verfahren (z. B. Browser/Canvas, Fingerprinting) JavaScript voraussetzen, kann das Blocken Skripten von *3rd-Parties* diese Mechanismen zumindest erschweren und insbesondere Canvas- und Browser-Fingerprinting (vgl. 3.3) verhindern. Cookie-Tracking wird allerdings nicht unterbunden. Zudem trägt die Verbreitung von *Content Delivery Networks* (CDNs) wie googleapis.com, die Skript-Bibliotheken für viele Webseiten bereitstellen, dazu bei, dass die Funktionalität von Webseiten eingeschränkt wird, wenn das Nachladen dieser Skripte verhindert wird.

87 Die Liste enthält die Adressen zu 11 Anbietern, die der Autor des Plugins für vertrauenswürdig hält. Vgl. HTTPS://NOSCRIPT.NET/FAQ#QA1_5 (letzter Zugriff 26.09.2016).

Nur für Expert_innen

NoScript offenbart durch seine Funktionsweise die Komplexität moderner Webseiten. Auch wenn ungeübten Nutzer_innen nicht klar ist, welches Skript genau welche Funktion erfüllt, erlaubt die Nutzung von NoScript doch zumindest das spielerische ausprobieren. Durch Aus- und Einschalten bestimmter Skripte lässt sich so auf Dauer in Erfahrung bringen, zu welchem Zweck sie verwendet werden. NoScript gibt allerdings keine Auskunft darüber, welche die Daten die Skripte erheben oder übermitteln.

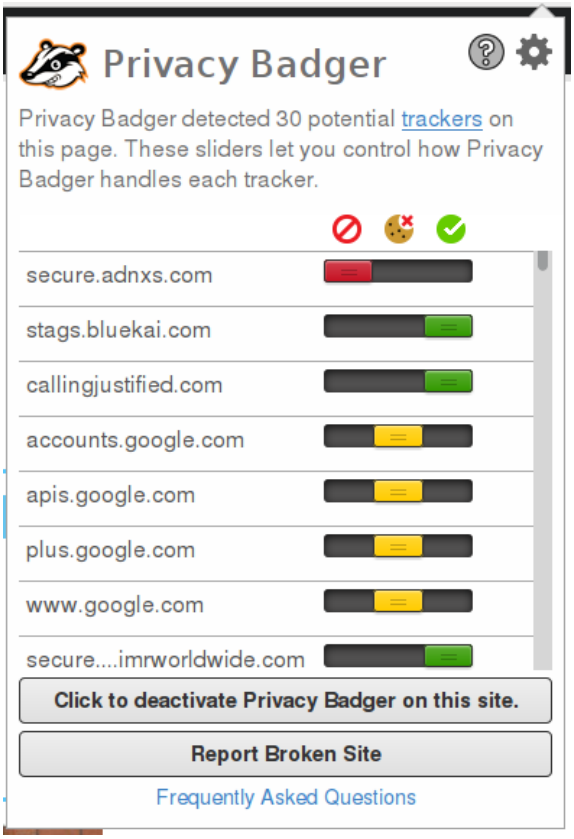


Abbildung 14: Hinweisfenster von Privacy Badger. Rot = blockieren aller Anfragen; gelb = Übertragung von Cookies deaktiviert.

Privacy Badger

Privacy Badger[88] wurde erst 2015 vorgestellt und ist einer der wenigen Tracking- und Ad-Blocker, der nicht auf Listen zu blockierender Domains basiert. Stattdessen analysiert das Browser-AddOn die Requests aller Seiten während einer Session. Werden Aufrufe an eine Dritt-Domain auf mehreren Seiten registriert wird diese Domain als verdächtig eingestuft und entweder geblockt oder die Übertragung von Cookies unterbunden (vgl. 14). Auf diese Weise können Probleme umgangen werden, die entstehen, wenn nützliche JavaScripte geblockt werden, obwohl sie zur Funktionalität der Seite beitragen.

Ausführliche Einführung

Die geblockten bzw. eingeschränkten Domains werden wie bei den übrigen Tools in einem Overlay-Bereich dargestellt und können von dem_der Nutzer_in ergänzend beeinflusst werden. Nach der Installation führt der Privacy Badger durch eine ausführliche Anleitung, die nicht nur die Funktionsweise des Tools selbst erklärt, sondern auch technische Hintergründe kurz darstellt. Sprechende Links („Click to deactivate Privacy Badger on this site“) und Verweise auf erklärende Texte tragen dazu bei, dass die Nutzung des Privacy Badger leichter zu erlernen ist als die der andere AdBlocker.

Zusammenfassung Ad- und Skript-Blocker

Den vorgestellten Werbe- und Tracking-Blockern ist gemein, dass sie auf Seiten der Nutzer_innen Wissen über die Funktionsweise von Online-Tracking voraussetzen. So ist die Darstellung von Trackern auf die Auflistung von Domains oder den Namen

88 Siehe HTTPS://WWW.EFF.ORG/PRIVACYBADGER (letzter Zugriff 26.09.2016).

des_der Anbieter_in beschränkt. Zudem unterscheiden sie sich stark in den Standardeinstellungen. Sowohl Privacy Badger als auch Ghostery blockieren in den aktuellen Versionen automatisch nichts, stattdessen müssen die Tracking-Dienste, die blockiert werden sollen, einzeln ausgewählt werden. Darüber hinaus sorgen sich die Werbetreibenden um die steigende Zahl der AdBlocker-Nutzer_innen, so dass verstärkt an Abwehrmechanismen (*AdBlocker-Blocker*) gearbeitet wird.[89] Es ist daher zu erwarten, dass das Schutzniveau durch AdBlocker in Zukunft sinkt. Tabelle 3 gibt einen Überblick über die Werkzeuge in Bezug auf die Anforderungen.

	Funktionalität		**Usability**		**Privacy Literacy**	
	Transparenz	Intervenierbarkeit	Erwartungskonformität	Lernförderlichkeit	Nachvollziehbarkeit	Folgenabschätzung
AdBlock	keine Transparenz über erfolgreiches Blockieren; Filterlisten selbst einsehbar	eigene Einstellungen möglich für Experten	funktioniert ab Installation ohne Interaktion	keine weiteren Funktionen	keine Erläuterungen	keine Unterstützung
Ghostery	zeigt geblockte und nicht geblockte Skripte als Overlay; Filterlisten einsehbar	Skripte können einzeln und pro Seite (de-)aktiviert werden	Hintergründe zu einzelnen Firmen erfahrbar	schlechte Standardeinstellungen, wenig Erläuterungen	Übersicht in Overlay	keine Information zum Inhalt/Umfang des Profilings
Disconnect	zeigt Anzahl geblockter Skripte und Typen; Filterlisten nicht einsehbar	Skripte können einzeln und pro Seite (de-)aktiviert werden	Links zu den Webseiten einzelner Firmen vorhanden	keine Einführung oder Hinweise bei der (ersten) Nutzung	Overlay zeigt gesparte Zeit und Bandbreite.	nur in Bezug auf Bandbreiteneinsparung
Privacy-Badger	zeigt Anzahl geblockter Skripte; „Selbstlernende“ Funktionen nicht nachvollziehbar	Skripte können einzeln in unterschiedlichen Stufen erlaubt werden	ausführliche Erläuterungen zur Funktionsweise von Tracking	Hinweise bei erstmaliger Nutzung	Übersicht im Overlay	keine Information zum Inhalt/Umfang des Profilings
NoScript	mehrere Varianten für Hinweise; Funktionsweise eindeutig	Skripte können einzeln und pro Seite (de-)aktiviert werden	schwierige Handhabung	keine Erläuterungen für unerfahrene Nutzer_innen	keine Hinweise auf übertragene Daten abseits des Requests selbst	keine Unterstützung

Tabelle 3: Vergleich von Werbe- und Tracking-Blockern.

89 Siehe dazu z. B. heise online vom 20.06.2015. *Google-Ex sagt Adblockern den Kampf an*. HTTP://HEISE.DE/-2718401 (Zugriff am: 26.09.2016).

3.5.5 Obfuscation

Wie in 3.4.2 beschrieben ist das Ziel von Obfuscation-Tools das Verschleiern der Profile der Nutzer_innen durch das Senden vieler zusätzlicher Anfragen. Hierdurch wird Tracking oder Profiling nicht blockiert, aber die sinnvolle Auswertung der Daten erschwert.

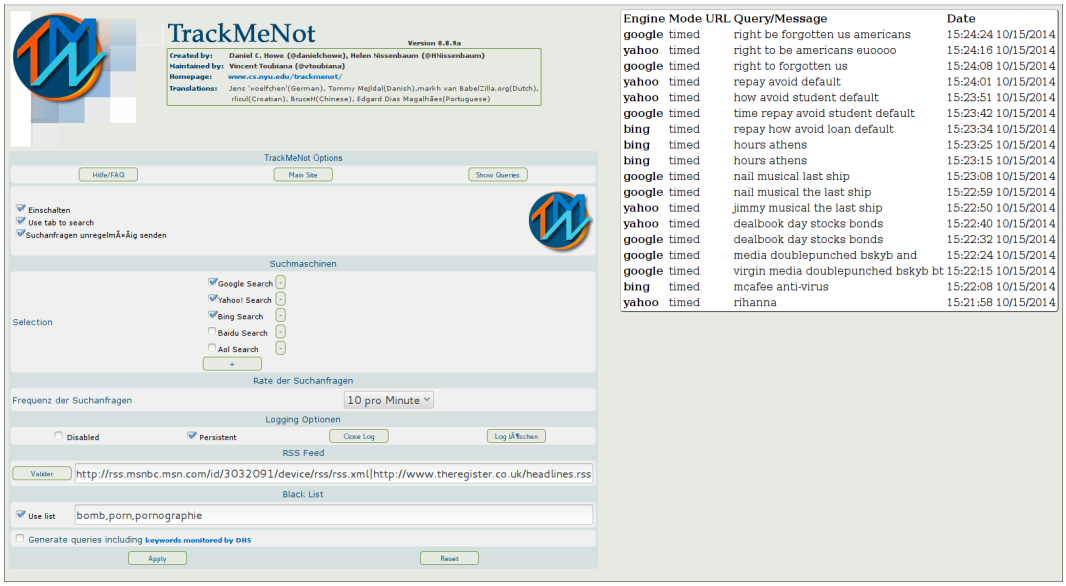


Abbildung 15: Screenshot der TrackMeNot-Einstellungen; auf der rechten Seite: automatisch ausgeführte Suchanfragen.

TrackMeNot

Eines der bekanntesten Obfuscation-Tools, das sich in der praktischen Verwendung befindet, ist TrackMeNot (Howe und Nissenbaum 2009; Toubiana, Subramanian und Nissenbaum 2011). Es hat zum Ziel, die Suchanfragen des_r Nutzer_in an unterschiedliche Suchmaschinen zwischen zufälligen Anfragen zu verbergen. Als Browser-Plugin (vgl. Abb. 15) für Firefox sendet TrackMeNot (TMN), während das Browserfenster geöffnet ist, Suchanfragen an eine oder mehrere Suchmaschinen. Die Suchbegriffe werden von unterschiedlichen Nachrichtenportalen gesammelt.

Diskussion um TMN

TrackMeNot hat viel Aufmerksamkeit in der wissenschaftlichen wie nicht-wissenschaftlichen Community erfahren und eine Diskussion um die Möglichkeiten und Grenzen von Obfuscation angestoßen. Darauf aufbauend wurde und wird TMN regelmäßig verbessert. Wesentliche Änderungen gegenüber der ersten Version basieren auf der Erkenntnis, dass die Suchbegriffe, die zur Verschleierung eingesetzt werden, nicht absolut zufällig sein sollten. Um zu verhindern, dass bestimmte Suchbegriffe, die ein schlechtes Bild auf den_die Nutzer_in werfen könnten, zufällig genutzt werden, besteht in der aktuellen Version die Möglichkeit, solche Suchbegriffe nicht zu

nutzen, die nachweislich unter Beobachtung amerikanischer Sicherheitsbehörden stehen.[90]

Obfuscation ohne Kenntnis des Profils

Das Problem bei TMN liegt darin, dass es ein vor allen Dingen theoretisch beschriebenes Problem behebt. Das Profil, das durch TMN beeinflusst werden soll, ist den Nutzer_innen nicht bekannt. Auch die Liste der Suchbegriffe, deren Anpassungsmöglichkeit – also die Intervention der Nutzer_innen in die Suchbegriffe – eine wesentliche Weiterentwicklung war, bezieht sich jeweils auf ein Profil, das nicht bekannt ist. Die Kritik, dass diese oder jene Suchbegriffe von den Nutzer_innen vermieden werden wollen, beruht nicht auf dem Wissen, inwieweit sie ein Profil in eine Richtung beeinflussen, sondern auf Vermutungen, zu welchem Profil sie führen könnten. Darüber hinaus ist die Nutzung von TMN für ungeübte Nutzer_innen nicht ohne Weiteres möglich. Es existiert zwar eine ausführliche Hinweisseite, aber die Anwendung an sich führt den_die Nutzer_in nicht an die Gebrauchsmöglichkeiten heran.

GoogleSharing

Tausch von Tracking Cookies

Unter dem Namen GoogleSharing (Kassner 2010) wurde 2010 bis 2013 ein Webdienst betrieben, deren Nutzer_innen über ein Browser-Plugin Google Tracking Cookies austauschten. Ziel dieser einfachen Obfuscation war es, die Profile mehrerer Nutzer_innen zu mischen. Der Dienst wurde 2013 eingestellt, eine Evaluation über den Erfolg ist dem Autor nicht bekannt. Möglich wäre etwa, dass Google die Nutzung eines Cookies durch einen anderen Browser (z. B. durch Browser-Fingerprinting) registriert und die Verschleierung so bemerkt. GoogleSharing hat keine besondere Verbreitung erfahren, was darauf zurückgeführt werden kann, dass es als technischer *proof of concept* konzipiert wurde und keine Einführung oder Erläuterung der Funktionsweise existierte.

AdNauseam

Klick auf jede angezeigte Werbung

AdNauseam[91] (vgl. Abbildung Error: Reference source not found) lässt sich als AddOn für Firefox installieren und dient der Verschleierung gegenüber Werbeanbieter_innen. Die Obfuscation wird erreicht, indem das AddOn versucht, alle Werbeanzeigen anzuklicken, die dem_der Nutzer_in präsentiert werden. Das geschieht auch, wenn Werbeblocker wie AdBlockPlus verwendet werden. Die Autor_innen begründen die Entwicklung von AdNauseam mit der mangelnden Bereitschaft der Online-Werbeindustrie datenschutzfreundlichere Standards umzusetzen. AdNauseam richtet sich explizit an

90 Vgl. Reuven Cohen 26.05.2012, HTTP://WWW.FORBES.COM/SITES/REUVENCOHEN/2012/05/26/DEPARTMENT-OF-HOMELAND-SECURITY-FORCED-TO-RELEASE-LIST-OF-KEYWORDS-USED-TO-MONITOR-SOCIAL-NETWORKING-SITES/2/ (letzter Zugriff 12.10.2015).

91 Siehe HTTPS://DHOWE.GITHUB.IO/ADNAUSEAM/ (letzter Zugriff 26.09.2016).

Menschen, die mit diesem Verhalten nicht einverstanden sind, hält aber trotzdem eine Erläuterung der Funktionsweise bereit. Darüber hinaus sind jegliche Aktionen des Add-Ons nachvollziehbar. Angeklickte Anzeigen sind im Nachhinein im „Ad Vault“ genannten Bereich einsehbar, so dass gegebenenfalls Muster und Profile identifiziert werden können.

Abbildung 16: Ad Nauseam Overlay. Zeigt die bisher gefunden und angeklickten Anzeigen

Kritisiert wird die Strategie von AdNauseam dafür, das sie dem Phänomen *ClickFraud* nicht unähnlich ist. ClickFraud, im Deutschen „Klick Betrug“, bezeichnet Strategien, automatisiert Werbeanzeigen anzuklicken, um dem_der Anzeigenschaltenden zu schaden, die in den gängigen Abrechnungsmodellen für jeden Klick zahlen müssen.

Undefined

Nur als Video veröffentlicht

undefined[92] ist ein Studienabschlussprojekt, das nur als Prototyp veröffentlicht wurde. Es kann zur Verschleierung von Profilen in sozialen Online-Netzwerken und gegenüber Suchmaschinen eingesetzt werden. Nach der Auswahl einer Strategie zum Umgang mit den eigenen Profilen (u. a. bestehende Interessen verstärken oder verschleiern) ruft das Tool entsprechende Webseiten auf und setzt automatisiert Postings ab. Die Dummy-Generation-Strategie beruht bei *undefined* auf einem Webservice, der automatisch die inhaltliche Ausrichtung bereits gemachter Postings oder Suchen analysierte und entsprechend gegenläufige oder unterstützende weitere Postings/Suchen vornimmt. Als Design-Experiment konzipiert liegen keine Auswertungen über die Effektivität dieser Verschleierungsstrategie vor.

Zusammenfassung Obfuscationtools

Es existieren nur zwei Obfuscation-Tools, die für Endnutzer_innen verwendbar sind. Während TrackMeNot mehrfach funktional verbessert wurde, ist die Usability mangelhaft. Außerdem lässt sich der Erfolg der Obfuscation nicht nachvollziehe. Das jüngere AdNauseam greift dagegen nachvollziehbar in das Profiling ein, indem es selbstständig Werbeanzeigen anklickt. Allerdings gibt es kaum Möglichkeiten, auf die Verschleierung Einfluss zu nehmen. Beide Tools sind vorbildlich bei der Darstellung ihrer eigenen Funktionsweise und erlauben Einsicht in Logs und getätigte Seitenaufrufe.

92 Siehe HTTP://VINCENTDUBOIS.FR/UNDEFINED.PHP; Quellcode HTTPS://GITHUB.COM/VAINZOU (letzter Zugriff 26.09.2016).

	Funktionalität		**Usability**		**Privacy Literacy**	
Name	Transparenz	Intervenierbarkeit	Erwartungskonformität	Lernförderlichkeit	Nachvollziehbarkeit	Folgenabschätzung
TrackMeNot	nicht über Tracking oder Profile; Log-Ansicht macht Obfuscation transparanet	soll verhindern, dass eindeutige Profile gebildet werden können; Überprüfung mit dem Tool nicht möglich	Gestaltung entspricht nicht aktuellen Erwartungen an eine Software	Nutzer_innen ohne Kenntnis des Themas haben Schwierigkeiten TMN zu bedienen	Obfuscation kann in separatem Fenster nachvollzogen werden; Profile allerdings nicht	Ausschlusslisten und Liste getätigter Suchanfragen helfen Folgen zu bedenken
Google Sharing	keine; agiert im Hintergrund und bezieht sich auf nicht-dargestellte Werbe-Profile	erlaubt den Austausch von Google-Tracking-Cookies zwischen Nutzer_innen	kaum Bedienelemente; nicht selbsterklärend	der geringe Funktionsumfang begründet das Fehlen weiterer Hilfen	die entstehenden, übernommenen oder übergebenen Profile, die an einen Cookie geknüpft sind, sind nicht einsehbar	keine Unterstützung
AdNauseam	zeigt geklickte und geblockte Werbung und ausführliche Logs	klickt unerwünschte Werbung zum Schaden der Werber	komplexe Systematik wird u. a. mit Video erklärt	Hilfeseite, Logs und „Ad Vault“ ermöglichen weitergehendes Verständnis	über die Anzeigen können mögliche Profile abgeschätzt werden.	unklar, wie sich das Klicken von Werbung auf die Profile auswirkt

Tabelle 4: Vergleich von Obfuscationtools.

3.5.6 Transparenz

Eine weitere Gruppe von Tools für Enduser_innen versucht die Transparenz von Tracking und Datenverarbeitung zu erhöhen. Dabei ist in der Regel nicht die Transparenz selbst das Ziel, sondern durch die Transparenz Aufmerksamkeit zu schaffen für die häufig unsichtbaren Datenflüsse. Wie bei den Obfuscation-Tools sind alle Produkte als Browsererweiterung konzipiert und werden als solche nicht von den Unternehmen, die Tracking und Profiling durchführen, angeboten, sondern von Dritten entwickelt.

Floodwatch

Floodwatch[93] ist eine Browsererweiterung, die über den Umfang von Werbung im Internet Transparenz schaffen soll und (in Ansätzen) die dabei entstehenden Profile darstellt. Das für Chrome entwickelte Browser-AddOn sammelt die Anzeigen auf Webseiten, während der_die Nutzer_in surft. Auf einer separaten Seite (vgl. Abbildung 17)

93 Siehe HTTP://FLOODWATCH.O-C-R.ORG/ zuletzt abgerufen am 26.09.2016.

werden die Anzeigen gebündelt dargestellt, um so einen Überblick über die Art, Form und den Inhalt der Anzeigen zu geben. Nutzer_innen sollen so in die Lage versetzt werden, oft angezeigte Marken und Werbekampagnen zu identifizieren, die in der Einzeldarstellung nicht auffallen. Es macht die Datenspeicherung durch Werbeanbieter insofern transparenter, als das der Umfang im zeitlichen Verlauf sowie ein Profil einschätzbar werden.

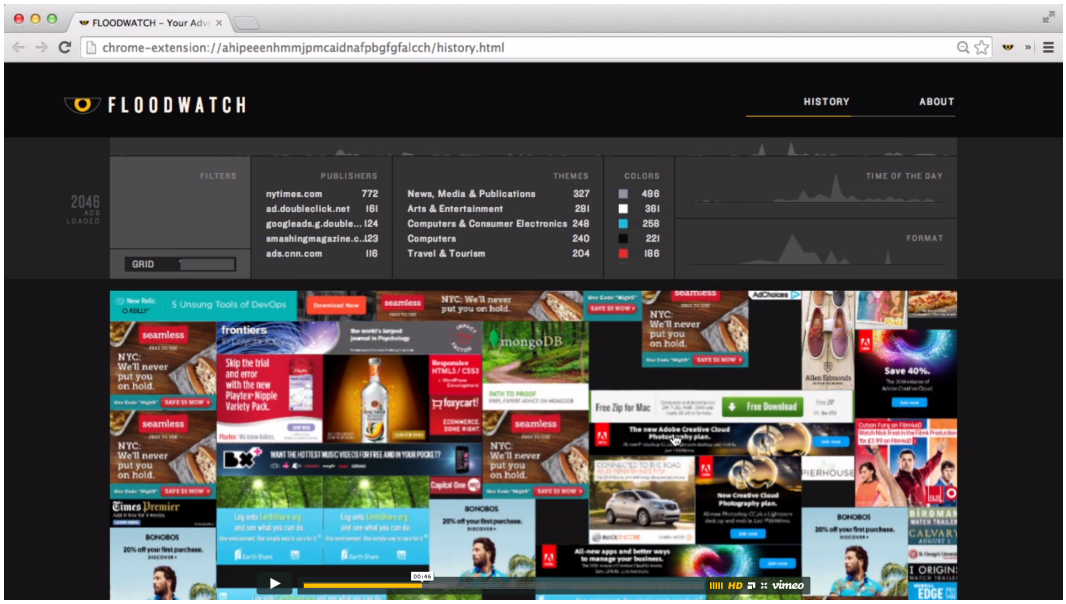


Abbildung 17: Screenshot aus einem Floodwatch Demonstrationsvideo. (11.10.2015)

Signature Mode

Zusätzlich bietet Floodwatch die Möglichkeit, die gesammelten Daten pseudonymisiert an einen zentralen Server zu übermitteln. Dort werden die Daten aggregiert und an die Nutzer_innen zurückgesendet. Diese können dann im sogenannten *Signature Mode* ihr Anzeigenprofil mit dem anderer Nutzer_innen vergleichen. Der Vergleich basiert auf der Anzahl der Werbeanzeigen, die bestimmten Interessenkategorien zugeordnet werden. Nutzer_innen sollen so die Einzigartigkeit ihres Anzeigenprofils abschätzen können.

about:profile

Das Firefox AddOn about:profile[94] kombiniert die Liste der besuchten Webseiten in einem Browser mit statistischen Daten aus verschiedenen Quellen, um Profilinformationen über den_die Benutzer_in des Browsers zu ermitteln. Dabei werden Informationen über die Kategorien von Webseiten (basierende auf dem Open Directory Project) und demografische Statistiken zusammengeführt und aggregiert. Der_die Nutzer_in bekommt so einen Einblick in den Umfang und die Genauigkeit des Profils, das theore-

94 Siehe: HTTPS://ADDONS.MOZILLA.ORG/EN-US/FIREFOX/ADDON/PROSPECTOR-ABOUTPROFILE/ letzter Zugriff 26.09.2016.

tisch über ihn oder sie erstellt werden könnte, wenn einem Profiling-Service alle besuchten Webseiten bekannt wären. Das Plugin Zusatzinformationen transparent, die durch die Zusammenführung mehrer Datenquellen entwickelt werden können. Allerdings wurde die Entwicklung nach der Vorstellung des Prototyps zu Gunsten des Interest Dashboards eingestellt.

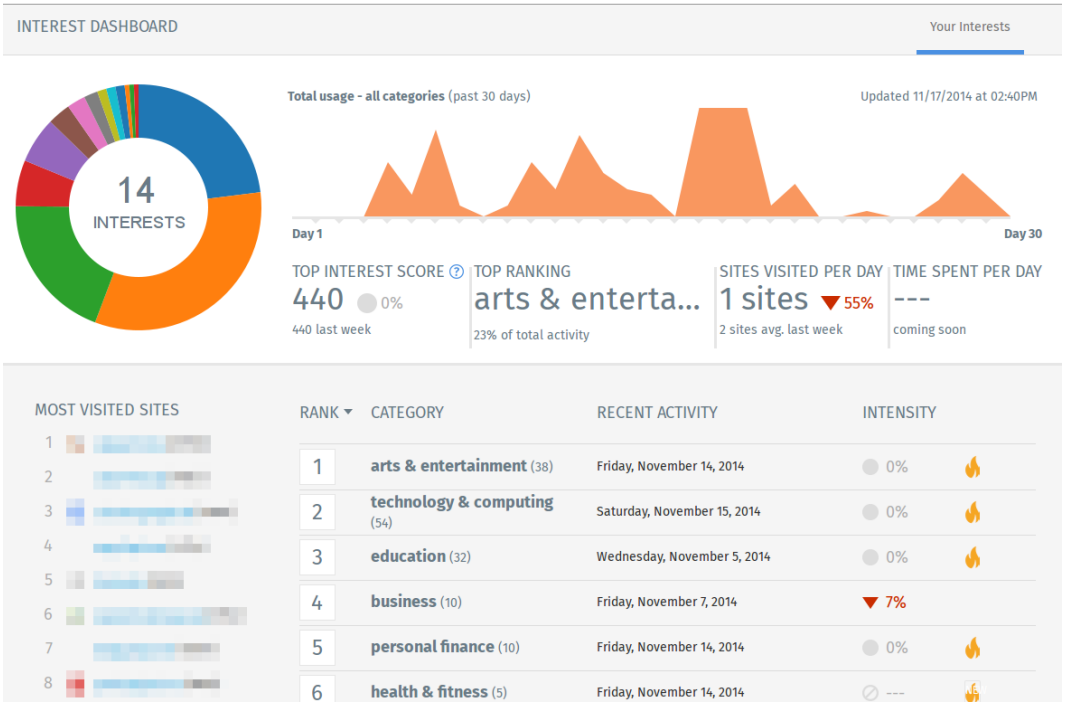


Abbildung 18: Screenshot des Firefox Interest Dashboard.

Mozilla Interest Dashboard

Das Interest Dashboard (ID, vgl. 18) ist eine von Mozilla entwickelte Erweiterung von *about:profile*. Ziel ist die automatische Kategorisierung von Webseiten in Interessen. Dazu zeigt das Plugin Rankings und Visualisierungen zum zeitlichen Verlauf der am häufigsten beobachteten Interessen. So soll Nutzer_innen ihr Profil transparent gemacht werden. Auch die Zuordnung von Seiten zu Interessen wird dargestellt. Sie erfolgt über eine automatische Auswertung der Seiteninhalte, der von den Betreiber_innen gesetzten Schlagworte und dem Seitentitel. Allerdings ist dieses Verfahren relativ fehleranfällig[95], nur wenige Seiten werden überhaupt in die Berechnung mit einbezogen und eine Korrektur oder eine Einsicht in das zugrundeliegende Kategoriensystem sind nicht möglich. Trotz der grafischen Aufbereitung sind einige Funktionen nicht selbsterklärend. Eine noch in der Entwicklung befindliche Funktion, die auf Basis der beobachteten Interessen Empfehlungen für weitere Seiten zu eben jenen Interessen

95 Im Selbstversuch wurden z. B. Webseiten zur Programmierung in Coffee-Script in die Kategorie „Food & Drinks“ einsortiert.

macht, deutet darauf hin, dass es ein weiteres Ziel des Plugins ist, auf neue und unbekannte Inhalte hinzuweisen.

Lightbeam

Lightbeam[96] soll die Transparenz von Tracking durch *3rd Parties* erhöhen. In einer Netzwerkgrafik (vgl. 19) wird dargestellt, welche Domains im Zuge des Aufrufs welcher Seiten aufgerufen wurden. So kristallisieren sich über die Zeit insbesondere solche 3rd Parties heraus, die in viele Seiten eingebettet sind. Diese Knoten können dann einzeln ausgewählt und blockiert werden, um so zentrale Tracking-Dienste gezielt auszuschalten. Lightbeam vergrößert so das Wissen über die Verknüpfungen, die zwischen Seiten bestehen, setzt allerdings auch voraus, dass der_die Nutzer_in bereits weiß, wie solche Netzwerke und das Tracking funktionieren.

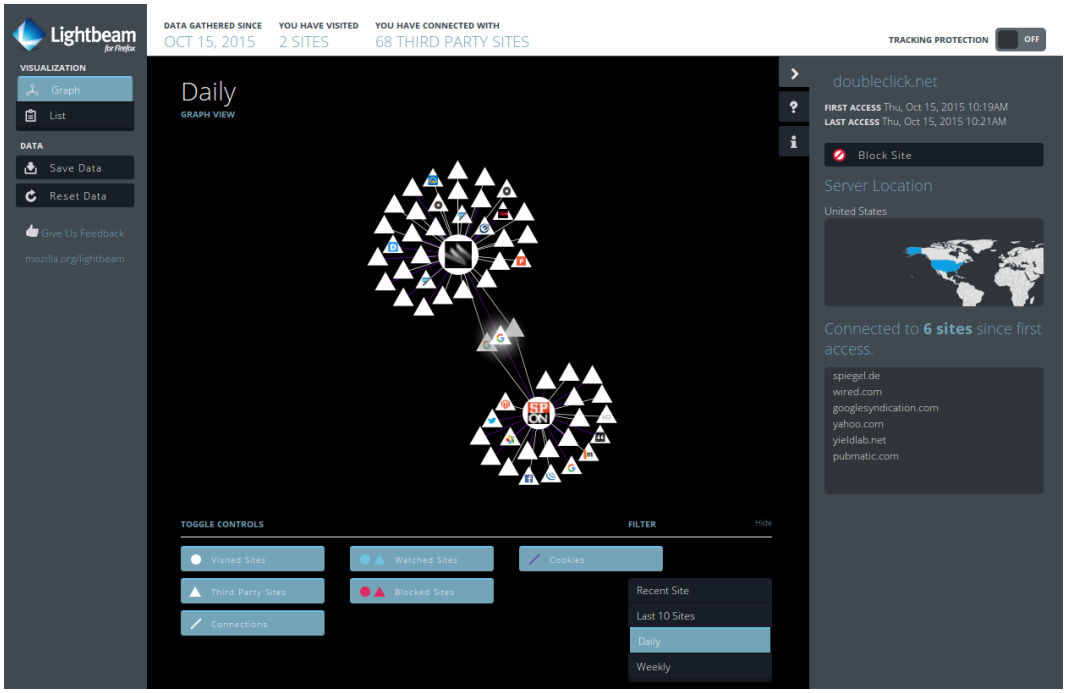


Abbildung 19: Screenshot von Lightbeam nach dem Aufruf von zwei Webseiten (große Kreise); die Dreiecke symbolisieren 3rd-Party-Skript, und die Verbindungen zeigen an, auf welcher Seite sie geladen wurden.

Xray

Xray ist ein Forschungsprototyp mit dem Ziel, den Zusammenhang zwischen Online-Werbung und Aktionen zu zeigen, die der_die Nutzer_in im Browser gemacht hat. Lécuyer u. a. (2014) haben in ihrer Studie den Zusammenhang zwischen Werbung in GMail, Amazon Produktplatzierungen sowie vorher besuchten Webseiten und empfan-

96 Siehe HTTPS://ADDONS.MOZILLA.ORG/EN-US/FIREFOX/ADDON/LIGHTBEAM/ (letzter Zugriff 26.09.2016).

genen E-Mails untersucht. Es ist bekannt, dass Google seinen E-Mail-Dienst GMail dazu nutzt, den Inhalt von Nachrichten zu kategorisieren, um über sein Werbenetzwerk entsprechende Anzeigen darzustellen. Das von den Autoren der Studie entwickelte Programm, das sich nicht an Endnutzer_innen richtet, soll die Verknüpfung zwischen E-Mails und Werbeanzeigen durch eine eigene Inhaltsanalyse sichtbar machen. Xray ist das einzige bekannte Werkzeug zur Schulung einer *critical literacy* der Kategorie *Datenanalyse als Technik und als Objekt*, da es Nutzer_innen ermöglicht, die versteckten Zusammenhänge von Googles Datenanalyse einzusehen und dabei selbst Datenanalyse einsetzt. Wie die Autor_innen, die dabei auftretenden Usability-Fragen angegangen sind, ist allerdings zum Zeitpunkt der Erstellung dieser Arbeit nicht bekannt. Daher wird Xray in der Übersicht nicht berücksichtigt.

	Funktionalität		**Usability**		**Privacy Literacy**	
	Transparenz	Intervenierbarkeit	Erwartungskonformität	Lernförderlichkeit	Nachvollziehbarkeit	Folgenabschätzung
Floodwatch	unterstützt die Selbsteinschätzung eines Werbeprofils; stellt die eigene Funktionsweise dar	keine Interventionsmöglichkeiten	beim Test traten unter bestimmten Umständen nicht behebbare Fehler auf; generell aber gute Erwartungskonformität	ausführliche Erläuterungen werden auf der Homepage gegeben	der *Signature Mode* ermöglicht eine bessere Einschätzung des eigenen Profils im Vergleich mit anderen	keine Hinweise auf mögliche Folgen; regt Auseinandersetzung an
Interest Dashboard (ID)	gewährt Überblick über das eigene Interessenprofil auf Basis eines nicht gut einsehbaren eigenen Algorithmus'	Empfehlungsfunktionalität für Seiten innerhalb oder außerhalb der eigenen Interessen noch in der Entwicklung	Metriken teilweise nicht nachvollziehbar; Grenzen des Tools werden nicht dargestellt	Fokus liegt auf der Erläuterung des erstellten Profils	das Interest Dashboard zeigt die Interessen, die einer Seite zugeordnet sind an	Grenzen des Tools werden nicht dargestellt
Lightbeam	visualisiert Verknüpfungen zwischen Webseiten die durch 3rd-Party-Tracking entstehen	Tracking kann insgesamt oder für verschiedene Anbieter unterbunden werden	grundsätzliches Verständnis von Tracking und Cookies wird vorausgesetzt	jenseits der Visualisierung ausführliche Erläuterungen zu den Funktionen	keine über das Tracking hinausgehenden Erläuterungen; Rohdaten und Logs einsehbar	keine Hinweise auf mögliche Folgen; regt Auseinandersetzung an

Tabelle 5: Vergleich von Transparenz-Tools. Xray und about:profile sind ausgenommen, da sie nicht (mehr) nutzbar sind.

Die vorgestellten Transparenz-Tools zeichnen sich dadurch aus, dass sie jeweils einen bestimmten Themenbereich (Interessen, Werbung, Tracking) betrachten und hier versuchen, den Nutzer_innen die komplexen Zusammenhänge darzustellen. Die in 5 ver-

glichenen Werkzeuge zeichnen sich dadurch aus, dass sie großen Wert auf die Visualisierung legen und so tatsächlich dazu beitragen können, die Transparenz in Bezug auf Vorgänge beim Tracking und Profiling zu verbessern.

3.6 ZUSAMMENFASSUNG

In diesem Kapitel wurden die ökonomischen (Abschnitt 3.1) und technischen (Abschnitt 3.3) Hintergründe von Online-Marketing und Web-Tracking erläutert, auf denen Online-Profiling basiert. Damit wurden unterschiedliche Aspekte der zweiten in der Einleitung gestellten Leitfrage beantwortet (*Was ist Online-Tracking und Online-Profiling? Wie funktioniert es, wie und zu welchem Zweck wird es eingesetzt?)*.

Privacy und Transparency Enhancing Technologies

Aufbauend auf der im vorherigen Kapitel dargestellten Notwendigkeit, Nutzer_innen mehr Möglichkeiten zur Beobachtung und Beeinflussung von Profilen zu geben wurden dann Privacy Enhancing Technologies vorgestellt, die versuchen, das Problem auf Systemebene anzugehen (Abschnitt 3.4). Für die Analyse existierender Transparenz-Werkzeuge wurden dann weitere Anforderungen formuliert (Abschnitt 3.5). Von den sechs Datenschutzzielen ist die Umsetzung von Transparenz und Intervenierbarkeit von größter Bedeutung. Diese kann sich, im Fall von P/TET sowohl auf die Profiling- und Tracking-Prozesse, als auch auf die Datenverarbeitung innerhalb der PET selbst beziehen. Darüber hinaus wird in der Forschung an vielen Stellen die Wichtigkeit von Usability für P/TET hervorgehoben, die sich unter anderem an der leichten Erlernbarkeit (Erwartungskonformität) sowie an den Funktionen zur Schulung der Nutzung (Lernförderlichkeit) messen lässt. Bereits im vorherigen Kapitel wurde die Idee entwickelt, dass die Auseinandersetzung mit Tracking und Profiling dazu genutzt werden kann, eine *critical data literacy* zu schulen. Konkret kann sich dies in technischer Unterstützung für die Nachvollziehbarkeit und Folgenabschätzung von Tracking und Profiling zeigen.

Vorstellung existierender Werkzeuge

Im Anschluss wurden eine Reihe von Softwarelösungen und Diensten betrachtet, die bereits in diesem Bereich vorhanden sind (Abschnitt 3.5). Es wurde überprüft, inwieweit die Lösungen der Tracking-Dienste selbst sowie solche in den Bereichen Blocking, Obfuscation und Transparenz den beschriebenen Anforderungen genügen. Dabei zeigte sich, dass die vorhandenen Anwendungen sich häufig auf eine Teilaufgabe fokussieren.

Hersteller sind nicht hilfreich

Die von den Tracking-Diensten angebotenen Webseiten bieten, wenn überhaupt, Transparenz über das Vorhandensein von Tracking und Opt-Out als einzige Möglichkeit der Intervention. Einzig Google bietet eingeloggten User_innen die Funktionen, das erstellte Profil einzusehen und zu verbessern. Die Logik ist hier, den_die

Einzelne_n von den positiven Aspekten der Technologie zu überzeugen und sie_ihn zur Mitarbeit zu überreden.

Technischer Wettlauf um AdBlocker

Die relativ große Anzahl von Tools zum Blocken von Werbeanzeigen und Tracking konzentriert sich wiederum auf Möglichkeiten, einen Widerspruch technisch durchzusetzen. Sie unterscheiden sich vor allem in der Art, wie sie Nutzer_innen an das Thema heranführen und ob sie weitergehende Möglichkeiten zur Einflussnahme bereitstellen. Während der überwiegende Teil mit bestehenden Listen bekannter Dienste arbeitet, bietet nur Ghostery die Möglichkeit, mehr über diese Dienste zu erfahren. Die Entwickler_innen des Privacy Badger wiederum geben sich besondere Mühe, die Funktionsweise ihrer Browsererweiterung zu erklären. Allerdings setzen alle Tools auf der technischen Ebene des Tracking an, was Profiling unmöglich macht, wenn denn das Blockieren funktioniert. Der Nachteil besteht einerseits darin, dass sich AdBlocker aktuell in einem Wettbewerb mit den Tracking-Diensten befinden, die wiederum AdBlocker-Blocker entwickeln. Andererseits bieten sie keine Möglichkeiten, das Profiling nachvollziehbar zu machen und so eine Folgenabschätzung zu ermöglichen.

Transparencytools in der Entwicklung

Eben dieser Aspekt steht bei einigen Transparency Enhancing Technologies im Vordergrund, die teilweise auch Obfuscation unterstützen. Allerdings sind viele Werkzeuge hier noch in der Entwicklung, wenig erprobt und die innere Logik ist nicht immer nachvollziehbar. So implementiert das Mozilla Interest Dashboard zwar einige gute Ideen, erstellt Profile allerdings auf einer eigenen, nicht nachvollziehbaren Logik. Die auf Werbung fokussierten Tools Floodwatch und AdNauseam machen zwar durch das Sammeln der (nicht) angezeigten Werbeanzeigen das Ausmaß sichtbar und erlauben in Ansätzen auch eine Einschätzung des entstehenden Profils, überlassen die Interpretation aber vollständig dem_der Nutzer_in. Die Werkzeuge, die Obfuscation erlauben sind entweder, wie GoogleSharing, selbst intransparent, oder verfügen über keine Möglichkeit nachzuvollziehen, ob und wie die Obfuscation sich auf das erstellte Profil auswirkt.

Offene Punkte

Während einige Tools gute Konzepte insbesondere mit Hinblick auf die Usability und Visualisierung entwickelt haben, sollen bei der Entwicklung einer eigenen Anwendung insbesondere die offenen Punkte berücksichtigt werden. Es soll untersucht werden, wie die Profile in der Praxis aussehen und inwiefern sie Nutzer_innen sichtbar und verständlich gemacht werden können. Eine Obfuscation soll sich zudem daran orientieren, dass ihr Erfolg praktisch sicht- und messbar werden kann.

4. EMPIRISCHE ANALYSE VON ONLINE-PROFILING UND OBFUSCATION

In den vorangegangenen Kapiteln wurde die Notwendigkeit einer Auseinandersetzung mit Mechanismen des Profiling hervorgehoben. Als notwendige Bestandteile einer *Privacy* und *Transparency Enhancing Technology* wurden die Funktionen zur Transparenz von und Intervenierbarkeit in Profiling festgelegt, die, wenn sie in einer angemessenen Weise nutzbar gemacht werden, User_innen in einem kritischen Umgang mit Profiling schulen sollen. Auch wenn einige Ansätze existieren, die diese Anforderungen umzusetzen versuchen, wurden bei bestehenden Werkzeugen in allen Bereichen Mängel festgestellt. Bisher existiert keine Möglichkeit für Nutzer_innen, ihr Profil ohne Mithilfe eines_r Anbieters_in einzusehen (Mangel an Transparenz). Die Einflussnahme beschränkt sich auf Obfuscation-Verfahren, deren Effektivität nicht nachgewiesen ist (Mangel an Intervenierbarkeit).

In diesem Kapitel werden diese Forschungslücken zur Konstruktion von Profilen im Bereich der Online-Werbung geschlossen. Zentral für die Konzeption der Untersuchung ist die Idee, dass eine praxisnahe Ermittlung entstehender Profile aus der Sicht eines_r Nutzer_in geschehen muss. Während die meisten Studien zu Online-Tracking eine große Menge von Webseiten besuchen und diese für sich genommen auf die eingesetzten Tracking-Arten hin untersuchen, wird in der vorliegenden Studie der Umfang von Tracking aus der Perspektive des Effekts, also der entstehenden Profile, betrachtet. Dazu wird die Methodik der automatisierten Profilanalyse vorgestellt und auf die dazu notwendige Datengewinnung eingegangen. Anschließend werden die Erkenntnisse genutzt, um mehrere Methoden zur Obfuscation zu evaluieren, die Intervention in die verschiedenen Formen der Profile ermöglichen.

Das Forschungsdesign besteht dabei aus drei wiederverwendbaren und anpassbaren Elementen der Datenerhebung (TrackTrack), Auswertung (TrackBack) und Nutzung (TrickTrack). Im Ergebnis zeigt sich, dass Profile effektiv und unabhängig von der Zuarbeit des_r Anbieters_in konstruiert und beeinflusst werden können.

4.1 ÜBERGEORDNETE ANFORDERUNGEN

Transparenz über tatsächliche Profile

Wie im vorherigen Kapitel deutlich wurde, mangelt es bei existierenden Werkzeugen insbesondere an zwei Punkten. Zum einen liegt der Fokus, insbesondere bei den Blocking-Tools, sehr stark auf der technischen Ebene, dem Tracking selbst. Die Art der Profile, die durch das Tracking generiert werden, ist hierbei weder einsehbar, noch werden die Vorgänge auf einer übergeordneten Ebene erläutert. Die wenigen

Anwendungen, die eine Einsicht ermöglichen, erstellen die Profile entweder selbst (Firefox ID) oder machen es zur Aufgabe der Nutzer_innen, diese herzuleiten (Floodwatch, AdNauseam). Wie aber am Beispiel der wenigen Informationsseiten der Tracking-Services gezeigt wurde, ist die Einsicht in (abstrahierte) Profile durchaus möglich. Eine Anforderung ist daher, die Informationen, welche die Tracking-Services selbst bereitstellen, für User_innen nutzbar zu machen, um die von ihnen erstellten Profile transparent zu machen.

Nachvollziehbarkeit der Obfuscation

Zum anderen konnte aus der Betrachtung der bestehenden Obfuscation-Tools gelernt werden, dass ihr Funktionsweise nur schwer nachzuvollziehen ist. Das ist insbesondere darauf zurückzuführen, dass es bisher keine Möglichkeit gibt den praktischen Effekt eines Obfuscation-Verfahrens messbar zu machen. Die zweite Anforderung ist daher der Entwurf von Obfuscation-Verfahren, deren Effekt für Endnutzer_innen nachprüfbar ist.

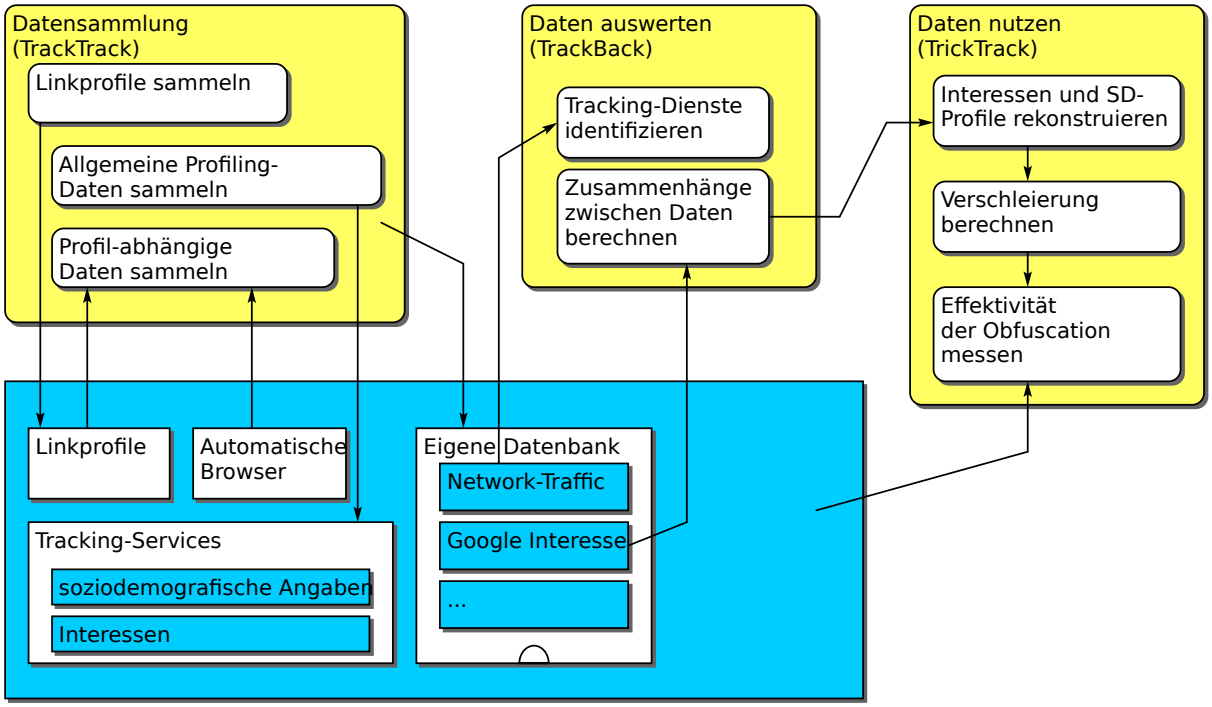


Abbildung 20: Übersicht der Module in TrackTrack

Diese beiden übergeordneten Anforderungen sind in einem mehrstufigen Erhebungs- und Analyse-Framework umgesetzt, das sich aus drei Komponenten zusammensetzt (vgl. 20):

Die drei entwickelten Komponenten

1. **TrackTrack** dient der Erhebung des Status Quo. In einem modularen System können unterschiedliche Datenquellen kombiniert werden, um Tracking-Netzwerke zu entdecken und die dabei entstehenden Profile zu erheben.

2. **TrackBack** ist eine Sammlung aus Programmen, welche die erhobenen Daten analysiert und die Ergebnisse des Profilings reproduzierbar macht.

3. **TrickTrack** setzt, basierend auf den Ergebnissen von TrackBack, eine Methode um, die eine Obfuscation des Profils erlaubt, deren Effekt mit TrackTrack überprüft wird.

Während TrackTrack also vor allen Dingen die erste Anforderung umsetzt und dazu dient, Daten zum Profiling zu erheben, umfasst TrickTrack verschiedene Verfahren der Obfuscation. Die Auswertung der Daten, des Profilings sowie der Obfuscation ist in TrackBack realisiert. Im nächsten Kapitel werden die gewonnenen Erkenntnisse und Verfahren für Nutzer_innen verfügbar gemacht und TrickTrack mit einer UI versehen.

4.2 METHODE ZUR AUTOMATISIERTEN ANALYSE VON PROFILING

Profiling als Black Box

Wie bereits in Kapitel 2 beschrieben, können Profiling-Dienste nur als Blackbox behandelt werden. Da die genaue Funktionsweise der Profilbildung als Teil deren Geschäftsmodells nicht öffentlich ist, ist die Blackbox-Methode die einzige Möglichkeit die Profile zu betrachten. Durch die Beobachtung des In- und Outputs der Blackbox kann versucht werden, etwas über ihre Funktionsweise in Erfahrung zu bringen. Die Erkenntnisse lassen sich später auch zum Design von Obfuscation-Verfahren nutzen. Einen ähnlichen Ansatz wählten Majumder und Shrivastava (2013) in ihrer Untersuchung zur Personalisierung bei Suchmaschinen und Datta, Tschantz und Datta (2015) bei ihrer *information flow* Analyse zu Online-Werbung.

Um eine in sich geschlossene Studie durchführen und auswerten zu können, müssen bestimmte definierte Start- und Endpunkte sowie eine definierte Zahl an Testprofilen gesetzt werden.

4.2.1 Werkzeuge

Zur Analyse der Funktionsweise und des Umfangs von Online-Profiling ist es notwendig mit einem Testset an Profilen zu arbeiten. Um eine relevante Zahl von Profilen erstellen und verwalten zu können, muss der Prozess dazu automatisiert werden. Möglichkeiten dazu bieten das skriptbasierte Crawling, Desktopautomatisierung und sogenanntes *Headless-Browsing*.

Crawling und Scraping

Beim skriptbasierten c*rawling* wird der HTML-Code einer Webseite beim Server abgefragt und im Anschluss analysiert. Vorteil dieses Verfahrens ist die relativ geringe Bandbreitenbenutzung, da keine Bilder oder Skripte abgerufen werden müssen. Wäh-

rend dies zur Analyse der Verbreitung von Trackern durchaus ausreicht, weil die URLs externer Tracking-Inhalte[97] ausgewertet werden können, wird allgemein angenommen, dass dieses Verfahren mit der Verbreitung aktiver Inhalte über JavaScript nicht mehr lange ausreichend ist[98]. Das Entnehmen einzelner Elemente von Webseiten wird *scraping* genannt und bietet sich zur Extraktion von statischen Informationen auf untersuchten Seiten an.

Automatisierte Browser

Da mittels JavaScript nicht nur weitere Skripte, sondern auch Werbeanzeigen geladen werden, ist *crawling* für die Simulation von User_innen keine Lösung bei der verlässliche Ergebnisse zu erwarten sind. Um das Tracking voll umfänglich zu untersuchen, müssen Cookies und das Ausführen von Skripten zum Browser-Fingerprinting durch die Tracker zugelassen werden. Daher sollte für die Analyse ein voll-funktionierender Browser eingesetzt werden. Zur Browser-Automatisierung existieren zwei unterschiedliche Ansätze. Einige Studien verwenden Desktop-Automatisierungssoftware, wie sie im Software-Testing zum Einsatz kommt. Hier wird häufig Firefox in Kombination mit Eigenentwicklungen[99] oder der Erweiterung Selenium[100] genutzt. Der Vorteil ist, dass ein nahezu unveränderter Browser genutzt werden kann, der auch durch Tracking-Skripte nicht von einem „realen“ User zu unterscheiden ist. Eine andere Möglichkeit stellt die Nutzung eines *Headless-Browser*s wie PhantomJS[101] oder SlimerJS, dar. Diese Systeme werden in der Regel benutzt, um automatisierte Software-Tests im Bereich Webentwicklung durchzuführen. Dazu bieten Headless-Browser Schnittstellen für programmatische Beschreibungen der zu erledigenden Schritte, so dass die Durchführung ressourcenschonend, ohne Desktop-Umgebung auskommt.

4.2.2 Grenzen

Ein Nachteil der vorgestellten Methode ist, dass keine kontinuierliche Beobachtung des Online-Profiling möglich ist. Sie wäre nötig, um weitere Effekte über die Zeit, unabhängig von Werbekampagnen und bei stetig wachsendem Umfang der Profile zu beobachten und gegebenenfalls Änderungen zu bemerken. Die Erhebung, die in dieser Arbeit vorgestellt wird, fand zwischen März 2014 und Februar 2015 statt. Unter Umständen sind nicht alle Ergebnisse reproduzierbar.

97 Vgl. dazu Barta (2014).

98 Als Zeichen der Anerkennung dieses Trends kann gewertet werden, das Google im Mai 2014 bekannt gab, bei der automatische Indexierung von Webseiten nun auch Inhalte zu berücksichtigen, die durch JavaScript erzeugt würden. HTTP://GOOGLEWEBMASTERCENTRAL.BLOGSPOT.COM/2014/05/UNDERSTANDING-WEB-PAGES-BETTER.HTML (letzter Zugriff: 26.09.2016).

99 Vgl. Mayer und Mitchell (2012).

100 Links zu Anwendungen und Bibliotheken finden sich im Anhang. Selenium wurde u.a. von Acar u. a. (2014) und Datta, Tschantz und Datta (2015) verwendet.

101 Genutzt u.a. von Acar u. a. (2013) und Purra (2015).

Nachteil der Automatisierung

Ein Nachteil des automatisierten Browsens ist, dass die Möglichkeit besteht, dass Webseitenbetreibende die Nutzung der Systeme erkennen, indem Sie ermitteln um welchen Browsertyp es sich handelt und ob er in einer Desktop-Umgebung ausgeführt wird (vgl. Shekyan 2015). Webseitenbetreiber, die die Nutzung ihrer Seiten durch Headless-Browser verhindern wollen, zum Beispiel um das automatisierte Auslesen von Inhalten zu verhindern, könnten diese Tests einsetzen, um verfälschte Inhalte an den Headless-Browser auszuliefern. Dies ist allerdings ein grundsätzliches methodisches Problem aller Studien und kann auch auf Browsererweiterungen zutreffen. Webseitenbetreiber_innen haben die Möglichkeit, Inhalte verändert anzuzeigen, wie dies auch häufiger bei der Nutzung von AdBlockern geschieht. Darüber hinaus kann die Art der Nutzung einer Webseite Aufschluss darüber geben, ob eine Seite automatisiert oder manuell aufgerufen wird. Dazu kann das Klick- und Scrollverhalten auf einer Webseite beobachtet werden.

Simulation eines regulären Browsers

Trotz dieser theoretischen Möglichkeiten hat keine der zitierten Studien eine negative Beeinflussung von Seiten bei der Nutzung automatisierter Browser festgestellt. Die vorliegende Studie wurde anfangs mit PhantomJS und bei der Evaluation der Obfuscation-Methoden schließlich mit SlimerJS durchgeführt. Der Wechsel begründet sich in der besseren Parallelisierbarkeit von SlimerJS, die beim Vergleich verschiedener Methoden eine wesentlich effizientere Ressourcennutzung ermöglichte. Zusätzlich wurden einige Maßnahmen ergriffen, um einen Teil der Verfahren zur Identifizierung von Headless-Browsern zu täuschen. Die Browserbeschreibung (BrowserID) sowie einige Browsereigenschaften wurden vor dem Einsatz so manipuliert, dass sie keine einfache Identifizierung als Headless-Browser erlauben und das System vorgibt, es handele sich um einen normalen Desktop-Browser. Nicht zu täuschen ist allerdings die Abfrage der Bildschirmauflösung mittels JavaScript, die zumindest bei PhantomJS immer eine Weite und Breite von 0 zurückliefert. Allerdings ist das kein eindeutiger Hinweis auf die Verwendung von PhantomJS, da auch der Chrome-Browser diese Abmessungen an Skripte übergibt, wenn eine Webseite in einem Tab geöffnet wird, der sich nicht im Fokus befindet.

4.3 DATENERHEBUNG MIT TRACKTRACK

Zur Untersuchung der Profile reicht es nicht, den Netzwerkverkehr allein als Quelle zu untersuchen. Es werden sekundäre Informationen über Kategorien-Systeme und konkrete Ausprägungen von Profilen benötigt, wie sie in 3.4.1 vorgestellt wurden. Um realistische Profile beobachten zu können, müssen daher in einem ersten Schritt die Daten bei den Tracking-Diensten erhoben werden. Hierzu können die in Tabelle 6 gelisteten Skripte genutzt werden, die zu den in Tabelle 7 beschriebenen Datenbanken führten. Zur Erhebung wurden zwei verschiedene Verfahren genutzt, die in TrackTrack umgesetzt sind.

Funktionsweise von TrackTrack

1. **Simulation von Nutzer_innen** und anschließendes *scrapen* der Profile bei den Tracking-Services, die diese offenlegen. Für dieses Verfahren bietet sich die Profilinformationsseite von Google an (vgl. Abb. 21).
2. **Scraping der Tracking-Services**, die keine personenbezogenen, sondern seitenbezogene Profile anbieten. Diese werden erhoben, um Profilinformationen zu schlussfolgern.

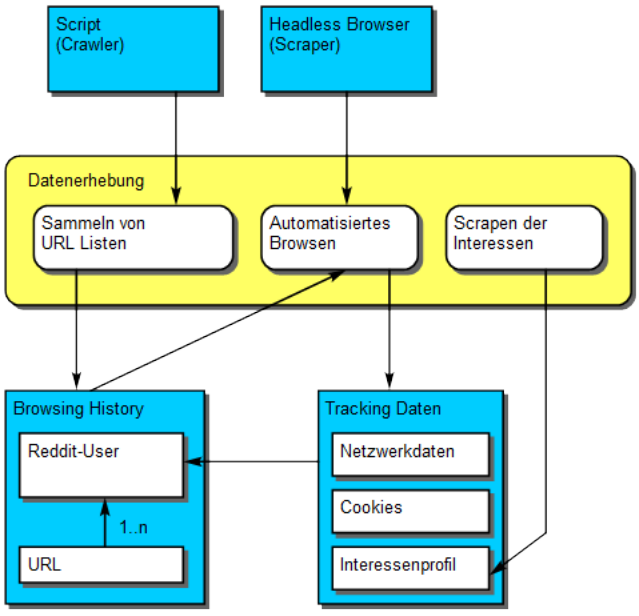


Abbildung 21: Modell der Datenerhebung

TrackTrack ist eine Sammlung von Skripten für die Kommandozeile, die Informationen über Online-Tracking und -Profiling aus verschiedenen Quellen erhebt und, für die Auswertung aufbereitet, in einer Datenbank ablegt. Bei der User_innen-Simulation (Option 1) werden mehrere Seiten hintereinander aufgerufen, vergleichbar mit einer Webbrowsing-Session. Wie auch reale Nutzer_innen nicht nur eine Seite aufrufen, surft TrackTrack automatisiert mehrere Seiten an und speichert dabei Informationen

über den Netzwerkverkehr, die ausgeführten Skripte, kontaktierte Server und abschließend das durch Google generierte Profil. Um Nutzer_innenverhalten zu simulieren, wurden nicht beliebige Seiten in einer Session aufgerufen, sondern Link-Empfehlungen von Nutzer_innen der Plattform reddit.com gesammelt (vgl. 4.3.1). Diese wurden in einer Session angesurft, dabei Daten über die beobachteten Tracker gesammelt und im Anschluss die von Google zugewiesenen Interessen zu diesem Profil erhoben (vgl. 4.3.2). Abbildung 21 veranschaulicht dieses Vorgehen. Das in Punkt zwei beschriebene Verfahren zu seitenbezogenen Profilen wurde auf die Dienste Alexa (4.3.4) und Quantcast (4.3.3), die im vorangegangenen Kapitel bereits erwähnt wurden, sowie den Dienstleister Compete (4.3.5) angewendet. Zusätzlich wurde im weiteren Verlauf zur Verschleierung eine zufällige Liste von Webseiten benutzt, die vom OpenDirectoryProject bezogen wurde (4.3.6).

Technische Rahmenbedingungen der Erhebung

Die Daten wurden in einem Zeitraum von 12 Monate zwischen März 2014 und Februar 2015 erhobenen. Dazu wurden zwischenzeitlich vier (virtualisierte) Server eingesetzt, die über IP-Adressen der Ruhr-Universität zugeordnet werden konnten. Alle Server wurden mit der Linux Distribution *Debian Wheezy* betrieben und führten kontinuierlich die überwiegend in JavaScript geschriebenen Skripte aus. Die Daten wurden in einer zentralen NoSQL-Datenbank (MongoDB) abgelegt, die das JSON Format zur Datenbeschreibung verwendet.

Bei der Entwicklung der Skripte wurde auf eine Reihe bestehender Bibliotheken zurückgegriffen, die im Anhang gelistet sind. Die Verwaltung der Anforderungen sowie des Codes wurde mit dem Versionskontrollsystem *git* sowie dem darauf basierenden Managementsystem *gitlab* vorgenommen. Alle entwickelten Skripte – exklusive der Daten – werden unter der GPL Version 3 Lizenz veröffentlicht.

Name	Datenarten	Ziel
Reddit	Listen mit 100 URLs von Profilseiten von reddit.com	generieren eines Sets von User/URL/Profil Kombinationen
User_innen Simulation	Netzwerkverkehr, Cookies, Screenshots	Umfangs des Trackings sowie einer Liste von Interessen auf der AdSettings-Seite
Google Interessenprofil	Liste von Interessen, die auf der Google-Interessenseite	extraktion der Interessen von der AdSettings-Seite
Quantcast	Liste von soziodemografischen Daten zu einer Domain	zur Aggregation über ein Profil
Compete	Liste von soziodemografischen Daten zu einer Domain	zur Aggregation über ein Profil
Alexa	Liste von soziodemografischen Daten und Kategorien	zur Aggregation über ein Profil
ODP Index	Liste von URLs	Input für die Verschleierung

Tabelle 6: Übersicht über die erhobenen Datensätze im Rahmen von TrackTrack

4.3.1 Datenquelle Reddit

Reddit ist eine Online-Plattform, auf der Nutzer_innen Links zu Webseiten in verschiedenen Kategorien (subreddits) veröffentlichen können. Die Nutzer_innen haben dann die Möglichkeit, die Links zu bewerten und zu diskutieren. Reddit arbeitet mit dem Prinzip des *collaborative filtering,* bei dem eine Community Wissen (hier in Form von Links) gemeinschaftlich organisiert. Reddit ist außerdem vollständig öffentlich einsehbar, auch wenn nur registrierte Nutzer_innen aktiv mitwirken können.

Die Nutzer_innen-Basis von Reddit gilt als sehr internetaffin und stammt in großen Teilen aus den USA.[102] Die Verkehrssprache in dem Netzwerk ist Englisch.

102 Es existiert keine wissenschaftliche Untersuchung zur Nutzer_innen-Basis, nur eine Befragung der Nutzer_innen aus dem Jahr 2011 vgl. HTTP://WWW.DAILYDOT.COM/SOCIETY/REDDIT-SURVEY-DEMOGRAPHICS/ (letzter Zugriff 26.09.2016).

Name	Umfang
Reddit_crawl	10000 User extrahiert, 800 nutzbar.
Browse	~ 30 GB Netzwerkdaten
GI Extract	Daten von 506 Nutzer_innen, die Links zu 7123 Domains enthalten
QuantcastInfos	für 4423 Domains
CompeteInfos	für 3178 Domains
AlexaInfos	für 1100 Domains
ODP Index	enthält Links zu ca. 4000 Seiten

Tabelle 7: Umfang der erhobenen Datenmengen durch TrackTrack

Erhebung

Reddit erlaubt Entwickler_innen über eine Schnittstelle[103] Zugriff auf bestimmte gespeicherte Daten. Genutzt wurde eine API, die das Profil eines_r zufälligen Nutzer_in übermittelt hat. Dieses Profil enthält die 100 zuletzt geposteten Links. Die abgerufenen Profile enthalten neben der Zuordnung der URLs zu einem User auch eine Kategorisierung auf Basis des Reddit-eigenen Kategorien-Systems, der *subreddits*. Subreddits sind moderierte Unterseiten innerhalb der Plattform zu bestimmten Themen (z. B. „Technology“ oder „History“, aber auch „What I learned today“).

Die Profile wurden mittels eines Skripts gecrawlt. Die API erlaubt allerdings keine strukturierte Analyse der kompletten Reddit-Daten. Stattdessen wurde über die Schnittstelle *random*, die alle 3 Minuten einen anderen, zufällig ausgewählten User_innennamen aus dem Datenbestand zurückliefert, der_die jeweilige Urheber_in ermittelt und über eine weitere Schnittstelle das vollständige Profil dieses_r User_in abgeruft. Das vollständige Profil wurde anschließend auf die Einhaltung der Anforderungen (s. u.) überprüft bevor, nach Ablauf von 3 Minuten wurde der Prozess automatisiert erneut gestartet.

103 Siehe HTTP://WWW.REDDIT.COM/API/ (letzter Zugriff 26.09.2016).

Datenaufbereitung und -auswahl

Finden geeigneter Reddit User_innen

Nur ein geringer Teil der Nutzer_innen von Reddit veröffentlicht URLs in einer Form, die für die Simulation einer Browsing-Session geeignet sind. Ein Großteil der Verweise, die User_innen bereitstellen, bezieht sich auf Reddit selbst oder eine Webseite, die eng mit Internetkultur verknüpft ist, wie der Bildhoster Imgur oder die Videoplattform Youtube. Ein Teil der Nutzer_innen verwendet Reddit allerdings auch, um für sie_ihn persönlich relevante Links zu sammeln. Eben diese Nutzer_innen sind für die vorliegende Studie interessant, da angenommen werden kann, dass diese Sammlung von Links ähnlich wie ein Browserverlauf behandelt werden kann.

Um diese Nutzergruppe und deren Linklisten zu sammeln wurden nur diejenigen Listen gespeichert, die eine gewisse Diversität aufwiesen. Es wurden nur solche Listen einbezogen, die mindestens 60 Links zu 20 verschiedene Domains enthielten. Von den insgesamt bis zu 100 Links pro Nutzer durften außerdem maximal 40 die folgenden Kriterien erfüllen:

- Verweise auf Dateien mit den Endungen *gif, pdf, jpg, jpeg*.
- Verweise auf die weitverbreiteten Bild- und Videoplattformen *imgur, youtu, me.me, meme*.
- Verweis auf Reddit-eigene Unterseiten, erkennbar an den Textelementen *reddit* und *self*.

Die Profildaten wurden darüber hinaus nicht weiter geändert, beim Import in die Datenbank wurden allerdings zusätzlich der Tag der Nutzung im Crawl-Vorgang und ein Parameter zur Art des Crawl-Vorgangs hinzugefügt.

Insgesamt wurden die Profile von 36980 Reddit Nutzer_innen überprüft. Davon erfüllten 9339 die Anforderungen. Diese URL-Profile dienen im Folgenden als Ausgangsbasis für die Erhebung der Interessenprofile und der Analyse des Umfangs von Tracking.

Einschränkungen

Reddit als Plattform und die Link-Profile haben sich während der Nutzung als gute Datenbasis herausgestellt. Nichtsdestotrotz ist es notwendig, die Annahme, dass es möglich ist, Reddit-Linkprofile wie Browserverläufe zu behandeln, zu problematisieren. Wie sich bei der Auswertung der Profile zeigt (siehe 4.5.1), entspricht die Art der Links, die bei Reddit gepostet werden, nicht dem gängigen Nutzungsprofil, etwa weil Internetstatistiken zeigen, dass im Verhältnis wesentlich mehr Seitenaufrufe in Social Networks passieren, als es die Reddit-Profile suggerieren. Allerdings ist dem Autor

trotz intensiver Recherche keine vergleichbare Datenbasis bekannt, die eine ähnliche Menge an Zuordnungen zwischen User_innen und Links zur Verfügung stellt. Die Nachteile der Reddit-Daten müssen daher bei der Auswertung berücksichtigt werden.

4.3.2 Datenquelle Google

Wie bereits beschrieben, wurden die bei Reddit erhobenen Link-Profile genutzt, um Nutzer_innen zu simulieren. Am Ende jeder User-Session, also dem Aufruf von ca. 100 URLs, wurde die in 3.4.1. beschriebene Informationsseite von Google aufgerufen, auf der zum Zeitpunkt der Studie auch nicht eingeloggten Nutzer_innen angezeigt wurde, welche Interessenkategorien für sie ermittelt wurden (siehe 3.5.2). Diese Interessen wurden durch *scraping* extrahiert.

Google Interest Categories

> /Mensch und Gesellschaft > Familie und Beziehungen >Familie > Eltern > Babys und Kleinkinder > Spielzeug für Babys und Kleinkinder
>
> /Haus und Garten > Küche und Esszimmer > Küchengroßgeräte > Geschirrspüler

Quelltext 1: Beispiele für eine Interessenhierarchie bei Google.

Der Nachteil der Darstellung auf der Informationsseite von Google ist, dass die Interessen ohne Kontext dargestellt werden. Über die Hilfe-Funktion kann man allerdings in Erfahrung bringen, dass Google eine Kategorisierung von Interessen[104] entwickelt hat, bei der 2500 Interessen in einer Baumstruktur organisiert sind. Quelltext 1 zeigt beispielhaft die Kategorienfolge des Interesses „Spielzeug für Babys und Kleinkinder“, das der Hauptkategorie „Menschen und Gesellschaft“ zugeordnet ist.

Etwas weniger als die Hälfte (867) der Interessen sind der Kategorie „Orte der Welt“ zugeordnet und beziehen sich auf geografische Regionen und Länder. Die übrigen 1095 Interessen sind in 24 Oberkategorien gegliedert, die im Folgenden *Google Interest Categories* (GIC) genannt werden. Tabelle 8 listet die Oberkategorien sowie die Anzahl der Unterkategorien.

104 Die vollständige Liste ist einsehbar unter HTTPS://SUPPORT.GOOGLE.COM/ADS/ANSWER/2842480 (letzter Zugriff 05.02.2015)

Interessenkategorie (Anzahl der Unterkategorie)	
Arts & Entertainment (147)	Travel (27)
News (21)	Autos & Vehicles (95)
Games (42)	Food & Drink (73)
Law & Government (36)	Beauty & Fitness (21)
Finance (50)	Jobs & Education (36)
Computers & Electronics (128)	Reference (30)
Internet & Telecom (34)	Online Communities (18)
Sports (69)	Pets & Animals (15)
Business & Industrial (121)	Books & Literature (9)
People & Society (40)	Home & Garden (48)
Science (25)	Hobbies & Leisure (30)
Shopping (71)	Real Estate (9)

Tabelle 8: Verteilung der Google Interest Categories (GIC)

Datenerhebung

Neben der Zuordnung von Interessen zu User_innen, die wie in Abbildung 21 dargestellt erhoben wurden, ist auch der Kategorien-Baum selbst in der Datenbank abgelegt, um die Hierarchie der Interessen automatisiert verarbeiten und so beliebige Interessen auf ihre Oberkategorien (GIC) zurückführen zu können.

Zusätzlich zu der User-zentrierten Erhebung wurde in weiteren Tests untersucht, in welchem Umfang Interessen direkt einer URL zuzuordnen sind, bzw. durch Google zugeordnet werden. Dazu wurden 10135 URLs aus den vorliegenden Reddit-Daten direkt einzeln angesurft und im Anschluss die Informationsseite aufgerufen. Wenn diesem, nur auf dem Besuch einer Seite basierendem, Profil ein Interesse zugewiesen wurde, wurde dies als Domain/Interesse-Zuordnung gespeichert.

Datenaufbereitung

Maß für die Breite eines Interessenprofils

Die Anzeigeneinstellungsseite von Google listet zugewiesene Interessen ohne weitere Informationen, z. B. der Gewichtung, die ausdrücken könnte, ob die Zuweisung des einen Interesses wahrscheinlicher ist als die eines anderen. Interessen unterschiedlicher Ebenen der Kategorien-Hierarchie werden zudem gleichwertig nebeneinander dargestellt. Wenn etwa das Interesse „Spielzeug für Babys und Kleinkinder“ angezeigt

wird, erscheint nicht automatisch auch „Menschen und Gesellschaft“ in der Liste. Gleichzeitig kann in anderen Fällen das Interesse „Familie“ in der Liste stehen, das in der Mitte der Hierarchie steht (vgl. Quelltext 1). Um die Hierarchie der Interessen trotzdem in der Auswertung berücksichtigen zu können, wurden die Interessen für die weitere Auswertung immer auf das jeweilige Basisinteresse (GIC) bezogen. Alle Interessen unterhalb von „Mensch und Gesellschaft“ („Familie“, „Eltern“ etc.) wurden dafür immer auf das Interesse an „Mensch und Gesellschaft“ bezogen. So kann ein etwas gröberes Interessenprofil (IP) für den User (U) definiert werden als Tupel $IP_U = (x_1, \ldots, x_n)$. Wobei $x_i = 1$ ist, wenn einer Subkategorie des GIC *i* zugewiesen wurde, und $x = 0$, wenn keine Subkategorie des GIC *i* zugewiesen wurden. Zum Vergleich verschiedener Interessenprofile kann dann ein Wert für die Breite eines Interessenprofils B als die Anzahl der zugewiesenen GIC herangezogen werden (vgl. Formel 1).

$$B_U = \sum IP_U$$

Formel 1: Berechnung der Breite eines Profils

Einschränkungen

Auffällig ist, dass die Kategorien nicht eine wertneutrale Auflistung aller vorstellbaren Interessen sind, sondern sich an dem Ziel der Kategorisierung orientieren, nämlich Personen Zielgruppen zuzuordnen. Diese wiederum orientieren sich an dem, was angeboten werden soll und worauf Kund_innen angesprochen werden sollen. Dementsprechend finden sich unter den Interessen vor allen Dingen solche, die sich in einem ausdifferenzierten Markt befinden. Während die Interessen an unterschiedlichen Automarken und damit zusammenhängenden Themen relativ breit gefächert gelistet werden (95 Interessen), finden sich andere Themen gar nicht wieder. So existieren für Personen, die sich für Computerspiele interessieren, 14 Unterkategorien, nur halb so viele Unterkategorien finden sich dagegen in der Kategorie *„/Mensch und Gesellschaft/Soziale Fragen und Fürsprecher“*. Eine Gewichtung zwischen den Interessen, auch wenn diese in einer Baumstruktur auf sehr unterschiedlichen Ebenen angesiedelt sein können, ist nicht ersichtlich.

Auskunft nicht vollständig

Nach einer Studie Datta u. a. (2015) kann allerdings nicht angenommen werden, dass dieses Interessenprofil vollständig dem entspricht, was auch bei Google intern verwendet wird, um passende Werbeanzeigen zu ermitteln. Datta und Tschantz haben nachgewiesen, dass die Liste weder vollständig ist, noch die Einflussnahme (das Löschen) durch den_die Nutzer_in immer einen Einfluss auf die angezeigte Werbung hat.

Blackbox Google

Darüber hinaus lässt sich über die Blackbox-Methode nicht ermitteln, auf Basis welcher Informationen Google die Interessen zuweist. Es ist bekannt, dass Google für seine Suche Inhaltsanalysen von Webseiten durchführt, und man kann sicherlich davon

ausgehen, dass diese Inhaltsanalysen zur Kategorisierung und Zuweisung von Interessen beitragen. Wie genau diese Zuweisung berechnet wird, ist allerdings Googles Geschäftsgeheimnis. Nicht ermittelt wurde außerdem, inwieweit die Erhebungsmethode, bei der Seiten nur aufgerufen werden, aber nicht mit Elementen darauf interagiert wird, gegebenenfalls das Ergebnis verändert. Die Tatsache, dass auch ohne Interaktion Interessen zugewiesen wurden, deutet darauf hin, dass die Interaktion keine notwendige Vorbedingung für das Profiling durch Google ist.

4.3.3 Datenquelle Quantcast

Quantcast ist Anbieter von *Audience Analytics*. Webseitenbetreiber_innen können die Tracking-Methoden von Quantcast nutzen, um Einsicht in die Nutzungsprofile ihrer Besucher_innen zu erhalten (siehe. 3.5.2). Quantcast veröffentlicht auf seinen Seiten detaillierte Statistiken über die Nutzer_innen von Webseiten (siehe auch 3.4.1). Es handelt sich um eine Übersicht von Zuordnungen zu bestimmten Kategorien im Verhältnis (siehe Tabelle 9) zum amerikanischen Durchschnitt von Internetnutzer_innen (vgl. 16).

Die Quantcast-Daten werden häufig im Marketing genutzt, um Webseiten zu identifizieren, die bestimmte Zielgruppen bedienen. Für die vorliegende Studie können die Daten genutzt werden, um aus dem Browserverlauf soziodemografische Eigenschaften zu schlussfolgern.

Attribute	Kategorien
Gender	männlich, weiblich
Alter	<18, 18-24, 25-34, 35-44, 45-54, 55-64, 65+
Einkommen	0-50.000, 50-100.000, 100-150.000, mehr als 150.000
Ethnizität	Other, Hispanic, African America, Caucasian, Asian
Ausbildung	Grad School, College, No College
Kinder	hat Kinder, Keine Kinder
Politische Orientierung	Demokraten, Republikaner, Unabhängig
Politisches Engagement	aktiv, teilweise aktiv, inaktiv

Tabelle 9: Übersicht über die Attribute und Kategorien, in die Quantcast Nutzer_innen einordnet. Die letzten beiden Kategorien wurden durch Quantcast erst nach Ende des Tests hinzugefügt und nicht ausgewertet.

Datenerhebung und Anpassung

Die Daten wurden mittels eines *Scrapers* von Quantcast extrahiert. Dazu wurden jeweils im Abstand von 10 Sekunden mit dem NodeJS Modul JSDOM die einzelnen Quantcastseiten[105] zu allen 7123 Domains aufgerufen, die im Reddit-Datensatz enthalten sind. Für 3430 dieser Domains hielt Quantcast Daten vor, die aus der HTML Struktur entnommen und im unten beschriebenen Format (siehe Quelltext 2) in der Datenbank gespeichert wurden. Die Daten wurden dabei in ein Format überführt, das sich für die weitere Analyse besser eignet. Um später Änderungen beobachten zu können, wurde im Feld „date“ der Zeitpunkt des Abrufs gespeichert. Die Felder „keywords“, „categories“ und „userstats.location“ sind spezifisch für die Daten von Alexa und blieben in dem Fall leer.

105 Informationen zu einer Domain sind jeweils abrufbar unter der Adresse HTTPS://WWW.QUANTCAST.COM/$DOMAIN, wobei „$DOMAIN“ für die jeweilige Toplevel-Domain steht.

```
1.      "src": [„Quantcast“, „Alexa“, „Compete“]
2.      "date": new Date().getTime() // Erhebungszeitpunkt
3.      "domain": String // Domain der Seite der die
Informationen zugeordnet sind
4.      "keywords":  Array //Stichworte der Seite (optional)
5.      "categories": Array //Kategorien (optional)
6.      "userstats": // soziodemografische Statistiken
7.              "gender": [] // m|f
8.              "age":          [] // versch. Kategorien
9.              "children":     [] // yes|no
10.             "income": [] // versch. Kategorien
11.             "education":    [] // versch. Kategorien
12.             "ethnicity":    [] // versch. Kategorien
13.             "location":[] // versch. Kategorien
```

Quelltext 2: Datenformat für alle Audience Analytics Dienste.

EINSCHRÄNKUNGEN

Ungenauigkeit der Daten

Die Datenerhebung fand über einen längeren Zeitraum statt, in dem sich die von Quantcast bereitgestellten Informationen über Zusammensetzung der *Audience* einer Seite ändern können. Aus Performance-Gründen wurden die Daten allerdings nur einmal erhoben und mit diesen wurde dann weitergearbeitet. Um die darauf aufbauenden Dienste in den Regelbetrieb zu nehmen, müssten die Daten möglichst häufig neu erhoben werden. Unabhängig davon hat bereits Kamerer (2013) gezeigt, dass die Zahlen, die Quantcast für Seiten ermittelt, die nicht QuantcastKund_innen sind, eine hohe Ungenauigkeit aufweisen. Im Folgenden wird dennoch mit den Daten weitergearbeitet, erstens da wir sie mit denen weiterer Anbieter_innen aufbereiten, und zweitens da auch Tracking-Provider beim Erstellen von Profilen mit diesen Ungenauigkeiten arbeiten.

Attribute	Kategorien
Gender	male, female
Ausbildung	Grad School, College, Some, College, No College
Location	Home, School, Work
Kategorien	eine Alexa-Eigene Kategorisierung
Keywords	Suchbegriffe, mit denen Nutzer_innen auf die Seite aufmerksam geworden sind

Tabelle 10: Soziodemografische Daten, die Alexa für eine Reihe von Domains bereit stellt.

4.3.4 Datenquelle Alexa

Wie in Abschnitt 3.4.1 beschrieben, vertreibt Alexa Informationen über die Reichweite (in Form von Besuchszahlen), die Herkunft sowie die als *Engagement* bezeichnete Anzahl wiederkehrender Besucher_innen von Webseiten. Darüber hinaus bietet Alexa, ebenso wie Quantcast, für eine Reihe von Webseiten Abschätzungen über soziodemografischen Eigenschaften der Besucher_innen an (siehe Tabelle 10).

Datenerhebung und Anpassung

Im Rahmen der Datenerhebung wurden vor allen Dingen die zuletzt genannten Informationen automatisiert abgerufen. Informationen über die Auslastung einer Seite sowie themenverwandte Seiten und Suchbegriffe wurden nicht verwendet. Die Daten wurden durch den *Scraper* den HTML Seiten entnommen und in das einheitliche JSON Format überführt.

Einschränkungen

Geringe Datenqualität

Kamerer (2013) stellt für die Metriken bei Alexa noch größere Schwankungen fest als bei Quantcast. Zudem formuliert der Autor Kritik an der Erhebungsmethode der soziodemografischen Daten über die Alexa-Toolbar. Da diese vor allen Dingen bei Webseitenbetreiber_innen und Dienstleister_innen beliebt ist, die Suchmaschinenoptimierung betreiben, müsse man von einer erheblichen Verzerrung in den Daten ausgehen. Darüber hinaus ist auch inhaltlich nicht klar, welchen Wert die Information über die „Location“ eines_r Besucher_in hat, die einmalig bei der Installation der Toolbar erhoben wird. Mit der Verbreitung von Laptops steigt die Anzahl der Nutzer_innen, die mit

derselben Voreinstellung auch an anderen Orten ihren Browser mit der Toolbar nutzen und damit die Ergebnisse verfälschen.

4.3.5 Datenquelle Compete

Compete.com ist Teil der Unternehmensgruppe *Millward Brown*, der weltweit zweitgrößten Marktforschungsorganisation, und verfügt nach eigenen Angabe über das größte US-amerikanische Panel für online Verhaltensanalyse. Dabei kombiniert und normalisiert das Unternehmen Daten aus einem Panel von 350.000 Nutzer_innen (s. u.) sowie verschiedenen weiteren Quellen[106]. Darauf aufbauend wird unter anderem eine Programmierschnittstelle angeboten, über die Entwickler_innen tagesaktuelle soziodemographische Zielgruppenprofile für Webseiten abrufen können.

Attribute	**Kategorien**
Gender	male, female
Income	60-100.000, under-30k, 30-60k, 100k-plus
Age	<18, 18-24, 25-34, 35-44, 45-54, 55-64, 65+

Tabelle 11: Soziodemografische Daten, die Compete anbietet.

Tabelle 11 zeigt die verfügbaren Datenpunkte pro Domain. Die Werte werden, anders als bei den übrigen Anbieter_innen, nicht im Verhältnis zu einem amerikanischen Mittel, sondern in Prozent angegeben.

Datenerhebung und Anpassung

Da die API von Compete nur einen eingeschränkten Zugriff erlaubt, wurden die Informationen für die 7123 in dem Reddit-Datensatz enthaltenen Domains über einen Zeitraum von 142 Tagen (bei 50 Anfragen am Tag) abgerufen. Von den 7123 Domains hielt Compete für 3178 Daten vor.

Einschränkung

Unbekannte Datenqulität

Auch für die von Compete bereitgestellten Schätzungen über Besuchszahlen stellte Kamerer (2013) eine hohe Varianz und Abweichung von direkt gemessenen Werten fest. Über die Genauigkeit der soziodemografischen Daten ist nicht viel bekannt. Allerdings ist Compete der einzige Anbieter, der genauer auf seine Datenquellen eingeht. Nach Kamerer bezieht Compete seine Daten über ein Panel von 350.000 Nut-

106 Vergleiche die Angabe auf HTTPS://WWW.COMPETE.COM/ABOUT-COMPETE/OUR-DATA/ und in Roskill (2010) (zuletzt abgerufen am 26.01.2015).

zer_innen, die sich über eine Webseite registrieren, im Anschluss eine Software zum Tracking ihres Surfverhaltens installieren und in regelmäßigen Abständen an zusätzlichen Befragungen teilnehmen.[107]

4.3.6 Datenquelle Open Directory Project

Das Open Directory Project[108] (ODP) ist eine von einer Community gepflegte öffentliche Datenbank in der Links zu Webseiten gesammelt werden. Das Projekt ist das nach eigenen Angaben größte Projekt seiner Art. Das ODP stellt sein Verzeichnis, das Webseitenadressen mit Kategorisierungen enthält, frei zur Verfügung. Die Daten des ODP werden im weiteren Verlauf genutzt, um zufällige Adressen zu generieren. Von Nachteil für diesen Zweck ist es, dass im ODP auch defekte Links manuell überprüft werden, so dass die Datenbank einige Links enthält, die auf nicht mehr existente Webseiten weisen.

4.4 ANALYSE DES TRACKINGS

4.5 Ziel einer ersten Auswertung der gesammelten Daten ist es, etwas über die Funktionsweise und den Umfang von Online-Profiling zu lernen. Die gewonnenen Informationen sollen dann im Weiteren dazu genutzt werden,, mit TrickTrack ein Werkzeug zu entwickeln, um Transparenz und Interventionsmöglichkeiten herzustellen. Im Folgenden werden einige der Erkenntnisse, bezogenen auf den Gesamtdatensatz und die als User-Profile zusammengefassten Teildatensätze, betrachtet.

4.5.1 Durchschnitt der Nutzer_innen

Wie in Abschnitt 4.2 beschrieben, wurden die *Linkprofile* von 506 Reddit-User_innen ausgewertet und automatisiert angesurft. Die Linklisten enthielten im Durchschnitt 96 URLs bei einer Standardabweichung (s) von 13,72 und verlinkten zu 44 unterschiedlichen Domains (s=14,85). Zudem machten die 10 meist verlinkten Seiten eines Profils 59,6 % (s=15,7) der Gesamtmenge der Links aus. Diese Diversität in den Linklisten ist von Vorteil, da so das entstehende Profil nicht zu sehr von einzelnen Webseiten abhängig ist. Diese Diversität entspricht auch in etwa dem Maß, das in der Abschlussevaluation bei tatsächlichen Browserverläufen beobachtet wurde und kann daher als nutzbar für die weitere Analyse angesehen werden.

---

107 Vgl. auch HTTPS://WWW.CONSUMERINPUT.COM/FAQ/ (letzter Zugriff 26.09.2016). Das Registrierungsformular für das Panel ist allerdings seit 2010 nicht mehr freigeschaltet. Zuletzt war eine Registrierung möglich im Juli 2010 (siehe HTTPS://WEB.ARCHIVE.ORG/WEB/20100706214117/HTTP://CONSUMERINPUT.COM/PANEL/REGISTER; letzter Zugriff 26.09.2016). Das dort dargestellte Formular entspricht bei den Angaben zur Ethnizität nicht den von Compete bereitgestellten Daten.

108 HTTP://DMOZ.ORG/ (letzter Zugriff 26.09.2016).

4.5.2 Aufgerufene Webseiten

Über alle Linkprofile aggregiert enthält der Datensatz 45829 Links zu 7123 Domains, wobei die Anzahl der Links zu unterschiedlichen Seiten stark variiert. Die 100 meist verlinkten Webseiten machten ca. 49 % der gesamten Links aus. Tabelle 12 zeigt die 12 meist verlinkten Domains und zum Vergleich den „Alexa Rank“, eine verbreitete Metrik, die die meist besuchten Webseiten weltweit enthält, siehe zum Vergleich Tabelle 13.

Aufgerufene Webseiten nicht repräsentativ

Der Vergleich zeigt, dass die Daten keinesfalls repräsentativ für alle Internetnutzer_innen sind, was mit der vorherigen Einschätzung über die Daten der Plattform Reddit übereinstimmt. Weltweit sind die meistbesuchten Webseiten in den Kategorien *Online-Netzwerk* (Facebook.com, Twitter.com), *Suche* (Google.com/ co.in, Live.com, Baidu.com, Yahoo.com) und *Online-Handel* (Taobao.com und Amazon.com). Der Vorteil des vorliegenden Datensatzes für die Profil-Analyse liegt darin, dass die Links weniger auf die geschlossenen Netzwerke oder Suchseiten verweisen, sondern auf Webseiten mit redaktionell bereitgestellten Inhalten wie Nachrichtenportalen. Diese Art von Seiten, auf denen *Publisher* etwas veröffentlichen und Einnahmen vor allem durch Werbung generieren, setzen wesentlich häufiger 3rd-Party-Tracking ein, als die in der Alexa-Toplist vorne liegenden Seiten. Anbieter_innen von Social Networks und Suchmaschinen hingegen – insbesondere die mit großer Reichweite – sind häufig selbst Betreiber_innen von Tracking und Werbenetzwerken. Dies trifft mindestens auf Google (inklusive Youtube), Facebook, Yahoo und Amazon zu. Zusätzlich stellt die Verlinkung auf *Publisher*-Seiten sicher, dass weniger Links auf geschlossene Webseiten verweisen. Während Einträge, zum Beispiel bei Facebook, teilweise nur für Nutzer_innen zugänglich sind, die im Netzwerk eingeloggt sind oder auch durch den_die Urheber_in gelöscht werden können, sind Artikel auf Nachrichtenseiten häufig länger online und werden weniger häufig verändert oder gelöscht.

Nr.	Domain	Nr.	Domain	Nr.	Domain	Nr.	Domain
1	Google.com	4	Yahoo.com	7	Wikipedia.org	10	Qq.com
2	Facebook.com	5	Baidu.com	8	Taobao.com	11	Google.co.in
3	Youtube.com	6	Amazon.com	9	Twitter.com	12	Live.com

Tabelle 12: Meist besuchte Webseiten nach Alexa.com.

Nr.	Domain	# Links	Alexa Rank
1	imgur.com	3173	49
2	youtube.com	2725	3
3	theguardian.com	1033	134
4	nytimes.com	854	115
5	reuters.com	686	297
6	bbc.co.uk	659	62
7	washingtonpost.com	587	289
8	huffingtonpost.com	554	68
9	en.wikipedia.org	480	6
10	news.yahoo.com	376	4
11	flickr.com	372	107
12	reddit.com	372	50

Tabelle 13: Meist verlinkte Webseiten und zum Vergleich der Alexa Rank.

4.5.3 HTTP-Requests

Wie stark die im Testset verlinkten Seiten mit den Werbenetzwerken verbunden sind, zeigt sich anhand der Analyse der Verbindungsdaten in Form von HTTP-Requests.

Insgesamt wurden 5.842.745 HTTP-Requests an unterschiedliche Domains gesendet. Dabei konnte beobachtet werden, dass ca. 20 % der Anfragen, die beim Aufbau einer Webseite ausgeführt werden, an Dritt-Server gingen. Diese Dritt-Server waren dabei nicht nur Werbenetzwerke und User-Tracking-Services, sondern auch Bilder, Videos oder sogenannte Social Plugins, die eine Verbindung zu sozialen Netzwerken (wie der Facebook Like Button) herstellen und ebenfalls zum Tracking eingesetzt werden.

50% aller Anfragen gehen an 50 Domains

Insgesamt geht ein Großteil der Anfragen an eine kleine Menge von Diensten. Etwa die Hälfte (49,28 %) der Requests ruft Inhalte von 50 Domains ab. Von diesen 50 Domains stehen 31 direkt mit Werbenetzwerken in Verbindung. Darunter, ähnlich wie in anderen Studien[109], die Werbe- und Analyse-Dienste von *Google*, die Social Plugins (Share- und Like-Buttons) von *Facebook* und *Twitter* sowie die Diskussionsplattform *Disqus*, aber auch die Marktforschungs- und -analysedienste *Quantcast*, *Scorecardresearch* und *Outbrain* sowie die Werbenetzwerke *Taboola* und *Adnexus*. Die ebenfalls

109 Siehe u. a. (Acar u. a. 2014; Gomez, Pinnick, und Soltani 2009; Mayer und Mitchell 2012).

häufig kontaktierte Domain *Betrad.com*[110] gehört allerdings nicht zu den Werbetreibenden, sondern zu Evidon, dem Hersteller des AdBlocker *Ghostery* (siehe 3.4.3), die über diese Domain das Einbinden des „AdChoices“ Icons ermöglicht. Aus der Statistik der Gesamtabfragen geht außerdem hervor, dass insgesamt 16,23 % aller Anfragen an Server von Google[111] gerichtet waren.

4.5.4 Cookies

Mehr als 600 Cookie pro User_in

Wie in 3.3 erläutert, sind HTTP-Cookies elementarer Bestandteil der Tracking-Infrastruktur. Im Durchschnitt waren nach den Durchläufen nach 96 angesurften Seiten pro User 645 Cookies vorhanden. Kontinuierliche Untersuchungen, die sich auf den Umfang des Cookie-Trackings spezialisieren, haben ermittelt, dass in etwa 17 Cookies[112] pro Domain gespeichert werden. Bei durchschnittlich 44 besuchten Domains pro User ist die vorliegende Untersuchung daher nur knapp unter dem erwarteten Wert von 745 Cookies. 22 zeigt die Verteilung von Cookies pro Linkprofil. Die dargestellte Anzahl an Cookies bezieht sich auf die am Ende einer Session vorhandenen Cookies, nicht auf die Anzahl der Anfragen, bei denen ein Cookie mit übermittelt wurde, oder eventuell gelöschte Cookies.

Cookies Syncing

In dem Datensatz konnte außerdem der Umfang von *Cookie Syncing* (siehe 3.2), bezogen auf einzelne Sessions, überprüft werden. Dazu wurden pro Linkprofil die HTTP -Requests untersucht und die dabei übermittelten Cookies, sortiert nach ihrem Namen und dem Wert, gesammelt.

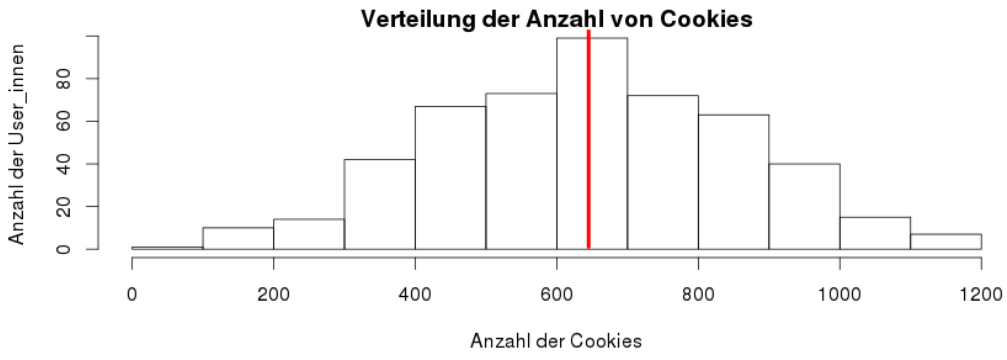


Abbildung 22: Anzahl von Cookies, die pro User gespeichert wurden; rote Linie zeigt den Durchschnitt

110 Vermutlich eine Verkürzung von „**Bet**te**r Ad**vertising“

111 Das schließt Anfragen an Youtube genauso mit ein, wie das Tracking Netzwerk Google Analytics und das Werbenetzwerk Doubleclick.

112 Siehe HTTP://COOKIEPEDIA.CO.UK/ (letzter Zugriff 26.09.2016).

Dabei zeichnet sich ab, dass vor allem Google und der SocialPlugin-Provider *AddThis* am Cookie-Syncing partizipieren. Der *AddThis*-Cookie *___atuvc* war pro Linkprofil im Durchschnitt von 30 Domains mit demselben Wert gesetzt, was eine Kooperation der Dienste voraussetzt. Ähnlich mitteilsam ist parse.ly, ein Unternehmen, das im Bereich *Audience Analytics* tätig ist, deren Dienste aber weniger weit verbreitet sind. Die ebenfalls viel genutzten Google Cookies *__ga* und *__gads* werden zwar auch zwischen Domains synchronisiert, tauchen aber auch genauso häufig mit singulären Werten auf. Im Allgemeinen bestätigen die Ergebnisse die Studie von Acar u. a. (2014), nach der Cookie-Syncing einzelnen Anbieter_innen erlaubt, bis zu 50 % der Besuche auf den 3000 bekanntesten Webseiten zu beobachten.

% des Link-profils	**Services/Domains**
>80 %	Google (alle Dienste, inkl. Fonts)
50-60%	Google (nur google-analytics.com; doubleclick.com/net)
40-50%	scorecardresearch.com (audience analytis), facebook.com, twitter.com
30-40%	quantserver.com (audience analytics; source of Quantcast.com)
10-20%	adnxs.com, taboola.com, outbrain.com, bluekai.com, disqus.-com, rubiconproject.com, addthis.com (advertisment networks)
5-10%	chartbeat.com, optimizely.com, amazon-adsystems.com, krxd.net

Tabelle 14: Umfang der beobachtbaren Webseitenaufrufe nach Anbieter_in.

4.5.5 Umfang des Trackings

Tabelle 14 zeigt eine Liste derjenigen Tracking-Dienste, deren Server sehr häufig beim Aufbau einer Seite direkt kontaktiert werden.Zudem lässt sich nachvollziehen, in welchem Umfang sie eine Browsersession beobachten können. Kombiniert man alle Google-Services, ergibt sich daraus, dass Google über 81,63 % aller aufgerufen Seiten informiert wird. Traut man den Google-eigenen Datenschutzbestimmungen, nach denen Dienste wie Google Fonts oder das Google-eigene CDN Google API kein User_innen-Tracking betreiben, liegt der Wert immer noch über 50 %.

Die Zahlen für die einzelnen Dienstleister_innen liegen bei dieser Auswertung etwas unter den Werten der anderen Studien. Die Marktdurchdringung von 81 % durch Google liegt dagegen nahe an den 88 %, die Gomez, Pinnick und Soltani (2009) ermittelt haben. Die Unterschiede lassen sich einerseits mit der unterschiedlichen Messmethode erklären, aber auch mit einem sich rasch ändernden Markt, der seit 2009 viele weitere Anbieter_innen, etwa im Bereich *Real Time Bidding,* hervorgebracht hat.

4.6 INTERESSENPROFILE

Interessen von unbekannten Nutzer_innen

Die mögliche Beobachtung von mehr als 80 % der Webseitenbesuche durch Google ist ein guter Indikator dafür, dass die generierten Interessenprofile für Googles Dienste repräsentativ sind.

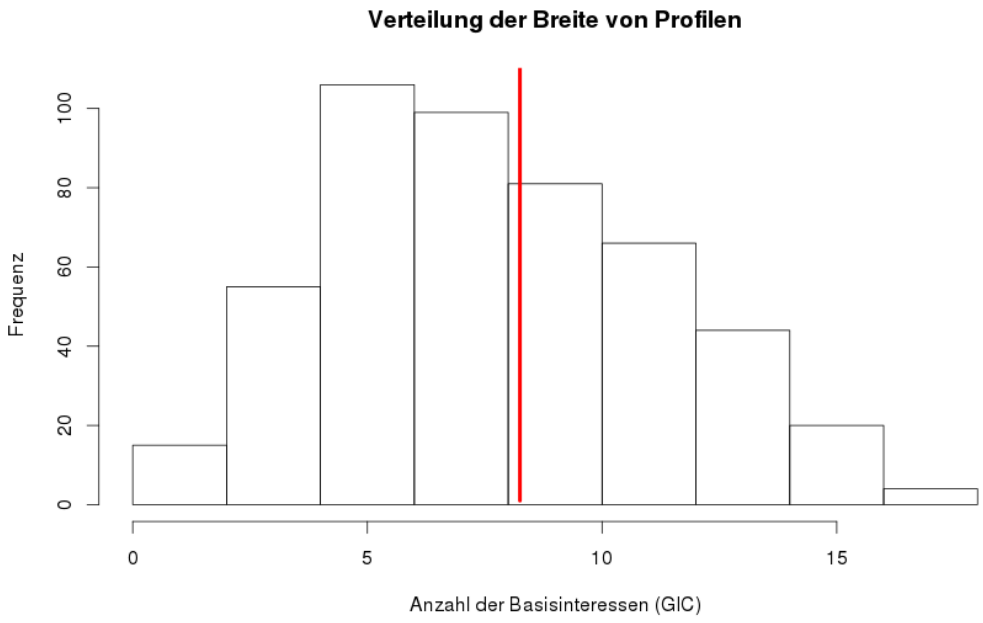


Abbildung 23: Verteilung der Anzahl von GIC

Die Auswertung des *scrapings* von der Seite Googles ergab, dass im Durchschnitt jeder_m Nutzer_in 16.34 (s=7.5) Interessen zugewiesen wurden. Die genaue Zahl lag dabei zwischen 1 und 37 der 1195 Interessen möglichen Interessen. Die Kategorisierung in Weltregionen wurde nicht beobachtet. Zum Zweck der Vergleichbarkeit wurden diese Interessen auf die Basis-Interessenkategorien (GIC) reduziert. So zeigt sich, dass jeder_m Nutzer_in im Schnitt 8,25 GIC (s=3,53) der 24 GIC zugewiesen werden. 23 zeigt die Verteilung dieser Profilbreite.

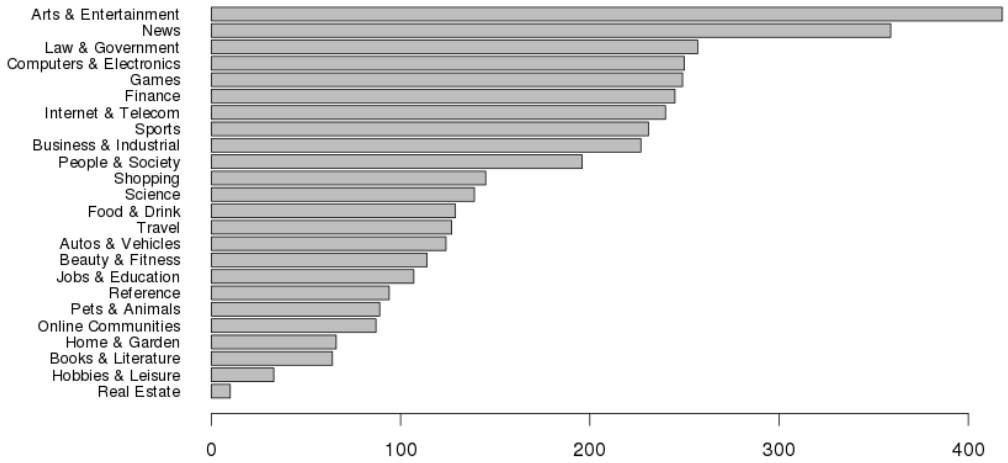


Abbildung 24: Häufigkeitsverteilung der einzelnen Interessenkategorien (GIC)

24 zeigt die Häufigkeit der einzelnen GIC, aggregiert über alle Profile. Aus der Übersicht lässt sich ablesen, zu welchen Themen bei Reddit häufig Links gepostet werden, die sich dann als Interesse im Profil niederschlagen können. Das vor allen Dingen Links zu den Kategorien „Unterhaltung“ und „News“ in den Daten enthalten sind, entspricht der Ausrichtung von Reddit auf Aktualität. Das hohe Vorkommen von Links mit Bezug zu „Computer/IT“ oder „Spielen“ entspricht darüber hinaus dem gängigen Cliché der Nutzer_innenbasis. Die Verteilung deutet aber auch darauf hin, dass auch abseits dieser Hauptthemen auf Reddit Communities existieren.

Direkte Zusammenhänge

Wie in 4.3.2 beschrieben, wurden die Interessen nicht nur in Bezug auf Linkprofile ermittelt, sondern auch direkte Zusammenhänge zwischen Domains und Interessen beobachtet. Tabelle 15 zeigt die Anzahl der Domains pro Interesse, wenn in einer Session nur eine URL besucht wurde. Zur Analyse dieser direkten Zusammenhänge wurden 101035 URLs von 6918 Domains abgerufen. Allerdings konnte bei Google hier nur für 1308 Domains direkt ein Interesse erhoben werden. Im Durchschnitt können so 4,43 GIC auf eine Domain bezogen werden. Die URL–GIC-Kombinationen lassen sich später auch als Input für die Obfuscation nutzen.

GIC	Domains	URLs	GIC	Domains	URLs
News	592	4167	Jobs & Education	137	427
Arts & Entertainment	575	3753	Shopping	122	342
Law & Government	382	2553	Travel	121	457
Business & Industrial	294	1433	Autos & Vehicles	107	305
People & Society	247	1058	Online Communities	97	352
Sports	243	1874	Pets & Animals	89	348
Computers & Electronics	242	1882	Food & Drink	88	355
Science	185	1210	Books & Literature	88	503
Games	176	814	Home & Garden	71	227
Internet & Telecom	159	1590	Beauty & Fitness	57	379
Reference	151	674	Hobbies & Leisure	52	122
Finance	148	834	Real Estate	26	45

Tabelle 15: Anzahl der Domains und URLs pro GIC, für die eine direkte Zuordnung beobachtet wurde.

4.7 SOZIODEMOGRAFISCHE PROFILE

Wie in Abschnitt 4.3 beschrieben, wurden neben den Interessenprofilen auch domainabhängige Informationen über das soziodemografische Profil der Nutzer_innen von verschiedenen Seiten erhobenen, da die Daten nicht direkt pro User_in von einem_r der Anbieter_innen bezogen werden können.

Die domainbezogenen Daten lassen sich nutzen, um in Kombination mit dem Linkprofil und durch die Aggregation der domainabhängigen Profile das Profil eines_r Nutzer_in zu berechnen. Die Annahme dabei ist, dass sich durch das Zusammenführen mehrerer Domainprofile bestimmte Eigenschaften in dem aggregierten Profil verstärken, so dass sich ein mehr oder weniger klares soziodemografisches Profil für eine_n einzelne_n Nutzer_in ergibt.

Gender	Age	Chil-dren	Income	Education	Ethnicity
Male: 49%	<18: 18%	No: 51%	0-50k: 51%	No College: 45%	Caucasian: 75%
Female: 51%	18-24: 12%	Yes: 49%	50-100: 29%	College: 41%	African America: 9%
	25-34: 17%		100-150k: 12%	Grad School: 45%	Asian: 4%
	35-44: 17%		>150k: 8%		Hispanic: 9%
	45-55: 17%				Other: 1%
	55-64: 10%				
	65+: 2%				

Tabelle 16: Vergleichswerte für die soziodemographischen Angaben über Internetnutzer_innen in den USA.

Umfang der Datensätze

Insgesamt wurden von den drei Anbieter_innen in folgendem Umfang Datensätze erhoben: Quantcast: 3430, Alexa: 1100, Compete: 3178. Die 7708 Datensätze beziehen sich dabei auf 6169 Domains. Dabei existieren für 5380 Domains Daten aus einer Quelle, für 709 aus zwei und für 80 wurden Daten bei allen drei Quellen extrahiert. Um aus diesen einzelnen Datensätzen mit ihrem jeweils spezifischen Aufbau (s. o.) Profile konstruieren zu können, müssen die Daten zusammenge- und in Wahrscheinlichkeitswerte überführt werden.

Daten Zusammenführung

Quantcast und Alexa geben Statistiken im Verhältnis zum US-amerikanischen Mittel an. Für die weitere Berechnung soll allerdings kein Vergleich mit den durchschnittlichen Internetnutzer_innen gezogen werden, sondern das wahrscheinlichste Attribut ermittelt werden. Um eine Nutzer_innen-zentrierte Wahrscheinlichkeit für die unterschiedlichen Attribute zu berechnen, ist es daher notwendig, sie in Verbindung zu setzen mit dem Mittelwert, auf den sie sich beziehen. Als Vergleichswert zum amerikanischen Mittel der Internetnutzer_innen verwendet Quantcast die Werte wie in Tabelle 16 dargestellt. Dieselbe Verteilung wird auch in Bezug zu den Alexa-Werten gesetzt.

Die Umrechnung (vgl. Formel 2) des Verhältniswertes S in eine Angabe über die Wahrscheinlichkeit P erfolgt durch die Multiplikation mit der Gesamtwahrscheinlichkeit des Auftretens des jeweiligen Attributes *Attr* in der amerikanischen Vergleichsgruppe USAVG.

$$P_{Attr} = S_{Attr} * USAVG_{Attr}$$

Formel 2: Zur Konvertierung der Verhältnis- in Prozentwerte.

Zusammenführen der Werte aus allen Quellen

Diese Werte lassen sich nun mit denen von Compete zusammenführen und mitteln. Zwei Kategorien sind allerdings in den Datensätzen unterschiedlich gestaltet. Alexa verwendet zur Unterscheidung der Ausbildung vier Attribute statt drei und Compete

andere Einkommensgruppen als Quantcast. Um für die weitere Verwendung eine größere Datenbasis nutzen zu können, wurde ein Datenmodell erstellt, das alle drei Quellen kombiniert. Tabelle 17 zeigt die entstandenen Kategorien nach der Datenzusammenführung.

Kategorie	**Quellen**	**Attribute**
Gender	Quantcast, Alexa, Compete	Male, Female
Age	Quantcast, Compete	<18, 18-24,25-34,35-44,45-54,55-64,65+
Income	Quantcast, Compete	0-50.000, 50-100.000, >100.000
Education	Quantcast, Alexa	No College, College, Grad School
Children	Quantcast	No Kids, Has Kids
Ethnicity	Quantcast	Caucasian, African America, Asian, Hispanic, Other
Location	Alexa	Home, School, Work

Tabelle 17: Kategorien des soziodemografischen Profils nach der Datenzusammenführung.

Für die Kombination der Datensätze wurden daher die Werte der Attribute „College“ und „Some College“ addiert und mit den „College“-Werten von Quantcast kombiniert. Die Einkommensklassen wurden neu aufgeteilt und ein kleinster gemeinsamer Nenner gebildet, bei dem 0–50 die CompeteWerte von 0–30 und 30-60k enthält sowie die Quantcast-Gruppen 100–150 und >150k zu >100k kombiniert.

Die so zusammengeführten Werte können dann in Kombination mit den Linkprofilen genutzt werden, um Nutzer_innenspezifische, soziodemografische Profile zu berechnen. Auf diese Weise können für 502 User-Profile soziodemografische Angaben gemacht werden. Für 291 liegen nach dieser Berechnung vollständige Datensätze mit Daten zu allen Kategorien vor. Dabei liegen im Durchschnitt zu 63,5% der besuchten Domains Informationen vor.

Aggregiert man die einzelnen Profile und zählt jeweils die Attribute, die am wahrscheinlichsten sind, ergeben sich – unter Annahme der Korrektheit der Basisdaten – für die User_innenbasis, die in Tabelle 18 dargestellten Durchschnittseigenschaften.

Gender	Age	Chil-dren	Income	Education	Ethnicity
Male: 485	<18: 0	No: 237	0-50k: 70	No College: 14	Caucasian: 147
Female: 11	18-24: 9	Yes: 0	50-100: 22	College: 256	African American: 2
	25-34: 115		100-150k: 279	Grad School: 55	Asian: 24
	35-44: 82				Hispanic: 10
	45-55: 44				Other: 0
	55-64: 6				
	65+: 3				

Tabelle 18: Anzahl der Linkprofile. die sich mehrheitlich der jeweiligen Kategorie zuordnen lassen.

Die ermittelten Profile sind demnach überwiegend von weißen Männern zwischen 25 und 34, die gut ausgebildet und gut verdienend sind. Auch diese Daten decken sich wieder mit den Eigenauskünften (vgl. Fußnote 102).

4.8 EVALUATION EIGENER PROFILING UND OBFUSCATION VERFAHREN

Nachdem die Kohärenz der gesammelten Daten betrachtet wurde, ist das Ziel, im nächsten Schritt zu klären, inwiefern die Erkenntnisse konkret eingesetzt werden können, um Transparenz und Intervenierbarkeit bei Online-Profiling zu ermöglichen. Um ersteres zu unterstützen, sollen die bisher gesammelten Daten genutzt werden, um für beliebige Nutzer_innen das jeweilige Interessenprofil und das soziodemografische Profil zu erstellen, in die diese sonst keinen Einblick haben. Möglichkeiten der Intervention schaffen dann Verfahren zur *Obfuscation*, deren Effektivität anhand der bestehenden Linkprofile in Kombination mit den bereits genutzten Methoden zum automatisierten Browsen getestet und evaluiert wird.

4.8.1 Berechnung eines Interessenprofils

TrickTrack soll für beliebige Nutzer_innen berechnen, welches Interessenprofil Google wahrscheinlich zuweisen würde, wäre das Profiling erfolgreich. Dieses kann dann einerseits zur Herstellung von Transparenz, andererseits aber auch zur Obfuscation verwendet werden, um gezielt solche Seiten anzusurfen, die dem Profil widersprechen. Dazu ist es notwendig auch für unbekannte Nutzer_innen und deren Browserverläufe ein Interessenprofil zu ermitteln, ohne dass die Seiten angesurft werden

müssen oder der_die Nutzer_in das Tracking erlaubt hat. Zwei Möglichkeiten das Interessenprofil zu generieren wurden getestet.

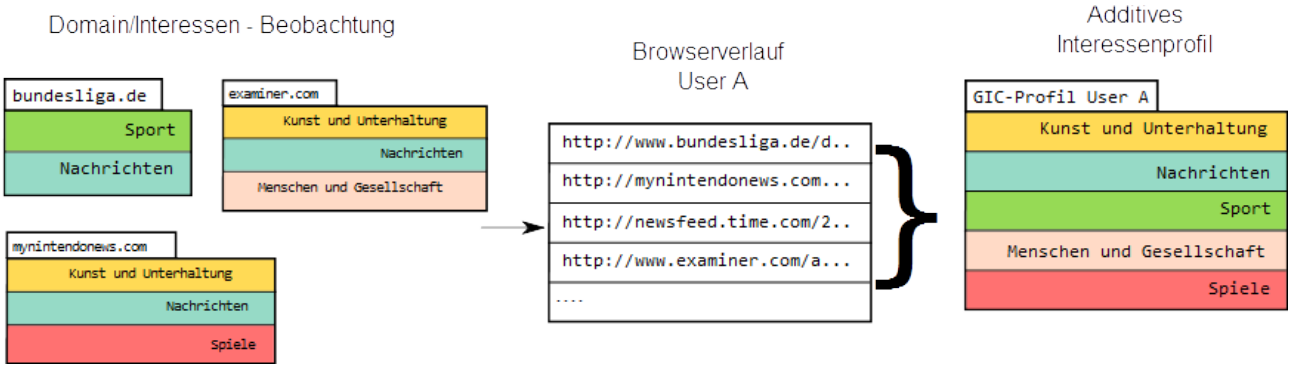


Abbildung 25: Berechnung des Interessenprofils über beobachtete Zusammenhänge.

Additionsverfahren

Beim *Additionsverfahren* wurden alle Basis-Interessen (GIC) gesammelt, die bei der Analyse der direkten Zusammenhänge zwischen Domains und Interessen erhoben wurden (vgl. 5.4). Ein GIC wird dann in das Profilinteresse aufgenommen, wenn eine der Domains aus dem Linkprofil diesem Interesse zugeordnet ist. Abbildung 25 zeigt beispielhaft das additive Verfahren.

Naive Bayes Classifier

Als zweites wurde ein *Naive Bayes Classifier* (Han 2011:350) erstellt. Dabei wird die Zuweisung von Interessen in einem Netzwerk trainiert, indem für jede beobachtete Domain die Wahrscheinlichkeit berechnet wird, ob sie mit der Zuweisung eines Interesses in Verbindung steht. Die Wahrscheinlichkeit P, dass eine Domain D_j dazu führt, dass ein Basisinteresse GIC_j zugewiesen wird, errechnet sich aus der Anzahl der beobachteten gemeinsamen Vorkommen $N_{i,j}$ geteilt durch die Anzahl aller Beobachtungen von GIC_j mit beliebigen Domains D (vgl. Formel 3).

$$P(D_i|GIC_j)=\frac{N_{i,j}}{D\times GIC_j}$$

Formel 3: Wahrscheinlichkeit, dass der Besuch der Domain D zu dem Interesse GIC führt.

In dem entstehenden Netzwerk werden also die Knoten (Domains) über die gewichteten Wahrscheinlichkeiten verknüpft. Je öfter das Interesse und die Domain zusammen auftreten, desto wahrscheinlicher ist es, dass diese Domain dazu führt, dass dieses Interesse dem Profil zugeordnet wird. Darauf aufbauend berechnet sich die Wahrscheinlichkeit, dass einem_r Nutzer_in U mit der Domainliste UD ein bestimmtes Interesse GIC_j zugeordnet wird wie in Formel 4.

$$U(GIC_j)=\prod_{i\in UD} P(D_i|GIC_j)$$

Formel 4: Wahrscheinlichkeit, dass ein GIC einem User U zugewiesen wird.

Im Gegensatz zu dem additiven Verfahren, das auf beobachteten direkten Zusammenhängen zwischen Seitenaufruf und Interesse beruht, tragen hierbei auch solche Domains dazu bei, ein Profil zu ermitteln, für die kein direkter Zusammenhang beobachtet wurde. Da auf Basis der Arbeit von Datta, Tschantz und Datta (2015) angenommen werden muss, dass die beobachteten Interessen auf der Google-AdSetting-Seite nicht alle zugewiesenen Interessen anzeigt, kann sich die Berechnung über Gewichtungen positiv auf die Effektivität des Zuweisungsverfahrens auswirken. Diese Möglichkeit zur Konstruktion eines Interessenprofils ist beispielhaft in Abbildung 26 dargestellt.

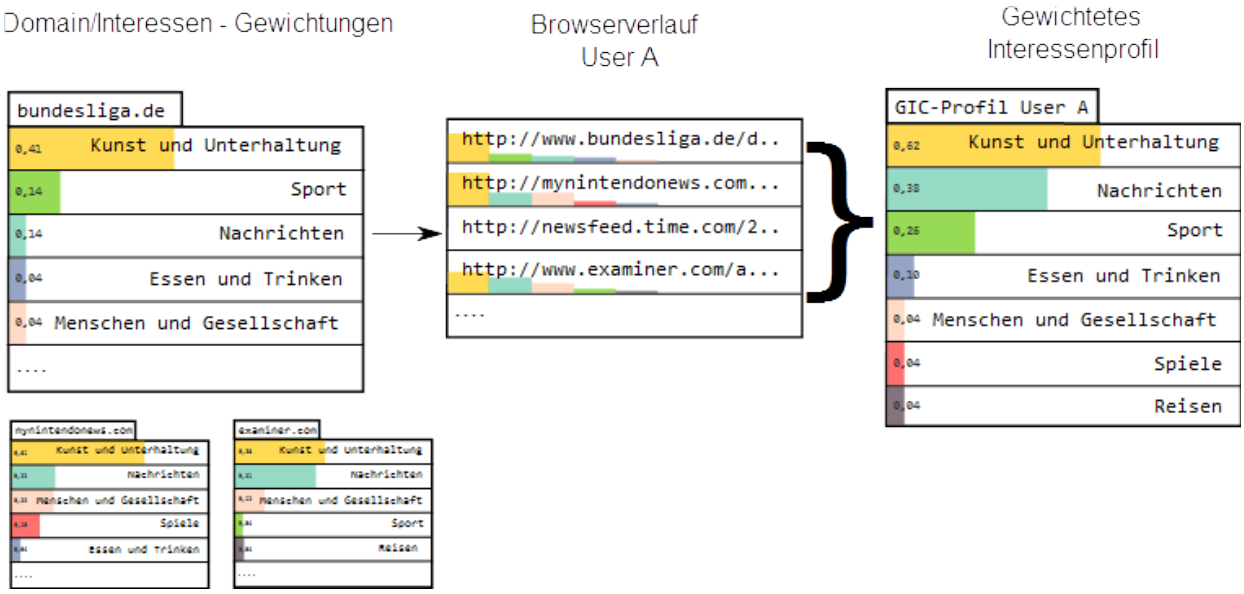


Abbildung 26: Berechnung eines Interessenprofils über die ermittelten Gewichtungen.

Nur kurz untersucht wurden Korrelationen zwischen Interessen von Nutzer_innen. Zwar gibt es einzelne Interessen, die häufig zusammen auftreten (vgl. Abbildung im Anhang), allerdings sind die Korrelationen meist schwach. Darüber hinaus ist das Rauschen in der Zuweisung, auf das in 4.8.3 noch näher eingegangen wird, derart hoch, dass dieses Verfahren als zu ungenau verworfen wird.

Vergleich der Verfahren

Kulcynski Koeffizient

Um zu überprüfen, welches der Verfahren besser dazu geeignet ist, das Interessenprofil einer bisher unbekannten Liste von URLs zu berechnen, wurden Tests mit dem vorhandenen Datenset durchgeführt. Als Vergleichsmaß soll unter anderem der Kulcynski-Koeffizient (Cheetham und Hazel 1969) dienen, der das Verhältnis zwischen Gemeinsamkeiten (der Schnittmenge zwischen den Interessenprofilen der beiden Durchläufe) und der Summe der *false positives* und *false negatives* zusammenfasst. Der Koeffizient relativiert also die Überschneidung zwischen den Mengen (korrekt

Klassifizierung) anhand der Anzahl der falschen Klassifizierungen. Je fehlerhafter die Klassifizierung ist, desto kleiner ist der Koeffizient.

$$K = \frac{Gemeinsamkeiten}{False\,Positves + False\,Negatives}$$

Formel 5: Berechnung des Kulcynski Koeffizienten

Beim Vergleich der Verfahren auf Basis der 506 erhobenen Profile ergibt sich K = 0,61 beim gewichteten Verfahren gegenüber K = 0,57 beim Additionsverfahren. Dies ist insbesondere auf die hohe Zahl von *False Positives* (16,50 gegen 7,50) beim Additionsverfahren zurückzuführen. Beim Bayes-Verfahren werden im Durchschnitt 5,50 der 8,24 GIC richtig zugewiesen. In TrickTrack werden die Interessenprofile daher anhand der besuchten Webseiten mit Hilfe des *Bayes Classifiers* berechnet.

Allerdings hat das Verfahren der Profilberechnung über die Gewichtungen seine Grenzen, insbesondere in Bezug auf seine Generalisierbarkeit. Während die Interessen ursprünglich auf Basis von Webseiten-Besuchen zu den Profilen hinzugefügt wurden, wurden zwei Vereinfachungen vorgenommen, die das System beeinflussen. Die Interessen wurden immer auf jeweilige Basisinteressen der Kategorie zurückgeführt und die Zuordnung dann zwischen Interessen und Domains (nicht vollständige URLs) gespeichert. Dies hat zur Folge, dass vor allem Webseiten, die vielfältige Inhalte bereitstellen, wie z. B. Nachrichtenportale, eine große Zahl von Interessen zugeordnet sind.

Cross-Validation

Zudem besteht bei einem Klassifizierungsmechanismus immer die Gefahr des *Overfittings* (siehe 2.3.2), da die Gewichtungen gegebenenfalls gut die Gegebenheiten des beobachteten Samples widerspiegeln, aber nicht übertragbar sind. Für die vorliegenden Daten wurde daher eine *Cross-Validation* durchgeführt. Dabei wurden die Gewichtungen aus 10 % der vorliegenden Daten berechnet (*training*) und damit die Profile der übrigen 90 % berechnet (*test*). Dieses Verfahren wurde 10-mal angewendet, so dass alle Datensätze einmal zum Training verwendet wurden. Auf diese Weise wurden jedem Profil 8 GIC zugewiesen, durch den Schwellwert sinkt die Zahl der Fehlzuweisungen deutlich. Dabei lag K im Durchschnitt bei 2,23, wobei 5,08 Interessen korrekt zugewiesen wurden bei einer *false positive* Rate von 2,57 und eine *false negative* Rate von 2,96.

Keine Generalisierbarkeit

Auch wenn die Daten für die weitere Studie eine ausreichende Genauigkeit aufweisen und vor allem zeigen, dass das Verfahren tauglich ist, ist die Datenbasis, im Vergleich zur Menge an möglichen Profilen, zu gering. Wie bereits oben erwähnt, ist die Interessenbreite der Reddit-Nutzer_innen, die für das Training des Classifier genutzt wurde, begrenzt, manche Interessen sind kaum vertreten. Es kann also angenommen werden, dass die Generierung des Profils in TrickTrack ausreichend genau ist für jede mögliche Kombination von Webseiten.

Inverses Profil

Um dieser Ungenauigkeit gerecht zu werden, wird im Folgenden nicht weiter derart mit Interessenprofilen gearbeitet, dass je Kategorien binär entschieden wird, ob sie Teil des Profils ist oder nicht. Stattdessen soll mit den Wahrscheinlichkeiten aller GICs gearbeitet werden. Das heißt, das Interessenprofil soll alle Interessen umfassen und für jedes die spezifische Wahrscheinlichkeit darstellen. In der später vorgestellten Art der Visualisierung ist es so möglich zu erkennen, wie wahrscheinlich die Zuweisung durch den vorliegenden Algorithmus ist. Außerdem erlaubt es diese Berechnung, ein *inverses Profil* zu ermitteln, indem die Liste der GIC so sortiert wird, dass die mit der geringsten Wahrscheinlichkeit an der Spitze stehen. Dieses inverse Profil kann dann genutzt werden, um solche Webseiten auszuwählen, die am besten geeignet sind, das Profil zu verschleiern.

4.8.2 Berechnung eines soziodemografischen Profils

Auf ähnliche Weise wie bei der Konstruktion des Interessenprofils lässt sich der in Abschnitt 4.7 konstruierte Datensatz (Domains sind Wahrscheinlichkeitswerten für soziodemografische Attribute zugeordnet) nutzen, um anhand der besuchten Seiten ein soziodemografisches Profil zu erstellen. Dies lässt sich auf die Linkprofile von Reddit genauso anwenden wie auf bisher unbekannte Listen von URLs, z. B. aus dem Browserverlauf. Dazu wird für jede Kategorie C jedes Attribut a mit der höchsten Wahrscheinlichkeit berechnet, indem für alle Domains D des Users U die Wahrscheinlichkeiten addiert werden (vgl. Formel 6).

$$\forall C : max(C); \vec{C} = \begin{pmatrix} \frac{\sum_{D_i \in U} P_{C_A}^{D_i}}{|D|} \\ \vdots \end{pmatrix}$$

Formel 6: Berechnung der wahrscheinlichsten Kategorien.

EINSCHRÄNKUNGEN

Anders als bei den Interessenprofilen lässt sich die Validität dieser Profilkonstruktion nicht testen, da kein Datensatz vorliegt, bei dem die Zuordnung durch eine_n Dritten vorgenommen wurde.

Kleiner Datensatz

Zudem ergeben sich Einschränkungen aus dem geringeren Umfang dieses Datensatzes. Dadurch, dass die Datenbasis der Reddit-Profile überwiegend aus männlichen, weißen Usern im Alter zwischen 25 und 44 Jahren besteht, sind auch die domainbezogenen Statistiken überwiegend für solche Domains vorhanden, die dieser Nutzer_innengruppe zugeordnet wird. Damit liegen weniger Daten für solche Domains vor, die nicht dieser Nutzer_innengruppe entsprechen. Werden Daten auf User-Profile angewendet, die mehr Domains enthalten, für die keine Daten vorliegen, ist die Aussage-

kraft des resultierenden Profils daher geringer. Um einer Generalisierbarkeit näher zu kommen, müsste die zugrundeliegende Datenbasis daher weiter ausgebaut werden. Für den vorliegenden Zweck, nämlich die grundsätzliche Anwendbarkeit und Nutzung von Transparenz- und Verschleierungsverfahren zu testen, lässt sich mit den Daten jedoch weiter operieren.

4.8.3 Obfuscation-Optionen

Die zweite Nutzung der Daten dient der Obfuscation von Profilen, also der Möglichkeit für Nutzer_innen, das über sie angelegte Profil (gezielt) zu beeinflussen. Die Effizienz der Obfuscation wiederum hängt von der *Dummy Generation Strategie (DGS,* vgl. Kapitel 3.4.1) ab, um das Profil des Tracking Providers derart zu beeinflussen, dass es möglichst weit von dem eigentlichen Profil abweicht (Balsa u. a. 2012). Da allerdings die Obfuscation durch das Hinzufügen von Seitenaufrufen und nicht durch das Weglassen geschieht, kann eine Überlappung zwischen dem erwarteten (das der Tracking-Service eigentlich ermitteln sollte) und dem verschleierten Profil nicht verhindert werden. Stattdessen ist das Ziel, die Profile so zu beeinflussen, dass sie möglichst viele Interessen enthalten, die das erwartete Profil nicht enthält. Die Maximierung der *false positive* und *false negative* resultiert daher in einem möglichst kleinen Wert für K.

Im Folgenden werden mehrere DGS vorgestellt, die auf Basis der erhobenen Daten arbeiten. Die Ermittlung der Effizienz einer DGS, wie sie hier dargestellt wurde, ist in der Forschung bisher nicht durchgeführt worden.

Obfuscation des Google Profils

Dummy Generation Strategy

Die Dummy-Generation-Strategien basieren darauf, über das Gewichtungsnetz die wahrscheinlichen Interessen eines Linkprofils zu ermitteln und zur Verschleierung solche Seiten aufzurufen, die Interessen zugeordnet sind, die im Profil gering gewichtet oder nicht vorhanden sind (*ambiguating obfuscation)*. Der Unterschied zwischen den DGS liegt darin, wie die Adressen ausgewählt werden, die das Profil verschleiern sollen.

Dynamische und statisch gemessene Zusammenhänge

Die erste Kategorie der DGS beruht darauf, den *Dummy Traffic* ebenfalls über das Wahrscheinlichkeitsnetz zu ermitteln. Das inverse Profil wird dabei genutzt, um die URLs für den Dummy Traffic auszuwählen. Das in Abbildung 26 (rechts) dargestellte Profil wird dazu umgekehrt sortiert, und aus der Gewichtungsdatenbank (Abbildung 26, links) werden dazu solche Domains ausgewählt, für die ein hoher Zusammenhang berechnet wurde. Da die Datenbank der Gewichtungen sich abhängig von den mit TrackTrack erhobenen Daten weiterentwickelt, wird dies DGS im Weiteren als „dynamische DGS“ gekennzeichnet.

Die zweite Kategorie von DGS basiert auf dem Datensatz, der die direkt messbaren Zusammenhänge zwischen Domains und Interessen repräsentiert, und wird daher im Folgenden als statischer Datensatz bezeichnet (vgl. Tabelle 15). Zwar sind diese Daten nicht besonders nützlich, um Interessen für vorher unbeobachtete Linkprofile zu ermitteln, unter anderem auch weil nur für 13 % der Domains aus dem Datensatz ein solch direkter Zusammenhang ermittelt wurde. Trotzdem lässt sich diese Liste als Quelle nutzen, um Seiten für den Dummy Traffic auszuwählen.

Neben den beiden DGS wurde außerdem getestet, in welchem Umfang Dummy Traffic generiert werden muss, um ein Profil zu verschleiern, d. h. wie viel Dummy Traffic nötig ist, um das erwartete Profil zu verschleiern. Das Evaluationsverfahren ist in Abbildung 27 dargestellt.

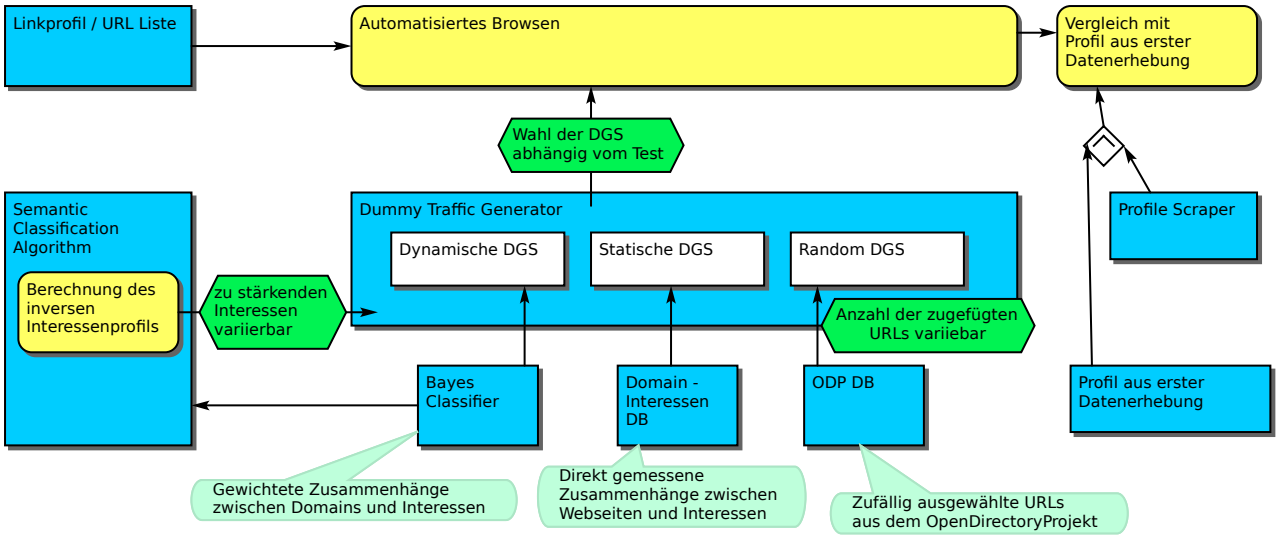


Abbildung 27: Übersicht des Verfahrens zum Vergleich verschiedener Dummy Generation Strategien.

Methode zum Vergleich von DGS

Bestimmen der notwendigen Anzahl an Tests pro DGS

Die Evaluation einer großen Anzahl an Seiten und User ist ein zeit- und ressourcenintensives Verfahren. Daher wurde vorab geschätzt, wie groß eine Testgruppe sein muss, auf Basis derer eine stabile Aussage über die Effizienz einer DGS getroffen werden kann. Dazu wurde die erste DGS mit 350 URL Profilen getestet und anschließend ermittelt, ab welcher Anzahl von Profilen sich der Durchschnitt der Kulcynski-Distanz nicht mehr signifikant ändert. Als nicht mehr signifikante Änderung werden solche Änderungen der Distanz angenommen, bei der sich der Durchschnitt durch das Hinzufügen eines weiteren Testprofils um nicht mehr als 0,01 ändert. K kann theoretisch zwischen 0 und 23 liegen, ist in der Praxis allerdings nicht größer als 13. Anhand dieser Rahmenwerte wurden die erhobenen GIC-Profile in 1000 Durchläufen zufällig auf-

addiert und jeweils ermittelt ab dem wievielten Profil die Änderung des Durchschnitts 0,01 nicht mehr überschreitet. Im Test war dies ab dem 218ten Profilen der Fall. Daher wurden für die folgenden Vergleiche der DGS jeweils mindestens 250 Profile untersucht

Tabelle 19 zeigt die verschiedenen Strategien und Variationen der Strategien im Vergleich. Sie basieren auf der Berechnung des Interessenprofils wie in 4.8.1 vorgestellt. Für die Messung wurden drei verschiedene Parameter verändert und auf ihren Effekt hin untersucht. Dazu zählen

1. die Dummy Generation Strategy (s. o.),
2. das Verhältnis zwischen den des realen Linksprofils und solchen, die über den Dummy Traffic Generator hinzugefügt wurden sowie
3. die Zahl der GIC, die versucht wurden zu stärken, weil sie im errechneten Profil besonders schwach vertreten waren.

Dumm Generation Strategien

Die verschiedenen Verfahren wurden innerhalb von 3 Wochen nacheinander mit SlimerJS durchgeführt. Die Dummy-Traffic-URLs wurden dazu vom Ende her in die URL-Listen eingefügt und zwar jeweils im Wechsel mit den URLs, die zum „realen“ Linkprofil gehörten. Die Bewertung der Strategien erfolgte in drei Runden. Zuerst wurden die drei DGS im Vergleich mit 25 Obfuscation-URLs getestet (Random, tt25dyn und tt25stat). Da sich die dynamische und statische Strategie als effektiver herausstellten, wurden diese nochmal im direkten Vergleich mit mehr Dummy Traffic getestet (tt50dyn, tt50stat). In einem letzten Durchlauf (tt75stat) wurde dann die in der zweiten Runde erfolgreichere Strategie mit mehr URLs getestet. Im Detail sind die in Tabelle 19 gelisteten Strategien wie folgt konstruiert gewesen:

Beschreibung der verglichenen Verfahren

Recrawl: Die URL-Profile wurden unverändert ein weiteres Mal aufgerufen (Kontrollgruppe).

Random: Es wurden 25 zufällige URLs aus dem ODP Datensatz ausgewählt und an das URL Profil angehängt.

tt25dyn: Der Bayes Classifier wurde genutzt, um die 5 unwahrscheinlichsten Interessen zu ermitteln. Auf dieselbe Weise wurden jeweils 5 URLs ermittelt, die diesen Interessen zugeordnet waren. 25 URLs wurden dem eigentlichen Linkprofil also hinzugefügt.

Kurz-form	DGS	Gestärk-te Inter-essen	Zusätz-liche URLs	⌀ K	⌀ GIC	⌀ GIC +	⌀ GIC -
Recrawl	Keiner	-	0	2,45 (1,93)	7,32 (2,12)	1,75 (1,63)	1,33 (1,33)
Random	Random	-	25	2,27 (1,80)	7,33 (2,02)	1,85 (1.80)	1,50 (1,31)
tt25dyn	Dyna-misch	5	25	1,90 (1,66)	7,27 (2,13)	2,11 (1,59)	1,76 (1,62)
tt50dyn	Dyna-misch	7	50	1,90 (1,73)	7,71 (1,91)	2,40 (1,76)	1,63 (1,47)
tt25 stat	Statisch	5	25	1,94 (1,80)	7,60 (1,93)	2,31 (1,76)	1,66 (1,48)
tt50stat	Statisch	10	50	1,83 (1,77)	7,95 (1,77)	**2,59** (1,71)	1,68 (1,39)
tt75stat	Statisch	15	75	**1,72** (1,62)	**8,18** (1,82)	2,93 (1,97)	**1,67** (1,34)

Tabelle 19: Vergleich verschiedener Dummy Generation Strategien (DGS). Standard-abweichung in Klammern.

tt50dyn: Die 7 unwahrscheinlichsten Interessen wurden ermittelt. Auf dieselbe Weise wurden jeweils 9 URLs ermittelt, die diesen Interessen zugeordnet waren. Durchschnittlich[113] wurden 50 URLs wurden hinzugefügt.

tt25stat: Der Bayes Classifier wurde genutzt, um die 5 unwahrscheinlichsten Interessen zu ermitteln. Die 5 URLs, die diese Interessen stärken sollten, wurden aus dem Datensatz entnommen, bei dem eine direkte Zuordnung zwischen Domain und Interesse möglich war. 25 URLs wurden so hinzugefügt.

tt50stat: Die 10 unwahrscheinlichsten Interessen wurden ermittelt. Die 5 URLs, die diese Interessen stärken sollten, wurden aus dem Datensatz entnommen, bei dem eine direkte Zuordnung zwischen Domain und Interesse möglich war. Im Durchschnitt wurden 50 URLs hinzugefügt.

tt75stat: Die 15 unwahrscheinlichsten Interessen wurden ermittelt. Die 5 URLs, die diese Interessen stärken sollten, wurden aus dem Datensatz entnommen, bei dem

113 Für einige wenige Interessen lagen keine 9 URLs vor, für die ein besonders hoher Zusammenhang ermittelt wurde, daher wurden teilweise weniger als 9 URLs hinzugefügt.

eine direkte Zuordnung zwischen Domain und Interesse möglich war. Im Durchschnitt wurden 75 URLs hinzugefügt.

Für den Vergleich der Methoden wurde wieder der Kulcynski-Koeffizient (vgl. Formel 7) herangezogen. Er zeigt hier das Verhältnis von den Gemeinsamkeiten zwischen dem ursprünglichen Interessenprofil P^{orig} und dem verschleierten Interessenprofil P^{DGS} sowie der Summe aus neuen und nicht mehr vorgefundenen GIC in den Profilen.

$$K = \frac{\left|P^{orig} \cap P^{DGS}\right|}{\left|P^{orig} \setminus P^{DGS}\right| + \left|P^{DGS} \setminus P^{orig}\right|}$$

Formel 7: Kulcynski-Koeffizient zum Vergleich der Profile

Je kleiner also der Kulcynski-Koeffizient, desto unähnlicher sind sich zwei Profile. Ein K-Wert von 1 bedeutete, dass die Profile genauso viele gleiche wie unterschiedliche Interessen aufweisen, die Obfuscation wäre optimal. Bei einem Wert kleiner 1 wäre das ursprüngliche Profil gegenüber dem verschleierten geringer ausgeprägt.

Ergebnisse

Starkes Rauschen

Tabelle 19 zeigt in den rechten Spalten die Ergebnisse der verschiedenen DGS. Deutlich sind die großen Schwankungen innerhalb des Profilings von Google selbst zu erkennen (Recrawl, Zeile 1). Nur etwa 60% der GIC sind beim zweimaligen Aufruf derselben Linkprofile gleich.

Steigerung der Obfuscation mit mehr hinzugefügten URLs

Dennoch führen die verschiedenen DGS zu einer Veränderung der Profile. Die Strategie, zufällige URLs aus der Datenbank des ODP hinzuzufügen, hat sich als am wenigsten effektiv erwiesen. Sie führt zwar zu Veränderungen im Profil (K = 2,27), allerdings nur kaum mehr als die Kontrolluntersuchung Recrawl (K = 2,45). Daher wurden in der zweiten Runde nur noch die beiden anderen Strategien getestet, wobei hier die Strategie, die auf direkt messbare Zusammenhänge zwischen URLs und Interessen zurückgriff (tt50stat), leicht besser abschnitt als die dynamische Strategie. Daher wurde in einer letzten Runde eine weitere Erhöhung des Dummy Traffic mit dem statischen DGS und einer höheren Anzahl der zu stärkenden GIC gewählt, die zur stärksten gemessenen einer Änderung des Profils (K = 1,72) führt.

Obfuscation funktioniert

Abbildung 28 zeigt die Änderung an dem GIC Profil über alle User. Auf der x-Achse wird die Anzahl der verlorenen GIC angezeigt, auf der y-Achse die Anzahl der gewonnen. Je heller ein Hexagon, desto mehr User_innen hatten diese Kombination von Änderungen. Je weiter rechts und je weiter oben ein Profil eingeordnet ist, desto stärker ist die Verschleierung. Am häufigsten wurden 2–3 GIC verloren und 2–3 weitere gewonnen. Das Profil ändert sich also um 4–6 GIC. Im Durchschnitt ergibt sich eine Änderung des Profils um 56 %. Abbildung 29 zeigt die Entwicklung der Breite und Änderungen der GIC, wenn die Anzahl der hinzugefügten URLs erhöht wird.

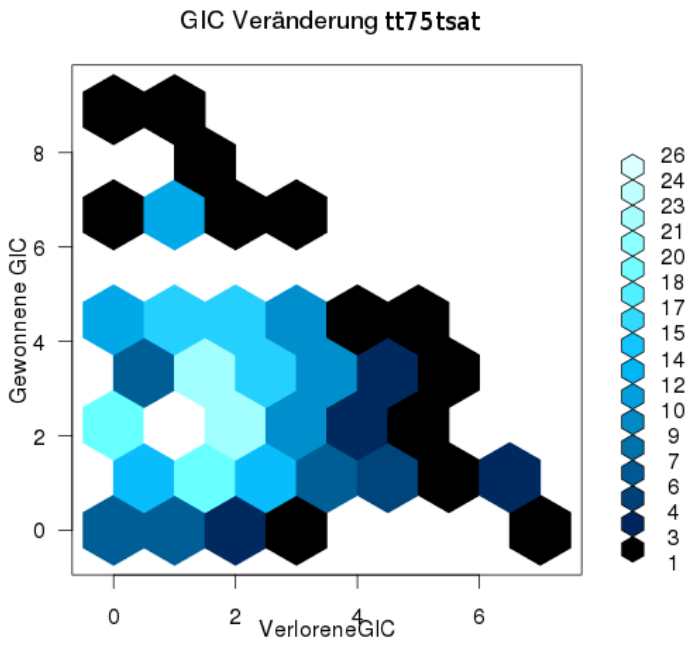


Abbildung 28: Vergleich der Veränderung bei einfacher Wiederholung (links) und der Obfuscation mit tt75stat. Ergebnisse aller Durchläufe im Anhang.

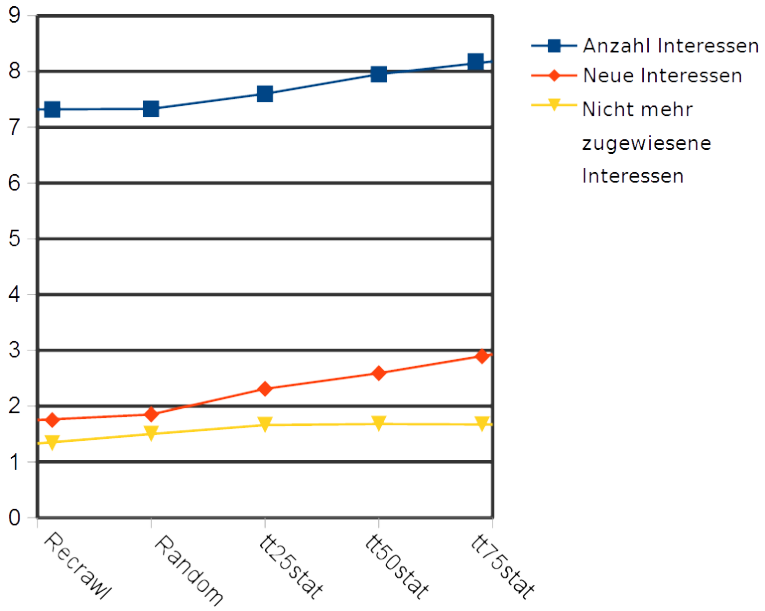


Abbildung 29: Entwicklung der zugewiesenen GIC bei Veränderung der Zahl der zur Verschleierung aufgerufenen URLs.

Bewertung

Hohe Schwankungen im Profil

Das Ergebnis der Kontrollgruppe (Zeile 1) bestätigt die Hypothese, dass die Zuordnung der Interessen durch Google alles andere als stabil ist, sondern eine hohe Unge-

nauigkeit aufweist, auch wenn die Ähnlichkeit mit K = 2,45 hier am höchstens ist. Im Durchschnitt erscheinen 1,29 neue GIC (s=1,72), während 1,84 (s=1,32) nicht mehr auftauchen, wenn die Seiten in derselben Reihenfolge im gleichen Verfahren aufgerufen werden. Das entspricht einer Änderung von 42 % des Interessenprofils.

Konstante Profilbreite

In der Gesamtsicht der Spalte **⌀ GIC** wird deutlich, dass die Anzahl der zugewiesenen GIC nicht direkt mitder Anzahl der besuchten Seiten zusammenhängt. Die Breite des GIC steigt bei keiner Strategie weit über 8. Eine Erklärungsmöglichkeit ist, dass auch aus Sicht von Tracking-Providern ein Profil unterscheidbar und aktuell bleiben muss. Es ist möglich, dass eher Interessen zugewiesen werden, die auch im Gesamtvergleich ein plausibles Profil ergeben, während die, die nicht so häufig zusammen auftreten, auch nicht so häufig zugewiesen werden. Wenn bereits früh die Interessen *Hobby und Freizeit* sowie *Kochen und Essen* ermittelt wurden, dauert es möglicherweise länger, bis der Besuch von Seiten der Kategorie *Finanzen* oder *Unternehmen und Industrie* Eingang in das Profil findet. Allerdings bestätigen die gemessenen Korrelationen zwischen den Interessen (siehe Anhang 2.2) dieses Argument nicht. Eine andere Erklärungsmöglichkeit ist, dass die Interessenprofile möglichst die Interessen des aktuell Surfenden darstellen sollen und nach einem bestimmten Maß Interessen, die schon länger bestehen oder schwächer sind, verworfen werden. Das würde die relativ stabile Breite des GIC erklären, deckt sich allerdings nicht mit der Beobachtung, dass die Anzahl der verlorenen GIC relativ stabil ist, während die Anzahl der zusätzlichen GIC mit der Anzahl der hinzugefügten URLs steigt.

Obfuscation funktioniert

In Bezug auf die drei Parameter Strategie, Menge des *Dummy Traffic* und Menge der zu stärkenden Interessen, lässt sich feststellen, dass Obfuscation am effektivsten ist, wenn versucht wird, mit möglichst vielen URLs möglichst viele Interessen zu stärken, die im Profil vorher nicht vertreten waren. Allerdings ergeben sich keine eindeutigen Hinweise, welcher der drei Parameter hier den größten Einfluss hat. Um den Einfluss der verschiedenen Parameter genauer zu bestimmen, wäre es notwendig die Tests bei Veränderung nur eines Parameters zu wiederholen. Die grundsätzliche Frage, ob die Verschleierung von Interessenprofilen mit Dummy Traffic möglich ist, kann allerdings auch mit den vorliegenden Daten beantwortet werden. Einen positiven Einfluss hat sicherlich die Erhöhung des Dummy Traffic im Verhältnis zum „realen“ Traffic.

Strategien auf Basis des errechneten Profils erfolgreicher

Zwar hat sich in der zweiten Runde die statische Strategie etwas effektiver gezeigt, allerdings war hier auf Grund eines Designfehlers in der Software nicht dieselbe Zahl zu stärkender Interessen gewählt worden. Die Entscheidung für die weitere Evaluation der statischen Strategie fiel unter der Annahme, dass die direkt messbaren Zusammenhänge zwischen Webseiten, bei der die statische Strategie genutzt wurden, zu weniger Dummy Traffic führen würde, der keinen Einfluss auf das Profil hat. Durch die Gewichtung im Classifier kann es durchaus passieren, dass Dummy Traffic generiert

wird, der zu Webseiten führt, die keine Tracking-Skripte enthalten und damit ohne Einfluss auf das Profil sind. Allerdings ist der Einfluss dieses Faktors in den vorliegenden Daten nicht messbar.

Ebenfalls nicht einzuschätzen ist, wie wichtig die Wahl des Schwellwerts bei den Interessen ist, von denen angenommen wird, dass sie bisher nicht Teil des Profils waren. Es gilt einerseits zu verhindern, dass Dummy Traffic eingestreut wird, der Interessen bestärkt, die bereits im Profil vorhanden sind, andererseits sollen möglichst viele zusätzliche Interessen die Verschleierung verbessern. Da die Profilbreite allerdings, wie oben beschrieben, relativ statisch war, lässt sich der Effekt nicht nachweisen.

Keine Dummy Classification durch Google

In Abschnitt 3.4.2 waren Angriffsszenarien auf Obfuscation-Verfahren beschrieben worden. Um den *Query Based* Angriffen entgegenzuwirken, waren Änderungen am Browser vorgenommen worden, die dessen Identifizierung erschweren sollten. In Bezug auf die *Profile Based*-Angriffe wurden allerdings keine Maßnahmen ergriffen. Die Tatsache, dass es dennoch möglich war, die Profile teilweise zu verschleiern, lässt daher den Schluss zu, dass Google bisher entweder keine Mechanismen implementiert hat, um die Verschleierung des Interessenprofils zu entdecken, oder diese nicht funktionieren.

In weiteren Tests sollte auch überprüft werden, inwiefern sich die Entscheidung, den Dummy Traffic im Wechsel mit dem Real Traffic zu sortieren, auf die Profile auswirkt. Weitere Arbeitshypothesen könnten sein, dass die Profile eher beim ersten „Bemerken“ eines_r Nutzer_in durch einen Tracker erstellt und vor allen Dingen bestätigt werden. Sollte Google zum Ziel haben, feste Persönlichkeitsprofile zu konstruieren, müsste Dummy Traffic zu Beginn einer Session ausgeführt werden. Dann würde das Profil beim ersten Einsatz eines Obfuscation-Werkzeugs zurückgesetzt werden.

Obfuscation des soziodemografischen Profils

Die Verschleierung des soziodemografischen Profils lässt sich, anders als die der Interessen, nicht an den Ergebnissen von tatsächlichen Tracking-Diensten messen. Die AdSetting-Seite von Google beinhaltet zwar die Möglichkeit, auch soziodemografische Profilinformationen anzuzeigen, allerdings wurde dies in den Tests nur sehr selten beobachtet, so dass eine vergleichende Analyse nicht effizient durchgeführt werden kann. Die Berechnung des soziodemografischen Profils wurde in Abschnitt 4.7 vorgestellt. Die Verschleierung kann analog zu der dynamischen DGS für das Interessenprofil geschehen. Es soll Dummy Traffic derart ausgewählt werden, dass er die Kategorien stärkt, denen im Ausgangsprofil die geringste Wahrscheinlichkeit zugewiesen ist. Die Auswahl von Domains, die angesurft werden können, um die jeweiligen Attribute zu stärken, kann aus den erhobenen Datensätzen von Quantcast, Alexa und Compete

erfolgen. Eine effiziente Obfuscation würde dazu führen, dass alle Kategorien pro Attribut gleich wahrscheinlich sind.

4.9 ZUSAMMENFASSUNG

In diesem Kapitel wurde die Datenerhebung mit TrackTrack und Analyse mit TrackBack vorgestellt. Dies diente der Beantwortung der dritten in der Einleitung genannten Leitfrage (*Wie sehen Profile aus, die durch Online-Tracking generiert werden, und wie werden diese ermittelt?)*.

Von der Plattform Reddit wurden Linkprofile extrahiert, diese automatisiert abgerufen und anhand der dabei mitgeschnittenen Daten der Umfang von Online-Tracking gemessen, dem User_innen ausgesetzt sind. Zusätzlich wurden bei bekannten Online-Trackern Daten gesammelt, um die Linkprofile durch interessenbezogene und soziodemografische Angaben anzureichern. Beide Datenarten wurden mit Bezug zu den Linklisten ausgewertet und der Umfang des Profilings analysiert. Zentrale Erkenntnisse sind, dass Google, als Provider zahlreicher Dienste, dazu in der Lage ist, einen Großteil (im Durchschnitt 81 %) des Surfverhaltens einzelner Nutzer_innen zu beobachten. Beim Besuch von 100 Seiten entstehen so Interessenprofile, die durchschnittlich 8 sogenannte *Google Interest Categories* (GIC) enthalten. Darüber hinaus wurden Algorithmen vorgestellt und evaluiert, mittels derer sich soziodemografische Profile und Interessenprofile auch für Nutzer_innen berechnen lassen, für die die Daten bei den Anbieter_innen nicht vorliegen. Auf Basis des selbst entwickelten Interessen-Profiling wurden dann Möglichkeiten getestet, die Profile zu verschleiern. Dazu wurden verschiedene Dummy Generation Strategies getestet und deren Effektivität wiederum am Google Dienst gemessen. Zwar weist das Interessen-Profiling durch Google bereits erhebliche Schwankungen auf, mit den vorgestellten Verfahren ließen sich die beobachteten Profile allerdings noch undeutlicher machen.

In diesem Kapitel wurden Verfahren entwickelt, welche die funktionalen Anforderungen an eine Privacy und Transparency Enhancing Technology erfüllen, wie sie im vorherigen Kapitel definiert wurden. Die Analyse der Funktionsweise (Abschnitt 4.3) und des Umfangs (4.4) von Profiling können zusammen mit der anbieterunabhängigen Rekonstruktion von Interessen- (4.6) und soziodemografischen (4.7) Profilen genutzt werden, um die Transparenz für Nutzer_innen zu erhöhen. Die evaluierten Verschleierungsmethoden (4.8) können Nutzer_innen helfen, in die Profile zu intervenieren und sie den eigenen Bedürfnissen anzupassen.

5. ENTWICKLUNG UND EVALUATION EINES TRANSPARENZ- UND OBFUSCATION-ADDONS

Nachdem in den vorherigen Kapiteln festgestellt wurde, dass es notwendig ist, Nutzer_innen Einfluss auf Profiling zu ermöglichen und es überhaupt erst transparaent zu machen, wird im vorliegenden Kapitel ein Prototyp vorgestellt und evaluiert, der diese Forderung umsetzt. Der Prototyp implementiert die in Kapitel 4 erprobten Methoden unter Berücksichtigung der in Kapitel 3 beschriebenen Anforderungen an eine solche Technologie. Er sollte also leicht zu benutzen und lernförderlich sein und so die Awareness für Online-Profiling steigern und ein Verständnis von Profilen fördern.

Im Folgenden werden zuerst die Designziele (Transparenz, Literacy, Usability) der TrickTrack genannten Browsererweiterung vorgestellt (5.1) sowie die konkrete Umsetzung beschrieben (5.2). Im Anschluss folgt eine Evaluation der Nutzbarkeit und Nützlichkeit von TrickTrack, die auf einer qualitativen Erhebung mit 10 Proband_innen basiert (5.3).

5.1 DESIGNZIELE

Das Browser-Plugin soll die Erkenntnisse, die durch die Sammlung (TrackTrack) und Analyse (TrackBack) von Online-Tracking und -Profiling gewonnen wurden, für User_innen nutzbar machen. Das Ziel ist, *transparent zu machen (5.1.1)*, wie Profile aussehen könnten, und über gezielte *Obfuscation* auf diese Profile Einfluss zu nehmen. Bekannte Usability-Probleme anderer Privacy Enhancing Technologies (vgl. 3.5) sollen möglichst vermieden und so die privacy literacy der Nutzer_innen gefördert werden.

5.1.1 Transparenz

Die Anforderungen an Transparency Enhancing Technologies wurden bereits in Abschnitt 3.5 diskutiert. Das Ziel des Prototyps ist es, den gesamten Prozess von der Datenerhebung (Tracking) über die Verarbeitung in Form von Profilen, einer Erläuterung der Nutzung bis zu den Möglichkeiten der Einflussnahme transparent zu gestalten (vgl. Abbildung 30).

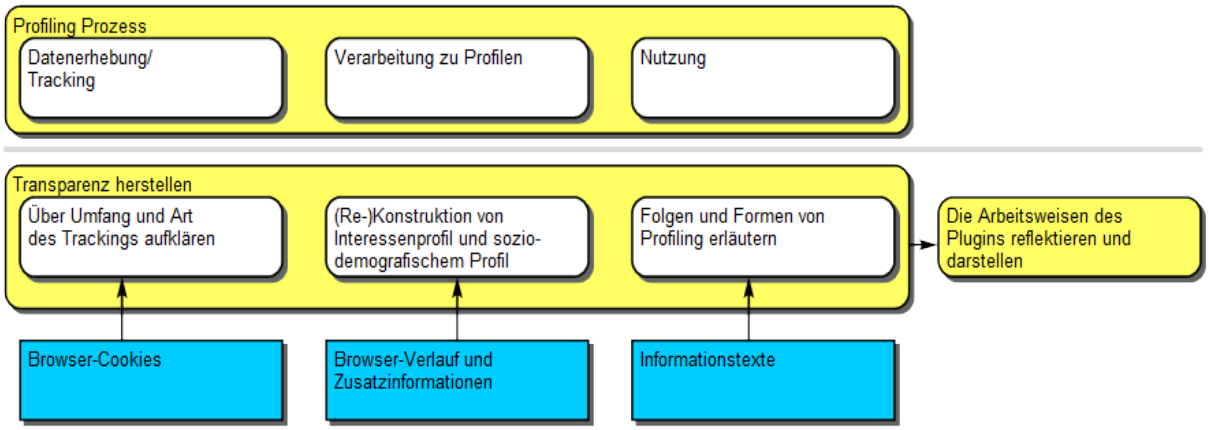


Abbildung 30: Elemente des Profiling-Prozesses und wie Transparenz zu diesen erzeugt werden kann.

Erstes Element dieses Prozesses ist der technische Vorgang des Trackings. Wie in 3.3 beschrieben, beruht dieser in vielen Fällen noch auf Browser-Cookies. Da diese über eine Browsererweiterung ausgelesen werden können, lässt sich der Umfang des Trackings, dem ein_e Nutzer_in ausgesetzt war, so im Browser rekonstruieren.

Unabhängigkeit von Anbietern

Da nicht davon ausgegangen werden kann, dass Anbieter_innen die Transparenz in Bezg auf die tatsächlich erstellten Profile mittelfristig erhöhen, soll das Plugin selbst Profile (re-)konstruieren und darstellen, die möglichst nah an denen liegen, die die Services generieren. Dazu werden die im vorherigen Kapitel gesammelten Daten genutzt, um auf Basis der im Browser vorliegenden Informationen potentielle Profile zu generieren. Die beiden implementierten Module erstellen interessenbezogene sowie soziodemografische Profile. Darauf aufbauend können gegensätzliche oder „Anti-Profile“ erzeugt werden, um die Verschleierung so zu steuern, dass die durch Tracker erstellbaren Profile weniger eindeutig und damit weniger nützlich werden. Eine Übersicht der funktionalen Elemente von TrickTrack ist in Abbildung 38 dargestellt.

Herausforderung: eigene Arbeitsweise reflektieren

Eine besondere Herausforderung stellt die gleichzeitige Vermittlung der Funktionsweise von Profiling und der Darstellung der Funktionsweise des Programms selbst dar. Um die Kompetenzen der Nutzer_innen in Bezug auf Profiling zu verbessern ohne die Grenzen des Programms zu verschweigen, müssen beide Ebenen ausreichend dargestellt werden. Darüber hinaus sollen die Erläuterungen zur Funktionsweise des Plugins auch dazu dienen klarzustellen, dass die Daten nicht in Kooperation mit den Tracking-Providern selbst generiert werden. Es sollen vielmehr die Kontingenz der Profile dargestellt, die Unterschiedlichkeit der Datenarten repräsentiert, ihre Herkunft und ihr Nutzen erläutert sowie Möglichkeiten angeboten werden, auf Basis der vorliegenden Informationen aktiv zu werden.

5.1.2 Privacy Literacy

In Bezug auf die privacy literacy soll TrickTrack insbesondere einen aktiven Umgang mit Profilen fördern und so einen Rollentausch zwischen Datenerhebenden und Datenauswertenden ermöglichen. Es soll die Ungenauigkeit und Variabilität des Profiling vermitteln, die darauf beruht, dass Profiling nicht den Zweck hat, die Persönlichkeit des_der Nutzer_in abzubilden (vgl. Kapitel 2). Die Ermutigung zum kreativen Umgang mit Profilen kann so die Autonomie der Nutzer_innen stärken. Berendt (2012) nennt sechs Prinzipien für Anwendungen, die das erfolgreiche Erlernen von critical data literacy fördern sollen, die eben auch zu einer privacy literacy beitragen. Werkzeuge zu diesem Thema sollen, nach Berendt, erstens unterschiedliche Nutzungskontexte unterstützen und sich dabei mit authentischen Fragestellungen beschäftigen. Zweitens sollen sie eine konstruktive und aktive Auseinandersetzung mit den vorliegenden Daten fördern, indem sie verschiedene Auswertungsmöglichkeiten bereitstellen. Drittens können Tools, die verschiedene Perspektiven, etwa auf einen Datensatz oder Funktionen der Auswertung, bieten, eine kritische Auseinandersetzung fördern. Viertens sollte die Anwendung sozial sein, also das Teilen von und gemeinsame Arbeiten mit den Ergebnissen fördern. Und zuletzt sollte Artikulation und Reflexion der Datenanalyse zur Internalisierung und dem besseren Verständnis der Materie beitragen.

Elemente konstruktiven Lernens

Für das TrickTrack Browser-Plugin ist die Umsetzung der Anforderungen wie folgt geplant:

- *Authentizität*: Die Darstellung möglicher Profile des_der Nutzer_in basierend auf dem Browserverlauf (basierend auf den Berechnungen in 4.8.1 und 4.8.2) stellen einen direkten Bezug zwischen der Datenauswertung (dem Profiling) und den Nutzer_innen her.
- *Konstruktivität*: Die in Abschnitt 4.8.3 vorgestellten Interventionsmöglichkeiten durch Obfuscation können genutzt werden, um aktiv und konstruktiv Einfluss auf Profile zu nehmen.
- *Multiperspektivisch*: Durch Begleittexte kann die Funktionsweise von Tracking und die Konstruktion von Profilen durch verschiedene Anbieter_innen dargestellt werden.
- *Artikulation und Reflexion:* Der Bezug auf Profile, die sich aus dem „eigenen“ Browserverlauf konstruieren lassen (siehe Authentizität), kann die Reflexion über die Ungenauigkeit von Profiling und den sich daraus ergebenden Konsequenzen fördern.

- Im initialen Design ist die Erweiterung ohne *soziale Komponente* implementiert. Der Prototyp ist ohne Internetanbindung und Datenaustausch funktionsfähig. Eine soziale Komponente würde die Implementierung eines Datenaustauschs voraussetzen, dessen Ausgestaltung mit unverhältnismäßig hohem Aufwand verbunden wäre.

5.1.3 Usability

Bei der Entwicklung von TrickTrack soll außerdem ein weiterer Schwerpunkt auf die Usability, insbesondere die leichte Erlernbarkeit und Erwartungskonformität, sowie auf die Klarheit bei der Vermittlung der Funktionsweise gelegt werden.

- *Erwartungskonformität*: TrickTrack soll ohne Vorwissen nutzbar sein. Daher soll sich an gängigen Gestaltungsmustern orientiert und die zur Verfügung stehenden Optionen sollen verständlich erläutert werden.
- *Lernförderlichkeit*: Neben der Herstellung von Transparenz in Bezug auf Profiling implementiert TrickTrack auch Verschleierungsoptionen. Das Interface muss einen Übergang zwischen diesen Funktionen schaffen, die den_die Nutzer_in nicht überfordern.
- *User Experience:* Statt nur das bestehende Unwohlsein mit Tracking zu bestätigen soll TrickTrack auch den spielerischen Umgang mit der vermeintlich komplizierten Technologie üben. Dadurch sollen die A*wareness* und der Wille, sich mit dem Thema auseinanderzusetzen, erhöht werden. Darüber hinaus werden auch in der Forschung zur Usability spielerische Elemente von Software verstärkt als Faktor betont (Hassenzahl 2010).

Evaluation der Designziele

Um zu erheben, inwiefern das vorgestellte Plugin die Designziele erreicht, werden mehrere Evaluationsmethoden eingesetzt (vgl. 5.3). Probleme mit der Usability sollen im Rahmen einer **beobachteten Nutzung** und einer abschließenden schriftlichen Bewertung anhand eines **standardisierten Fragebogens** erfolgen. Ob und welche Kompetenzgewinne (*privacy literacy*) die Nutzer_innen in Bezug auf den Umgang mit Profilen erfahren, wird in einer **leitfadengestützen Vor- und Nachbefragung** ermittelt werden.

5.2 TRICKTRACK: DAS BROWSER PLUGIN

Entwicklungsumgebung

TrickTrack wurde als AddOn für den Firefox Browser und mit dem Firefox AddOn-SDK[114] entwickelt. Nur durch die Integration in den Browser kann das AddOn Zugriff

114 Siehe: HTTPS://DEVELOPER.MOZILLA.ORG/EN-US/ADD-ONS (zuletzt abgerufen am 26.11.2014).

auf den Verlauf des Browsers und die gespeicherten Cookies erhalten. Darüber hinaus verarbeitet das AddOn Daten in HTML und JavaScript. Die verwendeten Bibliotheken sind im Anhang 10.5 gelistet.

Keine Serveranbindung

Die Datensätze mit den Gewichtungen zwischen soziodemografischen Daten und Interessen sind in JSON Dateien lokal abgelegt. So ist keine Kommunikation mit einem zentralen Server notwendig, bei der Nutzungsdaten anfallen würden. Gleichzeitig birgt dies den Nachteil, dass Änderungen an den Statistiken nicht zeitnah für den_die Nutzer_in einsehbar sind. Da im Rahmen der vorliegenden Arbeit mit dem Datenset eines Stichtags gearbeitet wurde, kann dies für die prototypische Umsetzung vernachlässigt werden.

Hauptkomponenten

TrickTrack[115] besteht aus zwei Hauptkomponenten: einem Informationsbereich, der sich im oberen Teil mit weißem Hintergrund befindet; (vgl. 31) und einem Interventionsbereich im unteren Bereich (vgl. 32). Beide Bereiche beziehen sich immer auf einen von drei Themenbereichen (Basis Informationen, Interessenprofil, soziodemografisches Profil), die in Form von Tabs auswählbar sind. Zusätzlich existiert im unteren Bereich eine statische Informationsleiste (grau hinterlegt).

5.2.1 Basis Informationen

In der ersten Übersicht (vgl. Abbildung 31) werden mehrere Zahlen dargestellt, die aus dem Browserverlauf extrahiert werden. Ziel dieser Übersicht ist es, durch die – meist hohen – Zahlen das Interesse für weitere Analysen zu wecken und ein Gefühl für den Umfang von Tracking zu vermitteln. McDonald und Cranor (2010) haben gezeigt, dass, obwohl zwei Drittel der Befragten ein grobes Verständnis von Cookies haben, eine große Ungewissheit über die Notwendigkeit und Umgangsmöglichkeiten mit Cookies besteht. Ziel ist es hier auch nicht, die technische Funktionsweise von Cookies zu erläutern. Vielmehr sollen die Notwendigkeit der Cookie-basierten Reidentifizierung für Tracking und das Verhältnis zwischen 3rd Party-Tracking und anderen Cookies dargestellt werden.

Dazu gehört die Anzahl

- der aufgerufen Seiten in den vergangenen (maximal 365) Tagen,
- der dazugehörigen Domains,
- der dabei im Browser hinterlegten Cookies,
- derjenigen Cookies, die von einer 3rd Party – also einer Seite, die nicht direkt aufgerufen wurde – hinterlegt wurden,

115 Das Plugin steht unter HTTPS://TRICKTRACKING.COM zum Download bereit.

- derjenigen *3rd Parties*, deren Domain mit den Buchstaben 'ad' beginnt und darauf schließen lässt, dass es sich um Werbenetzwerke handelt.

Abbildung 31: Screenshot der Übersicht über Basisinformationen, die aus dem Browserverlauf extrahiert werden.

Ziel dieser Darstellung ist es, einen kurzen Einblick in die Daten zu geben, die der Browserverlauf enthält und die von Trackern aber auch von TrickTrack ausgewertet werden. Zusätzlich wird ein Hinweisfeld eingeblendet, dass die Daten in einen Kontext rückt und beschreibt, dass im Durchschnitt 17 Cookies pro Webseite gespeichert werden und es großen Tracking-Provider, wie Google oder Facebook, damit möglich ist, bis zu 80 % der Webseitenaufrufe zu beobachten.

Unterhalb dieser Auflistung stellt TrickTrack einen Auszug aus der Liste der meist besuchten Webseiten sowie eine prozentuale Auswertung des Anteils der 10 meistbesuchten Seiten am gesamten Browserverlauf (siehe Abb. 32) dar. Daneben ist ein Hinweis eingeblendet, der auf das Problem der Filterbubble hinweist und zu einer weiteren Informationsseite (siehe 5.2.4) führt.

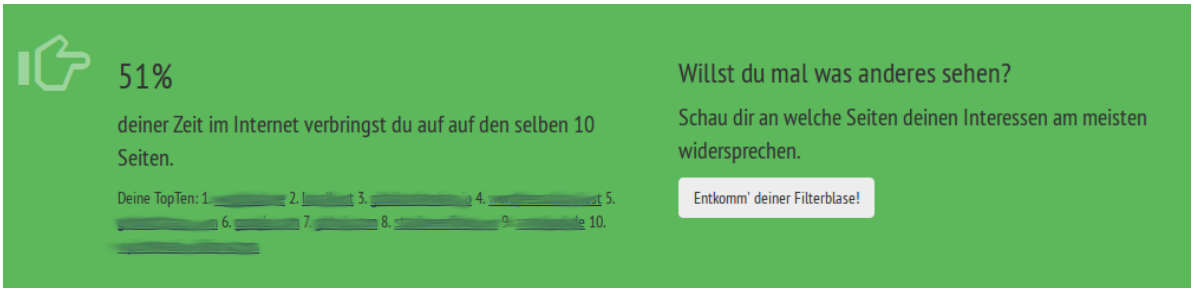


Abbildung 32: Hinweis auf den Anteil der meistbesuchten Seiten am Gesamtumfang und Hinweis auf Informationen zur Filterbubble.

5.2.2 Interessenprofil

Das Interessenprofil (siehe 33) ist eine Visualisierung der errechneten Google-Interessen. Hierzu werden die summierten Gewichtungen pro Kategorie normalisiert und in

Darstellung der Interessen im Verhältnis

Kreisform dargestellt. Die Größe des Kreises ist abhängig von der errechneten Wahrscheinlichkeit, dass ein Interesse auf den_die Benutzer_in zutrifft (vgl. 4.8.1). Über eine Zeitleiste lässt sich der Zeitraum eingrenzen, über den das Interessenprofil gebildet wird. Dabei werden nur solche Seiten mit in die Profilbildung einbezogen, die in diesem Zeitraum zuletzt besucht wurden.

Ein Klick auf einen Kreis zeigt die ermittelten Webseiten, deren Gewichtung am stärksten zu dem Interesse beiträgt. Ein Link unterhalb dieser Darstellung weist darauf hin, dass das Interessenprofil beeinflusst werden kann und führt zu dem entsprechenden Abschnitt auf der Seite „Anti-Filterbubble“.

Rechts neben dem Interessenprofil beschreibt ein Informationstext, wie die Interessenkategorien konstruiert sind und ordnet das Ergebnis zusätzlich ein. Es wird beschrieben, dass die Darstellung keine Abbildung der Persönlichkeit des_r Nutzer_in ist, sondern auf einen konkreten Zweck hin entwickelt wurde. Ein weiterer Link öffnet einen Dialog, der wiederum die Datenerhebung von TrickTrack erläutert und das Vorgehen zur Erstellung des Interessenprofils und der Crawl-Vorgänge beschreibt. Hier findet sich zudem ein Verweis auf die Profilinformationsseiten der Anbieter_innen Google und Bluekai.

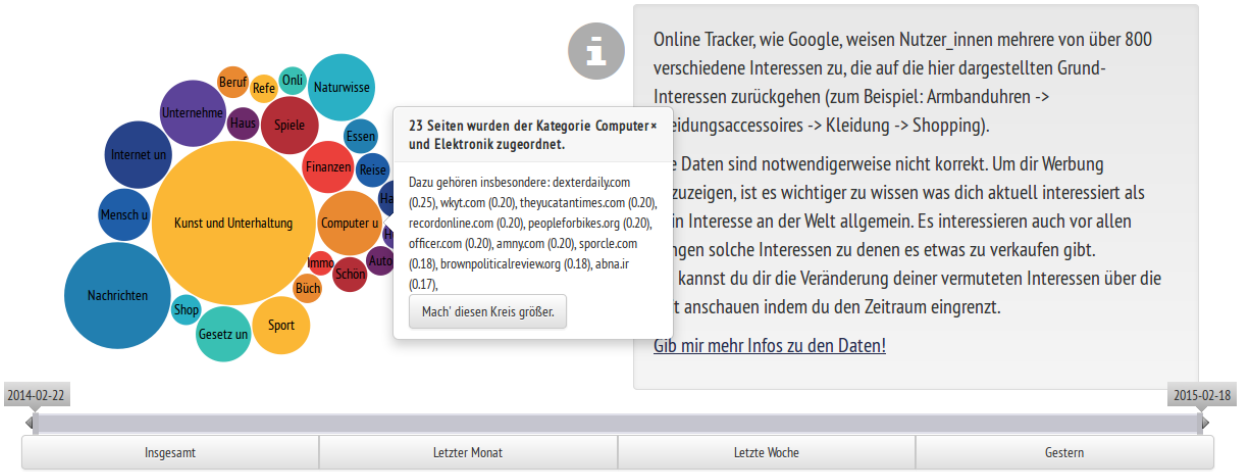


Abbildung 33: Darstellung der gewichteten Interessen

Obfuscation des Interessenprofils

Unterhalb der Profilinformation wird die Möglichkeit zur Verschleierung angeboten (Abb. 34, oben). Nach einem Klick öffnet sich ein weiterer Browser-Tab im Hintergrund, der Seiten zur Verschleierung aufruft. Dazu wird der_die Nutzer_in über die Funktionsweise der Obfuscation aufgeklärt und darauf hingewiesen, wie diese beendet werden kann (Abb. 34, unten).

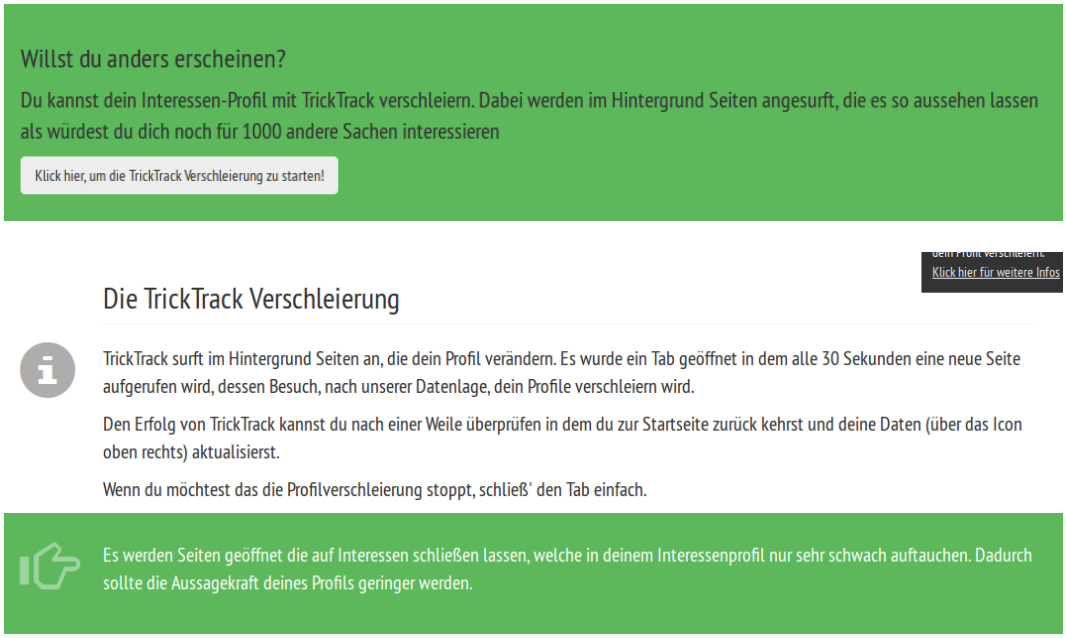


Abbildung 34: Hinweis auf die Möglichkeit der Verschleierung der Interessen (oben) und Informationsseite über die laufende Verschleierung (unten).

5.2.3 Soziodemografisches Profil

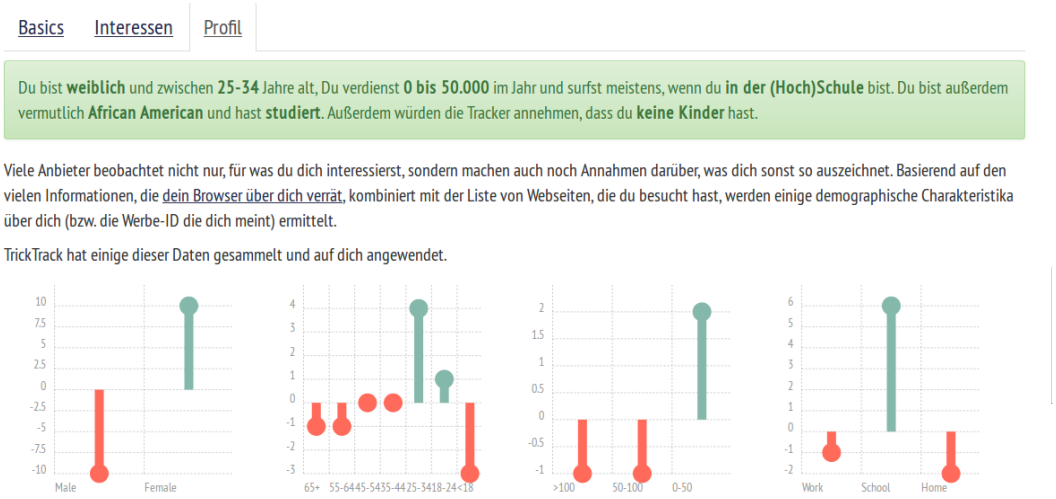


Abbildung 35: Darstellung des soziodemographischen Profils.

Absolute Darstellung des soziodemografischen Profils

Abbildung 35 zeigt die zwei Darstellungsarten des soziodemografischen Profils (vgl. 4.8.2). Im oberen, hellgrün hinterlegten Bereich ist das Ergebnis der Analyse in kurzen Sätzen zusammengefasst. Dabei sind die wahrscheinlichsten Kategorien je Attribut fett gedruckt. Im unteren Bereich findet sich eine Erläuterung der Herkunft dieser Daten. Die Graphen zeigen für jede Kategorie, wie stark sie vom Mittelwert abweicht, so dass die Nutzer_innen einen Eindruck davon bekommen können, wie genau die statistische Einschätzung ist. Nicht in der Abbildung zu sehen ist ein weiterer Hinweis-

text, der Informationen zur Genauigkeit der Analyse enthält und beschreibt, zu welchem Prozentsatz die Seiten in die Analyse mit eingeflossen sind.

Abbildung 36: Auswahlmenüs für die Verschleierung des soziodemografischen Profils.

Obfuscation des soziodemografischen Profils

Unterhalb dieser Informationen befindet sich wieder die Möglichkeit zur Intervention. In mehreren Auswahlfeldern kann die_der Nutzer_in ein Profil zusammenstellen. Mit einem Klick startet dann die Verschleierung des Profils, bei der Webseiten aufgerufen werden, deren soziodemografische Profile eine besonders starke Ausprägung bei den ausgewählten Attributen aufweisen.

5.2.4 Weiterführende Informationen

Anti-Filterbubble

Wie oben beschrieben, wird an unterschiedlichen Stellen auf die Informationsseite zur Filterbubble verwiesen. Diese Unterseite (vgl. Abb. 37) enthält eine kurze Beschreibung des Filterbubble-Problems. Darunter befindet sich eine Auflistung der GIC, sortiert mit aufsteigender Gewichtung der Wahrscheinlichkeit der Zuweisung im Interessenprofil und kombiniert mit einer zufälligen Auswahl an Links zu Seiten, die dem jeweiligen GIC zugeordnet wurden.

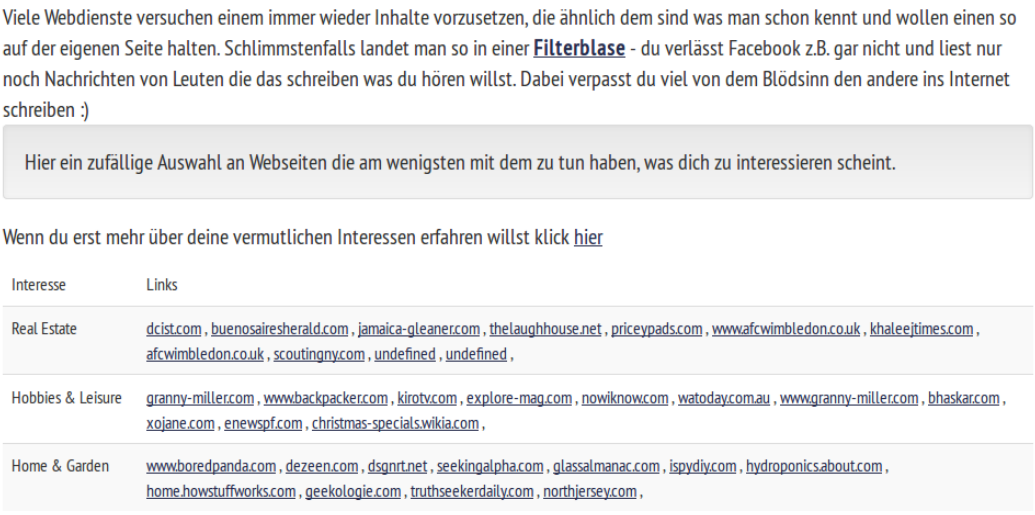


Abbildung 37: Informationen zur Filterblase und zu Webseiten, die dem ermittelten Profil widersprechen.

Weitere Hinweise

Zusätzlich zur Darstellung des angenommenen Profils des_r Nutzer_in informiert TrickTrack in begleitenden Texten über den Umfang der Datensätze und die Funktionsweise von Online-Tracking und Profiling. Dabei erläutern die Texte anhand der in Kapitel 4 erhobenen Daten, auf welcher Datenbasis die Anbieter_innen operieren und welche Ziele sie verfolgen. Verschiedene Links verweisen auf weitergehende Informationen, wie eine übersichtlichere Darstellung der Interessenliste von Google[116] oder der Profilinformationsseite von Bluekai. Das Ziel der Beschreibung ist dabei, die Instabilität der Profile sowie deren Funktion als Element der Dividualisierung zu erläutern (Kapitel 2).

5.2.5 Abwägungen und Zusammenfassung

Bei der Gestaltung von TrickTrack mussten an einigen Stellen Entscheidungen getroffen werden, um die gesetzten Designziele zu erreichen. Genauigkeit bei der Auswertung des Tracking und der Beschreibung der Daten stehen dem Ziel, ohne große Vorkenntnisse ein Verständnis für die Probleme des Profiling zu erzeugen, gegenüber. Die Beschreibungen sind in der Regel bewusst einfach gehalten mit dem Ziel, die Funktionen und Funktionsweisen auf eine Art zu erläutern, die auch ohne technischen Jargon auskommt.

Auch bei der Implementierung der Funktionen war es nicht möglich, alle Designziele vollständig zu erreichen. Die Struktur der zur Verschleierung genutzten Daten macht es unvermeidlich, dass eine Seite auf mehrere Interes

sen Einfluss nimmt. Die Verschleierung ist zwar im statistischen Mittel korrekt, in der individuellen Anwendung kann aber nicht sichergestellt werden, dass das resultierende neue Profil exakt dem von dem_der Nutzer_in gewählten entspricht (siehe Kapitel 4). Anders als bei der Obfuscation der Interessen wurde beim soziodemografischen Profil darauf verzichtet, eine „komplette“ Verschleierung zu automatisieren. Wie in Abschnitt 4.8.3 dargestellt, existiert für jede Seite ein eigenes soziodemografisches Profil, das für mehrere Attribute verschiedene Werte vorsieht und zudem regelmäßigen Änderungen unterworfen ist. Aus diesen komplexen Informationen lassen sich daher weder ein inverses Profil noch Webseiten bestimmen, die dieses inverse Profil in Gänze zur Verschleierung nutzbar machen können. Wie bei der Verschleierung des Interessenprofils, führen außerdem Abhängigkeiten innerhalb der soziodemografischen Profile der Seiten dazu, dass eine gezielte Verschleierung nicht möglich ist.

Abbildung 38 stellt die Komponenten des Browser-Plugins TrickTrack im Zusammenhang dar.

---

116 Siehe HTTP://TRICKTRACKING.COM/GOOGLES_INTERESSENLISTE.HTML (letzter Zugriff 13.02.2015).

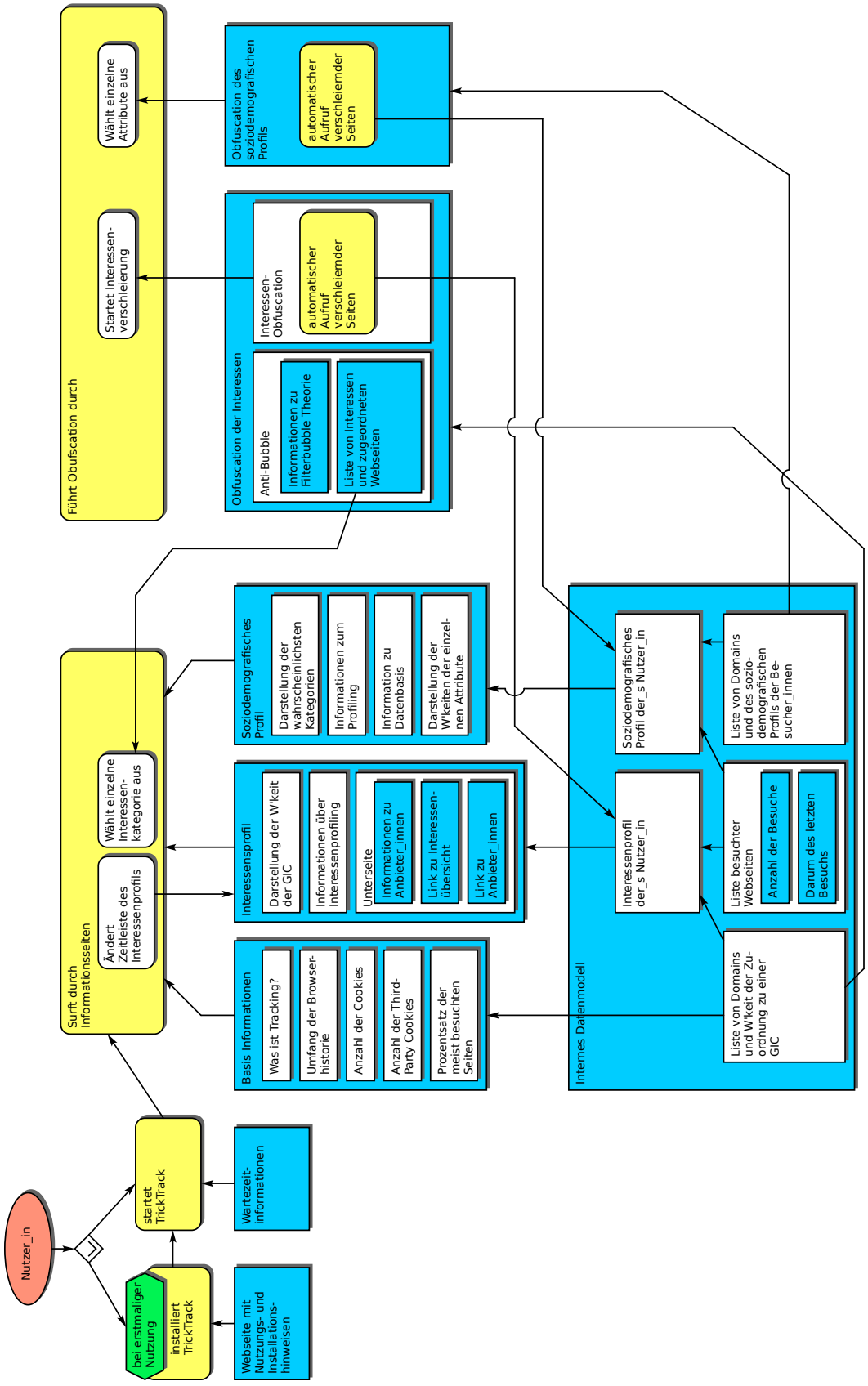


Abbildung 38: Übersicht der TrickTrack Komponenten

5.3 EVALUATION

Leitfragen der Evaluation

Zur Bewertung der Nützlichkeit und Nutzbarkeit von TrickTrack wurde eine überwiegend qualitative Evaluation basierend auf Interviews und teilnehmender Beobachtung durchgeführt. Diese Methoden sind vergleichbar mit ähnlich gelagerten Studien zum Thema Online Behavioural Advertising (Agarwal u. a. 2013; McDonald und Cranor 2010) und der Usability von Privacy Enhancing Technologies mit Bezug zu OBA (Leon u. a. 2012; Ur u. a. 2012). Entsprechend der Designziele ist die Evaluation an drei Leitfragen ausgerichtet:

1. Macht TrickTrack Profiling derart transparent, dass die Funktionsweisen nachvollziehbar sind?
2. Kann TrickTrack das Verständnis und die Reflexion von Online-Profiling und dessen Funktionsweise verbessern bzw. anstoßen?
3. Ist TrickTrack derart gestaltet, dass die Nutzung leicht erlernbar ist und das Programm und dessen Funktionsweisen von Benutzergruppen, die sich bislang nicht mit Profiling auseinandergesetzt haben, verstanden werden?

5.3.1 Durchführung

Aufbau der Evaluation

Die Evaluation basiert im Wesentlichen auf einer Expert_innenbefragung. Expert_innen meint im vorliegenden Fall erfahrene Internetnutzer_innen. Dabei wird einerseits impliziert, dass diese Gruppe regelmäßig mit Online-Werbung als der sichtbaren Folge von Online-Tracking in Kontakt kommt, und andererseits, dass sie für die Bewertung der Usability von TrickTrack aufgrund ihrer Erfahrung im Umgang mit gängigen Benutzungsmustern von Browsern, Webseiten und Browser-AddOns geeignet ist.

Vorbefragung

Zur Beantwortung der ersten beiden Leitfragen soll vorab geklärt werden, welchen Wissensstand und welche Einstellung zum Thema Online-Tracking die Befragten besitzen. Dazu wurde ein Interview-Leitfaden in Anlehnung an eine Studie zur Einstellung gegenüber Online Behavioral Advertising (Ur u. a. 2012) übersetzt. Der vollständige Fragebogen befindet sich im Anhang (siehe 10.9.3), die 16 Fragen richteten sich auf

- die Meinung des_der Befragten gegenüber Online-Werbung allgemein („Was denkst Du/ denken Sie allgemein über Werbung im Internet?“, Fragen 1,2.1,2.2,2.3),

- das Wissen um und die Wahrnehmung von auf Online-Tracking basierender Werbung („Inwieweit findest du/finden Sie, dass Werbung im Internet relevant ist für dich/Sie, bzw. das sie deinen/Ihren Interessen entspricht?“, Fragen 3,4,4.1,4.2,4.3,4.4,6,6.1),
- das Wissen über die Funktionsweise von Online-Tracking und entsprechender Privacy Enhancing Technologies (z.B. „Hast Du/haben Sie schon einmal davon gehört, dass Du/Sie verhindern kannst/können, dass Dir Werbung angezeigt wird? Wenn ja: Welche Tools kennst Du?“, Fragen 5,7,8)

Darüber hinaus wurden einige Rahmendaten erhoben. Dazu gehörten einerseits Informationen und Selbsteinschätzungen, die dem Vergleich mit dem von TrickTrack ermittelten Profil dienen (Interessen bei der Internetnutzung sowie soziodemografische Daten). Andererseits wurden Details zur Internetnutzung und zu den verwendeten Geräten und Browsern erfragt, die die technischen Aspekte des Tracking beeinflussen.

Nutzungsbeobachtung

Die Nutzungsphase von TrickTrack zur Klärung der dritten Leitfrage ist offen (explorativ) gestaltet. Der_m Teilnehmer_in wurde ein Link zu einer Webseite genannt, über den das Browser-AddOn bezogen werden konnte, mit der Bitte, dieses zu installieren und zu nutzen. Bei der Beobachtung wurden insbesondere die folgenden Schritte der Nutzung beobachtet:

Schritte der Nutzungsphase

1. Installation und Start von TrickTrack,
2. Nutzung der drei Informationsreiter: Basic, Interessen, Profil,
3. Nutzung der verschiedener Obfuscation-Vverfahren: Interessen-Neutralisierung, Charakteristika-Stärkung, Anti-Filterbubble.

Vor Beginn des Nutzungstests wurden die Teilnehmer_innen gebeten, während der Nutzung ihr Vorgehen, ihre Fragen und Probleme durch „Thinking Aloud“ (Boren und Ramey 2000) zu artikulieren.

Abschlussinterview und Fragebogen

Nach Abschluss der 20-30 minütigen Nutzungsphase sollte eine Nachbefragung dazu dienen, Dinge abzufragen, die gegebenenfalls bei der Nutzung aufgefallen sind, sowie sicherzugehen, dass bestimmte Dinge angesprochen werden. Die vier Fragen (siehe Anhang 10.9.2) beziehen sich auf die Wahrnehmung der Inhalte von TrickTrack (z. B. „Inwieweit würdest Du/würden Sie sagen, entspricht das Profil, das TrickTrack ermittelt hat dem, was dir/Ihnen an Werbung angezeigt wird?“) sowie die Nützlichkeit des

Tools („Würdest du TrickTrack oder ein ähnliches Werkzeug regelmäßig nutzen, um dein Profil zu überprüfen?“).

Abschließend wurden die Teilnehmer_innen gebeten, die Usability anhand des standardisierten QUIS-Fragebogens (Sarodnick und Brau 2011:190 ff) zu bewerten.

Auswahl der Proband_innen

Die Auswahl der Expert_innen für ein Interview geschah durch E-Mail-Einladungen an ca. 30 Personen, die dem Autor bekannt waren und deren Hintergrund als nicht-technisch eingestuft wurde. Voraussetzung für die Teilnahme war die regelmäßige Nutzung von Firefox (bzw. Firefox-basierten Browsern) auf einem Gerät mit Desktop-Betriebsystem. Darüber hinaus wurden die Angefragten ob über das Thema des Interviews (Onlinewerbung und Tracking) sowie die ungefähre Dauer informiert.

Interviewsituation

Die Gespräche wurden in Privaträumen durchgeführt und nach Einwilligung der Teilnehmer_innen (TN) aufgezeichnet. Im Anschluss wurden die Vor- und Nachbefragungen transkribiert. Während der Beobachtung wurden zudem Notizen angefertigt. Tabelle 20 zeigt in einer Übersicht die TN-Codes sowie Einschränkungen, die sich während der Evaluation ergaben.

Bei der Auswahl wurde darauf geachtet, dass die Teilnehmenden keinen expliziten IT- oder Informatik-Hintergrund haben. Keine_r der interviewten Personen hat einen entsprechenden Abschluss, obwohl eine Person im IT-Bereich tätig ist. Neben der geringen Gruppengröße[117] gibt es einige Gründe, warum die Auswahl der TN nicht repräsentativ ist. Alle TN besitzen einen Hochschulabschluss, überwiegend aus den Geistes- und Gesellschaftswissenschaften, und auch die Altersstruktur (20-34 Jahre) ist relativ homogen.

117 Nach (Nielsen 1994) ist Anzahl der Testpersonen zumindest für eine Usabiltity-Analyse ausreichend, um 80 % der möglichen Probleme zu identifizieren.

TN	Dauer Interview	Anmerkung
1RF4	21:06	Keine Browser-Chronik, daher keine Tests
6GL3	50:45	Nutzt AdBlockPlus
BNT1	70:33	Nutzt NoScript und AdBlockEdge
E312	63:03	Nutzt NoScript und weiteres.
GF9A	42:16	Nutzt AdBlockPlus (nicht selbst installiert)
JAR7	58:10	Nutzt AdBlockPlus
N3PP	61:16	Nutzt AdBlockPlus, Disconnect.me und Header Modifier
R2D2	67:25	Nutzt AdBlockPlus
RR75	52:30	Nutzt AdBlockPlus
Z0ZJ	53:19	Nutzt AdBlockPlus

Tabelle 20: Liste der Teilnehmer_innen und erhobenen Daten.

5.3.2 Auswertung der Interviews

Internetnutzung und Zugang

90 % der Teilnehmer_innen schätzen sich selbst als internetaffin bzw. sehr internetaffin ein, nur eine Person beschrieb sich selbst als „normal“ in Bezug auf ihre Internetnutzung. Dieses Ergebnis deckt sich mit denen der breiter angelegten Online-Studie von ARD und ZDF.[118] Die meisten Nutzer_innen derselben Altersgruppen sind täglich mehrere Stunden online – wobei mit der Verbreitung von „Always-On“-Geräten wie Smartphones eine genaue Abgrenzung der Zeiträume kaum noch möglich und sinnvoll ist.

Am häufigsten nutzen die Teilnehmer_innen Internetdienste an ihrem Laptop. Nur ein Test fand an einem Desktop-PC statt. Auch wenn alle Befragten über ein Smartphone sowie zwei Personen über ein Tablet verfügten, war der Laptop das am häufigsten genutzte Gerät, das in der Regel auch nur von der jeweiligen Person genutzt wurde.

118 Laut einer Studie von ARD/ZDF nutzen 97,6 bis 100 % der 14 bis 39-jährigen das Internet und zwar mindestens 6 Tagen die Woche, täglich im Durchschnitt 200 Minuten - vgl. HTTP://WWW.ARD-ZDF-ONLINESTUDIE.DE/INDEX.PHP?ID=503 (zuletzt abgerufen am 26.019.2016).

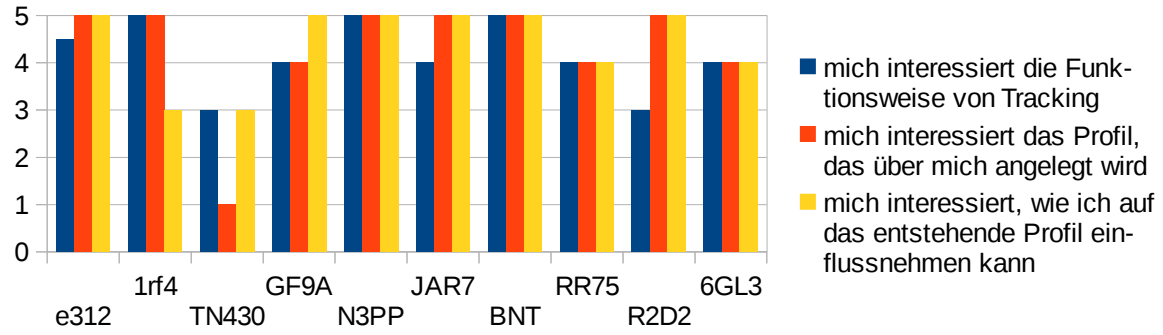


Abbildung 39: Zustimmungswerte zu Fragen, die Hintergründe von Online-Tracking und Profile betreffend.

Hohes Interesse am Thema

Zuletzt wurde in drei Varianten abgefragt, inwiefern sich die Personen für das Thema interessieren (siehe Abbildung 39). Ohne dass das genaue Ziel oder die Funktionsweise von TrickTrack vorher erklärt worden wären, aber sicherlich durch die Situation der Befragung bedingt, äußern die Teilnehmer_innen ein sehr hohes Interesse daran, die Funktionsweise von Tracking, die Profile sowie die Beeinflussung der Profile zu verstehen. Dabei ist das Interesse an den Profilen bzw. der Beeinflussung etwas höher als das an der Funktionsweise von Online-Tracking.

Vorkenntnisse

Beispiele für Targeted Advertising

Die Frage „Hast Du/haben Sie schon einmal davon gehört, dass du/Sie beeinflussen kannst/können, nach welchen Interessen dir Werbung angezeigt wird? Wenn ja: Welche Möglichkeiten kennst du?“ beantworten nur zwei Personen positiv. Beide sind Nutzer_innen von Google Mail und haben dort die Personalisierung der Werbung in einer Anzeigenleiste, die nicht von AdBlock entfernt wurde, deaktiviert. Eine TN hat außerdem bereits mehrfach bei Facebook Werbung als unpassend markiert, wenn ihm_r diese zu anstößig erschien – etwa bei Werbung für Partnerbörsen. Die Werbeeinstellungsseite von Google sowie die Opt-Out-Möglichkeit von AdChoices war allen TN vor der Befragung nicht bekannt.

AdBlocker weit verbreitet

Alle Teilnehmer_innen haben bereits ein AddOn installiert und, da es bei den meisten ein AdBlocker war (siehe Tabelle 20), kann von einer Awareness für das Thema ausgegangen werden. Zudem berichteten zwei der Befragten, dass sie das AddOn nicht selbst installiert hätten, sondern ein_e Bekannte_r es ihnen eingerichtet habe. Ein_e Teilnehmer_in hat neben einem AdBlocker, wie sich im Test herausstellte, unfreiwillig ein weiteres Plugin installiert, dass Werbefenster öffnet und damit den AdBlocker konterkarierte. Gleichzeitig berichtete ein_e TN, dass er_sie auch Freund_innen empfiehlt, AdBlocker zu nutzen und dabei hilft, sie zu installieren. 1RF4 hat einen AdBlo-

cker installiert und zusätzlich, um Tracking zu vermeiden, das Anlegen der Browser-Chronik deaktiviert.

Die Nutzung eines AdBlockers liegt damit über der von Pagefair (2015) ermittelten Rate von 25 %, wobei andere Statistiken darauf hindeuten, dass die Nutzung bei Firefox-Nutzer_innen jüngeren Alters höher liegt[119].

Funktionsweise von Cookies bekannt

Auf die Frage, was sie über die technischen Vorgänge beim Tracking und Profiling wissen, nannten vier Personen das Prinzip der Cookies zur Wiedererkennung von Nutzer_innen. Auch hier fehlt aber Wissen über die genauen technischen Hintergründe:

> Ich könnte es dir jetzt nicht explizit beschreiben, aber ich weiß, dass es damit zusammenhängt, dass es Cookies gibt, die Webseiten halt bei mir speichern. Das in diesen Cookies natürlich auch vorhergegangenen Suchanfragen enthalten sind, die dann wiederum natürlich ausgewertet werden können von diesen Webseiten. (RR75)

Wenig Kenntnisse über die Funktion von Profiling

Wenig konkretes Wissen ist zu Online-Profiling und Personalisierung vorhanden. Viele kennen das Prinzip von Personalisierung, bei dem die Zusammenhänge leicht nachvollziehbar sind, zum Beispiel bei Empfehlungen auf Amazon oder auf Profilinformationen bezogene Werbung bei Facebook. Nur grobe Vorstellungen haben die TN in Bezug auf die Funktionsweise des übergreifenden Profilings von Nutzer_innen durch Tracking.

> Ich glaube […], dass die Daten sammeln und dass sich daraus eine Errechnung ergibt. Also ich weiß nicht genau […] wie das dann tatsächlich berechnet wird, also wie das System das macht. Aber ich glaube, dass die Daten sammeln und ausdeuten. (GF9A)

> Es muss ja so funktionieren […], dass man auf Grund bestimmter festzustellender Größen ein Raster erstellt. Also sich überlegt, dass ein Mann Anfang 30, der Sportseiten guckt, vielleicht dieses [Interesse hat][…], so stell ich mir das vor, dass man es durch Rasterbildung macht. (E312)

Welche „Daten“ und „festzustellenden Größen“ zum Profiling herangezogen werden, können 9 von 10 TN nicht beantworten. Nur eine Person (N3PP) wusste „das […] mein Browser erzählt, was ich für ein Betriebssystem habe, wo ich ungefähr sitze.“ Die Informationsseiten der Werbeanbieter_innen, die in 3.5.2 ff. dargestellt sind, waren allen TN nicht bekannt. Dies deckt sich mit den Ergebnissen von Leon et al. (2012).

---

119 Verlässliche Zahlen hierzu liegen nicht vor, einzelne Statistiken weisen aber darauf hin HTTP://ONLINEMARKETING.DE/NEWS/INFOGRAFIK-WIE-ADBLOCKER-DAS-INTERNET-ZERSTOEREN-WERDEN (zuletzt aufgerufen am 26.09.2016).

EINSTELLUNG GEGENÜBER ONLINE-WERBUNG

Als Einstiegsfrage wurden die TN gebeten, ihre erste Assoziation zu dem Thema „Online-Werbung“ zu benennen. Die Antworten lassen sich in zwei Kategorien einteilen. Die einen nannten Formen von Werbung („Popups“ oder „Banner“), die anderen ihren Umgang damit („AdBlocker“).

Positive Aspekte von Werbung

Bezogen auf die guten Aspekte, die sie mit Werbung im Internet verbinden, finden sich viele Aussagen, die sich auf die Möglichkeit der Finanzierung von Internetdiensten beziehen, z. B. „Ich hätte die meiste Zeit meines Lebens keine E-Mail-Adressen, wenn nicht irgendwelche Leute das kostenlos anbieten, weil sie wieder mit Werbung Geld machen“ (BNT). Auf die Werbung selbst bezogen wird vor allen Dingen der Nutzen klassischer Werbung genannt: das „Aufmerksamgemacht-Werden“ auf ein Produkt oder einen Dienst, insbesondere auf Seiten, die sich einem speziellen Thema widmen und dort passende Werbung einblenden.

> Es gibt tatsächlich eine einzige Seite, bei der ich bewusst den AdBlocker ausgestellt habe, und das ist [XXX], weil ich mich da informiere darüber, in meinem Bereich, was gerade aktuell passiert, was ich sonst wahrscheinlich mit Mehraufwand herausfinden müsste auf vielen Einzelseiten. (N3PP)

Solche TN, denen bei der Frage nach guten Aspekten ein konkretes, positives Beispiel für personalisierte Werbung einfiel, fügten meist hinzu, dass es ihnen aber in der Regel missfalle, wenn so etwas geschieht.

Störende Aspekte von Werbung

Am meisten stört die Befragten vor allen Dingen solche Online-Werbung, die versucht, die Aufmerksamkeit auf sich zu lenken (durch Popups oder automatisch startende Multimedia-Elemente). Außerdem missfällt neben stark personalisierter Werbung auch solche, die als *Native Advertising*[120] bezeichnet wird. Dabei handelt es sich um Werbung, die nicht das redaktionelle Angebot umrahmt, sondern sehr stark mit den regulären Inhalten einer Webseite verwoben ist. Als häufigstes Beispiel wurde hier die Werbung genannt, die in den Facebook Newsfeed eingebunden ist und sich optisch kaum von den Beiträgen anderer Nutzer_innen unterscheidet.

ERFAHRUNG MIT TARGETED ADVERTISING

Fast alle Befragten haben eigene Nutzungserfahrung mit Targeted Advertising. Sei es mit den – von AdBlock ausgenommen – Werbeanzeigen bei Google, den eingebetteten Anzeigen bei Facebook, Amazon-Empfehlungen, personalisierten Newslettern oder

120 Siehe z. B. den Beitrag von Even (2014) für den Bundesverband der digitalen Wirtschaft: „Native Advertising für nahtlose Werbung im Content„ HTTP://WWW.BVDW.ORG/MEDIEN/NATIVE-ADVERTISING-FUER-NAHTLOSE-WERBUNG-IM-CONTENT?MEDIA=5905 (letzter Zugriff 26.09.2016).

aber auch mit klassischen Banneranzeigen. In der letzten Kategorie sind Re-Targeting Anzeigen für die Nutzer_innen am auffälligsten. Hierbei werden Anzeigen für Produkte in Online-Shops geschaltet, die vor einiger Zeit – die Angaben der TN reichen von Tagen bis Wochen – angeschaut, aber nicht gekauft wurden. Diese Form des Targeting scheint ihre Wirkung nicht zu verfehlen. Zwei TN berichten, dass sie zwar nicht genau dieses Produkt gekauft hätten, aber auf Grund der Werbeanzeige wieder den Shop besucht und ein anderes Produkt gekauft hätten. Allerdings wird diese Erfahrung im Anschluss dann nicht unbedingt als positiv wahrgenommen:

> Mich ärgert es eigentlich immer dann besonders, wenn ich merke, dass es eben doch funktioniert. (Z0ZJ)

Viele empfinden Personalisierung allerdings vor allem als aufdringlich.

> Aber Amazon ist zum Beispiel einfach anmaßend, was die glauben, was mit einem Thema zu tun hat. (GF9A)

> Ich finde das eher gruselig. (JAR7)

Unbehagen gegenüber Personalisierung

Die Beschreibung als gruselig impliziert, dass von den TN wahrgenommen wird, das Personalisierung stattfindet, aber nicht klar ist, wie und auf Basis welcher Informationen sie durchgeführt wird. Dabei erscheint vielen, auch gerade wenn sie um den Zweck des Manipulationsversuchs wissen, die Werbung zu invasiv.

> Ich finde es schwierig, weil eigentlich finde ich das problematisch, weil es gibt eben diese [Marke], und die kriegen das sehr, sehr gut hin, das Kaufverhalten auszuwerten und eben zu gucken […] was könnte diesem Kunden noch gefallen, und das funktionierte in diesem Fall durchaus schon ein-, zweimal. (Z0ZJ)

> Ich glaube, ich fühle mich weniger […] spezifisch zum Objekt der Werbung degradiert, wenn das Werbung ist, die jede_r andere auch gerade sehen würde, in dem Moment. (R2D2)

Andersrum formulieren es zwei weitere Befragte:

> Ich fände es schöner, wenn ich alleine auf die Produkte komme. (1RF4)

> So konsumiere ich nicht. (BNT)

Während 1RF4 den Wunsch äußert, sich autonom im Warenangebot zu bewegen und eigenständig Produkte zu entdecken, ist sich BNT sicher, ihr_sein Konsumverhalten gegenüber neuen Werbestrategien behaupten zu können.

Jenseits des Unbehagens gegenüber den Versuchen der Beeinflussung äußerten die Teilnehmer_innen auch den Wunsch nach nicht-personalisierter Werbung aus Gründen, die eigentlich im Sinn der Werbenden sind.

> Eigentlich finde ich es tatsächlich interessant, […] von Dingen zu wissen, die es so gibt, für die ich aber vielleicht keinen Bedarf habe. (N3PP)

Einige berichteten, dass sie ja 'eigentlich' wüssten, dass sie nur das richtige Plugin benötigten würden, um dem zu entgehen, aber in der Regel die Dringlichkeit nicht gegeben ist, die Aktivitäten an dieser Stelle zu verstärken.

Personalization Reactance

Interessant ist, wie die Befragten zwar einerseits Beispiele nennen können, in denen sie von personalisierter Werbung „profitiert“ haben, andererseits alle Befragten diese trotzdem ablehnen. Diese Abwehrhaltung beschreiben White u. a. (2007) als *Personalization Reactance*. In einer Studie zu E-Mail-Werbung zeigten sie, dass die Einstellung gegenüber Personalisierung abhängig ist von der wahrgenommen Nützlichkeit des durch die Personalisierung angebotenen Produkts oder Dienstleistung sowie der Begründung der Personalisierung. Wird eine Dienstleistung als nicht besonders nützlich wahrgenommen, was bei Werbung häufig der Fall ist, wird (übermäßige) Personalisierung als nicht gerechtfertigt wahrgenommen. Ähnliche beschreiben Malheiros u. a. (2012) den Effekt von Personalisierung mit Fotos.

BEFÜRCHTETE FOLGEN VON PROFILING

An die Befragung zum Thema Online-Werbung schlossen sich einige Fragen nach der Einschätzung persönlicher und gesellschaftlicher Folgen von Profiling an. Teil der Einleitungsfragen war, welche Szenarien mit negativem Ausgang sich die_der TN vorstellen könnte, die mit dem Profiling bei Online-Werbung zusammenhingen. Dabei sollten einerseits konkrete Szenarien genannt, andererseits eine langfristige Perspektive betrachtet werden.

Unfreiwillige Preisgabe

Die konkreten individuellen Folgen, die die TN beschrieben, bezogen sich in weiten Teilen auf unfreiwillige Datenpreisgaben, z. B. weil andere Personen den eigenen Browser benutzten. So könnten Dritten Dinge bekannt werden, die nicht vollständige Profile, aber doch sensitive Informationen enthalten [„Also mal angenommen du bestellst […] irgendeinen Ratgeberbuch zu irgendeiner Problemlage oder so“ (Z0ZJ)]. Eine zweite Sorge bezog sich auf die Gefahr, dass die Daten zweckentfremdet werden, etwa zur Überwachung:

> Ich würde das sehr negativ einschätzen, weil, um es mal ein bisschen runter zu brechen, im Grunde Unternehmen, die diese Dienste anbieten und diese Daten sammeln, um diese Werbung anzubieten, auf kurz oder lang zu den Instrumenten von Staatsorganen oder möglicherweise Kriminellen werden können, die den Leuten schaden wollen. (R2D2)

Überwachung und Informationelle Selbstbestimmung

Dabei bezieht sich BNT auf eine allgemeinere Beschreibung dessen, was als *informationelle Selbstbestimmung* bezeichnet werden kann:

> Also, wenn diese Datenverbünde irgendwie einmal da sind, kriegt man die erst einmal nicht mehr weg. Die können dann überall wieder auftauchen, in jedem Kontext und an der Stelle wird es dann gesellschaftlich. Es stellt sich dann irgendwann die Situation ein, dass ich eigentlich erwarten muss, dass, wenn irgendjemand etwas über mich weiß oder irgendwie über mich akkumuliert hat, dass ich davon ausgehen muss, dass jeder das hat. […] Ich fühle mich aber noch nicht so. (BNT)

> [Es ist] zum einen eine Souveränitätsgeschichte […], wenn man dann viel Online-Shopping betreibt und dadurch irgendwie Werbung und Daten generiert, die dann für SCHUFA-Auskünfte und so was […] eine Rolle spielen könnten. Und […], wenn man dann eher politische Seiten konsumiert […], dass man dann in verschiedene Verdachtssphären geraten kann. (JAR7)

Veränderung des Zusammenlebens

Darüber hinaus befürchten GFA9 und Z0ZJ eine Veränderung des Zusammenlebens:

> Dadurch, dass wir immer individueller betrachten werden, dass die Leute das auch glauben […], ich glaube, dass das eine totale Gefahr ist […], weil […] es gibt dann so was wie eine Form der Gemeinschaft nicht mehr. Weil für jeden die Werbung […] total zugeschnitten [ist]. Und mit dieser Haltung geh ich dann raus in Kontakt […]: „wenn du mich nicht total individuell behandelst, so wie das selbst mein Computer schon kann, wenn ich im Internet surfe oder so“. Und ich glaube einfach, dass dieses Tempo, das da vorgelegt wird und das, wie die reale Welt eben aussieht, echt auseinanderklafft. (GFA9)

> Ansonsten kann ich mir schon vorstellen, […][dass] sich ein Bezug zur Welt auch verändert. […] „Ich beweg mich jetzt gerade hier, was habe ich zur Verfügung an diesem Ort, wo kann ich shoppen gehen“ oder wie auch immer, das verändert natürlich auch meinen Blick auf etwas wie Stadt […]. (Z0ZJ)

ZUSAMMENFASSUNG UND DISKUSSION

Die Interviews bestätigen in weiten Teilen die Erkenntnisse anderer Studien (Agarwal u. a. 2013; Malheiros u. a. 2012; Ur u. a. 2012). Online-Werbung wird häufig als aufdringlich wahrgenommen und insbesondere solche Werbung, die auf Profiling basiert, wie *targeted advertising*, stellt für Betroffene einen Grenzübertritt dar, ab dem der invasive Charakter nicht mehr toleriert wird. Die Verbreitung von AdBlockern ist unter den Befragten unter anderem deswegen so hoch, weil sie um die Zwecke der Beeinflussungsversuche, nämlich der Manipulation des Konsumverhaltens, wissen und teilweise auch schon bemerkt haben, dass sie funktioniert.

Im Rückblick auf die Diskussion in Kapitel 2, kann dieses Verhalten als einen Versuch gewertet werden, Autonomie zu erhalten. Einige Teilnehmer_innen äußerten, dass ihnen durchaus bewusst sei, dass die Algorithmen häufig nicht sonderlich komplex sind (der Empfehlungs-Algorithmus von Amazon ist für viele transparent und nachvollziehbar), dennoch missfällt ihnen schon der Versuch der Beeinflussung. Dazu kommt die

Sorge vor dem Erfolg der Personalisierung und den möglichen Folgen einer Internalisierung der Techniken, denen zugetraut wird, das gesellschaftliche Zusammenleben als Ganzes zu verändern.

Darin gründet auch der Wunsch, die Technik besser zu verstehen und insbesondere die Folgen, also die Profile selbst, beeinflussen zu können. Gleichzeitig sind bereits existierende Strategien und Privacy Enhancing Technologies, wie sie in Kapitel 3 vorgestellt wurden, nicht bekannt.

5.3.3 Auswertung Beobachtung

Nach Abschluss des Interviews wurden die Teilnehmer_innen über die Vorgehensweise beim „Thinking Aloud“ aufgeklärt und anschließend gebeten, die Seite HTTP://TRICKTRACKING.COM aufzurufen sowie das AddOn von dort zu installieren. Bis auf eine Ausnahme, bei der ein Programmierfehler die Funktionsfähigkeit von TrickTrack unter bestimmten Umständen einschränkte, wurden ab diesem Moment keine weiteren Hinweise gegeben. Kamen die Nutzer_innen in diese Situation, hat der Beobachter interveniert, um den Fehler zu umgehen. Im Folgenden werden die einzelnen Elemente der Nutzung durchgegangen. Dabei wird zuerst das Vorgehen und dann Beobachtungen und Probleme beschrieben.

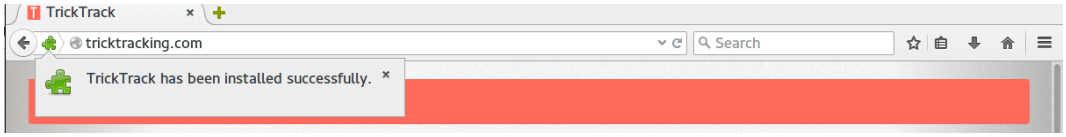


Abbildung 40: Bestätigungsnachricht über die erfolgreiche Installation. Erst im Anschluss erscheint das Icon auf der äußeren rechten Seite.

Installation

Installation erfolgreich

Die Installation wurde von alle Teilnehmer_innen erfolgreich ausgeführt. Die Installation des AddOns startete automatisch nach dem Klick auf den entsprechenden Link. Im Anschluss waren zwei Bestätigungen notwendig, unter anderem da das AddOn nicht signiert war, was zu einem Warnhinweis führte.

Probleme beim Starten

Das erste Problem trat nach dem Abschluss der Installation auf. Für einige Teilnehmer_innen war nicht klar, wie das AddOn zu starten ist. Nach der erfolgreichen Installation, die am linken oberen Rand des Fensters bestätigt wird (vgl. Abb. 40), erscheint das Icon in der AddOn Leiste am oberen rechten Bildschirmrand. Die Veränderung am rechten Rand wurde durch den größeren Hinweis am linken Rand übersehen. Hauptsächlich diejenigen, die bisher keine AddOns selbst installiert hatten oder deren AddOn Leiste nur AdBlockPlus enthielten, waren sich nicht darüber im Klaren, wie das

AddOn zu starten ist. AdBlockPlus startet nach der Installation automatisch und es braucht keine weitere Interaktion durch den_die Nutzer_in. Es dauert daher in einigen Fällen bis zu zwei Minuten bis TrickTrack gestartet wird.

Nach dem Start zeigt TrickTrack zuerst einen Dialog an, der darüber informiert, dass der Browserverlauf ausgelesen werden wird und deswegen der Browser möglicherweise eine Weile nicht reagieren könnte. Dieser Hinweis wurde positiv aufgenommen.

Basisinformationen

Ausnahmslos kritisiert wurde die automatisiert ablaufende Kurzeinführung durch kleine Informationsfelder, die neben Elementen, die sie beschrieben, angezeigt wurden. Die Animation der Informationen verlaufe zu schnell und lenke ab von dem Text, den sie gelegentlich überdecke.

Unklarheit über Begriffe

Nachdem TrickTrack Daten eingelesen hat (siehe Tabelle 21), zeigt eine Übersicht die Auswertung des Browserverlaufs und der gespeicherten Cookies. Obwohl, wie sich in den Interviews gezeigt hat, die Teilnehmenden über die Existenz und den Nutzen von Cookies informiert sind, ist die Unterscheidung zwischen *1st*- und *3rd-Party*-Cookies den meisten nicht bekannt. Sie zeigten sich überrascht über den Umfang der im Browser hinterlegten Informationen.

TN	Webseiten-aufrufe im letzten Jahr	Domains	Cookies	Davon 3rd-Party-Cookies	Verhältnis TopTen
6GL3	5527	219	358	174	75 %
BNT1	6979	526	278	74	52 %
E312	3328	390	832	530	51 %
GF9A	30941	919	1492	718	60 %
JAR7	10975	3087	1299	331	49 %
N3PP	79587	2709	1448	426	48 %
R2D2	47990	2665	1012	265	61 %
RR75	21729	2037	1245	355	44 %
Z0ZJ	87001	1653	1264	296	77 %

Tabelle 21: Übersicht über den Umfang der analysierten Daten pro TN, die in der Basis-Übersicht angezeigt wurde.

Unklarheit über Begriffe

> Es hat mich schon überrascht, dass da dann doch so viele Cookies sind, die man dann doch nicht will […]. (JAR7)

Vergleichbarkeit gewünscht

Die Aufmerksamkeit für die Statistiken blieb in der Regel kurz. Mehrfach wurde der Wunsch geäußert, die Zahlen in einen Kontext zu setzen, um einschätzen zu können, ob z. B. die Anzahl der Cookies im Vergleich zu anderen Nutzer_innen hoch ist.

> Diese ganzen Zahlen bei den Basics sagen mir erst mal nicht so viel, weil ich nicht so den Vergleich hab. Ich weiß nicht wie viel so der Durchschnittsbürger durchschnittlich in 84 Tagen surft. (6GL3)

Mehr Zeit verbrachte ein Teil der Nutzer_innen bei der Auflistung der besuchten Domains nach Häufigkeit des Aufrufs.

> Das ist absurd, weil ich anhand der Seiten, die da stehen, selber meine Geschichte erzählen kann. (RR75)

Anders als die Übersicht über die Anzahl von Cookies und Webseitenaufrufen, kann die Liste der vielbesuchten Seiten offensichtlich dazu dienen, eine Beziehung herzustellen zwischen den aus dem Browserverlauf abgeleiteten Daten und der Realität des_der Nutzer_in.

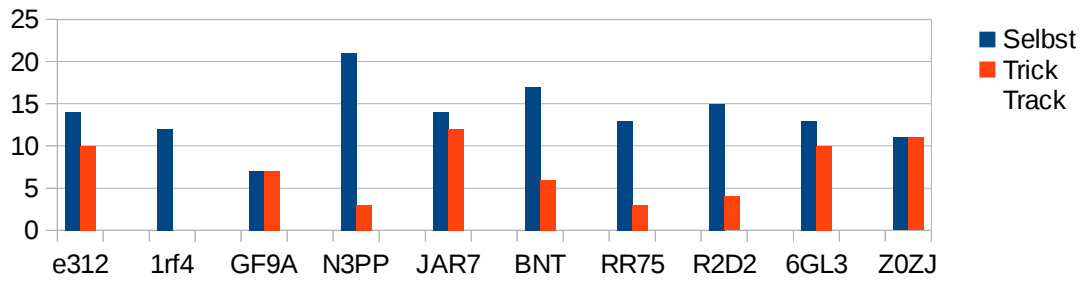
Vergleich von Selbsteinschätzung und Interessenprofil

Vergleich der Selbsteinschätzung und des Interessenprofils

In der Vorabbefragung wurden die TN gebeten, aus dem Gedächtnis heraus die Webseiten, die sie in den letzten Wochen besucht haben, in das Kategoriensystem von Google einzusortieren. Abbildung 41 zeigt den Vergleich dieser Selbsteinschätzung mit dem von TrickTrack ermittelten Interessenprofil, der Breite der Profile anhand der Anzahl der selbst zugewiesenen Interessen und den Interessen, die TrickTrack mit mehr als 10 gewichtet hat. TrickTrack stellt zwar alle Interessen dar, für die sich eine Gewichtung größer als Null ergibt, allerdings ist erst ab einer Gewichtung von etwa 10 die Darstellung derart, dass man erkennen kann, welches Interesse durch einen Kreis repräsentiert wird.

TrickTrack stößt Reflexion an

Auch wenn die Methodik und TrickTrack mit den Selbsteinschätzungen nicht vergleichbar ist, haben gerade die Widersprüche zu einer Reflexion geführt: „[…] dafür interessiere ich mich also. Das hab ich ganz gut angegeben, oder?“ (6GL3). Den Daten von TrickTrack wurde darüber hinaus mehr Objektivität zugesprochen.



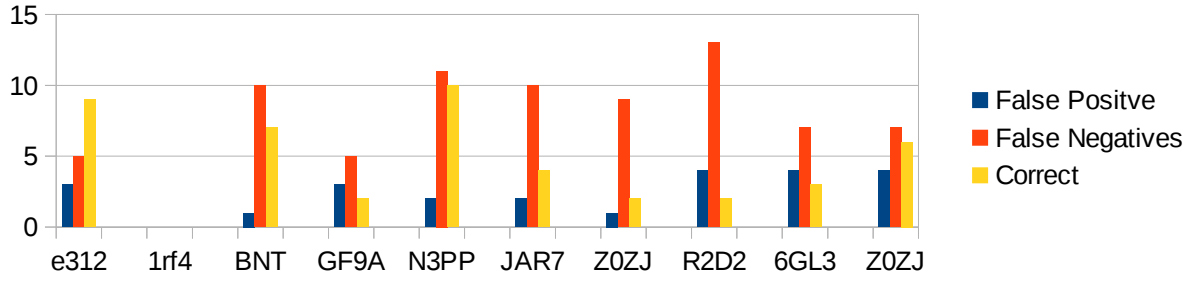


Abbildung 41: Oben: Anzahl der selbst zugewiesenen Interessen (blau) und der von TrickTrack ermittelten (orange; Wert > 10); unten: Vergleich der False Negatives, False Positives und korrekt zugewiesenen Interessen.

Ein_e TN, die nach eigenen Angaben auf Wohnungssuche war, wunderte sich über das als gering dargestellte Interesse an „Immobilien“.

> [Die Zuweisung von] „Immobilien“ ist tatsächlich komisch, weil das ist super klein, das wundert mich tatsächlich ein bisschen. […] das kommt hier immer nur als superkleiner Fleck vor, das wundert mich ein bisschen. Vielleicht aber, weil ich immer nur auf einer Seite war, oder auf zwei. (Z0ZJ)

Die Möglichkeit, den Zeitraum der Berechnung für das Interessenprofil anzupassen, wurde hier nicht wahrgenommen. Allerdings führte auch hier der Unterschied zwischen Selbsteinschätzung und dargestelltem Interessenprofil nicht zum Widerspruch, sondern zur Reflexion über die Funktionsweise des Profilings. Ein anderer TN begann nach einiger Zeit dann auch die Kategorisierung an sich zu hinterfragen.

> Ich frag mich schon irgendwie, wie diese Kategorien zustanden kommen. Zum Beispiel *Gesetze und Regierungen*, ich hab das zwar angekreuzt, aber ich frag mich das schon. Also liegt das an einem Zeitungsartikel, den ich lese? Wobei, hier gibt es auch noch „Nachrichten“. (R2D2)

Fehler in der Darstellung

Bei der Nutzung des Informationsfensters, das beim Klick auf die Interessenkreise angezeigt wurde, zeigten sich zwei Probleme. Das erste bezieht sich auf einen Fehler in der Implementierung, der dazu führte, dass für alle Interessen die Zahl der Seiten mit Null angegeben wird. Auf diesen Fehler wurden die TN hingewiesen. Ein weiterer Fehler findet sich in der Liste der Seiten, die eine besonders hohe Gewichtung in Bezug auf das jeweilige Interesse haben. Diese Liste enthält in vielen Fällen überwiegend solche Domains, die relativ vielfältige Inhalte darstellen und daher zu vielen Interessen in Beziehung stehen. Da die Liste nur jeweils 10 Domains darstellt, nahmen die Nutzer_innen an, es würden zu allen Interessen dieselben Seiten in Beziehung gesetzt.

Soziodemografisches Profil

Vergleich des Profils mit Angaben

Die Überprüfung der Korrektheit des berechneten soziodemografischen Profils ließe sich im Prinzip einfach ermitteln. Allerdings stießen sich einige der TN an den möglichen Ausprägungen der Kategorien. Das betraf neben der binären Angabe des Genders ebenso die Festlegung auf eine Ethnizität, der sich nicht alle Teilnehmer_innen zuordnen konnten und die sich darüber hinaus auf einen US-amerikanischen Kontext bezieht. Ebenso konnte die bei Alexa vorgenommene örtliche Zuordnungen nicht von allen TN gemacht werden, da, wie beschrieben, die meisten unabhängig vom Ort mit einem Laptop im Internet surfen.

Abbildung 42 zeigt die soziodemografischen Profile, die von TrickTrack errechnet wurden, im Vergleich mit den von den Befragten selbst angegebenen Daten. Im Durchschnitt liegt TrickTrack bei 4,37 der Eigenschaften richtig, das auf den soziodemografischen Daten zu durchschnittlich 8 % der besuchten Webseiten basiert. Die Übereinstimmungen könnten allerdings mit den in 4.8.2 beschriebenen Einschränkungen in den Daten zusammenhängen. Die Befragten zählten sich zum Großteil zu den Hauptgruppen der Internetnutzer_innen, auf die sich auch die Statistiken beziehen. Zudem deckt sich das durchschnittliche Profil der Befragten auch zu großen Teilen mit dem durchschnittlichen Profil der Reddit-User (siehe 18).

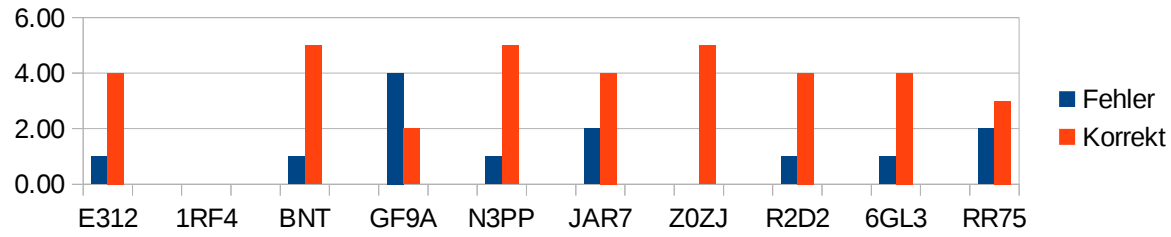


Abbildung 42: Vergleich der Eigenangaben mit dem von TrickTrack ermittelten soziodemografischen Profil (ausführliche Tabelle im Anhang 3).

Im Vergleich zum Interessenprofil führt die explizierte Darstellung des soziodemografischen Profils bei den Testpersonen zu klareren Positionen, insbesondere wenn mehrere Attribute fehlerhaft zugewiesen wurden.

> Beim Profil hat mich wirklich überrascht, dass da so wenig zustimmt. (JAR7)

Fehler werden deutlicher wahrgenommen

Das Zitat zeigt, dass bei dieser Darstellung die Fehler wesentlich deutlicher wahrgenommen wurden. Für JAR7 wurden nur zwei Eigenschaften (Geschlecht und Ethnie) falsch bestimmt, während 4 richtig ermittelt wurden (Einkommen, Bildungsabschluss, Alter, Kinder). Dennoch wurde die Fehlerquote als sehr hoch wahrgenommen. Auf der anderen Seite wurden auch zutreffende Attribute bemerkt, worüber E312 sein Unbehagen ausdrückte

> Wenn sie treffen, fühlt man sich irgendwie ertappt, beziehungsweise hat irgendwie das Gefühl, dass man wegen so minimaler Parameter so berechenbar ist, dass die den Geschmack von einem kennen, und umgekehrt ist es dann auch so: „ihr Trottel“ (E312)

Das Unbehagen steht hier im Zusammenhang mit der Zusatzinformation, das TrickTrack das Profil nur auf Basis von 8 % der besuchten Domains relativ genau berechnen konnte.

Weiterführende Informationen

Filterblase

Für Verwirrung sorgte bei einigen Nutzer_innen die „Anti-Filterblase“ (siehe 5.2.4), auf der die GIC sowie Domains, die stark mit den jeweiligen Interessen verknüpft sind, gelistet werden. Das Ziel der Liste ist, Hinweise auf Webseiten zu geben, die Interessen zugeordnet sind, die dem Profil des_der User_in vielleicht widersprechen. Mehrfach äußerten Teilnehmer_innen stattdessen (und korrekt) aber, keine der Seiten aufgerufen zu haben. Zu der Seite gelangten die Nutzer_innen entweder über den Link „Entkomm' deiner Filterblase“ oder beim Klick auf ein einzelnes Interesse im Interessenprofil über „Mach' diesen Kreis größer“. Erwartet werden hier zusätzliche Informationen über das Profil in Bezug auf die Filterblase oder das jeweilige Interesse und nicht direkt Hinweise, um der Filterblase zu „entkommen“. In den meisten Fällen wurde allerdings früher oder später der Hinweis am oberen Rand der Seite gelesen und das Missverständnis aufgeklärt.

Obfuscation

Eigene Erklärungen zu Obfuscation

Alle Nutzer_innen fanden nach einiger Zeit den Weg zu einer der Verschleierungsmethoden. Dabei gab es keine Missverständnisse über die Art und Weise, wie die Verschleierung funktioniert. Die Nutzer_innen erklärten sich das Prinzip selbst.

> Es geht also TrickTrack tatsächlich darum, wahrscheinlich auch unter dem Wissen, dass man dem nie ganz entfliehen kann - oder nur schwerlich […] –, eine bestimmte Seite in eine bestimmte Richtung zu lenken. Jetzt kann ich hier verändern, was TrickTrack [meinem Profil] dann näher bringen möchte. (Z0ZJ)

> Das ist natürlich eine interessante Option, ich kann einstellen, welche Teile meiner Datensätze ich beeinflussen möchte, damit ich in bestimmte Kategorien eingeteilt werde. (6GL3)

Die Teilnehmer_innen entwickelten zum Teil Spaß daran, dem automatisierten Browsen zuzuschauen, regelmäßig die Einstellungen bei der Profilverschleierung zu ändern und wieder abzuwarten.

> Das könnte eine lustige Nebenbeschäftigung werden […], ist ein bisschen wie ein Bildschirmschoner. (6GL3)

Skepsis gegenüber automatisiertem Surfen

Allerdings zeigten viele auch eine Skepsis gegenüber dem Prinzip der Verschleierung. Mehrfach angemerkt wurde, dass die ausgewählten URLs zur Obfuscation ausschließlich englisch bzw. amerikanisch waren. Dies lässt sich auf die Datenerhebung beim überwiegend englischsprachigen Reddit zurückführen.

> Aber ich rufe ja jetzt einen Haufen US Newsseiten auf, weiß ich auch nicht, was das [Programm da] macht. Aber ich hab jetzt eine Ethnie, […] das wollte ich aber gar nicht. (JAR7)

Insbesondere bei der Verschleierung des soziodemografischen Profils waren die Effekte relativ schnell sichtbar. Durch die geringe Datenbasis, auf der die Profile errechnet wurden, führten schon wenige Seitenaufrufe zu einer Änderung der Kategorien. Da allerdings zu jeder Seite Wahrscheinlichkeiten für mehrere Kategorien vorlagen, änderte sich in der Regel nicht nur ein Attribut, was einigen Nutzer_innen aufgefallen ist.

5.4 ERGEBNIS DER EVALUATION

Das Ergebnis der Evaluation von TrickTrack ist grundsätzlich positiv. TrickTrack wurde als gut verständlich wahrgenommen und Nutzer_innen bestätigten, nicht nur etwas über ihre eigenen Profile, sondern auch über die Funktionsweise von Profiling und Tracking im Allgemeinen gelernt zu haben. In Bezug auf die zu Beginn von Abschnitt 5.3.1 genannten Leitfragen für die Evaluation ergibt sich folgendes Bild.

*Leitfrage 1: Macht TrickTrack **Profiling derart transparent**, dass die Funktionsweise nachvollziehbar ist?*

Wie oben beschrieben sind keiner der Testpersonen die Transparenzangebote der Tracking-Dienste bekannt gewesen. Die Darstellung der beiden Profiltypen wurde daher positiv aufgenommen. Durch die Fokussierung auf die Profile anstelle von Werbung, macht TrickTrack die Effekte von Tracking transparent. Das AddOn adressiert dabei das Problem, dass die TN schon dazu veranlasst hatte, einen AdBlocker zu installieren, der von diesem aber nur indirekt und meistens unsichtbar behandelt wird.

> **[Ich habe gelernt,] dass einen „AdBlock“ nicht vor Tracking schützt;** und dass [Profiling] relativ einfach […] zu beeinflussen [ist]. (6GL3)

Verbesserungsbedarf ergibt sich bei der Darstellung des Zusammenhangs von Tracking und Profiling. Insbesondere die Terminologie zur Beschreibung in der ersten „Basis“-Übersicht ist nicht allgemein verständlich.

*Leitfrage 2: Kann TrickTrack das **Verständnis und die Reflexion von Online-Profiling** und deren Funktionsweise verbessern bzw. anstoßen?*

Wie oben beschrieben führt insbesondere die Darstellung der Profile zu einer Reflexion über die Möglichkeiten und Funktionsweisen von Profiling. In den Beobachtungen äußerten die TN sowohl die Kritik an dem Kategoriensystem, das dem Profiling zugrunde liegt als auch an der begrenzten Aussagekraft und Fehlerhaftigkeit. Durch die Erläuterung der Funktionsweise von TrickTrack ist es den TN auch möglich, die Profiling-Prozesse nachzuvollziehen.

> Ihr macht im Grunde das gleiche, was die großen Konzerne machen und dadurch, dass ihr auf die gleiche Datenbank zugreift, könnt ihr abschätzen, wie die Profile aussehen, die erstellt werden (6GFA).

Im Vergleich der Darstellungsweisen der zugewiesenen Profile (Verhältnis-Darstellung beim Interessenprofil, absolute Darstellung beim soziodemografischen Profil) zeigt sich, dass die Darstellung im Verhältnis zu längeren Auseinandersetzungen führt. Durch die absolute Formulierung des soziodemografischen Profils als Satz wurden die Zuweisungen dagegen als „eindeutig“ wahrgenommen. Dies entspricht einerseits nicht der Praxis und sorgte andererseits dafür, dass sich bei falsch ermittelten Attributen entweder Erleichterung einstellte, dass Tracking doch nicht so „gut“ sei, oder Zweifel an der Funktionsfähigkeit von TrickTrack geäußert wurden. Während letzteres in einem größeren Kontext möglicherweise unproblematisch ist, könnte ersteres dem Ziel entgegenwirken, was TrickTrack eigentlich verfolgt, nämlich die Awareness für Profiling zu steigern, statt dieses auf Grund der wahrgenommenen Fehler als unwichtig abzutun.

Das soziodemografische Profil bekam wesentlich mehr Aufmerksamkeit, da es einfacher zu verstehen und einzuschätzen zu sein scheint. Gleichzeitig ist die Darstellung der Wahrscheinlichkeiten pro Kategorie in Form des Balkendiagramms zu abstrakt und wurde von den meisten TN übergangen. Die Interesseneinschätzung war für die Teilnehmer_innen allerdings schlüssiger und hat häufiger zu einer Reflexion darüber geführt, wie diese oder jene Einschätzung wohl zu Stande kommt und auch darüber ,was die eigenen Wünsche in Bezug auf Privatheit online sind:

> Ich glaube, was ich mir immer in der Vorstellung wünsche, ist natürlich, dass es am besten wenig Informationen über mich gibt und nicht die falschen Informationen. Aber das ist wahrscheinlich einfach nicht machbar. (Z0ZJ)

Bei beiden Darstellungsformen waren es vor allem die *false positives*, die die Aufmerksamkeit auf sich lenkten: wenn etwa ein Interesse an Autos unterstellt wurde, die Personen sich aber selbst nicht so einschätzten, oder eben ein falsches Gender zugeordnet wurde. *False negatives* – also zu kleine Interessenkreise oder nicht-Zuordnungen (wie „Other“ bei der Ethnizität) wurden dagegen eher positiv aufgenommen und nur wenig diskutiert.

Auch die Möglichkeiten zur Obfuscation wurden positiv aufgenommen beziehungsweise haben dazu beigetragen, ein Verständnis von Profiling herauszubilden, das um dessen Grenzen weiß.

> Das hätte ich nicht gedacht, dass das so leicht geht, beziehungsweise dass man dieselben Datensätze benutzen kann, um herauszufinden was da die interessanten Webseiten sind [und] das mit denselben Mechanismen [die Tracker benutzen]. Je genauer man Sachen analysieren kann, desto genauer kann man auch sagen, **wenn du diese Seite besuchst, wird das dein Profil so oder so ändern. Das ist ziemlich plausibel, aber da hab ich so noch nie drüber nachgedacht**. (BNT)

Der TN formuliert eine der in 2.5.3 beschriebenen Möglichkeiten, kritisch mit Daten umzugehen (*Datenanalyse als Methode, Verhalten in Data Spaces als Objekt*) und beschreibt die sich daraus ergebenen Möglichkeiten für die Transparenz von Profiling gegenüber den Nutzer_innen. Auch wenn einige Schwächen in der Umsetzung aufgedeckt wurden, so ist doch zumindest das Prinzip klar geworden.

*Leitfrage 3: Ist TrickTrack derart gestaltet, dass die Nutzung **leicht erlernbar** ist und das Programm und dessen Funktionsweisen von Benutzergruppen, die sich bislang nicht mit Profiling auseinandergesetzt haben, verstanden werden?*

Insgesamt wurde die Usability des Prototyps positiv beurteilt (siehe Abbildung 43). Alle Teilnehmer_innen fanden sich ohne Einführung oder zusätzliche Hilfestellung in dem System zurecht. Die wesentlichen Funktionalitäten wurden von allen entdeckt und ausprobiert. Einzig einige Nebeninformationen (ein Dialog zu den Details des statischen Durchschnitts der amerikanischen Internet-Community sowie die Seite, auf der alle möglichen Interessen gelistet waren) wurden nicht von allen ausprobiert.

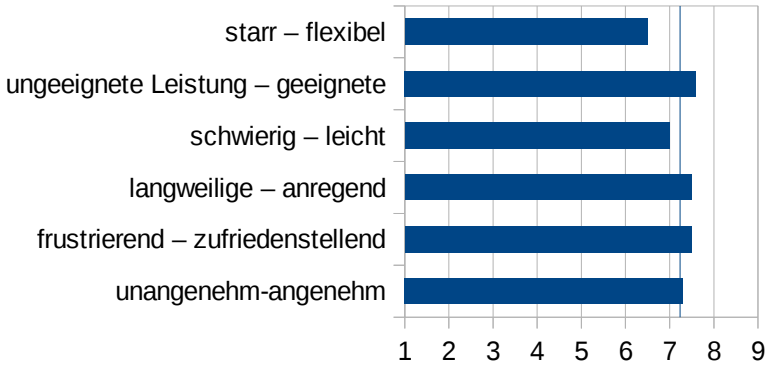


Abbildung 43: Generelle Einschätzung von TrickTrack; die Linie zeigt das Mittel.

Kritik aus Usability-Perspektive wurde vor allen Dingen an vier Punkten geäußert:

- Häufig wurde die **Einführungsanimation** kritisiert, die automatisch und ungefragt startete.

- Mehrfach wurde festgestellt, dass TrickTrack **viel Text** zur Erläuterung enthalte. Einige sahen dies aber auch als notwendig an.
- An einigen Stellen gab es Unklarheit bezüglich der **Terminologie**. Unbekannt war der Begriff der *3rd-Parties*, der weiterer Erläuterung bedarf. Auch die Kategorie „Location“ als Profilattribut erschloss sich den meisten nicht, sowie die Abkürzung „k“ für Kilo (=1000) bei der Einkommenskategorisierung war nicht allen bekannt.
- Die **Grafiken** zur Darstellung, welches Attribut je Kategorie des soziodemografischen Profils die stärkste Gewichtung erfahren hat, erschloss sich den Wenigsten (Abb. 44). Einerseits lässt sich dies darauf zurückzuführen, dass die Grafiken unterhalb der Auswahlboxen für die Profilveränderung angezeigt wurden, obwohl sie sich auf die darüber liegenden Aussagen bezogen. Andererseits wurde das Fehlen einer Erklärung der Skalen kritisiert.

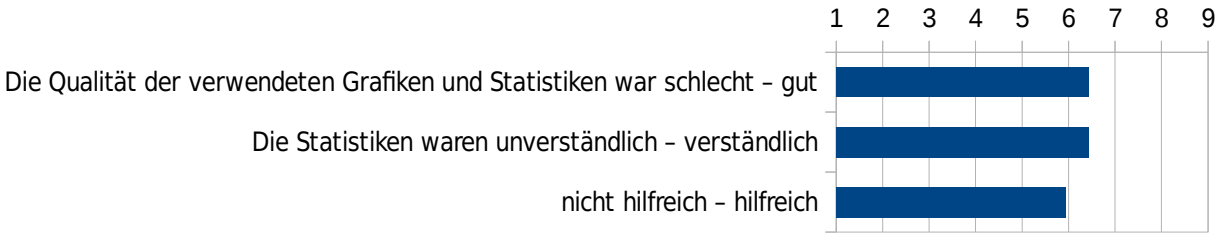


Abbildung 44: Ergebnisse des Abschnitts zu Grafiken im QUIS-Fragebogen.

5.5 VERBESSERUNGSPOTENTIALE

Aus der Evaluation ergeben sich einige Verbesserungspotentiale, die für eine Weiterentwicklung von TrickTrack in Betracht gezogen werden sollen.

1. Eine **kontinuierliche Erweiterung der Datenbasis** ist nötig, um die Genauigkeit der ermittelten Profile auch für Menschen anderer Sprach- und Nutzungsgruppen zu erhöhen. Dies betrifft vor allem die in Kapitel 4 vorgestellten Werkzeuge. Auf diese Weise kann möglicherweise auch die Ungenauigkeit, die TrickTrack durch die Verwendung Daten Dritter enthält, verringert werden.

2. Die Nutzung von TrickTrack beschränkte sich bei der Evaluation auf eine einmalige Anwendung. Damit eine **regelmäßige Nutzung**, die zu einer weiteren Förderung der privacy literacy beitragen kann, nicht nur von der Aufmerksamkeit der Nutzer_innen abhängt, muss TrickTrack (a) kontinuierlich verschleiern und (b) die Änderung der Profile über einen längeren Zeitraum verfolgbar machen können.

3. Die **Transparenz** kann sowohl im Hinblick auf Profiling als auch auf TrickTrack selbst verbessert werden. Vorstellbar ist, dass vor dem Start der Obfuscation anzeigt

wird, welche Seiten aufgerufen werden und welcher konkrete Effekt davon erwartet wird. Dies käme auch der Sorge eines TN entgegen, dass TrickTrack ungewollt Seiten aufruft, deren Besuch negative Effekte haben könnte (z. B. Bombenbauanleitungen). Diese Funktion ist bei TrackMeNot (3.5.5) implementiert.

4. Zur **Förderung von Reflexion und Artikulation** des Profiling kann vor dem Start, etwa während der Wartezeit, eine Selbsteinschätzung des Profils durchgeführt werden. Die explizite Darstellung im Vergleich zum automatisiert ermittelten Profil kann eine Diskussion über die Verlässlichkeit von Profiling anstoßen. Außerdem soll auch für die Darstellung des soziodemografischen Profils eine Verhältnis-Darstellung gewählt werde.

5. Eine **soziale Komponente** (siehe 5.1) enthält der Prototyp von TrickTrack nicht. Eine Funktion ähnlich des Signaturenvergleichs bei Floodwatch (3.5.6) kann eine bessere Einschätzung, insbesondere über den Umfang des Trackings, vereinfachen. Darüber hinaus ist auch vorstellbar, dass über ein zentrales System Surfprofile ausgetauscht werden, so dass die Datenbasis, die für die Obfuscation genutzt werden kann, vergrößert wird.

5.6 ZUSAMMENFASSUNG

In diesem Kapitel wurde die Entwicklung und Evaluation von TrickTrack beschrieben, um die letzte der, in der Einleitung gestellten, Leitfragen zu beantworten (*Wie kann das gewonnene Wissen genutzt werden, um von Profiling Betroffenen zu helfen, Profiling zu verstehen und zu beeinflussen?)*.

TrickTrack ist ein Browser-AddOn, das zum Ziel hat, eine critical data literary in Bezug auf Online-Tracking zu schulen und Möglichkeiten der Profil-Verschleierung anzubieten. TrickTrack beruht auf den in Kapitel 3 identifizierten Komponenten, die eine Privacy und Transparency Enhancing Technology zu Profiling enthalten sollte, und arbeitet mit den in Kapitel 4 erhobenen Daten und getesteten Verfahren. Die Designziele (5.1) sind, die Effekte von Online-Profiling transparent zu machen und die Grenzen zu veranschaulichen. Bekannte Fehler ähnlicher Werkzeuge, insbesondere im Hinblick auf die Usability, sollten vermieden werden.

Der Prototyp besteht aus drei Hauptelementen mit Zusatzinformationen und Verschleierungstaktiken (5.2). Das AddOn wurde für Firefox-Browser entwickelt und in einer qualitativen Studie mit 10 Teilnehmer_innen evaluiert (5.3).

Die TN äußerten sich bereits in der Vorbefragung eher kritisch gegenüber Online-Tracking und personalisierter Werbung. Sie äußerten vielfach die Bedenken, auf die bereits in Kapitel 2 eingegangen wurde.

Die Evaluation ergibt, dass der Prototyp in weiten Teil den gesetzten Zielen entspricht. Die Nutzung von TrickTrack führt zu einer Reflexion der Praxis des Online-Tracking (5.4). Einerseits dadurch, dass die Praxis des Profiling durch die Darstellung der Profile des_r Nutzer_in transparent gemacht wurde, und andererseits indem die TN durch die Nutzung der Obfuscation die Flüchtigkeit der Profile erfahren konnten. Auch aus Sicht der Usability entspricht TrickTrack in weiten Teilen den Anforderungen. Aus der Evaluation ergeben sich allerdings auch Verbesserungspotentiale (5.5), deren Umsetzung dazu beitragen kann, den Nutzen von TrickTrack in der Praxis zu erhöhen. Durch das Hinzufügen einer weiteren Komponente, die dem Austausch zwischen Nutzer_innen dienen kann, soll zudem ein Interpretationskontext für bestimmte Daten geschaffen werden.

6. FAZIT

Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit dem Phänomen des datenbasierten Profilings insbesondere zu Marketingzwecken im Internet. Dazu wurden verschiedene Anwendungsbeispiele von Profiling beschrieben und die Funktionsweise und Auswirkung auf Privatheit und informationelle Selbstbestimmung in Anschluss an philosophische, medienwissenschaftliche und datenschutzrechtliche Debatten diskutiert (Kapitel 2). Anhand mehrerer Privacy Enhancing Technologies wurde die informatische Perspektive auf das Thema untersucht und daraus Anforderungen und Verbesserungspotenziale abgeleitet. Diese beziehen sich auf die Punkte Verschleierung, Transparenz, Usability und die Förderung einer privacy literacy (Kapitel 3). In einer user_innenzentrierten Studie wurden die Ausgestaltung und der Umfang von Profilen deutlich, die auf Basis der Beobachtung des Surfverhaltens erstellt werden können. Anschließend wurden verschiedene Methoden zur Herstellung von Transparenz und Verschleierung dieser Profile erfolgreich getestet und evaluiert (Kapitel 4). Diese Verfahren wurden abschließend in einer Browsererweiterung umgesetzt und die Usability und Auswirkung auf die privacy literacy der Nutzer_innen des Tools in einer qualitativen Studie evaluiert (Kapitel 5). Als Ergebnis kann festgehalten werden, dass das Browser-AddOn TrickTrack dazu geeignet ist, Transparenz von und die Intervenierbarkeit in Profiling zu verbessern und so die Kompetenz der Nutzer_innen in Bezug auf Online-Profiling fördert.

6.1 BEANTWORTUNG DER LEITFRAGEN

Die Ergebnisse der vorliegenden Arbeit können anhand der in Abschnitt 1.3 vorgestellten Leitfragen reflektiert werden.

1. *Inwiefern werden Privatheit und Autonomie individuell wie strukturell durch technologische Entwicklungen wie Profiling beeinflusst?*

Beschreibung von Profiling

Beim Profiling werden einzelne Personen als Menge von Attributen beschrieben, die entweder direkt gemessen oder aus dem Verhältnis weiterer Attribute zueinander hergeleitet werden (siehe 2.3). Profiling ist stark verknüpft mit informationstechnischen Entwicklungen im Bereich des Data-Mining und wird vor allem zur Risikobewertung und Verhaltensvorhersage, etwa im Internet, eingesetzt. Die Folgen von Profiling sind nicht nur die Beeinflussung, sondern häufig auch Bevormundung und Diskriminierung (siehe 2.4). Profiling kann beschrieben werden als Technik des *panoptic sort* (siehe 2.4.2), an dem eine Vielzahl von Dienstleister_innen und Institutionen beteiligt ist, die Personen in jeder Situation jeweils neu anhand ihrer Profile kategorisieren und klassifizieren.

Kritik an Profiling

So wie Profiling eingesetzt wird steht es im Widerspruch zu einer Gesellschaftsordnung, die sich auf die *liberale Hypothese* beruft. In westlichen Demokratien wird jede_r Einzelne als liberales, autonomes Subjekt aufgefasst, dessen_deren Entscheidungsfreiheit sich im Recht auf informationelle Selbstbestimmung ausdrückt, welches vorsieht, dass jede_r selbst bestimmen können soll, wer was wann über einen weiß (siehe 2.1). Zwar gibt es Kritik an diesem Autonomiebegriff, insofern jener der Komplexität moderner Gesellschaften mit ihren asymmetrischen Machtbeziehungen nicht gerecht werde (siehe 2.2), aber auch unter Berücksichtigung dieser Widersprüche bleibt die Beschreibung von Profiling unvollständig. Um die Funktionsweise von Profiling besser verstehen zu können, ist eine Distanzierung von der individuellen Perspektive notwendig, die in aktuellen (datenschutzrechtlichen) Debatten unterrepräsentiert ist.

Die kybernetische Hypothese

In einer konsequent vom Profiling her gedachten Gesellschaft existieren keine einzelnen Subjekte, stattdessen werden alle als de-personalisierte, aber adressierbare Systeme wahrgenommen, deren innere Eigenschaften anhand weniger, sich ändernder, aber in jedem Fall messbarer Verhaltensweisen immer neu bestimmt werden. Dabei dient das Eigenschaften-Profil als Ausgangspunkt für Versuche der Einflussnahme durch Steuerung der Signale, die an diese Systeme (z. B. in Form von Werbung) gesendet werden. Dadurch, dass einzelne Aktionen der Systeme die Wahrnehmung der Eigenschaften und damit des nächsten Signals beeinflussen können, kann man hier von einem kybernetischen System sprechen. Die Sicht von Profiling-Diensten auf die Welt basiert auf der *kybernetischen Hypothese* (siehe 2.4.5).

Konflikt mit Privatheit

In der vorliegenden Arbeit wurde daher der Konflikt herausgearbeitet, der zwischen dem kybernetischen Profiling und einem liberal gedachten Verständnis von Privatheit entsteht. Die Grenzen der juristischen Regulierung von Profiling wurden aufgezeigt (siehe 2.2.3) und davon ausgehend die Idee entwickelt, dass Internetnutzer_innen eine Unterstützung benötigen, deren Fokus nicht auf der (Wieder-)Herstellung von Kontrolle beruht, sondern die Existenz multipler Profile sichtbar macht und Einfluss auf diese gewährt (siehe 2.5).

2. *Was sind Online-Tracking und Online-Profiling? Wie funktionieren sie, wie und zu welchem Zweck werden sie eingesetzt?*

Profiling zur Werbezwecken

Online-Tracking ist ein im Internet weit verbreitetes Verfahren, auf dessen Basis Profiling vor allen Dingen zum Zweck des behavioural oder targeted advertising betrieben wird (siehe 3.1). Dabei wird versucht, das Verhalten von Internetnutzer_innen im Browser, zum Beispiel anhand der aufgerufenen Webseiten, zu beobachten (3.2). Zur Reidentifizierung werden Cookie-Tracking, Browser-Fingerprinting und andere Techniken eingesetzt (siehe 3.3). Auf Basis der beim Tracking erhobenen Daten werden dann Profile zu den Interessen oder soziodemografischen Charakteristika der

Nutzer_innen erstellt, um auf Webseiten geschaltete Anzeigen dementsprechend anzupassen.

TrackTrack misst Umfang von Profiling

Bisherige Untersuchungen haben sich häufig auf die technischen Aspekte von Tracking beschränkt und gemessen, in welchem Umfang eine bestimmte Technik eingesetzt wird. In einer eigenen Studie wurde anhand von 506 Surfprofilen das Ausmaß des Trackings, dem ein_e einzelne_r User_in ausgesetzt ist, untersucht. Mit *TrackBack* konnte ermittelt werden, dass große Tracking-Provider wie Google über 80 % der Webseitenbesuche einer Browsersession beobachten können. Dabei ist Tracking mit Cookies immer noch die am weitesten verbreitete Technik, deren Reichweite durch Cookie Syncing erhöht wird. Um zu diesem Ergebnis zu kommen, wurde ein Verfahren entwickelt, das automatisiert Browser simuliert und die angelegten Profile abfragt. Für die Surfprofile wurden im Schnitt 94 Webseiten aufgerufen und das Tracking beobachtet (siehe 4.3).

3. *Wie sehen Profile aus, die durch Online-Tracking generiert werden und wie werden diese ermittelt?*

Analyse mit TrackBack

Für die Analyse der angelegten Profile wurde *TrackTrack* entwickelt. Der Fokus der untersuchten Anbieter_innen liegt auf dem Interessen- und dem soziodemografischen Profiling. Letzteres enthält Daten über das vermutete Geschlecht, Einkommen, Alter oder der ethnischen Zugehörigkeit von Besucher_innen einer Webseite (siehe 4.3).

Umfang des Profiling

Anhand der Surfprofile wurden die erhobenen Interessenprofile dargestellt, die Google den nicht-eingeloggten Nutzer_innen zuweist. Die Profile basieren auf einer Auswahl von 1800 Interessen, die sich auf 24 Basisinteressen, die *Google Interest Categories* (GIC), zurückführen lassen. Die Zahl der GIC wurde als Metrik verwendet, um die Veränderung der Profile unter unterschiedlichen Bedingungen zu beobachten. Google weist in der Untersuchung im Durchschnitt 8 der 24 GIC zu. Allerdings änderten sich die konkret zugewiesenen GIC bei Wiederholungen unter gleichen Ausgangsbedingungen um 40 %. Es konnte auch gezeigt werden, dass durch gezieltes Ansurfen von Webseiten, die nicht dem bisherigen Profil entsprechen, dass Profil beeinflusst werden kann. Neben den Interessenprofilen wurden soziodemografische Daten zu verschiedenen Webseiten von drei Anbieter_innen abgerufen und die Datensätze zusammengeführt, um auch hieraus Profile zu ermitteln. Eine ausführliche Auswertung ist in Kapitel 4 beschrieben.

4. *Wie kann das gewonnene Wissen genutzt werden, um von Profiling Betroffenen zu helfen, Profiling zu verstehen und zu beeinflussen?*

Entwicklung von TrickTrack

Die entwickelten Verfahren wurden angepasst und in dem Browser-AddOn TrickTrack zusammengeführt (siehe 5.2). TrickTrack zeigt Nutzer_innen an, wie die Profile aussehen, die über sie angelegt werden. Neben der Einsicht, die Nutzer_innen damit in ihre

Profile bekommen, können Sie TrickTrack auch nutzen, um ihre Profile gezielt zu beeinflussen. Ziel war es eine Möglichkeit zu schaffen, mit Profilen zu interagieren und sie durch *informed Obfuscation* zu verschleiern (siehe 4.8).

Förderung von privacy literacy

Bei der Entwicklung von TrickTrack wurden drei Designebenen besonders berücksichtigt. Erstens sollten die oben beschriebenen funktionalen Elemente der Transparenz und Verschleierung umgesetzt werden, zweitens sollten diese möglichst leicht nutzbar beziehungsweise erlernbar sein (*Usability*) und drittens sollte die Nutzung die privacy literacy der Nutzer_innen fördern (siehe 5.1). Diese Designziele wurden in einer qualitativen Evaluation mit 10 Testpersonen überprüft. Dabei bestätigte sich die Kritik an Profiling, die theoretisch hergeleitet worden war, da die meisten Befragten sich durch Profiling in ihrer Autonomie eingeschränkt sehen. In Bezug auf TrickTrack konnte außerdem nachgewiesen werden, dass die Ziele Nützlichkeit und Nutzbarkeit erreicht wurden und die Teilnehmer_innen einen Erkenntnisgewinn aus der Darstellung und Beeinflussung ihrer Profile ziehen konnten (siehe 5.4).

6.2 ZENTRALE FORSCHUNGSBEITRÄGE

Widerspruch zwischen Privatheit und Profiling

Die Beiträge dieser Arbeit zur Forschungsdiskussion liegen vor allem in zwei Bereichen. Zum einen wurde die Diskussion um Profiling über verschiedene Diskurse und Disziplinen hinweg verfolgt und die Ergebnisse der überwachungskritischen Forschung mit den Diskursen um Privatheit und informationelle Selbstbestimmung zusammengebracht. Die interdisziplinäre Auseinandersetzung hat, wie auch die Veröffentlichungen des Autors in unterschiedlichen Fachbereichen zeigen, in die verschiedenen Diskurse zurückgewirkt. So wird im juristischen/datenschutzrechtlichen Bereich die Debatte unterstützt, die eine Wiederorientierung des Datenschutzes an systemischen Fragen fordert, ohne individuelle Rechte aus dem Blick zu verlieren.

Profiling als Teil der liquid surveillance

Zweitens, konnte diese Arbeit in der medien- und kulturwissenschaftlichen Auseinandersetzung dazu beitragen, Profiling als Element derjenigen Technologien zu begreifen, die ihre Wurzeln in der Kybernetik haben. Für die Forschung, die sich mit Fragen der Überwachung beschäftigt, lässt sich aus dieser Arbeit der Schluss ziehen, dass Profiling eine (weitere) *liquide* Überwachungstechnik (siehe 2.4.4) ist, die nicht nur disziplinarisch agiert und Ausschlüsse produziert, sondern sich kontinuierlich anpasst und diejenigen, über die Profile erstellt werden, versucht zu beeinflussen. Dies macht einerseits eine aktive individuelle Gegenmaßnahme schwierig, eröffnet aber Möglichkeiten der Intervention, die hier auch vorgestellt wurden.

Nutzer_innen-zentrierte Analyse für mehr Transparenz

Im Bereich der Informatik hat diese Arbeit Beiträge vor allem im Forschungsbereich der Privacy Enhancing Technologies (PETs) geleistet. Die in Kapitel 4 vorgestellte Un-

tersuchung von Online-Profiling am Beispiel der Interessen- und der soziodemografischen Profile ist die erste dem Autor bekannte Studie, die Tracking nicht nur in Bezug auf einzelne Webseiten untersucht, sondern nutzer_innen-zentriert betrachtet. Zudem zählt sie zu wenigen Untersuchungen, die nicht die technischen Verfahren des Trackings, sondern die darauf aufbauend erstellten Profile ins Zentrum der Analyse stellt. So gelingen Einblicke in die ansonsten sehr geschlossene Blackbox des Online-Profilings, die in einer Transparency Enhancing Technology umgesetzt wurden.

Science of Obfuscation

Die vorgestellten Verfahren liefern Ansätze für eine *science of obfuscation* (siehe 3.4.2), da erstmals eine Metrik entwickelt wurde, um den Erfolg der Verschleierungsstrategie in der Praxis durch die Simulation von Profilen messbar zu machen. Darüber hinaus zeigt die Arbeit, dass eine PET, die sich mit Online-Tracking beschäftigt, so gestaltet werden kann, dass sie zeitgemäße Anforderungen an die Usability berücksichtigt und dabei auch ihre eigene Funktionsweise transparent darstellen kann. So ermöglicht die Visualisierung der Profile eine Auseinandersetzung mit dem Thema Profiling, das auch über die Verschleierung hinaus eine critical data literacy schulen kann.

6.3 ANKNÜPFUNGSPUNKTE FÜR WEITERE ARBEITEN

Anwendung auf weitere Profiling-Bereiche

Das Hauptaugenmerk dieser Arbeit lag in der Analyse des Profilings im Internet und der Verschleierung des Profilings durch Google und ähnliche Anbieter_innen. Um ein breiteres und allgemeineres Verständnis der Funktionsweise von Online-Profiling zu erhalten, müssen weitere Anbieter_innen betrachtet werden. Die erarbeiteten Verfahren sind dazu geeignet, für weitere Anwendungsbereiche angepasst zu werden und sind öffentlich zugänglich.

Verschleierung optimieren

Ziel sollte es sein, das Vorgehen beim Profiling transparenter zu machen und die Möglichkeiten der Intervenierbarkeit in Profiling durch die Nutzer_innen selbst zu erweitern. Dabei sollten die Verfahren und Metriken insbesondere zur Betrachtung von Obfuscation weiter verbessert und in der Praxis erprobt werden.

Profiling einschränken

Darüber hinaus wäre es im Sinne der Stärkung der liberalen Gesellschaft, die Ergebnisse dieser Arbeit zu nutzen, um grundsätzlich die Grenzen der Zulässigkeit von Profiling zu diskutieren, die notwendig sind, um die informationelle Selbstbestimmung auch unter den heutigen technischen Bedingungen nicht nur zu erhalten, sondern auch zu stärken.

7. LITERATURVERZEICHNIS

90. Konferenz der unabhängigen und Datenschutzbehörden des Bundes und der Länder. 2015. *Das Standard-Datenschutzmodell*. Darmstadt. Abgerufen 10. Oktober 2015 (https://www.datenschutz-mv.de/datenschutz/sdm/SDM-Handbuch_V09a.pdf).

360pi. 2014. *Approaches to Price Dynamism - When should you change your prices?* Abgerufen 4. Mai 2015 (http://discover.360pi.com/acton/attachment/9666/f-01bf/1/-/-/-/-/360RP14_Approaches_to_Price_Dynamism_1407a.pdf).

Aaltonen, Janne und Timo Ahopelto. 2015. „Method and System for Delivering Advertisments to Mobile Terminals". Abgerufen 22. Juli 2015 (http://www.freepatentsonline.com/y2015/0199725.html).

Acar, G. u. a. 2013. „FPDetective: Dusting the Web for Fingerprinters". in *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*. Berlin. Abgerufen 28. Juli 2015 (http://homes.esat.-kuleuven.be/~gacar/fpdetective/).

Acar, G. u. a. 2014. „The Web Never Forgets: Persistent Tracking Mechanisms in the Wild". in *Proceedings of the 21st ACM Conference on Computer and Communications Security*. Abgerufen 4. November 2015 (https://securehomes.esat.kuleuven.be/~gacar/persistent/the_web_never_forgets.pdf).

Acquisti, A. und J. Grossklags. 2005. „Privacy and Rationality in Individual Decision Making". *Security & Privacy, IEEE* 3(1):26–33.

Advertising Research Foundation. 2007. *The Online AdvertisingPlaybook: Proven Strategies and Tested Tactics From the Advertising Research Foundation*. herausgegeben von J. T. Plummer. Hoboken, N.J: Wiley.

Agarwal, Lalit, Nisheeth Shrivastava, Sharad Jaiswal, und Saurabh Panjwani. 2013. „Do Not Embarrass: Re-examining User Concerns for Online Tracking and Advertising". S. 8:1–8:13 in *Proceedings of the Ninth Symposium on Usable Privacy and Security*, *SOUPS '13*. New York, NY, USA: ACM. Abgerufen 10. Februar 2015 (http://doi.acm.org/10.1145/2501604.2501612).

Albrecht, Jan Philipp. 2013. *Bericht über den Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr (allgemeine Datenschutzverordnung)*. Brüssel: Ausschuss für bürgerliche Freiheiten, Justiz und Inneres. Abgerufen 3. November 2014 (http://www.europarl.europa.eu/sides/getDoc.do?type=REPORT&reference=A7-2013-0402&language=DE&mode=XML).

Alexander, James und Jonathan Smith. 2010. „Disinformation: A Taxonomy". *Technical Reports (CIS)*. Abgerufen (http://repository.upenn.edu/cis_reports/920).

American City Business Journals. 2000. „Bezos calls Amazon experiment ‚a mistake'". *Puget Sound Business Journal*. Abgerufen 24. Oktober 2014 (http://www.bizjournals.com/seattle/stories/2000/09/25/daily21.html).

Ante, Spencer E. 2013. „Online Ads Can Now Follow You Home". *Wall Street Journal*, April 29. Abgerufen 28. Juli 2015 (http://online.wsj.com/article/SB10001424127887324482504578453223207072376.html).

Arendt, Hannah. 1960. *Vita activa oder Vom tätigen Leben*. Stuttgart: Kohlhammer.

Article 29 Data Protection Working Party. 2013. *Advice paper on essential elements of a definition and a provision on profiling within the EU General Data Protection Regulation*. Abgerufen 31. Mai 2013 (http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2013/20130513_advice-paper-on-profiling_en.pdf).

Article 29 Data Protection Working Party. 2014. *Opinion 05/2014 on „Anonymisation Techniques“*. Abgerufen 7. Juli 2014 (http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf).

Ash, Tim, Maura Ginty, und Rich Page. 2012. *Landing Page Optimization: The Definitive Guide to Testing and Tuning for Conversions*. John Wiley & Sons.

Ayenson, Mika, Dietrich James Wambach, Ashkan Soltani, Nathan Good, und Chris Jay Hoofnagle. 2011. *Flash Cookies and Privacy II: Now with HTML5 and ETag Respawning*. Rochester, NY: Social Science Research Network. Abgerufen 12. September 2014 (http://papers.ssrn.com/abstract=1898390).

Backstrom, Lars und Jon Kleinberg. 2014. „Romantic Partnerships and the Dispersion of Social Ties: A Network Analysis of Relationship Status on Facebook“. S. 831–841 in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, *CSCW '14*. New York, NY, USA: ACM. Abgerufen 7. November 2014 (http://doi.acm.org/10.1145/2531602.2531642).

Balsa, E., C. Troncoso, und C. Diaz. 2012. „OB-PWS: Obfuscation-Based Private Web Search“. S. 491–505 in *2012 IEEE Symposium on Security and Privacy (SP)*.

Banse, Christian, Dominik Herrmann, und Hannes Federrath. 2012. „Tracking Users on the Internet with Behavioral Patterns: Evaluation of its Practical Feasibility“. S. 235–248 in *Information Security and Privacy Research*. Springer.

Barbrook, Richard und Andy Cameron. 1996. „The Californian Ideology“. *Science as Culture* 6(1):44–72.

Barnett, Arnold. 2004. „CAPPS II: The Foundation of Aviation Security?“ *Risk Analysis* 24(4):909–916.

Barta, Maximilian. 2014. „Tracking und Rekonstruktion von Browsinghistorien durch Betreiber von sozialen Netzwerken“. Masterarbeit, Ruhr-Universität Bochum, Bochum.

Beisenherz, Gerhard und Marie-Theres Tinnefeld. 2011. „Aspekte der Einwilligung“. *Datenschutz und Datensicherheit - DuD* 35(2):110–15.

Benhabib, Seyla. 1991. „Modelle des öffentlichen Raums: Hannah Arendt, die liberale Tradition und Jürgen Habermas“. *Soziale Welt* 147–165.

Berendt, Bettina. 2012. „Data Mining for Information Literacy“. S. 265–97 in *Data Mining: Foundations and Intelligent Paradigms*, *Intelligent Systems Reference Library*, herausgegeben von D. E. Holmes und L. C. Jain. Springer Berlin Heidelberg. Abgerufen 19. November 2014 (http://link.springer.com/chapter/10.1007/978-3-642-23151-3_12).

Berry, Michael J. und Gordon Linoff. 1997. *Data Mining Techniques: For Marketing, Sales, and Customer Support*. New York, NY, USA: John Wiley & Sons, Inc.

Bilton, Ricardo. 2012. „Ghostery: A Web Tracking Blocker that Actually helps the Ad Industry“. *VentureBeat*. Abgerufen 25. Juni 2013 (http://venturebeat.com/2012/07/31/ghostery-a-web-tracking-blocker-that-actually-helps-the-ad-industry/).

Bizer, Johann. 2007. „Sieben Goldene Regeln des Datenschutzes“. *Datenschutz und Datensicherheit - DuD* 31(5):350–56.

Boda, Károly, Ádám Máté Földes, Gábor György Gulyás, und Sándor Imre. 2012. „User Tracking on the Web via Cross-Browser Fingerprinting“. S. 31–46 in *Information Security Technology for Applications*, *Lecture Notes in Computer Science*, herausgegeben von P. Laud. Springer Berlin Heidelberg. Abgerufen 24. Juni 2013 (http://link.springer.com/chapter/10.1007/978-3-642-29615-4_4).

Boren, T. und J. Ramey. 2000. „Thinking Aloud: Reconciling Theory and Practice“. *IEEE Transactions on Professional Communication* 43(3):261–78.

Born, Achim. 2015. „Wir wollen DICH“. *iX*, Januar, 96–100.

Bosco, Francesca, Elena D’Angelo, und Elise Vermeersch. 2014. *Comparative Report on Automated Profiling in the 28 EU Member States and Switzerland*. Abgerufen 19. August 2015 (http://profiling-project.eu/wp-content/uploads/2015/01/Profiling_final_report_20141.pdf).

boyd, danah. 2010. „Making Sense of Privacy and Publicity“. Abgerufen 28. Juli 2015 (http://www.danah.org/papers/talks/2010/SXSW2010.html).

boyd, danah und Alice E. Marwick. 2011. „Social Privacy in Networked Publics: Teens Attitudes, Practices, and Strategies“. S. 1–29 in. University of Oxford. Abgerufen 18. September 2014 (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1925128).

Bozdag, Engin und Jeroen van den Hoven. 2015. „Breaking the Filter Bubble: Democracy, Design and Ethics (forthcoming)“. Abgerufen 11. November 2015 (http://www.researchgate.net/publication/276272469).

Brin, David. 1999. *The Transparent Society*. First Trade Paper Edition. New York: Basic Books.

Broderick, Mark. 2015. „What’s the Price Now?“ *Communications of the ACM* 58(4):21–23.

Brosche, Kolja. 2014. „Echtzeit-Daten werden Treibstoff digitaler Werbung“. S. 187–97 in *Realtime Advertising*, herausgegeben von O. Busch. Springer Fachmedien Wiesbaden. Abgerufen 7. Februar 2015 (http://link.springer.com/chapter/10.1007/978-3-658-05358-1_16).

Brunst, Phillip W. 2009. *Anonymität im Internet - rechtliche und tatsächliche Rahmenbedingungen*. Berlin: Duncker & Humblot. Abgerufen 4. Mai 2012 (http://digitool.hbz-nrw.de:1801/webclient/StreamGate?folder_id=0&dvs=1336110540061~835).

Brunton, Finn und Helen Nissenbaum. 2011. „Vernacular Resistance to Data Collection and Analysis: A Political Theory of Obfuscation“. *First Monday* 16(5). Abgerufen 8. Januar 2014 (http://firstmonday.org/ojs/index.php/fm/article/view/3493).

Brunton, Finn und Helen Fay Nissenbaum. 2015. *Obfuscation: a User’s Guide for Privacy and Protest*. Cambridge, Massachusetts: MIT Press.

Bull, Hans P. 2009. *Informationelle Selbstbestimmung - Vision oder Illusion?: Datenschutz im Spannungsverhältnis von Freiheit und Sicherheit*. 1. Auflage. Mohr Siebeck Gmbh & Co. K.

Busch, Oliver. 2014. *Realtime Advertising - Digitales Marketing in Echtzeit: Strategien, Konzepte und Perspektiven*. Springer Fachmedien Wiesbaden. Abgerufen (http://link.springer.com/book/10.1007%2F978-3-658-05358-1).

Butler, Eric. 2010. „Firesheep“. *Codebutler*. Abgerufen 23. September 2014 (http://codebutler.com/firesheep/).

BVDW e.V. 2014. *Realtime Advertising Kompass 2014/2015*. Abgerufen 18. Februar 2015 (http://www.bvdw.org/mybvdw/media/download/kompass-realtime-advertising-2014-2015.pdf).

BVerfG. 1983. „65, 1 - Volkszählungsurteil“. Abgerufen 20. Mai 2009 (https://dejure.org/dienste/vernetzung/rechtsprechung?Text=BVerfGE%2065,%201).

Calo, M.Ryan. 2013. *Digital Market Manipulation*. Rochester, NY: Social Science Research Network. Abgerufen 7. November 2013 (http://papers.ssrn.com/abstract=2309703).

Camenish, Jan, Pierangela Samarati, Simone Fischer-Hübner, und Maren Raguse. 2009. *First report on mechanisms*. Prime Life Project. Abgerufen 3. Oktober 2014 (http://www.primelife.eu/images/stories/deliverables/d2.1.1-first_report_on_mechanisms-public.pdf).

Campbell, Fred B. 2014. „The Slow Death of ‘Do Not Track’“. *The New York Times*, Dezember 26. Abgerufen 8. Januar 2015 (http://www.nytimes.com/2014/12/27/opinion/the-slow-death-of-do-not-track.html).

Caulkins, Jonathan P. 2004. „CAPPS II: A Risky Choice Concerning an Untested Risk Detection Technology“. *Risk Analysis* 24(4):921–924.

Cheetham, Alan H. und Joseph E. Hazel. 1969. „Binary (Presence-Absence) Similarity Coefficients“. *Journal of Paleontology* 43(5):1130–36.

Chen, Jianqing und Jan Stallaert. 2010. *An Economic Analysis of Online Advertising Using Behavioral Targeting*. Rochester, NY: Social Science Research Network. Abgerufen 15. Juli 2015 (http://papers.ssrn.com/abstract=1787608).

Chittaranjan, G., J. Blom, und D. Gatica-Perez. 2011. „Who's Who with Big-Five: Analyzing and Classifying Personality Traits with Smartphones“. S. 29–36 in *2011 15th Annual International Symposium on Wearable Computers (ISWC)*.

Coll, Sami. 2014. „Power, Knowledge, and the Subjects of Privacy: Understanding Privacy as the Ally of Surveillance“. *Information, Communication & Society* 17(10):1250–63.

Cranor, L. F. 2012. „Can Users Control Online Behavioral Advertising Effectively?“ *IEEE Security Privacy* 10(2):93–96.

D'Angelo, Gabriele, Fabio Vitali, und Stefano Zacchiroli. 2010. „Content Cloaking: Preserving Privacy with Google Docs and other Web Applications“. *Proceedignгs of 25th ACM Symposium on Applied Computing (SAC) 2010*. Abgerufen 8. Juni 2014 (http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.156.9798).

Dankar, Fida Kamal und Khaled El Emam. 2013. „A Theoretical Model for Obfuscating Web Navigation Trails“. S. 137–144 in *Proceedings of the Joint EDBT/ICDT 2013 Workshops*, *EDBT '13*. New York, NY, USA: ACM. Abgerufen 4. April 2014 (http://doi.acm.org/10.1145/2457317.2457341).

Danna, Anthony und Oscar H. Gandy. 2002. „All That Glitters is Not Gold: Digging Beneath the Surface of Data Mining“. *Journal of Business Ethics* 40(4):373–86.

Datta, Amit, Michael Carl Tschantz, und Anupam Datta. 2015. „Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination“. *Proceedings on Privacy Enhancing Technologies* 2015(1):92–112.

De Bock, Koen W. und Dirk Van den Poel. 2010. „Predicting Website Audience Demographics for Web Advertising Targeting Using Multi-Website Clickstream Data“. *Fundamenta Informaticae* 98(1):49–70.

Degeling, Martin. 2014. „Profiling, Prediction und Privatheit: Über das Verhältnis eines liberalen Privatheitbegriffs zu neueren Techniken der Verhaltensvorhersage“. S. 69–92 in *Medien und Privatheit*, *Medien, Texte, Semiotik*, herausgegeben von S. Garnett, S. Halft, M. Herz, und J. M. Mönig. Passau: Verlag Karl Stutz. Abgerufen (http://martin.degeling.com/pubs/Profiling_Prediction_Privatheit_Degeling_2014.pdf).

Degeling, Martin. 2015a. „Meet Your Online Tracking Profiles with TrickTrack“. Berlin. Abgerufen (martin.degeling.com/pubs/Meet_your_Online_Tracking_Profiles_with_TrickTrack_-_Martin_Degeling.pdf).

Degeling, Martin. 2015b. „On The Vagueness Of Online Profiling“. in *„Profile, Predict, Prevent“ Blockchain Workshops*, *Blockchain Workshops*. Paris, France. Abgerufen (martin.degeling.com/pubs/on_the_vagueness_of_online_profiling.pdf).

Degeling, Martin. 2015c. „TrickTrack - Verschleierung für Online-Profiling (Poster)“. in *Die Zukunft der informationellen Selbstbestimmung*. Berlin.

Degeling, Martin. 2016. „Googles Interessenprofiling (in Druck)“. in *Profile. Interdiszplinäre Beiträge*, *Digital Cultures*. Braunschweig: meson press. Abgerufen (martin.degeling.com/pubs/Googles_Interessenprofiling.pdf).

Degeling, Martin und Jan Nierhoff. 2013. „Privacy-By-Design am Beispiel einer Anwendung zur Unterstützung kollaborativer Reflexion am Arbeitsplatz“. S. 2060–71 in *Proceedings INFORMATIK 2013*, vol. P-220, *Lecture Notes in Informatics (LNI)*. Koblenz: Köllen Druck+Verlag GmbH.

Deleuze, Gilles. 1990. „Postskriptum über die Kontrollgesellschaften“. *L'autre journal* 1(1). Abgerufen 28. Juli 2015 (http://www.nadir.org/nadir/archiv/netzkritik/postskriptum.html).

Dixon, Pam. 2013. *What Information Do Data Brokers Have on Consumers, and How Do They Use It?* Abgerufen 17. Oktober 2014 (http://www.worldprivacyforum.org/2013/12/testimony-what-information-do-data-brokers-have-on-consumers/).

Domingos, Pedro. 2012. „A few Useful Things to Know About Machine Learning“. *Communications of the ACM* 55(10):78.

Duhigg, Charles. 2012. „How Companies Learn Your Secrets“. *The New York Times*, Februar 16. Abgerufen 31. Januar 2013 (https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html).

Dwork, Cynthia und Adam Smith. 2009. „Differential Privacy for Statistics: What we Know and What we Want to Learn“. *Journal of Privacy and Confidentiality* 1(2). Abgerufen (http://repository.cmu.edu/jpc/vol1/iss2/2).

Epp, Clayton, Michael Lippold, und Regan L. Mandryk. 2011. „Identifying Emotional States Using Keystroke Dynamics“. S. 715–724 in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, *CHI '11*. New York, NY, USA: ACM. Abgerufen 7. November 2014 (http://doi.acm.org/10.1145/1978942.1979046).

Esposito, Elena. 2007. *Die Fiktion der wahrscheinlichen Realität*. Frankfurt am Main: Suhrkamp.

EU-Parlament. 1995. „Richtlinie 95/46/EG“. *Amtsblatt* (L 281):31–50.

Farahat, Ayman und Michael C. Bailey. 2012. „How Effective is Targeted Advertising?“ S. 111–120 in *Proceedings of the 21st International Conference on World Wide Web*, *WWW '12*. New York, NY, USA: ACM. Abgerufen 22. Februar 2015 (http://doi.acm.org/10.1145/2187836.2187852).

Ferraris, Valeria, Francesca Bosco, G. Cafiero, Elena D'Angelo, und Y. Suloyeva. 2013. *Defining Profiling*. Rochester, NY: Social Science Research Network. Abgerufen 10. August 2015 (http://papers.ssrn.com/abstract=2366564).

Ferraris, Valeria, Francesca Bosco, und Elena D'Angelo. 2013. *The Impact of Profiling on Fundamental Rights*. Rochester, NY: Social Science Research Network. Abgerufen 6. September 2015 (http://papers.ssrn.com/abstract=2366753).

Floridi, Luciano. 2006. „The Ontological Interpretation of Informational Privacy“. *Newsletter ACM SIGCAS* 36(1). Abgerufen (http://www.springerlink.com/content/u72834q5105m257n/).

Foucault, M. 1977. *Überwachen und Strafen: die Geburt des Gefängnisses*. Suhrkamp, Frankfurt am Main.

Foucault, Michel. 1982. „The Subject and Power“. *Critical Inquiry* 8(4):777–95.

Fredrikson, M. und B. Livshits. 2011. „RePriv: Re-imagining Content Personalization and In-browser Privacy“. S. 131–46 in *2011 IEEE Symposium on Security and Privacy (SP)*.

Galloway, Alexander R. 2004. *Protocol: How Control Exists After Decentralization*. Cambridge, MA: MIT Press.

Galloway, Alexander R. 2010. „Black Box, Black Bloc.“ Abgerufen 10. Februar 2014 (http://cultureandcommunication.org/galloway/pdf/Galloway,%20Black%20Box%20Black%20Bloc,%20New%20School.pdf).

Gandy, Oscar H. 1993. *The Panoptic Sort - A Political Economy of Personal Information*. Boulder u.a.: Westview Press.

Gandy Jr, Oscar H. 1996. „Coming to Terms with the Panoptic Sort Oscar H. Gandy Jr.“ *Computers, surveillance, and privacy* 132.

Geambasu, Roxana, Tadayoshi Kohno, Amit Levy, und Henry M. Levy. 2009. „Vanish: Increasing Data Privacy with Self-Destructing Data“. in *Proceedings of the 18th USENIX Security Symposium*. Abgerufen 22. Juli 2009 (http://vanish.cs.washington.edu/pubs/usenixsec09-geambasu.txt).

Geiselberger, Heinrich. 2013. *Big Data: das neue Versprechen der Allwissenheit*. Berlin: Suhrkamp.

Geuss, Raymond. 2001. *Public Goods, Private Goods*. Princeton University Press.

Gola, Peter, Christoph Klug, Barbara Körffer, und Rudolf Schomerus. 2012. *BDSG Bundesdatenschutzgesetz*. 11. Auflage. Beck Juristischer Verlag. Abgerufen (http://beck-online.beck.de/default.aspx?vpath=bibdata%2fkomm%2fGolaSchomerusKoBDSG_10%2fcont%2fGolaSchomerusKoBDSG.htm).

Gomez, Joshua, Travis Pinnick, und Ashkan Soltani. 2009. *KnowPrivacy*. Berkeley, CA: UC Berkeley, School of Information. Abgerufen 25. Januar 2015 (http://www.knowprivacy.org/).

Greenhalgh, Sam. 2015. „HSTS Super Cookies“. Abgerufen 7. Januar 2015 (http://www.radicalresearch.co.uk/lab/hstssupercookies/).

Grosskreutz, Henrik, Benedikt Lemmen, und Stefan Rüping. 2010. „Privacy-Preserving Data-Mining“. *Informatik-Spektrum* 33(4):380–83.

Guha, Saikat, Bin Cheng, und Paul Francis. 2011. „Privad: Practical Privacy in Online Advertising“. S. 169–182 in *Proceedings of the 8th USENIX Conference on Networked Systems Design and Implementation*, *NSDI'11*. Berkeley, CA, USA: USENIX Association. Abgerufen 3. September 2014 (http://dl.acm.org/citation.cfm?id=1972457.1972475).

Gürses, S., B. Preneel, und Bettina Berendt. 2009. „PETs under Surveillance: A Critical Review of the Potentials and Limitations of the Privacy as Confidentiality Paradigm“. Seattle, Washington, United States. Abgerufen 28. Juli 2015 (http://www.cosic.esat.kuleuven.be/publications/article-1302.pdf).

Gutwirth, Serge. 2008. „Beyond Identity?“ *Identity in the Information Society* 1(1):123–33.

Gutwirth, Serge und Paul De Hert. 2008. „Regulating Profiling in a Democratic Constitutional State“. S. 271–302 in *Profiling the European Citizen*, herausgegeben von M. Hildebrandt und S. Gutwirth. Springer Netherlands. Abgerufen 23. Oktober 2014 (http://link.springer.com/chapter/10.1007/978-1-4020-6914-7_14).

Gutwirth, Serge und Mireille Hildebrandt. 2010. „Some Caveats on Profiling“. S. 31–41 in *Data Protection in a Profiled World*, herausgegeben von S. Gutwirth, Y. Poullet, und P. D. Hert. Springer Netherlands. Abgerufen 13. Oktober 2014 (http://link.springer.com/chapter/10.1007/978-90-481-8865-9_2).

Haggerty, Kevin D. und Richard V. Ericson. 2000. „The Surveillant Assemblage“. *The British Journal of Sociology* 51(4):605–622.

Han, Jiawei. 2011. *Data mining: concepts and techniques*. 3rd ed. Burlington, MA: Elsevier.

Hannak, Aniko, Gary Soeller, David Lazer, Alan Mislove, und Christo Wilson. 2014. „Measuring Price Discrimination and Steering on E-commerce Web Sites“. in *Proceedings of the IMC'14*. Vancouver, BC, Canada.

Hass, Berthold H. und Klaus W. Willbrandt. 2011. „Targeting von Online-Werbung: Grundlagen, Formen und Herausforderungen“. *MedienWirtschaft: Zeitschrift für Medienmanagement und Kommunikationsökonomie* 8(1):12–21.

Hassenzahl, Marc. 2010. *Experience Design: Technology for all the right reasons*. Morgan & Claypool.

Hasso-Plattner-Institut. 2012. „Schufa-Forschungsprojekt gekündigt - Hasso-Plattner-Institut“. Abgerufen 21. Februar 2015 (http://hpi.de/pressemitteilungen/2012/schufa-forschungsprojekt-gekuendigt.html).

Heller, Christian. 2011. *Post Privacy: Prima leben ohne Privatsphäre*. 1. Aufl. Beck.

Herrmann, Thomas. 2006. „SeeMe in a nutshell.“ Abgerufen 28. Juli 2015 (http://www.imtm-iaw.rub.de/imperia/md/content/seeme/seeme_in_a_nutshell.pdf).

Hildebrandt, Mireille. 2006a. „Privacy and identity“. S. 61–104 in *Privacy and the criminal law*, herausgegeben von E. Claes, A. Duff, und S. Gutwirth. Antwerp/Oxford: Intersentia.

Hildebrandt, Mireille. 2006b. „Profiling: From Data to Knowledge“. *Datenschutz und Datensicherheit - DuD* 30(9):548–52.

Hildebrandt, Mireille. 2008. „Defining Profiling: A New Type of Knowledge?“ S. 17–45 in *Profiling the European Citizen*, herausgegeben von M. Hildebrandt und S. Gutwirth. Springer Netherlands. Abgerufen 23. Oktober 2014 (http://link.springer.com/chapter/10.1007/978-1-4020-6914-7_2).

Hildebrandt, Mireille. 2012. „The Dawn of a Critical Transparency Right for the Profiling Era“. *Stand Alone* 41–56.

Howe, Daniel C. und Helen Nissenbaum. 2009. „TrackMeNot: Resisting surveillance in web search“. *Lessons from the Identity Trail: Anonymity, Privacy, and Identity in a Networked Society* 417–436.

Kamerer, David. 2013. „Estimating Online Audiences: Understanding the Limitations of Competitive Intelligence Services“. *First Monday* 18(5). Abgerufen 14. Februar 2015 (http://firstmonday.org/ojs/index.php/fm/article/view/3986).

Kamishima, Toshihiro, Shotaro Akaho, Hideki Asoh, und Jun Sakuma. 2012. „Enhancement of the Neutrality in Recommendation.“ S. 8–14 in *Decisions@ RecSys*.

Kammerer, Dietmar. 2014. „Das Ende des Privaten. Geschichten eines Diskurses“. S. 243–58 in *Medien und Privatheit*, *MTS*. Passau: Karl Stutz.

Kamp, Meike und Martin Rost. 2013. „Kritik an der Einwilligung“. *Datenschutz und Datensicherheit - DuD* 37(2):80–84.

Kamp, Meike und Thilo Weichert. 2005. *Scoringsysteme zur Beurteilung der Kreditwürdigkeit - Chancen und Risiken für Verbraucher -*. Abgerufen (https://www.datenschutzzentrum.de/scoring/2005-studie-scoringsysteme-uld-bmvel.pdf).

Kang, Jerry. 1997. „Information Privacy in Cyberspace Transactions“. *Stanford Law Review* 50:1193.

Kassner, Michael. 2010. „GoogleSharing: A way to prevent tracking by Google“. *TechRepublic*. Abgerufen 7. Februar 2014 (http://www.techrepublic.com/blog/it-security/googlesharing-a-way-to-prevent-tracking-by-google/).

Kleinz, Torsten. 2015. „Interactive Advertising Bureau sagt Adblockern den Kampf an“. *heise online*. Abgerufen 1. Oktober 2015 (http://heise.de/-2835640).

Konferenz der Datenschutzbeauftragten des Bundes und der Länder. 2015. „Entschließung zur Verfolgung des Nutzerverhaltens im Internet“. Abgerufen (https://ssl.bremen.de/datenschutz/sixcms/detail.php?gsid=bremen236.c.9759.de).

Kosinski, Michal, David Stillwell, und Thore Graepel. 2013. „Private Traits and Attributes Are Predictable from Digital Records of Human Behavior“. *Proceedings of the National Academy of Sciences* 110(15):5802–5.

Kramer, Adam D. I., Jamie E. Guillory, und Jeffrey T. Hancock. 2014. „Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks“. *Proceedings of the National Academy of Sciences* 111(24):8788–90.

Kurz, Constanze und Frank Rieger. 2011. *Die Datenfresser: Wie Internetfirmen und Staat sich unsere persönlichen Daten einverleiben und wie wir die Kontrolle darüber zurückerlangen*. 2. Aufl. Fischer (S.), Frankfurt.

Lambrecht, Anja und Catherine Tucker. 2011. *When does Retargeting Work? Information Specificity in Online Advertising*. Rochester, NY: Social Science Research Network. Abgerufen 15. Mai 2013 (http://papers.ssrn.com/abstract=1795105).

Lécuyer, Mathias u. a. 2014. „XRay: Enhancing the Web's Transparency with Differential Correlation". S. 49–64 in *23rd USENIX Security Symposium (USENIX Security 14)*. San Diego, CA: USENIX Association. Abgerufen (https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/lecuyer).

Lee, Lara und Daniel Sobol. 2012. „What Data Can't Tell You About Customers". *Harvard Business Review*. Abgerufen 2. September 2012 (http://blogs.hbr.org/cs/2012/08/what_data_cant_tell_you_about.html).

Leon, Pedro u. a. 2012. „Why Johnny Can't Opt out: A Usability Evaluation of Tools to Limit Online Behavioral Advertising". S. 589–598 in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, *CHI '12*. New York, NY, USA: ACM. Abgerufen 14. April 2014 (http://doi.acm.org/10.1145/2207676.2207759).

Lessig, Lawrence. 2006. *Code: And Other Laws of Cyberspace, Version 2.0*. New York: Basic Books.

Linden, G., B. Smith, und J. York. 2003. „Amazon.com recommendations: item-to-item collaborative filtering". *IEEE Internet Computing* 7(1):76–80.

Loser, Kai-Uwe und Martin Degeling. 2014. „Security and Privacy as Hygiene Factors of Developer Behavior in Small and Agile Teams". Turku, finland. Abgerufen (martin.degeling.com/pubs/Security and Privacy as Hygiene Factors of Developer Behavior in Small and Agile Teams - Loser Degeling.pdf).

Loser, Kai-Uwe, Martin Degeling, und Thomas Herrmann. 2012. „Power and Transparency: Asymmetries and Symmetries in Cooperation". Seattle, Washington, United States: ACM.

Lyon, David. 1998. „The World Wide Web of Surveillance: The Internet and Off-World Power-Flows". *Information, Communication & Society* 1(1):91–105.

Lyon, David. 2002. *Surveillance as Social Sorting: Privacy, Risk and Automated Discrimination*. Routledge.

Lyon, David und Zygmunt Bauman. 2013. *Daten, Drohnen, Disziplin: ein Gespräch über flüchtige Überwachung*. Berlin: Suhrkamp.

Machanavajjhala, A., J. Gehrke, D. Kifer, und M. Venkitasubramaniam. 2007. „L-Diversity: Privacy Beyond k-Anonymity". *ACM Transactions on Knowledge Discovery from Data (TKDD)* 1(1).

Majumder, Anirban und Nisheeth Shrivastava. 2013. „Know Your Personalization: Learning Topic Level Personalization in Online Services". S. 873–884 in *Proceedings of the 22Nd International Conference on World Wide Web*, *WWW '13*. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee. Abgerufen 24. Oktober 2014 (http://dl.acm.org/citation.cfm?id=2488388.2488464).

Malheiros, Miguel, Charlene Jennett, Snehalee Patel, Sacha Brostoff, und Martina Angela Sasse. 2012. „Too Close for Comfort: A Study of the Effectiveness and Acceptability of Rich-media Personalized Advertising". S. 579–588 in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, *CHI '12*. New York, NY, USA: ACM. Abgerufen 5. Mai 2014 (http://doi.acm.org/10.1145/2207676.2207758).

Matwin, Stan. 2013. „Privacy-Preserving Data Mining Techniques: Survey and Challenges". S. 209–21 in *Discrimination and Privacy in the Information Society*, *Studies in Applied Philosophy, Epistemology and Rational Ethics*, herausgegeben von B. Custers, T. Calders, B. Schermer, und T. Zarsky. Springer Berlin Heidelberg. Abgerufen 23. Juli 2013 (http://link.springer.com/chapter/10.1007/978-3-642-30487-3_11).

Mayer, J. R. und J. C. Mitchell. 2012. „Third-Party Web Tracking: Policy and Technology". S. 413–27 in *2012 IEEE Symposium on Security and Privacy (SP)*.

Mayer, Jonathan. 2014. „How Verizon's Advertising Header Works". *Web Policy Blog*. Abgerufen 28. Oktober 2014 (http://webpolicy.org/2014/10/24/how-verizons-advertising-header-works/).

Mayer, Jonathan. 2015. „The Turn-Verizon Zombie Cookie". *Web Policy Blog*. Abgerufen 15. Januar 2015 (http://webpolicy.org/2015/01/14/turn-verizon-zombie-cookie/).

Mayer-Schönberger, Viktor und Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work and Think*. London: John Murray.

McDonald, Aleecia M. und Lorrie Faith Cranor. 2010. „Americans' Attitudes About Internet Behavioral Advertising Practices". S. 63–72 in *Proceedings of the 9th Annual ACM Workshop on Privacy in the Electronic Society*, *WPES '10*. New York, NY, USA: ACM. Abgerufen 1. Februar 2015 (http://doi.acm.org/10.1145/1866919.1866929).

McDonald, Aleecia und Jon M. Peha. 2011. *Track Gap: Policy Implications of User Expectations for the „Do Not Track" Internet Privacy Feature*. Rochester, NY: Social Science Research Network. Abgerufen 6. Oktober 2015 (http://papers.ssrn.com/abstract=1993133).

McStay, Andrew. 2011. *The Mood of Information: A Critique of Online Behavioural Advertising*. London: The Continuum International Publishing Group.

Meckel, Miriam. 2012. *Vielfalt im digitalen Medienensemble - Medienpolitische Herausforderungen und Ansätze*. St. Gallen. Abgerufen 25. September 2012 (http://www.i-comp.org/en_us/resources/resources/download/1364).

Mönig, Julia Maria. 2015. „Das Private in der politischen Philosophie Hannah Arendts. Vom ‚oikos' zum Cyberspace". Dissertation, Universität Passau, Passau.

Morozov, Evgeny. 2013. *To Save Everything, Click Here: Technology, Solutionism, and the Urge to Fix Problems That Don't Exist*. London: Allen Lane.

Morris, Adam. 2012. „Whoever, Whatever: On Anonymity as Resistance to Empire". *Parallax* 18(4):106–20.

Mowery, Keaton und Hovav Shacham. 2012. „Pixel perfect: Fingerprinting canvas in HTML5". *Proceedings of W2SP*.

Müller, Martin U., Marcel Rosenbach, und Thomas Schulz. 2013. „Die gesteuerte Zukunft". *Der Spiegel*, Mai 18, 64–74.

Murray, Dan und Kevan Durrell. 2000. „Inferring Demographic Attributes of Anonymous Internet Users". S. 7–20 in *Web Usage Analysis and User Profiling*, *Lecture Notes in Computer Science*, herausgegeben von B. Masand und M. Spiliopoulou. Springer Berlin Heidelberg. Abgerufen 7. November 2014 (http://link.springer.com/chapter/10.1007/3-540-44934-5_1).

Narayanan, Arvind. 2013. „Personalized coupons as a vehicle for perfect price discrimination". *33 Bits of Entropy*. Abgerufen 3. September 2014 (http://33bits.org/2013/06/25/personalized-coupons-price-discrimination/).

Nguyen, Tien T., Pik-Mai Hui, F.Maxwell Harper, Loren Terveen, und Joseph A. Konstan. 2014. „Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity". S. 677–686 in *Proceedings of the 23rd International Conference on World Wide Web*, *WWW '14*. New York, NY, USA: ACM. Abgerufen 23. November 2015 (http://doi.acm.org/10.1145/2566486.2568012).

Nielsen, Jakob. 1994. „Heuristic Evaluation". S. 25–62 in *Usability inspection methods*, herausgegeben von J. Nielsen und R. . L. Mack. Wiley.

Norberg, Patricia A., Daniel R. Horne, und David A. Horne. 2007. „The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors". *Journal of Consumer Affairs* 41(1):100–126.

O A. 1990. *Bundesdatenschutzgesetz*. Abgerufen 2. Mai 2010 (http://www.gesetze-im-internet.de/bdsg_1990/BJNR029550990.html).

O A. 2007. *Telemediengesetz*. Abgerufen 28. Juli 2015 (http://www.gesetze-im-internet.de/tmg/BJNR017910007.html).

O A. 2016. *Datenschutz-Grundverordnung*. Abgerufen (http://www.europarl.europa.eu/sides/getDoc.do?type=TA&language=EN&reference=P8-TA-2016-0125).

Odlyzko, Andrew. 2003. „Privacy, Economics, and Price Discrimination on the Internet“. S. 355–366 in *Proceedings of the 5th international conference on Electronic commerce*, *ICEC ’03*. New York, NY, USA: ACM. Abgerufen 24. Mai 2013 (http://doi.acm.org/10.1145/948005.948051).

Olejnik, Lukasz, Tran Minh-Dung, und Claude Castelluccia. 2013. „Selling Off Privacy at Auction“. Abgerufen 12. September 2014 (http://hal.inria.fr/hal-00915249).

Pagefair. 2015. *The 2015 Ad Blocking Report*. Abgerufen 10. August 2015 (http://blog.pagefair.com/2015/adblock-explorer/).

Pariser, Eli. 2012. *Filter Bubble: Wie wir im Internet entmündigt werden*. Carl Hanser Verlag GmbH & Co. KG.

Pedreshi, Dino, Salvatore Ruggieri, und Franco Turini. 2008. „Discrimination-aware Data Mining“. S. 560–568 in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, *KDD ’08*. New York, NY, USA: ACM. Abgerufen 31. Januar 2013 (http://doi.acm.org/10.1145/1401890.1401959).

Pekárek, Martin und Stefanie Pötzsch. 2009. „Requirements and Concepts for Privacy-Enhancing Access Control in Social Networks and Collaborative Workspaces“. Abgerufen 18. Januar 2009 (http://www.primelife.eu/results/documents).

Petri, Thomas B. 2008. „Das Urteil des Bundesverfassungsgerichts zur ‚Online-Durchsuchung‘“. *Datenschutz und Datensicherheit - DuD* 32(7):443–48.

Pew Research Center. 2014. „Public Perceptions of Privacy and Security in the Post-Snowden Era“. *Pew Research Center’s Internet & American Life Project*. Abgerufen 25. November 2014 (http://www.pewinternet.org/2014/11/12/public-privacy-perceptions/).

Pfitzmann, Andreas und Marit Hansen. 2008. „Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management - A Consolidated Proposal for Terminology“. Abgerufen (http://dud.inf.tu-dresden.de/Anon_Terminology.shtml).

Pias, Claus. 2003. „Zeit der Kybernetik–eine Einstimmung“. S. 9–41 in *Cybernetics-Kybernetik: The Macy-Conferences 1946-1953*, vol. 1, herausgegeben von C. Pias. Diaphanes.

Pohle, Jörg. 2014. „Kausalitäten, Korrelationen und Datenschutzrecht“. S. 85–105 in *Fundationes I: Geschichte und Theorie des Datenschutzes*. Münster: Monsenstein und Vannerdat.

Pohle, Jörg. 2016. „Transparenz und Berechenbarkeit vs. Autonomie- und Kontrollverlust: Die Industrialisierung der gesellschaftlichen Informationsverarbeitung und ihre Folgen“. *Mediale Kontrolle unter Beobachtung* 5(1). Abgerufen (http://www.medialekontrolle.de).

Purra, Joel. 2015. *Swedes Online: You Are More Tracked Than You Think*. Abgerufen 22. Oktober 2015 (http://www.-diva-portal.org/smash/record.jsf?pid=diva2%3A807623&dswid=2625).

Quantcast. o. J. *Understanding Digital Audience Measurement*. Abgerufen 7. November 2014 (http://www.quantcast.com/white-papers/quantcast-methodology.pdf).

Ritter, Martina. 2008. *Die Dynamik von Privatheit und Öffentlichkeit in modernen Gesellschaften*. Springer Berlin / Heidelberg. Abgerufen 6. Oktober 2014 (http://www.springer.com/springer+vs/soziologie/gender+studies/book/978-3-531-14649-2).

Rosa, Hartmut. 2014. *Beschleunigung: die Veränderung der Zeitstrukturen in der Moderne*. 10. Aufl. Frankfurt am Main: Suhrkamp-Taschenbuch-Verl.

Roskill, Damian. 2010. „Compete Data Methodology White Paper“. Abgerufen 14. Februar 2015 (https://blog.compete.com/2010/03/15/compete-data-methodology-white-paper/).

Rössler, B. und D. Mokrosinska. 2013. „Privacy and Social Interaction“. *Philosophy & Social Criticism* 39(8):771–91.

Rössler, Beate. 2001. *Der Wert des Privaten*. Frankfurt: Suhrkamp. Abgerufen (http://www.suhrkamp.de/buecher/der_wert_des_privaten-beate_roessler_29130.html).

Rössler, Beate. 2003. „Der Wert des Privaten“. S. 15–32 in *Privat!* dpunkt. Abgerufen (https://www.dpunkt.de/leseproben/1888/Kapitel%201.pdf).

Rost, Martin. 2013. „Zur Soziologie des Datenschutzes“. *Datenschutz und Datensicherheit* 37(2):85–91.

Rost, Martin und Andreas Pfitzmann. 2009. „Datenschutz-Schutzziele — revisited“. *Datenschutz und Datensicherheit - DuD* 33(6):353–58.

Sarodnick, Florian und Henning Brau. 2011. *Methoden der Usability Evaluation: wissenschaftliche Grundlagen und praktische Anwendung*. 2. Auflage. Bern: Huber.

Scahill, Jeremy und Glenn Greenwald. 2014. „The NSA's Secret Role in the U.S. Assassination Program“. *The Intercept*. Abgerufen 21. Februar 2015 (https://firstlook.org/theintercept/2014/02/10/the-nsas-secret-role/).

Schaar, Peter. 2009. *Das Ende der Privatsphäre: Der Weg in die Überwachungsgesellschaft*. Goldmann Verlag.

Schallaböck, Jan. 2014. *Verbraucher-Tracking*. Abgerufen 25. Juli 2014 (http://www.gruene-bundestag.de/fileadmin/media/gruenebundestag_de/themen_az/digitale_buergerrechte/Tracking-Bilder/Verbraucher_Tracking.pdf).

Scherer, Michael. 2012. „Inside the Secret World of the Data Crunchers Who Helped Obama Win“. *Time*, Juli 12. Abgerufen 11. Dezember 2014 (http://swampland.time.com/2012/11/07/inside-the-secret-world-of-quants-and-data-crunchers-who-helped-obama-win/).

Scherffig, Lasse. 2009. „The Human Being as a Servo. Von Feedback Control zur Kybernetik.“ S. 766–76 in *GI Jahrestagung*.

Schuler-Harms, Margarete. 2005. „Die kommerzielle Nutzung statistischer Persönlichkeitsprofile als Herausforderung für den Datenschutz.“ S. 5–37 in *Living by numbers – Leben zwischen Statistik und Wirklichkeit*. Düsseldorf.

Schwartz, Paul M. und Daniel J. Solove. 2011. „The PII Problem: Privacy and a New Concept of Personally Identifiable Information“. *New York University Law Review* 86:1814.

Seemann, Michael. 2014. *Das neue Spiel: Strategien für die Welt nach dem digitalen Kontrollverlust*. Freiburg im Breisgau: orange-press.

Shekyan, Sergey. 2015. „Detecting Headless Browsers“. Abgerufen 21. Oktober 2015 (http://www.slideshare.net/SergeyShekyan/shekyan-zhang-owasp).

Simitis, Spiros. 1987. „Reviewing Privacy in an Information Society“. *University of Pennsylvania Law Review* 135(3):707–46.

Simitis, Spiros, Ulrich Dammann, und Anne Arendt, Hrsg. 2011. *Bundesdatenschutzgesetz*. 7. Auflage. Baden-Baden: Nomos-Verlags-Gesellschaft.

Singel, Ryan. 2012. „Online Analytics Firm Settles Suit Over Unstoppable User Tracking“. *WIRED*. Abgerufen 12. September 2014 (http://www.wired.com/2012/10/kissmetrics-tracking/).

Singer, Natasha. 2012. „Acxiom, the Quiet Giant of Consumer Database Marketing“. *The New York Times*, Juni 16. Abgerufen 23. November 2012 (https://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html).

Sofsky, Wolfgang. 2009. *Verteidigung des Privaten. Eine Streitschrift*. München: C.H.Beck. Abgerufen (http://www.chbeck.de/downloads/Leseprobe_Verteidigung%20des%20Privaten%20978-3-406-58359-9.pdf).

Soltani, Ashkan, Shannon Canty, Quentin Mayo, Lauren Thomas, und Chris Jay Hoofnagle. 2009. „Flash Cookies and Privacy“. in *AAAI Spring Symposium: Intelligent Information Privacy Management*. Abgerufen 11. September 2014 (dx.doi.org/10.2139/ssrn.1446862).

Stalder, Felix. 2010. „Autonomy and Control in the Era of Post-Privacy“. *OPEN* (19). Abgerufen 6. Juni 2013 (http://www.skor.nl/_files/Files/OPEN!%20Key%20Texts_Stalder.pdf).

Steidle, Roland und Ulrich Pordesch. 2008. „Im Netz von Google. Web-Tracking und Datenschutz“. *Datenschutz und Datensicherheit - DuD* 32(5):324–29.

Steinmüller, W. u. a. 1971. *Grundfragen des Datenschutzes, Gutachten im Auftrag desm Bundesministeriums des Inneren*. Bundesministeriums des Innern.

Steinmüller, Wilhelm und Adalbert Podlech. 2007. „Das informationelle Selbstbestimmungsrecht Wie es entstand und was man daraus lernen kann“. *FIfF-Kommunikation* 2015(1):15–19.

Stelter, Brian. 2008. „Finding Political News Online, the Young Pass It On“. *The New York Times*, März 27. Abgerufen 30. Oktober 2014 (http://www.nytimes.com/2008/03/27/us/politics/27voters.html).

Stiftung Warentest. 2010. „Auskunfteien - Fehler über Fehler“. *Finanztest*. Abgerufen 22. Februar 2015 (https://www.test.de/Auskunfteien-Fehlerhafte-Daten-gespeichert-4047751-0/).

Sweeney, L. 2002. „k-Anonymity: a Model for Protecting Privacy.“ *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems* 10(5):557–70.

Sweeney, Latanya. 2013. *Discrimination in Online Ad Delivery*. Rochester, NY: Social Science Research Network. Abgerufen 28. Mai 2013 (http://papers.ssrn.com/abstract=2208240).

Symantec Corporation. 2015. *State of Privacy Report 2015*. Reading, UK. Abgerufen 24. Februar 2015 (http://www.symantec.com/content/en/us/about/presskits/b-state-of-privacy-report-2015.pdf).

Tang, Q. 2010. *From Ephemerizer to Timed-Ephemerizer: Achieve Assured Lifecycle Enforcement for Sensitive Data*. Enschede: Centre for Telematics and Information Technology, University of Twente. Abgerufen (http://eprints.eemcs.utwente.nl/17095/).

Tiqqun. 2007. *Kybernetik und Revolte*. Zürich; Berlin: Diaphanes.

Toubiana, Vincent, Arvind Narayanan, Dan Boneh, Helen Nissenbaum, und Solon Barocas. 2010. „Adnostic: Privacy Preserving Targeted Advertising.“ in *Proceedings of the 17th Network and Distributed System Security Symposium*. San Diego, CA.

Toubiana, Vincent, Lakshminarayanan Subramanian, und Helen Nissenbaum. 2011. „TrackMeNot: Enhancing the privacy of Web Search“. *arXiv:1109.4677*. Abgerufen 28. Mai 2012 (http://arxiv.org/abs/1109.4677).

Trojanow, Ilija und Juli Zeh. 2010. *Angriff auf die Freiheit: Sicherheitswahn, Überwachungsstaat und der Abbau bürgerlicher Rechte*. München: Dt. Taschenbuch-Verl.

Turow, Joseph. 2012. *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth*. Yale University Press.

Turow, Joseph, Michael Hennessy, und Nora Draper. 2015. *The Tradeoff Fallacy - How Marketers Are Misrepresenting American Consumers And Opening Them Up To Exploitation*. Philadelphia, Pennsylvania, United States: Annenberg School for Communication - University of Pennsylvania. Abgerufen 9. Juni 2015 (https://www.asc.upenn.edu/sites/default/files/TradeoffFallacy_1.pdf).

Ur, Blase, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, und Yang Wang. 2012. „Smart, Useful, Scary, Creepy: Perceptions of Online Behavioral Advertising“. S. 4 in *Proceedings of the Eighth Symposium on Usable Privacy and Security*.

Valentino-DeVries, Jennifer, Jeremy Singer-Vine, und Ashkan Soltani. 2012. „Websites Vary Prices, Deals Based on Users' Information". *Wall Street Journal*, Dezember 24. Abgerufen 6. Mai 2014 (http://online.wsj.com/news/articles/SB10001424127887323777204578189391813881534).

Vissers, Thomas, Nick Nikiforakis, Nataliia Bielova, und Wouter Joosen. 2014. „Crying Wolf? On the Price Discrimination of Online Airline Tickets". *HotPET Symposium*. Abgerufen (http://www.securitee.org/files/pdiscrimination_hotpets2014.pdf).

W3C Technical Architecture Group. 2015. „Unsanctioned Web Tracking" herausgegeben von M. Nottingham. Abgerufen 18. Juli 2015 (http://www.w3.org/2001/tag/doc/unsanctioned-tracking/#why-unsanctioned-tracking-is-harmful).

Wanying Luo, Qi Xie, und Urs Hengartner. 2009. „FaceCloak: An Architecture for User Privacy on Social Networking Sites". S. 26–33 in. Vancouver, BC. Abgerufen (http://www.cs.uwaterloo.ca/~uhengart/publications/pas-sat09.pdf).

Warren, S. und L. Brandeis. 1890. „The Right to Privacy". *Harvard Law Review* 4(5):193–220.

Weich, Andreas. 2016. „Sich profilieren und profiliert werden – zur (Medien-)Genealogie zweier Seiten einer Medaille (in Druck)". in *Profile. Interdiszplinäre Beiträge*, *Digital Cultures*. meson press.

Weichert, Thilo. 2006. „Kredit-Scoring und Datenschutz". *Unabhängiges Landeszentrum für Datenschutz Schleswig-Hiolstein*. Abgerufen 27. September 2010 (https://www.datenschutzzentrum.de/scoring/060404-kreditscoring.htm).

Westin, Alan. 1967. *Privacy and Freedom*. New York: New Jork Atheneum.

Whitaker, Reg. 1999. *Das Ende der Privatheit. Überwachung, Macht und soziale Kontrolle im Informationszeitalter*. München: Kunstmann.

White, Tiffany Barnett, Debra L. Zahay, Helge Thorbjørnsen, und Sharon Shavitt. 2007. „Getting Too Personal: Reactance to Highly Personalized Email Solicitations". *Marketing Letters* 19(1):39–50.

Youyou, Wu, Michal Kosinski, und David Stillwell. 2015. „Computer-Based Personality Judgments Are More Accurate than Those Made by Humans". *Proceedings of the National Academy of Sciences* 201418680.

Yuan, Shuai, Jun Wang, und Xiaoxue Zhao. 2013. „Real-time Bidding for Online Advertising: Measurement and Analysis". S. 3:1–3:8 in *Proceedings of the Seventh International Workshop on Data Mining for Online Advertising*, *ADKDD '13*. New York, NY, USA: ACM. Abgerufen 23. März 2014 (http://doi.acm.org/10.1145/2501040.2501980).

Zuckerberg, M. u. a. 2010. „Dynamically Providing a News Feed About a User of a Social Network". Abgerufen (https://www.google.com/patents/US7669123).

Zurawski, Nils. 2014. „Essay: Schluss mit der Privatsphäre?" *Surveillance Studies.org*. Abgerufen 30. September 2014 (http://www.surveillance-studies.org/2014/09/essay-schluss-mit-der-privatsphaere/).

Zwick, Detlev und Nikhilesh Dholakia. 2004. „Whose Identity Is It Anyway? Consumer Representation in the Age of Database Marketing". *Journal of Macromarketing* 24(1):31–43.

8. DANKSAGUNG

Ein Dissertationsprojekt ist immer auch eine Gemeinschaftsaufgabe. Zunächst möchte ich meinen beiden Betreuer_innen Thomas Herrmann und Bettina Berendt danken. Deren Unterstützung und Feedback diese Arbeit überhaupt ganz grundsätzlich erst möglich gemacht haben.

Darüber hinaus sei all denen gedankt, mit denen ich mich über die Jahre ausgetauscht und Ideen entwickelt habe. Ganz besonders gilt mein Dank hier Jasmin und meinen Kolleg_innen Jan, Michael, Kai-Uwe, Axel, Nina, Moritz, Sebastian und Chris.

Für die Korrekturarbeit, die beim Erstellen einer umfangreichen Textarbeit unerlässlich sind, danke ich ganz herzlich Mona, Marlen, Marita und Felix, ohne die bis zuletzt die Fertigstellung nicht zu leisten gewesen wäre.

Darüber hinaus danke ich für freundliche Gespräche und gutes Feedback dem Graduierten Kolleg „Privatheit“ in Passau, Mirjam, Moritz, Klaas, Kirsten, Henning, Anna-Lena, Mary, Emi, Jannes, Jascha, Anna, Florian, Maximilian und Claudio.

9. EIGENSTÄNDIGKEITSERKLÄRUNG

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbstständig angefertigt habe. Es wurden von mir ausschließlich die angegebenen Quellen und Hilfen in Anspruch genommen.

Eine Promotionsarbeit über dieses Thema liegt noch nicht vor.

_____

10. ANHANG

10.1 ABBILDUNGSVERZEICHNIS

Abbildung 1: Schematische Darstellung von Online-Tracking mit 2 Webseiten (A und B) und einem Trackingservice 10
Abbildung 2: Funktionsweise von Online Tracking und Profiling inklusive weitere Akteure. 71
Abbildung 3: Screenshot des Browsers „Chrome“ der die Cookies darstellt, die beim Aufruf der Webseite http://spiegel.de auf dem Rechner gespeichert werden. 74
Abbildung 4: Funktionsweise von Tracking Obfuscation nach Balsa, Troncoso und Diaz (2012). Eigene Darstellung. 82
Abbildung 5: Datenschutzschutzziele; Darstellung nach Rost (2012) 85
Abbildung 6: Beispiel für eine Zielgruppe für welche Werbung geschaltet werden kann. Quelle Google AdWords (letzter Zugriff 07.11.2014). „Shutterbug“ be- zeichnet nach Wikipedia die Gruppe der „enthusiastischen Amateurfotografen“ 88
Abbildung 7: Anzeigeneinstellungsseite von Google. (Screenshot von https://www.google.com/settings/ads vom 08.Juli 2015; die Seite wurde im Spätsommer 2015 neugestaltet) 89
Abbildung 8: Beispiel der Informationen die Quantcast über Besucher_innen einer Webseite bereit stellt (Stand 16.07.2015) 91
Abbildung 9: Screenshot der Alexa-Informationsseite zu wired.com (Stand 14.07.2015) 93
Abbildung 10: Logo der Kampagne „AdChoices“. Quelle: http://www.youradchoices.com/ (letzter Zugriff 05.02.2015) 95
Abbildung 11: Logo (oben) und Informationsoverlay (unten) von Ghostery. 99
Abbildung 12: Disconnect.me Overlay im Browser (oben), Privacy Icons (unten) 100
Abbildung 13: Optionsmenü zum Blockieren und Erlauben von Skripten auf theguardian.com 101
Abbildung 14: Hinweisfenster von Privacy Badger. Rot = blockieren aller Anfragen; gelb = Übertragung von Cookies deaktiviert 102
Abbildung 15: Screenshot der TrackMeNot-Einstellungen; auf der rechten Seite: automatisch ausgeführte Suchanfragen 104
Abbildung 16: Ad Nauseam Overlay. Zeigt die bisher gefunden und angeklickten Anzeigen 105
Abbildung 17: Screenshot aus einem Floodwatch Demonstrationsvideo. (11.10.2015) 108
Abbildung 18: Screenshot des Firefox Interest Dashboard 109
Abbildung 19: Screenshot von Lightbeam nach dem Aufruf von zwei Webseiten (große Kreise); die Dreiecke symbolisieren Third Party Skript und die Verbindungen zeigen an auf welcher Seite sie geladen wurden 110
Abbildung 20: Übersicht der Module in TrackTrack 115
Abbildung 21: Modell der Datenerhebung 120
Abbildung 22: Anzahl von Cookies die pro User gespeichert wurden; rote Linie zeigt den Durchschnitt 135
Abbildung 23: Verteilung der Anzahl von GIC 136
Abbildung 24: Häufigkeitsverteilung der einzelnen Interessenkategorien (GIC) 137
Abbildung 25: Berechnung des Interessenprofils über beobachtete Zusammenhänge 142

Abbildung 26: Berechnung eines Interessenprofils über die ermittelten Gewichtungen....143
Abbildung 27: Übersicht des Verfahrens zum Vergleich verschiedener Dummy Generation Strategien....147
Abbildung 28: Vergleich der Veränderung bei einfacher Wiederholung (links) und der Obfuscation mit tt75stat. Ergebnisse aller Durchläufe im Anhang....151
Abbildung 29: Entwicklung der zugewiesenen GIC bei Veränderung der Zahl zur Verschleierung aufgerufener URLs....151
Abbildung 30: Elemente des Profiling Prozesses und wie Transparenz zu diesen erzeugt werden kann....156
Abbildung 31: Screenshot der Übersicht über Basis Informationen die aus dem Browserverlauf extrahiert werden.....160
Abbildung 32: Hinweis auf den Anteil der meistbesuchten Seiten am Gesamtumfang und Hinweis auf Informationen zur Filterbubble....160
Abbildung 33: Darstellung der gewichteten Interessen....161
Abbildung 34: Hinweis auf die Möglichkeit der Verschleierung der Interessen (oben) und Informationsseite über die laufende Verschleierung (unten)....162
Abbildung 35: Darstellung des soziodemographischen Profils....162
Abbildung 36: Auswahlmenüs für die Verschleierung des soziodemografischen Profils....163
Abbildung 37: Informationen zur Filterblase und Webseiten die dem ermittelten Profil widersprechen....164
Abbildung 38: Übersicht der TrickTrack Komponenten....166
Abbildung 39: Zustimmungswerte zu Fragen die Hintergründe von Online-Tracking und Profile betreffend....171
Abbildung 40: Bestätigungsnachricht über die erfolgreiche Installation. Erst im Anschluss erscheint das Icon auf der äußeren rechten Seite....177
Abbildung 41: Oben: Anzahl der selbst zugewiesenen Interessen (blau) und der von TrickTrack ermittelten (orange; Wert > 10); Unten: Vergleich der False Negative, False Positives und korrekt zugewiesenen Interessen...180
Abbildung 42: Vergleich der Eigenangaben mit dem von TrickTrack ermittelten soziodemografischen Profil. (Ausführliche Tabelle im Anhang 3)....181
Abbildung 43: Generelle Einschätzung von TrickTrack; Linie zeigt Mittel....186
Abbildung 44: Ergebnisse des Abschnitts zu Grafiken im QUIS Fragebogen....186

10.2 TABELLENVERZEICHNIS

Tabelle 1: Vergleich der Tracking-Anbieter_innen in Bezug auf die Anforderungen....94
Tabelle 2: Vergleich der Anforderungen in Bezug auf Formen der Selbstregulierung....97
Tabelle 3: Vergleich von Werbe- und Tracking-Blockern....103
Tabelle 4: Vergleich von Obfuscationtools....107
Tabelle 5: Vergleich von Transparenz Tools. Xray und about:profile sind ausgenommen, da sie nicht (mehr) nutzbar sind....111
Tabelle 6: Übersicht über die erhobenen Datensätze im Rahmen von TrackTrack....121
Tabelle 7: Umfang der erhobenen Datenmengen durch TrackTrack....122
Tabelle 8: Verteilung der Google Interest Categories (GIC)....125
Tabelle 9: Übersicht über die Attribute und Kategorien in die Quantcast Nutzer_innen einordnet. Die letzten beiden Kategorien wurden durch Quantcast erst nach Ende des Tests hinzugefügt und wurden nicht ausgewertet.....127

Tabelle 10: Soziodemografische Daten die Alexa für eine Reihe von Domains bereit stellt....129
Tabelle 11: soziodemografische Daten, die Compete anbietet....130
Tabelle 12: Meist besuchte Webseiten nach Alexa.com....133
Tabelle 13: Meist verlinkte Webseiten und zum Vergleich der Alexa Rank....133
Tabelle 14: Umfang der beobachtbaren Webseitenaufrufe nach Anbieter_in....135
Tabelle 15: Anzahl der Domains und URLs pro GIC, für die eine direkte Zuordnung beobachtet wurde....138
Tabelle 16: Vergleichswerte für die soziodemographischen Angaben über Internetnutzer_innen in den USA....139
Tabelle 17: Kategorien des soziodemografischen Profils nach der Datenzusammenführung....140
Tabelle 18: Anzahl der Linkprofile die sich mehrheitlich der jeweiligen Kategorie zuordnen lassen....141
Tabelle 19: Vergleich verschiedener Dummy Generation Strategien (DGS). Standardabweichung in Klammern....149
Tabelle 20: Liste der Teilnehmer_innen und erhobenen Daten....170
Tabelle 21: Übersicht über den Umfang der analysierten Daten pro TN die in der Basis-Übersicht angezeigt wurde....178
Tabelle 22: Vergleich der durch TrickTrack bestimmten sozio-demografischen Profile mit den Angaben der TN. Grüne markiert sind korrekte Berechnungen, gelb gekennzeichnet sind, bei falsche Berechnung, die tatsächlichen Werte wobei rot die Werte mit der höchsten Wahrscheinlichkeit anzeigt....215

10.3 QUELLTEXTVERZEICHNIS

Quelltext 1: Beispiele für eine Interessenhierarchie bei Google....124
Quelltext 2: Datenformat für alle Audience Analytics Dienste....128

10.4 FORMELVERZEICHNIS

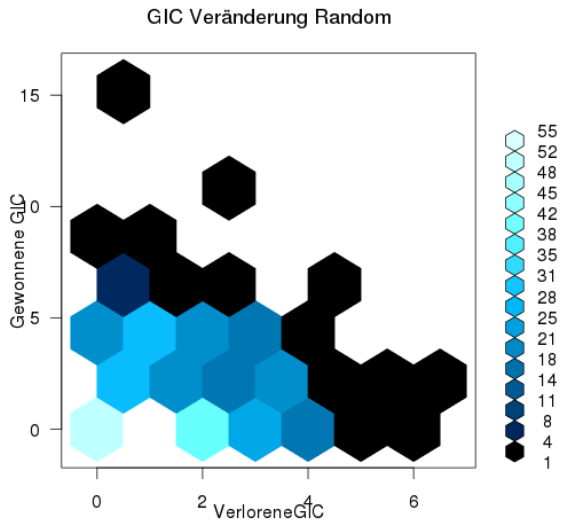
Formel 1: Berechnung der Breite eines Profils....126
Formel 2: Zur Konvertierung der Verhältnis- in Prozentwerte....139
Formel 3: Wahrscheinlichkeit, dass der Besuch der Domain D zu dem Interesse GIC führt....142
Formel 4: Wahrscheinlichkeit, dass ein GIC einem User U zugewiesen wird....142
Formel 5: Berechnung des Kulcynski Koeffizienten....144
Formel 6: Berechnung der wahrscheinlichsten Kategorien....145
Formel 7: Kulcynski-Koeffizient zum Vergleich der Profile....150

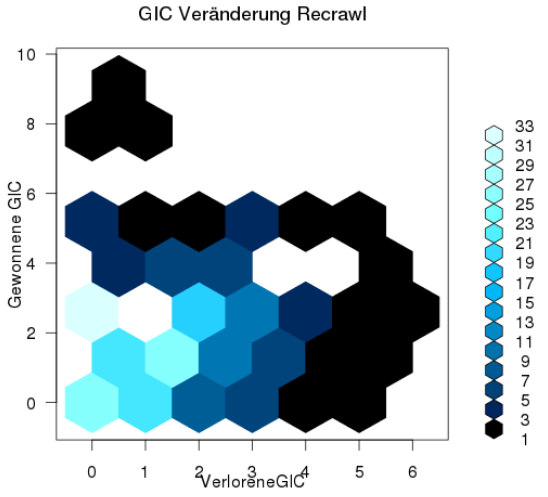
10.5 VERWENDETE SOFTWARE UND BIBLIOTHEKEN

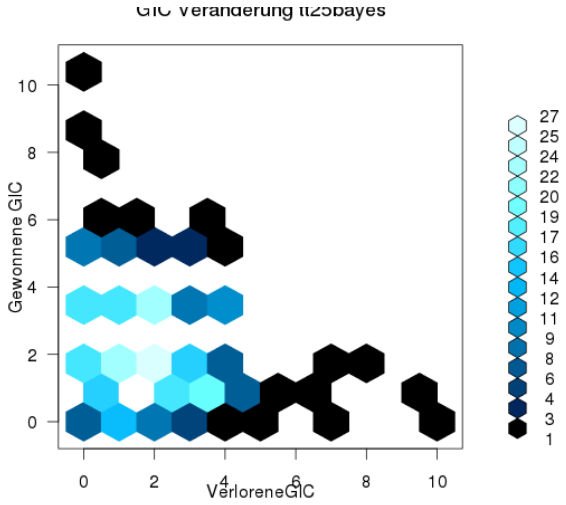
Name	Link	Copyright	Lizenz	Verwendet in
underscore.js	http://git.io/vch2o	© 2009–2014 Jeremy Ashkenas	MIT Lizenz	TrickTrack, TrackBack, TrackTrack
backbone.js	http://git.io/vch2H	© 2010–2013 Jermey Ashkenas	MIT Lizenz	TrickTrack
jquery	HTTPS://JQUERY.ORG	© jQuery Foundation	MIT Lizenz	TrackTrack/Back
Mongoosejs	http://mongoosejs.com/	© 2011 LEARNBOOST	MIT Lizenz	TrackTrack/Back
Bluebird	http://git.io/oA74uA	© 2014 Petka Antonov	MIT Lizenz	TrackTrack/Back
bcrypt	https://www.npmjs.com/package/bcrypt	© 2010 Nicholas Campbell	MIT Lizenz	TrackTrack
fs-extra	http://git.io/vch2L	© 2011–2015 JP Richardson	MIT Lizenz	TrackTrack
jsdom	http://git.io/vch2O	© 2010 Elijah Insua	MIT Lizenz	TrackTrack
nightmare	http://git.io/pWHpmw	© 2015 Segment.io	MIT Lizenz	TrackTrack
sleep	http://git.io/vch2E	© 2015 Erik Dubbelboer	MIT Lizenz	TrackTrack
request	https://github.com/request		Apache 2.0	Track/Back
node.js	HTTPS://NODEJS.ORG	© 2015 Node.js Found.	MIT Lizenz	Track/Back
phantomJS	http://phantomjs.org/	© 2010–2015 ARIYA HIDAYAT	BSD Lizenz	TrackTrack
SlimerJS	http://slimerjs.org/	© 2013–2015 Laurent Jouanneau	MPL 2.0 Lizenz	TrackTrack
CasperJS	HTTP://CASPERJS.ORG/	© 2011–2013 Nicolas Perriault	MIT Lizenz	TrackTrack
CoffeeScript	http://coffeescript.org/	© 2015 Jeremy Ashkenas	MIT Lizenz	Trick/Track/Back
R Studio	https://www.rstudio.com/	© 2015 RStudio	AGPL v3	TrackBack
Debian	https://www.debian.org/	© 1997–2015 SPI	GPL v2	
MongoDB	https://www.mongodb.org/	© 2015 MongoDB, Inc	AGPL v3.0	TrackTrack/Back

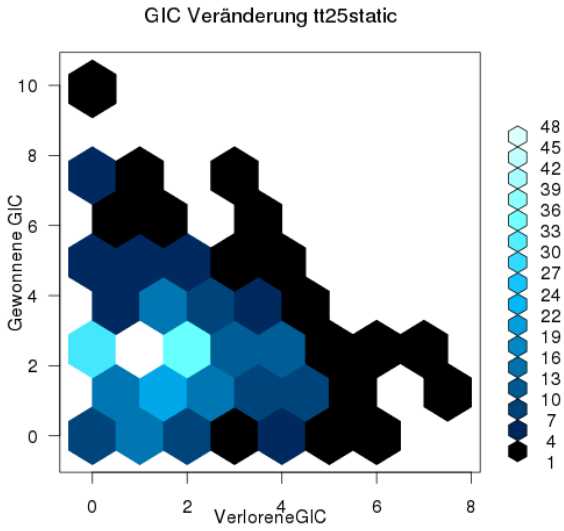
10.6 VERGLEICH ALLER OBFUSCATIONMETHODEN

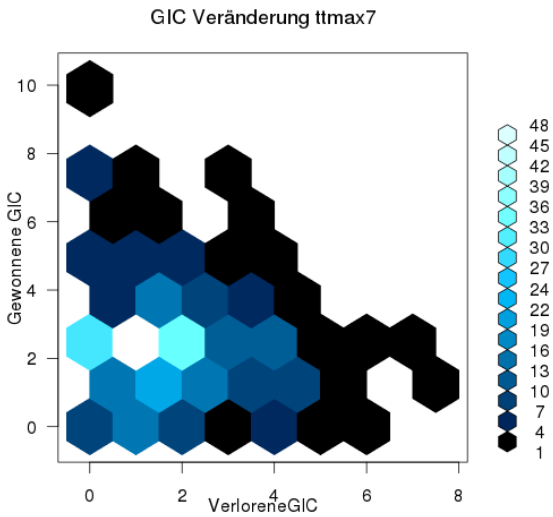
Die Graphen zeigen (analog zu Abbildung 28) die Veränderungen in den Profilen bei verschiedenen Strategien im Vergleich zur Ursprungserhebung.

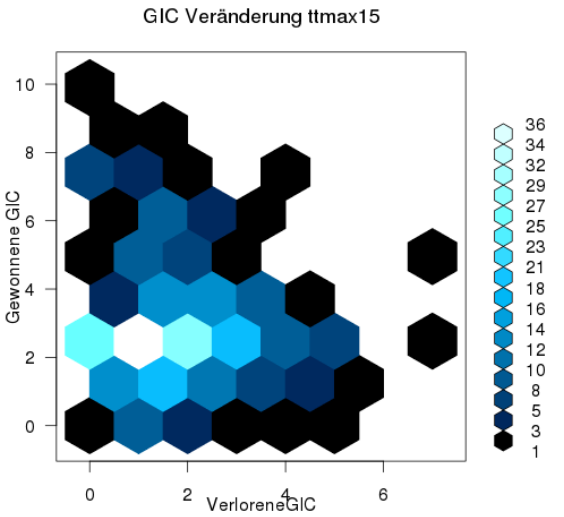






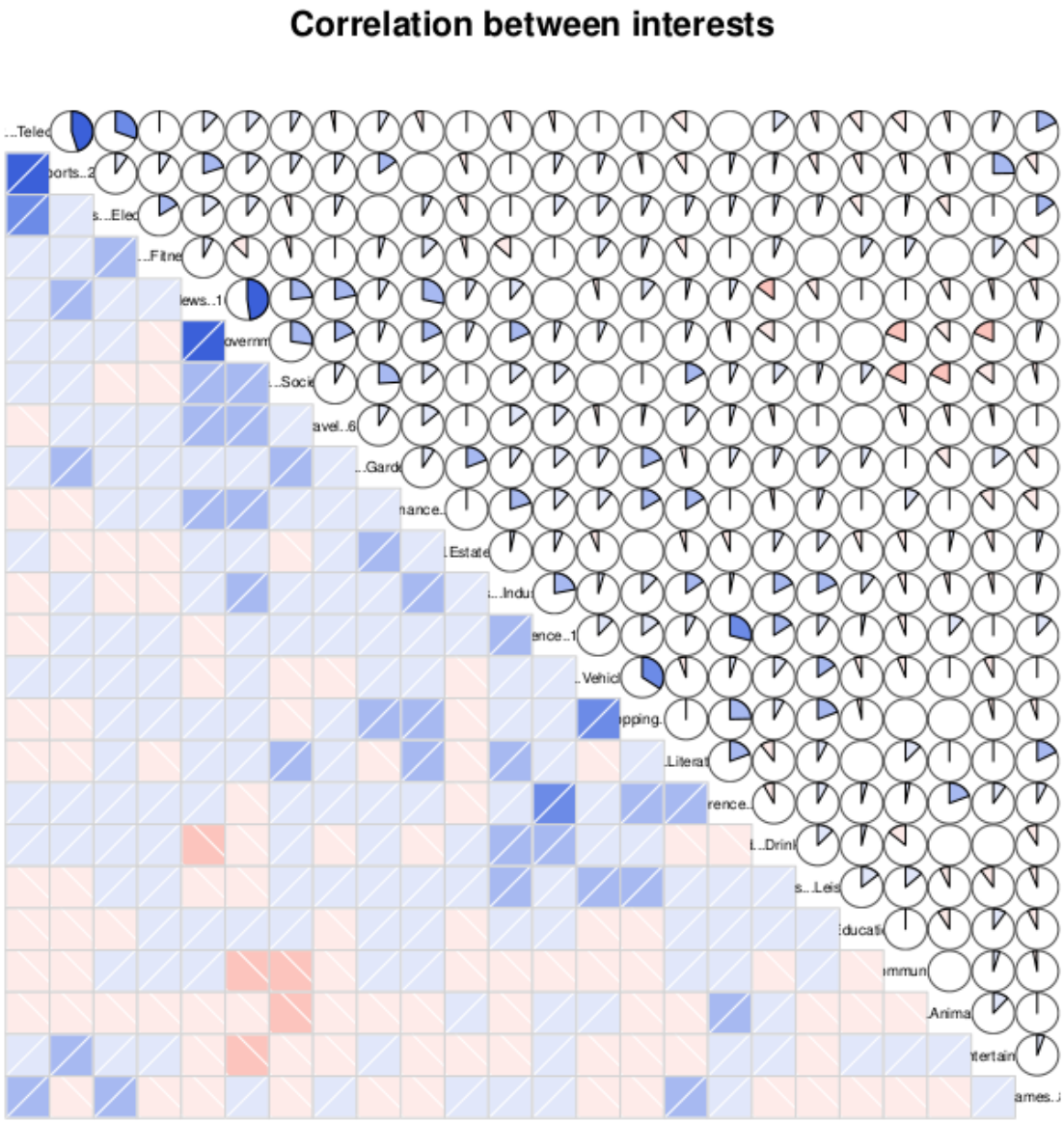






10.7 KORRELATIONEN VON INTERESSEN

Korrelation der Interessen in der ersten Erhebung. In der Diagonale sind die Interessen notiert. Die Kreise zeigen positive (blau) und negative (rot) Korrelationen an. Wesentliche Korrelationen finden sich nur zwischen „Telekommunikation und Internet“ und „Sport“ sowie „Nachrichten“ und „Gesetze und Regierungen“.



10.8 KORREKTHEIT DER SOZIODEMOGRAFISCHEN PROFILE IM TESTS

		E312		1RF4		BNT		GF9A		N3PP		JAR7		Z0ZJ		R2D2		6GL3		RR75	amerikanischer Durchschnitt
männlich	52.35	3.35			56.22	7.22	54.97	5.97	55.13	6.13	57.19	8.19	56.22	7.22	60.70	11.70	56.54	7.54	64.71	15.71	49
weiblich	47.65	-3.35			43.78	-7.22	45.03	-5.97	44.87	-6.13	42.81	-8.19	43.78	-7.22	39.30	-11.70	43.46	-7.54	35.29	-15.71	51
0-50k	51.43	0.43			54.42	3.42	49.73	-1.27	50.77	-0.23	56.02	5.02	54.08	3.08	53.98	2.98	51.22	0.22	41.63	-9.37	51
50-100k	26.78	-2.22			25.12	-3.88	31.71	2.71	27.53	-1.47	24.76	-4.24	26.74	-2.26	25.50	-3.50	26.89	-2.11	28.18	-0.82	29
mehr als 100k	21.76	1.76			20.46	0.46	18.56	-1.44	21.70	1.70	19.22	-0.78	19.18	-0.82	20.58	0.58	21.91	1.91	30.18	10.18	20
	99.97																				
Kaukasisch	81.96	6.96			75.30	0.30	57.19	-17.81	78.56	3.56	75.25	0.25	78.79	3.79	54.65	-20.35	75.79	0.79	78.55	3.55	75
African American	4.13	-4.87			8.63	-0.37	5.92	-3.08	8.02	-0.98	7.83	-1.17	5.82	-3.18	14.86	5.86	7.68	-1.32	7.41	-1.59	9
Asian	2.57	-1.43			5.01	1.01	3.15	-0.85	3.23	-0.77	4.48	0.48	4.73	0.73	7.77	3.77	4.15	0.15	3.94	-0.06	4
Hispanic	5.70	-3.30			7.93	-1.07	6.48	-2.52	7.21	-1.79	9.11	0.11	7.33	-1.67	19.37	10.37	9.24	0.24	7.12	-1.88	9
Other	0.89	-0.11			1.02	0.02	0.73	-0.27	1.01	0.01	0.98	-0.02	0.99	-0.01	1.03	0.03	0.99	-0.01	0.95	-0.05	1
Keinen	34.86	-10.14			34.23	-10.77	40.74	-4.26	34.05	-10.95	31.27	-13.73	26.81	-18.19	41.04	-3.96	44.91	-0.09	29.08	-15.92	45
Hochschulabschluss	49.92	8.92			43.84	2.84	42.65	1.65	55.29	14.29	53.58	12.58	55.06	14.06	47.89	6.89	43.74	2.74	58.62	17.62	41
Akademischer Titel	15.11	-29.89			21.55	-23.45	15.78	-29.22	10.67	-34.33	15.17	-29.83	18.04	-26.96	11.78	-33.22	11.30	-33.70	12.34	-32.66	45
Kinder	42.26	-6.74			25.59	-23.41	33.32	-15.68	32.27	-16.73	36.22	-12.78	34.10	-14.90	33.95	-15.05	30.72	-18.28	41.02	-7.98	49
keine Kinder	54.76	3.76			74.74	23.74	41.57	-9.44	67.95	16.95	63.60	12.60	65.94	14.94	66.31	15.31	69.64	18.64	59.06	8.06	51
<18	12.79	-5.21			17.76	-0.24	10.26	-7.74	16.94	-1.06	17.98	-0.02	13.29	-4.71	11.89	-6.11	17.97	-0.03	13.13	-4.87	18
18-24	11.90	-0.10			15.26	3.26	9.90	-2.10	17.51	5.51	14.14	2.14	14.74	2.74	17.09	5.09	18.63	6.63	10.07	-1.93	12
25-34	18.22	1.22			22.35	5.35	16.62	-0.38	26.15	9.15	21.35	4.35	25.46	8.46	22.39	5.39	25.17	8.17	21.72	4.72	17
35-44	19.71	2.71			18.79	1.79	16.15	-0.85	18.67	1.67	18.72	1.72	21.68	4.68	18.37	1.37	16.71	-0.29	22.15	5.15	17
45-54	17.85	0.85			13.55	-3.45	11.82	-5.19	12.20	-4.80	15.05	-1.95	14.28	-2.72	16.78	-0.22	12.69	-4.31	17.91	0.91	17
55-64	9.95	-0.05			6.92	-3.08	6.10	-3.90	4.58	-5.42	7.15	-2.85	6.00	-4.00	7.76	-2.24	5.00	-5.00	8.94	-1.06	10
> 64	5.51	3.51			4.16	2.16	3.21	1.21	2.51	0.51	3.94	1.94	2.80	0.80	4.45	2.45	2.32	0.32	5.13	3.13	2
																				Durchschnitt:	
Domains		390				526		919		2709		3087		1653		2665		219		2037.00	1.578,33
Basis		33.00				78.00		50.00		149.00		96.00		84.00		319.00		30.00		206.00	116,11
Prozent		0.08				0.08		0.05		0.06		0.03		0.05		0.12		0.14		0.10	0,08
Fehler		1.00				1.00		4.00		1.00		2.00		0.00		1.00		1.00		2.00	1,44
Korrekt		4.00				5.00		2.00		5.00		4.00		5.00		4.00		4.00		3.00	4,00

Tabelle 22: Vergleich der durch TrickTrack bestimmten soziodemografischen Profile mit den Angaben der TN. Grüne markiert sind korrekte Berechnungen, gelb gekennzeichnet sind, bei falsche Berechnung, die tatsächlichen Werte wobei rot die Werte mit der höchsten Wahrscheinlichkeit anzeigt.

10.9 DOKUMENTE DER NUTZER_INNEN BEFRAGUNG

10.9.1 Einverständniserklärung

„Hiermit willige ich, _____, ein, dass während des Experimentes sowohl Video- als auch Tonaufnahmen von mir gemacht werden dürfen. Diese Aufnahmen dienen rein wissenschaftlichen Zwecken und werden nicht an Dritte weitergegeben. Alle Aufnahmen werden pseudonymisiert ausgewertet und nicht zu meiner Person zugeordnet. Sobald die Aufnahmen nicht mehr für die wissenschaftliche Auswertung benötigt werden, spätestens jedoch am 31.12.2015, werden diese vernichtet.

Mir ist bewusst, dass ich dieser Einverständniserklärung jederzeit nachträglich mit einer E-Mail an MARTIN.DEGELING@RUB.DE widersprechen kann.“

_____ _____

(Datum) (Unterschrift)

10.9.2 Einleitung in die Befragung

Einleitung: Das Experiment ist drei geteilt. Erst führen wir ein kleines Interview zum Thema Werbung im Internet. Im Anschluss sollst Du/sollen Sie ein Tool benutzen – TrickTrack – das Informationen zu Online-Tracking aufbereitet. Dabei sollst Du/sollen Sie sich durch das Tool klicken und dabei laut darüber sprechen was Du/Sie denkst/denken.

10.9.3 Vorabinterview

1. Was ist das erste, das dir/ihnen einfällt, wenn Sie/Du an “Werbung im Internet” denken?

2. Was denkst Du/ denken Sie allgemein über Werbung im Internet?

2.1 Welche guten Aspekte siehst Du/sehen Sie an Werbung im Internet?

2.2 Ist Werbung im Internet für (Dich/Sie) nützlich? Wenn ja, können Sie ein Beispiel geben?

2.3 Stört Dich/Sie Werbung im Internet? Können Sie ein Beispiel geben?

3. Inwieweit findest du/finden Sie, das Werbung im Internet im relevant ist für dich/sie, bzw. das sie deinen/ihren Interessen entspricht?

4. Hast Du/Haben Sie das Gefühl, das Werbung die dir/ihnen angezeigt wird, wenn du/sie im Web surfst auf deine/ihre Interessen oder Eigenschaften von dir/ihnen, wie dein/ihr Geschlecht, Alter oder Wohnort zugeschnitten sind?

4.1 Wenn Ja: Findest du/Finden Sie es nützlich, dass die Werbung auf dich/sie zugeschnitten ist?

4.2 Wie denkst Du/denken Sie machen die Firmen das?

4.3 Wenn nein: Fändest Du/Fänden Sie es sinnvoll, wenn die Werbung stärker auf dich/Sie zugeschnitten wäre?

4.3.1Welche Informationen wärst du/wären Sie bereit preiszugeben, damit Werbung besser auf Dich/Sie zugeschnitten werden kann? Und an wen?

5. Hast du/Haben Sie schon mal von „targeted advertising“ oder „behavioural targeting“ (also gezielter Werbung und verhaltensbezogen Werbung) gehört? Wenn ja: Was heißt in dem Kontext „gezielt“ oder „verhaltensbezogen“ für Dich/sie

6. Kannst du dir Umständen vorstellen in denen du/sie es besonders schlecht fändest/finden, wenn verfolgt würde auf welchen Seiten du im Internet surfst, um dir dazu passende Werbung anzuzeigen?

6.1 Kannst du Online-Tracking auf gesamtgesellschaftlicher Ebene einordnen und mögliche Folgen beschreiben?

7. Hast Du/haben Sie schon einmal davon gehört, dass Du/Sie verhindern kannst/können das dir Werbung angezeigt wird? Wenn ja: Welche Tools kenst du da?

8. Hast Du/haben Sie schon einmal davon gehört, dass du/Sie beeinflussen kannst/können nach welchen Interessen dir Werbung angezeigt wird? Wenn ja: Welche Möglichkeiten kennst du?

9. Weitere Kommentare?

10. Bitte ordne auf einer Skala von 1 bis 5 eni

10.1 Ich bin interessiert an der Funktionsweise von Online-Tracking

10.2 Ich bin interessiert an dem Profil, das beim Online-Tracking über mich entsteht

10.3 Ich bin interessiert daran, dass Profil zu beeinflussen

10.9.1 Nutzung

Auf tricktracking.com gehen, installieren, benutzen → thinking out loud

10.9.2 Nachbefragung oder Beobachtung

I. Zu Online-Tracking/Lerneffekt

Stats: Gender, Einkommen, Ethnicity, Schulabschluss, Kinder (freiwillig)

Fandest Du/Fanden Sie, dass die Informationen über dich korrekt waren? Was denkst Du/denken Sei darüber (wenn wohl, wenn nicht)?

Was hast du empfunden, als du dein Profil gesehen hast (erwartbar, überraschend)?

Inwieweit würdest Du/würdne Sie sagen entspricht das Profil das TrickTrack ermittelt hat dem, was dir/Ihnen an Werbung angezeigt wird?

Hast Du/Haben Sie durch die Nutzung etwas über Online-Tracking gelernt? Hat sich deine Einstellung gegenüber Online-Tracking verändert?

II. Zu TrickTrack

Wie war allgmein dein Eindruck von TrickTrack?

Fehlen dir wichtige Informationen oder Funktionen?

Würdest du TrickTrack oder ein ähnliches Werkzeug regelmäßig nutzen, um dein Profil zu überprüfen?

Fragen nach Screenshots von Google/Bluekai; Daten export

10.10 QUIS FRAGEBOGEN (AUSWAHL)

10.10.1 Gesamteindruck

Bitte markieren Sie die Zahl, die Ihren Eindruck von diesem System am ehesten widerspiegelt.

Keine Angaben = KA

3.1	**Ihr Gesamteindruck von diesem System**	**unangenehm**		**angenehm**	
		1 2 3 4 5 6 7 8 9			KA
3.2		**frustrierend**		**zufrieden stellend**	
		1 2 3 4 5 6 7 8 9			KA
3.3		**langweilig**		**anregend**	
		1 2 3 4 5 6 7 8 9			KA
3.4		**schwierig**		**leicht**	
		1 2 3 4 5 6 7 8 9			KA
3.5		**ungeeignete Leistung**		**geeignete Leistung**	
		1 2 3 4 5 6 7 8 9			KA
3.6		**starr**		**flexibel**	
		1 2 3 4 5 6 7 8 9			KA

10.10.2 Darstellung

4.3	**Die Bildschirmlayouts waren hilfreich.**				**niemals**				**immer**			
					1	2	3	4	5	6	7	8	9	KA
	4.3.1	**Die Menge der auf dem Bildschirm dargestellten Informationen war**			**nicht ausreichend**						**ausreichend**			
					1	2	3	4	5	6	7	8	9	KA
	4.3.2	**Die Anordnung der Informationen auf dem Bildschirm war**			**unlogisch**						**logisch**			
					1	2	3	4	5	6	7	8	9	KA

4.4	**Die Reihenfolge der Bildschirmbilder war**			**verwirrend**					**klar**				
				1	2	3	4	5	6	7	8	9	KA
	4.4.1	**Der Nächster Bildschirm in der Reihenfolge war**		**nicht vorhersehbar**							**vorhersehbar**		
				1	2	3	4	5	6	7	8	9	KA
	4.4.2	**Das Zurückgehen zum vorherigen Bildschirm war**		**unmöglich**							**leicht**		
				1	2	3	4	5	6	7	8	9	KA
	4.4.3	**Die Reihenfolge der für die Aufgabe lösung notwendigen Schritte war**		**verwirrend**							**klar strukturiert**		
				1	2	3	4	5	6	7	8	9	KA

10.10.3 Terminologie und System-Informationen

5.1	**Die einheitliche Ausdrucksweise im Gesamtsystem war**		**inkonsistent**		**konsistent**	
			1 2 3 4 5 6 7 8 9			KA
	5.1.2	**Die aufgabenbezogene Ausdrucksweise war**	**inkonsistent**		**konsistent**	
			1 2 3 4 5 6 7 8 9			KA
	5.1.3	**Die verwendeten Begriffe waren**	**inkonsistent**		**konsistent**	
			1 2 3 4 5 6 7 8 9			KA

5.2	**Die Ausdrucksweise passte gut zu der auszuführenden Arbeit**		**immer**		**niemals**	
			1 2 3 4 5 6 7 8 9			KA
	5.2.1	**Die Verwendung von Fachbegriffen war**	**zu häufig**		**angemessen**	
			1 2 3 4 5 6 7 8 9			KA
	5.2.2	**Die Terminologie auf dem Bildschirm war**	**unklar**		**präzise**	
			1 2 3 4 5 6 7 8 9			KA

5.3	**Die Bildschirmmeldungen waren**	**inkonsistent**		**konsistent**	
		1 2 3 4 5 6 7 8 9			KA

	5.3.1	**Die Positionen der Anweisungen auf dem Bildschirm waren**	**inkonsistent**		**konsistent**	
			1 2 3 4 5 6 7 8 9			KA

5.4	**Die Bildschirmmeldungen waren**		**verwirrend**		**klar**	
			1 2 3 4 5 6 7 8 9			KA
	5.4.1	**Die Anweisungen für Kommandos und Funktionen waren**	**verwirrend**		**klar**	
			1 2 3 4 5 6 7 8 9			KA
	5.4.2	**Die Anweisungen zur Fehlerbehebung waren**	**verwirrend**		**klar**	
			1 2 3 4 5 6 7 8 9			KA

Bitte notieren Sie Ihre Kommentare über die verwendete Terminologie und die System Informationen hier:

_____

_____

_____

10.10.4 Lernfortschritt

6.1	**Der Lernaufwand, um das System bedienen zu können war**		**hoch**		**niedrig**		
		1 2 3 4 5 6 7 8 9					KA
	6.1.1	**Das Erlernen der Grundfunktionen war**	**schwierig**		**leicht**		
		1 2 3 4 5 6 7 8 9					KA
	6.1.2	**Das Erlernen der erweiterten Funktionen war**	**schwierig**		**leicht**		
		1 2 3 4 5 6 7 8 9					KA
	6.1.3	**Der Zeitraum, um die die Bedienung des Programms zu erlernen war**	**lang**		**kurz**		
		1 2 3 4 5 6 7 8 9					KA

6.2	**Das Ausprobieren von Eigenschaften und Funktionen durch Probieren war**		**entmutigend**		**ermutigend**		
		1 2 3 4 5 6 7 8 9					KA
	6.2.1	**Das Ausprobieren von Eigenschaften und Funktionen war**	**risikoreich**		**sicher**		
		1 2 3 4 5 6 7 8 9					KA
	6.2.2	**Das Entdecken neuer Eigenschaften und Funktionen war**	**schwierig**		**leicht**		
		1 2 3 4 5 6 7 8 9					KA

10.10.5 Systemeigenschaften

7.1	**Die System Geschwindigkeit war**		**zu langsam**		**schnell genug**		
			1 2 3 4 5 6 7 8 9				KA
	7.1.1	**Die Antwortzeit bei den meisten Operationen war**	**zu langsam**		**schnell genug**		
			1 2 3 4 5 6 7 8 9				KA
	7.1.2	**Die Anzeige der Informationen über die Verarbeitungsgeschwindigkeit war**	**zu langsam**		**schnell genug**		
			1 2 3 4 5 6 7 8 9				KA

7.5	**Es gilt: je mehr Erfahrung Sie besitzen, desto leichter können Sie mit dem System umgehen.**		**niemals**		**immer**		
			1 2 3 4 5 6 7 8 9				KA
	7.5.1	**Sie konnten Aufgaben ausführen, auch wenn Sie nur wenige Operationen kannten.**	**schwierig**		**einfach**		
			1 2 3 4 5 6 7 8 9				KA
	7.5.2	**Die Benutzung von speziellen Funktionen und Shortcuts war**	**schwierig**		**einfach**		
			1 2 3 4 5 6 7 8 9				KA

Bitte notieren Sie Ihre Kommentare über die Systemeigenschaften hier:

_____

10.10.6 Statistiken

10.1	**Die Qualität der verwendeten Grafiken und Statistiken war**		**schlecht**		**gut**	
			1 2 3 4 5 6 7 8 9			KA
	10.1.1	**Die Statistken waren**	**Unverständlich**		**Verständlich**	
			1 2 3 4 5 6 7 8 9			KA
	10.1.2		**Nicht hilfreich**		**hilfreich**	
			1 2 3 4 5 6 7 8 9			KA

10.4	**Die Farbverwendung war**		**unnatürlich**		**natürlich**	
			1 2 3 4 5 6 7 8 9			KA
	10.4.1	**Die Menge der verwendbaren Farben war**	**nicht angemessen**		**angemessen**	
			1 2 3 4 5 6 7 8 9			KA

Bitte notieren Sie Ihre Kommentare über die eingesetzten Multimedia Komponenten hier:
