

# GOOGLES INTERESSENPROFILING

Insbesondere Online Werbung, und damit die Finanzierungsbasis der meisten Internetdienste, ist ohne Profiling nicht mehr denkbar. Seit dem ersten „Klick-Mich“-Banner 1994<sup>1</sup> wächst die Industrie und mit dem Umsatz wuchsen auch die Komplexität der Struktur zur Werbevermittlung, die Anzahl der beteiligten Akteure und die dabei verwendeten Daten.

In folgendem Beitrag möchte ich die technischen Rahmenbedingungen von Online-Tracking und Profiling vorstellen. Die Beschreibung geschieht mit Blick auf meine eigene Analyse der Praxis von Google als größtem und wichtigsten Akteur, der ein Interessenprofiling (siehe Abb. 1) für einen Großteil der Internetnutzer\*innen durchführt. Dabei habe ich mich auf die Gruppe der User\*innen konzentriert, die nicht aktiv durch einen Login bei Google in das Profiling eingewilligt hat. Im Rahmen meines Dissertationsprojekts habe ich 506 Internetnutzer\*innen simuliert, um das Profiling zu analysieren. Dabei zeigt sich einerseits wie gut umfangreich die Profile sind die erstellt werden - nach 100 Seitenaufrufen hat Google im Durchschnitt 18 Interessen ermittelt sind - und andererseits die hohe Ungenauigkeit dieser Verfahren. Dieselben 100 Seitenaufrufe führen zu teils sehr unterschiedlichen Profilen.

Doch bevor ich einige Details meiner Untersuchung vorstelle, möchte ich kurz den Kontext erläutern, in dem das Profiling stattfindet. Ziel von Google und andere Anbieter\*innen<sup>2</sup> ist die Personalisierung von Werbung. Dem liegt die Annahme zu Grund, dass auf Profile zugeschnittene oder an vergangenem Verhalten ausgerichtete Anzeigen häufiger angeklickt und letztendlich mehr Vertragsabschlüsse oder Käufe produziert (*Conversion Rate*).<sup>3</sup> Operativ bedeutet das, dass Werbeanzeigen nicht einer möglichst großen Zahl von potentiellen Kund\*innen dargestellt werden soll, sondern vor allen Dingen diejenigen zum Kauf animiert werden, denen ein tatsächliches Interesse an einem Produkt oder eine Dienstleistung unterstellt wird. Ein Teil der Praktiken zur Verbesserung der „Conversion Rate“ wird unter dem Begriff *Online Behavioural Advertising*<sup>4</sup> zusammengefasst. Damit gemeint sind verschiedene Technologien, die das Surfverhalten von Internetnutzer\*innen verfolgen (Tracking) und auswerten, um

---

1 Zur Entwicklung des Online Werbemarktes seit seinen Anfängen siehe (Turow 2011)

2 Ghostery, ein AdBlocking und Online Tracking Analyse Dienst, listete zuletzt 2104 entsprechende Anbieter\*innen auf [HTTPS://WWW.GHOSTERY.COM/EN/APPS/](https://www.ghostery.com/en/apps/) (letzter Zugriff 05.06.2015)

3 Für eine weitergehende Kritik dieser Annahme siehe (McStay 2011)

4 Siehe zur Erläuterung dieses und anderer Begriffe (Fokusgruppe Targeting im BVDW 2014)



## Einstellungen für Werbung

### Einstellungen für Google Anzeigen

Durch Werbung können Webdienste und -inhalte kostenlos bereitgestellt werden. Mit diesen Einstellungen können Sie festlegen, welche Arten von Google Anzeigen für Sie eingeblendet werden.

Werbung bei Google		Google Anzeigen im Web <sup>?</sup>	
Geschlecht	Nicht verfügbar	Männlich <a href="#">Bearbeiten</a>	Basierend auf den von Ihnen besuchten Websites
Alter	Nicht verfügbar	35-44 <a href="#">Bearbeiten</a>	Basierend auf den von Ihnen besuchten Websites
Sprachen	Nicht verfügbar	Deutsch <a href="#">Bearbeiten</a>	Basierend auf den von Ihnen besuchten Websites
Interessen	Nicht verfügbar	Autos und Fahrzeuge und 11 weitere	Basierend auf den von Ihnen besuchten Websites

Abb 1: Googles Seite zur Einstellung für Werbeanzeigen, vgl.

<https://www.google.com/settings/ads>

Profile zu erstellen, die wiederum dazu dienen passende Werbeanzeigen auszuwählen. Dabei werden anhand der Liste aufgerufener Webseiten Aussagen über die Interessen und Charakteristika einer Person gemacht und Annahmen zum zukünftigen Verhalten getroffen.

## 1. ONLINE TRACKING UND PROFILE

Um die Art der Profile die durch Google erstellt werden genauer betrachten zu können möchte ich kurz verschiedene Formen von Profilen abgrenzen, die durch Online Tracking erzeugt werden<sup>5</sup>. „Profil“ meint im Folgenden eine Menge von Attributen, die idealerweise eine einzelne Person beschreiben. Die Person wird dabei häufig gleichgesetzt mit dem Gerät das oder dem Browser der zum Internetsurfen benutzt. Online Tracking Profile können eine unterschiedliche Zahl an Attribute umfassen, die wiederum verschiedene Abstraktionsniveaus (von der groben Genderkategorie bis zum Psychogram) enthalten. Zudem geht man davon aus, dass eine größere Datenbasis auch zu einer größeren Genauigkeit oder Aussagekraft führt.

Aus technischer Sicht entstehen die einfachsten Profile bereits auf der Netzwerkebene. Transaktionsprofile werden erstellt wenn ein Browser eine Verbindung zu einem Server mittels HTTP aufbaut. Bei jedem Seitenaufruf entsteht eine Art Profil des\*der Surfenden das Informationen über seine\*ihre IP Adresse und die Konfi-

5 Die Beschreibung erfolgt analog zur [Kategorisierung](#) von Pseudonymen in (Pfitzmann and Hansen 2010, 25ff)

guration des Browsers enthält. So können etwa Geräteabhängige Seiten ausgeliefert werden oder Zugriffe gewährt oder verweigert werden (*Geoblocking*). Solche Profile sind in ihrer einfachen Ausführung darauf beschränkt mehrere Seitenaufrufe einer IP-Adresse mit einem Server zu beinhalten. Diese, auf HTTP-Protokollebene vorgesehene, Form des Trackings wird von Webseitenbetreiber\*innen angewendet um Nutzungsstatistiken zu erstellen, welche Inhalte besonders beliebt sind oder beim *A/B Testing*, um verschiedene Varianten eines Online Dienstes zu vergleichen. Eine Zuordnung zu einer konkreten Person ist ohne Zusatzwissen (wer wann welche IP hat) nicht möglich. Es lassen sich nur Geräte und Browser unterscheiden.

Weiter verbreitet sind Profile die analog zu Rollen- oder Beziehungspseudonymen beschrieben werden können. Diese klassischen Online Tracking Profile aggregieren einen großen Teil der Webseitenbesuche eines\*r Nutzer\*in. Der Umfang hängt wesentlich davon ab in welcher Beziehung Nutzer\*innen zu bestimmten Unternehmen stehen. Wer kontinuierlich in den eigenen Facebook oder Google Account eingeloggt ist und das auf allen Geräten (Laptop, Tablet, Smartphone) mit er\*sie das Internet nutzen, über den\*die können wesentlich umfangreichere Profile erstellt. Facebook und Google sind damit in der Lage jegliche Webseitenaufrufe, die einen Element der Dienste enthalten, zu registrieren und mit dem Profil zu verbinden.

Technisch geschieht das Tracking hier nicht mehr nur auf Serverebene, sondern wird in Richtung der User verlagert indem Eigenschaften des Browsers zur Reidentifizierung genutzt werden.<sup>6</sup> Dabei nutzen die großen Netzwerke die Funktionsweise von Webseiten um die Reichweite ihres Trackings zu erhöhen. Durch Facebooks „Like“-Button und Googles Werbenetzwerk können auch Besuche auf Webseiten registriert werden, die Nutzer\*innen nicht über den Umweg der Suchmaschine oder des Newsfeeds erreicht haben.

Letztendlich arbeiten die Unternehmen daran ein vollständiges Personenprofil zu erstellen. Aktuell sind dem allerdings technische Grenzen gesetzt, etwa weil unterschiedliche Geräte genutzt werden (*Cross-Device-Tracking*), Nutzer\*innen Tracking durch Ad-Blocker unterbinden oder einfach gespeicherte Cookies löschen. Mittels verschiedener Techniken wird aber daran gearbeitet diese Lücken

---

6 Technische Verfahren dazu sind Cookie Tracking, aber auch Browser- oder Canvas Fingerprinting und nicht zuletzt die direkte Zusammenarbeit der Tracking Unternehmen beim Cookie Syncing. Siehe dazu ausführlich (Acar et al. 2014) bzw. [HTTPS://SECUREHOMES.ESAT.KULEUVEN.BE/~GACAR/PERSISTENT/INDEX.HTML](https://securehomes.esat.kuleuven.be/~gacar/persistent/index.html) (letzter Zugriff 10.06.2015)

zu schließen. Verizon<sup>7</sup>, ein amerikanischer Telekommunikationsprovider, nutzt zum Beispiel die Kontrolle über die Infrastruktur (auf IP-Ebene), um jedem\*r Nutzer\*in eine eindeutige ID zuzuweisen. Die dabei erstellten Profile umfassen auch Webseitenbesuche, die selbst keine Trackingsskripte enthalten.

## 2. NUTZUNG VON PROFILEN, ANREICHERUNG FÜR BESTIMMTE ZWECKE

Die so entstehenden Profile basieren auf Primärdaten, also solchen die eine Webseite oder ein\*e Netzwerkbetreiber\*innen direkt über den\*die Nutzer\*in in Erfahrung bringen kann. Einen tatsächlichen Nutzen haben diese Profile erstmal nur in einem eng begrenzten, eher technischen Rahmen wie zur Messung der Beliebtheit von bestimmten Webseiten. Für das Marketing nutzbar werden sie erst, wenn eine auf den Zweck bezogene Ebene hinzugefügt wird und Segmentierung möglich wird. Um Werbung zu schalten müssen Zielgruppen definierbar sein, die eine Kategorisierung und Sortierung nach marketingspezifischen Kriterien möglich machen.<sup>8</sup> Für Google und viele weitere Anbieter\*innen sind hier zwei unterschiedliche Arten von Kategorisierungen üblich. Erstens wird eine eher beschreibende Liste von soziodemografischen Kriterien, wie Altersgruppe, (binäre) Genderkategorien und verwendete Sprachen, erstellt und zweitens eine eher inhaltliche Ebene, die der vermuteten Interessen, abgeleitet.

Daten der ersten Kategorie werden meist über bekannte Korrelationen aus anderen Datenquellen ermittelt. Google kann hier mit den freiwilligen Eingaben arbeiten, die Nutzer\*innen mit Google Konto (etwa über GMail oder Google+) gemacht haben. In dem Vertrauen darauf, dass diese Eingaben, zumindest im statistischen Mittel, korrekt sind, lassen sie sich auf nicht registrierte Nutzer\*innen übertragen. Andere Unternehmen der Werbebranche können solche Informationen zu kaufen. Unternehmen wie QUANTCAST, ALEXA oder COMPETE, die im Bereich *Audience Analytics* tätig sind, verkaufen Datensätze (oder den Zugriff auf Datensätze) die Durchschnittswerte für die oben genannten Kategorien für die Nutzer\*innen von Webseiten enthalten. Die Mittelwerte für unterschiedliche Webseiten und die Stereotype der Nutzer\*innen dieser Seiten werden dann für ein konkretes Profil aggregiert und Wahrscheinlichkeiten ermittelt. Ein einfaches Beispiel auf Basis der Daten von Alexa.com: Wer vermehrt SPORT1.DE und HEISE.DE besucht für

---

7 Siehe dazu (Mayer 2014), zuletzt war Nutzer\*innen die Möglichkeit des Opt-Out angeboten worden, um das Tracking zu beenden.

8 Eine Praxis die im Internet stark optimiert und automatisiert ist, aber bereits im frühen 20 Jhd. einsetzt (vgl. Gandy 1993).

den wird angenommen, dass sie\*er männlich ist, einen akademischen Abschluss hat und auf der Arbeit surft.

Die zweite Kategorie von Informationen ist eine Liste von vermuteten Interessen. Bei Google basiert das Interessenprofil auf einer hierarchisierten Liste von 867 Interessen, die auf 24 Basisinteressen zurückgehen.<sup>9</sup> Quelltext 1 zeigt ein Element dieser Hierarchie

```
/Mensch und Gesellschaft > Familie und Beziehungen > Familie  
> Eltern > Babys und Kleinkinder > Spielzeug für Babys und  
Kleinkinder
```

Quelltext 1: Beispiel für eine Interessenhierarchie bei Google.

<b>Interessenkategorie (Anzahl der Unterkategorie)</b>	
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 1: Verteilung der Basisinteressen

Wie genau Google diese abgeschlossene Liste von möglichen Interessen herausgebildet hat ist nicht bekannt. Sie dient aber als gemeinsame Beschreibungssprache für Benutzer\*innen, die diese auf der Einstellungsseite ändern können, wie auch für Werbetreibende, die anhand der Liste Zielgruppen definieren können. Der Aufbau dieser Taxonomie ist also eng mit dem Ziel verknüpft und zeigt sich auch darin, dass die gelistete Zahl der Subkategorien pro Oberkategorie (siehe Tabelle 1), nicht jegliche Wissensbereiche abdeckt, sondern sich die Sortierung danach ausrichtet wofür Produkte angeboten und damit Werbung geschaltet werden könnte.

---

9 Weitere 1000 Interessen sind ortsspezifisch. Die vollständige Liste ist einsehbar unter [HTTPS://SUPPORT.GOOGLE.COM/ADS/ANSWER/2842480](https://support.google.com/ads/answer/2842480) (letzter Zugriff 04.06.2015)

Die Zuweisung von Interessen zu Nutzer\*in auf Basis von einer besuchten Webseite geschieht mit Hilfe von Inhaltsanalysen auf Webseiten, die für die Suchmaschinenpartei entwickelt wurden. Darüber hinaus haben Nutzer\*innen aber auch die Möglichkeit die Liste selbst einzusehen und zu verändern. Zuletzt konnte allerdings nachgewiesen werden, dass manuelle Änderungen an der Liste nicht zwangsläufig zu einer Veränderung in der Werbeschaltung führen.<sup>10</sup>

### 3. DER UMFANG DER INTERESSENPROFILE

Im Rahmen einer Studie habe ich einige Analysen dazu durchgeführt wie die Interessenprofile entstehen und welchen Umfang sie erreichen. Dazu habe ich mehr als 506 User (automatisiert) simuliert für die jeweils 100 Webseiten besucht wurden<sup>11</sup>. Im Anschluss wurden die auf der Einstellungsseite von Google angegebenen Interessen ausgewertet. Im Durchschnitt wurden durch Google jeder\*m Nutzer\*in 16,34 ( $s=7,5$ ) Interessen zugewiesen. Die Zahl lag dabei zwischen 1 und 37 der möglichen 867 Interessen. Bezogen auf die Basisinteressen zeigt sich, dass jedem\*r Nutzer\*in im Schnitt 8,25 ( $s=3,53$ ) der 24 Verfügbaren zugewiesen werden (siehe Abb. 2).

Nachweisen lässt sich, dass Googles Server im Durchschnitt über knapp 60% der Seitenaufrufe der simulierten User unterrichtet werden (vgl. Tabelle 2). Sei es durch die direkte Einbindung von Werbung (über Doubleclick) oder durch Google-Analytics<sup>12</sup>.

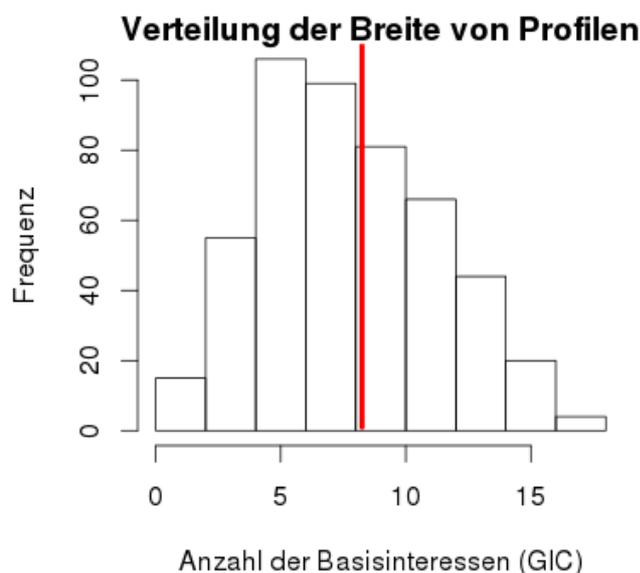


Abb 2: „Breite“ meint die Anzahl der Basisinteressen, die nach 100 Seitenaufrufen in einem Profil enthalten sind. Die rote Linie markiert den Durchschnitt.

en.

12 Dieser Anteil steigt, wenn auch die Einbindung weitere Dienste wie Googles Content Delivery Network (CDN), das Schriftarten und Skripte bereit gestellt wird.

Häufig wird angenommen, dass Google, als das Unternehmen mit der größten Internetreichweite, auch die genauesten und umfangreichsten Profile erstellt. Dabei wird impliziert, dass deren Genauigkeit über die Zeit und die Datenmenge wächst. Unter der Annahme, dass Google nicht andere Profile vorhält, die nicht beauskunftet werden, kann ich feststellen, dass insbesondere der letzte Punkt nicht nachweisbar ist. In den von mir beobachteten Fällen sind die Profile oft ungenau. Das Interesse an einem Thema wird dem Profil zugewiesen, wenn nur eine einzelne Seite dazu besucht wird, und sich andere Webseiten ausschließlich auf andere Basisinteressen beziehen. Es wird auch keine Unterscheidung getroffen zwischen übergeordneten Interessen die allgemein an eine Plattform gekoppelt sind und solchen die konkret mit einem Artikel zusammenhängen. So führt der Besuch von einem Artikel über ranziges Fett in Pommesbuden in einem Nachrichtenportal manchmal zur Zuweisung des Interesses an „Nachrichten“ und in anderen Fällen zum Interesse an „Essen und Trinken>Nahrungsmittel>Speisefette und -öle“. Unabhängig von einer, möglicherweise nicht dargestellten, Google-internen Priorisierung konnte ich beobachten, dass sowohl Basisinteressen als auch spezifische Interessen ähnlich schnell aus dem Profil „verschwinden“, wenn sie sich nicht bestätigen. Gleichzeitig konnte ich keine Regelmäßigkeit feststellen auf welcher zeitlichen Basis Interessen aus dem Profil gelöscht werden.

---

% des Link- Domains/Dienstleister profils	
50-60%	Google (google-analytics.com; doubleclick.com/net)
40-50%	scorecardresearch.com (Marktforschung), facebook.com, twitter.com
30-40%	quantserver.com (Marktforschung; Quelle für Quantcast.com)
10-20%	adnxs.com, taboola.com, outbrain.com, bluekai.com, Disqus.com, rubiconproject.com, addthis.com (Werbenetzwerke)
5-10%	chartbeat.com, optimizely.com, amazon-adsystems.com, krx.net

---

*Tabelle 2: Übersicht über Tracker und den Umfang der durch sie beobachtbaren Seitenaufrufe*

Allerdings macht die Ungenauigkeit im Profiling für Werbezwecke auch Sinn. Es geht eben nicht darum eine umfängliche und korrekte Persönlichkeitsbeschreibung vorzunehmen, sondern darum möglichst viele aktuelle Interessen zu kennen, für welche dann Werbung angezeigt werden kann. Da die Werbeschaltenden selbst auswählen müssen welchen Zielgruppen ihre Werbung angezeigt werden soll, ist es Googles vorderstes Ziel für möglichst viele Gruppen (und Kombinationen) eine möglichst große Menge an Zielprofilen vorzuhalten. Eben in der Hoffnung, dass mein Interesse an der Pommesbude auch bedeuten könnte, dass ich

mich über die Probleme eines Konkurrenten informiert habe und mich deswegen Angebote für Mengenrabatte bei der Palminbestellung ansprechen. Profile, wie Google sie erstellt, sind vom Zweck her auf die Zukunft gerichtet, da darauf basierende Werbung Bedürfnissen entsprechen soll, die noch nicht erfüllt sind oder noch gar nicht existieren.

Insofern sind die Profile die Google erstellt nicht klassische Personenprofile oder *Data Doubles*, im Sinne von digitalisierten und vermessenden Abbildungen, sondern Teilabbildungen, die zum Zweck der Werbung auch in die Zukunft projizieren. Die Ungenauigkeit, die vielleicht mal technisch begründet (war), weil keine genaue Interessenextraktion auf Basis eines Webseitenbesuchs möglich ist, streut nun auch notwendige Kontingenz ein, da Bedürfnisse auch für Produkte und Dienstleistungen geweckt werden sollen.

Abschließend lässt sich sagen, dass die vorstellte Form vom Profiling durch Google sich nicht dazu eignet die Ächtung des Konzerns als „Big Brother“ zu unterstützen. Dazu sind die Profile zu spezifisch auf den einen Zweck ausgerichtet und vor allem zu ungenau. Viel eher sind darin, und das gilt auch allgemein für Online Tracking im Werbereich, viele Elemente vereint die Lyon mit Bauman als *Liquid Surveillance*<sup>13</sup> identifiziert haben. Die Daten aus dem Browserverlauf ergeben keine vollständige Beschreibung, sondern dienen als *data particles* dazu, für den spezifischen Zweck neu sortiert und angereichert zu werden. Dabei versuchen die Dienstleister ihre Beobachtungen zu globalisieren; in Konkurrenz zueinander, aber eben durch Kooperationen<sup>14</sup> in Netzwerken. Die Allgegenwärtigkeit des Profiling zeigt sich nicht zuletzt darin, dass sich auch die nicht-Teilnahme, z.B. durch Ad-Blocker, registrieren und diese Nutzer\*innen als Zielgruppe adressieren lassen.

## 4. REFERENZEN

Acar, G., C. Eubank, S. Englehardt, M. Juarez, A. Narayanan, and C. Diaz. 2014. "THE WEB NEVER FORGETS: PERSISTENT TRACKING MECHANISMS IN THE WILD." In *Proceedings of the 21st ACM Conference on Computer and Communications Security*.

Datta, Amit, Michael Carl Tschantz, and Anupam Datta. 2014. "AUTOMATED EXPERIMENTS ON AD PRIVACY SETTINGS: A TALE OF OPACITY, CHOICE, AND DISCRIMINATION." *arXiv:1408.6491 [cs]*.

---

13 Damit beschreiben sie die zeitgenössische Überwachungskultur die, in aller Kürze, disziplinarische und kontrollgesellschaftliche Elemente vereint (Lyon 2010; Lyon und Bauman 2013)

14 Siehe Cookie-Syncing; Fußnote 6

#9 - Martin Degeling – Googles Interessenprofiling

Fokusgruppe Targeting im BVDW. 2014. "TARGETING BEGRIFFE UND DEFINITIONEN."  
BVDW e.V.

Gandy, Oscar H. 1993. *The Panoptic Sort - A Political Economy of Personal Information*. Boulder u.a.: Westview Press.

Lyon, David. 2010. "LIQUID SURVEILLANCE: THE CONTRIBUTION OF ZYGMUNT BAUMAN TO SURVEILLANCE STUDIES" *International Political Sociology* 4 (4): 325-38.  
doi:10.1111/j.1749-5687.2010.00109.x.

Lyon, David, and Zygmunt Bauman. 2013. *Daten, Drohnen, Disziplin: ein Gespräch über flüchtige Überwachung*. Berlin: Suhrkamp.

Mayer, Jonathan. 2014. "HOW VERIZON'S ADVERTISING HEADER WORKS." *Web Policy*. October 24. .

McStay, Andrew. 2011. *The Mood of Information: A Critique of Online Behavioural Advertising*. A&C Black.

Pfitzmann, Andreas, and Marit Hansen. 2010. *A TERMINOLOGY FOR TALKING ABOUT PRIVACY BY DATA MINIMIZATION: ANONYMITY, UNLINKABILITY, UNDETECTABILITY, UNOBSERVABILITY, PSEUDONYMITY, AND IDENTITY MANAGEMENT*.

Turow, Joseph. 2011. *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth*. New Haven: Yale University Press.