

Data, Trust, and Transparency in Personalized Advertising

by

Darren M. Stevenson

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
(Communication)  
in the University of Michigan  
2016

Doctoral Committee:

Professor Christian E. Sandvig, Chair  
Associate Professor Sonya Dal Cin  
Professor Nicole B. Ellison  
Professor Shane M. Greenstein, Harvard Business School

© Copyright Darren M. Stevenson, 2016

All Rights Reserved

For my father  
who sacrificed much so that I might have opportunities he did not  
who always encouraged me to do my best

## **Acknowledgements**

I was extremely privileged to pursue doctoral studies. I was also quite fortunate that my particular experience included working alongside so many remarkable people across several stimulating departments and institutions. At each point along the way I received generous intellectual, professional, social, and financial support. As a result, I am indebted to a great number of individuals, academic departments, and organizations willing to support me in various ways during this time. Attempting to recall and thank all those who substantively contributed to my doctoral education and the research reported in this document underscores the collective effort it is to complete a Ph.D., rather than an individualistic endeavor. Additionally, while this dissertation builds directly on the theories, ideas, and in many cases the precise suggestions of numerous colleagues, any errors herein are my own.

First and foremost, I am immensely grateful to Christian Sandvig, my Ph.D. supervisor. Christian guided my doctoral studies and this entire research project. Unbeknownst to us at the onset, these efforts would span multiple universities. Still, Christian stuck with me and with the project from beginning to end. I am thankful for his unwavering support over the past few years. Christian essentially introduced me to social science research and then led me to most of the specific ideas, literatures, and methods from which my own work would build upon. For years he routinely encountered my ideas and writing in their rawest form, moments when he would skillfully separate the wheat from the chaff. Often the chaff was quite dense. Yet, he engaged

this onerous task with patience, reforming and elevating my work with rigor and sophistication. Christian offered me immense freedom in my research program. He tolerated my repeated mistakes. He allowed me to mature as a scholar at my own pace. For these reasons and many others, any successes in this work are largely to his credit, the result of Christian's dedication as a scholar, mentor, and friend. Additionally, despite the unique intermediary role played by research-intensive universities in relation to people, government, private enterprise, and other organizations, that academic inquiry proceeds first in the interest of the public, above all other interests, is not a given. So it is my hope that Christian's unfaltering commitment to this ideal, to pursuing research motivated by and conducted in the public's interest, has permeated both my thinking and the pages of this work.

I am also intellectually indebted to the other members of my superb dissertation committee: Sonya Dal Cin, Nicole Ellison, and Shane Greenstein. Each shaped my thinking, research, analysis, and writing in their own unique ways. Along with Christian, this eclectic group demonstrated a richness in thought possible when individuals divergent in expertise, training, and even intellectual commitments converge on a problem of mutual interest. Intermingling these diverse, and occasionally conflicting, viewpoints and advice gave rise to a project that was both more colorful and certainly more enjoyable to work on than what may have emerged under a more uniform team of mentors. I am grateful to these scholars for locating common ground in my research project, for their generous time over the past few years, and for their steadfast dedication to this work and to my broader graduate education.

Individually, Sonya Dal Cin was an instrumental guide and considerate sounding board, always willing to offer an ear while keenly articulating my blind spots, the latter of which there

were many. Sonya always encouraged me to keep an open mind. She gently pulled me back towards her own form of smart objectivity whenever I veered off various academic cliffs.

Nicole Ellison provided incisive criticism, perhaps in a way only she can. Nicole offered meticulous feedback and added rigor precisely where I was lacking it. Her contributions not only enhanced this dissertation greatly but, more generally, improved how I came to think about social scientific inquiry and knowledge production itself.

Shane Greenstein offered much-needed extramural guidance, doing so with an insider perspective given the subject matter at hand. Shane's guidance and feedback helped to ground this work in reality. Shane challenged me to consider important questions, considerations I would have otherwise overlooked completely. He is also partially to blame for some of my optimism in regards to the topic under investigation here, which hopefully shows itself at different points in this work.

In addition to my committee, I am grateful to a host of additional informal mentors and colleagues who generously provided input and guidance on this work. For some this involved giving regular feedback on my thinking, research, and writing throughout my doctoral studies. Others offered less regular but equally influential feedback at various points along the way. Attempting to enumerate each of these individuals guarantees only that some will be omitted, and a big "Thank you, too!" to those I have unintentionally forgotten. Nevertheless, I wish to thank the following people for their generous time, feedback, and support on various parts of this dissertation project: Alessandro Acquisti, Elaine Adolfo, Van Anderson, Mark Andrejevic, Melissa Aronczyk, Michelle Ash, Cheol Gi Bae, Solon Barocas, James Barrett, Rajeev Batra, Joe Bayer, Catherine Blake, André Brock, Geoff Bowker, Matt Burton, Ergin Bulut, Ryan Calo, Chris Cascio, Jonathan Cave, Melissa Chalmers, Ajaay Chandrasekaran, Cliff Christians, Matt

Crain, Mathias Crawford, Leif Dahlberg, Norm Denzin, Mark Deuze, Steven Doran, Susan Douglas, Nora Draper, Allison Earl, Amy Eaton, Fernando Elichirigoity, Lorenz Engell, Fredrik Enoksson, Sarah Erickson, Ted Faust, Joe Ferrar, Christian Fuchs, Thomas Gambill, Martha Garcia-Murillo, Kelly Gates, Roland Gau, Adam Gerard, Lauren Guggenheim, Kristen Guth, Kevin Hamilton, David Hauser, Nicky Hentrich, James Hay, Bryce Henson, Muzammil Hussain, Annemarie Iddins, Jennifer Ihm, Matt Jackson, Mo Jang, Krishna Jayakar, Joy Yang Jiao, Leslie John, Steve Jones, Karrie Karahalios, Dam Hee Kim, Art Kramer, Ozan Kuru, Nojin Kwak, Tom Levin, Nina Li, Tim Libert, Mike Lorenc, Amanda Lotz, Shantel Martinez, Bob McChesney, Robin Means Coleman, Robert Mejia, Jessica Moorman, Yioryos Nardis, Annemarie Navar-Gill, John Nerone, Russ Neuman, Helen Nissenbaum, Jonathan Obar, J.P. Obley, Didem Ozkul, Karla Palma, Josh Pasek, Ben Pearson, Jean-Christophe Plantin, Alison Powell, Aswin Punathambekar, Charles Ransom, Julia Raz, Colin Rhinesmith, Aimee Rickman, Scott Robinson, John Rogers, Travis Ross, Anita Say Chan, Paddy Scannell, Amit Schejter, Barbara van Schewick, Doc Searls, Scott Selberg, Ashkan Soltani, Stuart Soroka, Paula Treichler, Joe Turow, Ellen Wagner, Daniel Weber, Christo Wilson, Siva Vaidhyanathan, Derek Vaillant, Anghy Valdivia, Dora Valkanova, Yue Wang, ShinJoung Yeo, and Rebecca Yu, along with others who aided this work but whom I am regretfully and unintentionally forgetting.

I have also benefited from the support of numerous academic departments and organizations during the past few years. Most notably, these were the Department of Communication Studies, Institute of Social Research (ISR), Media Psychology Laboratory, and Horace Rackham School of Graduate Studies all at the University of Michigan; Center for Internet & Society (CIS) at Stanford Law School; School of Computer Science & Communication at the Royal Institute of Technology (KTH) Stockholm; Institute of

Communications Research (ICR), College of Media, Center for People & Infrastructures, and Beckman Institute for Advanced Science & Technology all at the University of Illinois at Urbana-Champaign; Donald Bren School of Information & Computer Science at the University of California, Irvine; International Research Institute for Cultural Technologies & Media Philosophy (IKKM) at Bauhaus-Universität, Weimar; Social Science Research Council's (SSRC) Study Group on Communications and Public Policy; Télécom Ecole de Management at the Institut Mines-Télécom, Paris; Italian Society for Social Studies of Science and Technology; Swedish Foundation for International Cooperation in Research & Higher Education (STINT); Open Technology Institute at the New America Foundation; Searle Center on Law, Regulation, & Economic Growth at Northwestern University's Pritzker School of Law. These institutions, departments, and groups provided collegial environments for learning and in many cases generous financial support towards my research and scholarship.

Studying in various departments and organizations also allowed me to engage with a multitude of thoughtful classmates and colleagues, benefitting from these relationships both at work and at play. Often there was too much of the former and too little of the latter, though I am thankful balance managed to prevail on occasion. It would be difficult to thank so many people individually here, so a blanket thank you will have to suffice. Regardless, I am extremely grateful for the *social support* these individuals offered. I believe normal people simply call this “friendship.” It has made all the difference.

I am also grateful for my many loving friends and family members outside the academy for their unwavering support and encouragement. My utmost thanks for listening to my ideas, opinions, and occasional rants during this time, especially when this ventured into the nuances



(nonsense?) of topics you may not have cared about per say but cared about me and therefore chose to engage.

Finally, I am tremendously thankful for my partner in life and exceedingly generous wife, Jennifer (Fetzer) Stevenson. Jen, you more than anyone bore the brunt of my Ph.D. experience. You lived your life in sacrifice mode so that I could attend to my pursuits, not the least of which included an unexpected relocation to Ann Arbor. Thanks is not enough; still, thank you. You offered me perspective and motivation when I needed it the most. You endured my “conversations” (speeches) on the minutiae of academia, my colleagues, and whatever topic I was overly preoccupied with at the time. You worked tirelessly to financially support our family while simultaneously being a wonderful mother. You endured my travel schedule, my late nights and weekends spent working, and all the times I allowed pursuing a Ph.D. to engulf not just my own life but also yours. When one of us needed to be present, physically or mentally, you were always there. You made it such that my life would work during this period at the expense of your own needs and ambitions. And you did all of this with love and grace. May I somehow be as good to you in the years and decades to come as you have been to me during this busy period of life.

## Table of Contents

Dedication .....	ii
Acknowledgements .....	iii
List of Figures .....	x
List of Appendices .....	xi
Abstract .....	xii
Chapter 1	
Introduction .....	1
Chapter 2	
Context of the dissertation studies .....	12
Chapter 3	
Third-party personal data and personalized advertising up close and personal .....	41
Chapter 4	
It acts like it knows me: How consumers think about personalized online advertising .....	88
Chapter 5	
The trouble with transparency: Experimental evidence for the negative effects of transparency on trust and preference for personalized advertising .....	150
Chapter 6	
Conclusion .....	221
Appendices .....	225
References .....	290

## List of Figures

### Figure

2.1	Simplified internet advertising diagram .....	18
2.2	Real-time bidding process .....	23
2.3	Data privacy and trust in organizations .....	33
4.1	Examples of participant drawings of personalized advertising .....	112
4.2	Three styles of drawings resembling network diagrams .....	113
4.3	Example of basic linear model .....	113
4.4	Example of linear cycle model .....	114
4.5	Example of interconnected model .....	115
5.1.	Average approval ratings of consumer data use in experiment 2 .....	182
5.2.	Results of manipulation check for experiment 2 .....	184
5.3.	Results of manipulation check for experiment 3 .....	199
5.4	Results of manipulation check for experiment 4 .....	205

## **List of Appendices**

### Appendix

A	Focus group moderator question guide .....	225
B	Participant word associations for personalized online advertising .....	227
C	Participant drawings of personalized online advertising .....	228
D	Manipulation for experiment 1 .....	267
E	Dependent measures for all experiments .....	272
F	Study-specific acknowledgements .....	288

## **Abstract**

Advertising wields the power to change the way we see and experience the world and how we perceive those around us. Though marketing practices have long been characterized by information asymmetry, with individuals unable to see the extent to which data describing them is held by various organizations nor how it is used, recent developments have intensified this arrangement. For both firms and individuals, the personalization of online advertising content is justified by increased efficiencies. Marketers benefit from cost savings by reducing expenditure wasted on reaching individuals who fall outside the target audience, providing better return on advertising investment. Individuals benefit from advertising personalization by encountering marketing content they are measurably more likely to be interested in, filtering the cacophony of advertising marketers seek to distribute to different audiences. And yet the opaque processes by which advertising content is selectively presented online prevents individuals from making reasonable judgments about contemporary media systems and practices. In light of these challenges, in this dissertation I investigate how today's evolving, digital marketing system organizes interaction between marketers and individuals. Three empirical studies are presented offering insights into how marketers envision their audiences, how audiences envision marketing practices, and, together, how each have come to understand and use the data, information, and communication technologies that now bind them together. In the first study, using participant observation I examine the nature and dynamics of third-party personal data available to

marketers on digital ad-buying platforms. In the second study, drawing on a series of focus groups I uncover how individuals reason about advertising personalization, focusing on the mental models people rely on when interacting with advertising they perceive to be personalized. In the third study, across four experiments I examine the causal influences of transparency and trust on how individuals make judgments about personalized advertising. Viewed together, findings from this work may be of interest to marketing managers who rely on advertising personalization techniques, designers and developers of technologies leveraging consumer data collection, policymakers who oversee advertising and digital privacy matters, and academic researchers in the fields of communication, marketing, management, public policy, and human-computer interaction.

## Chapter 1

### Introduction

Consider the following hypothetical scenario. John is an accountant. He lives in Minneapolis. Last year around this same time, along with his partner, John spent a week in Tuscany. A dream vacation on all accounts. In fact, on a wall in the couple's home office hangs a small framed photograph from the trip depicting the two predictably "supporting" the notorious tower in Pisa. But today, while sitting at his computer before heading off to work, John skims a friend's post on a social media platform, reading about her troubles with her boss. Scrolling on, interspersed between friends' updates about politics and toddler photos, John glances at an advertisement in his social media news feed: *CHEAP FLIGHTS to Italy!* Briefly, his mind wonders. He looks up at the photograph on the wall from his trip and then back at the advertisement on the computer screen. He wonders whether he is seeing this ad because of the Tuscany trip he took, or if it is simply a coincidence? He wonders if other people looking at their news feed on this same platform are also seeing this ad, or is it just him? Maybe his partner, eager to take another trip, has been looking up flights on this shared computer. But would that affect the ads appearing in John's own news feed? He is not sure. He then recalls previously seeing an ad in on this platform for a local barbershop just down the street. So there must be a way to show ads to people according to where they live, right? He concludes, yes, it wouldn't make sense if everyone else around the world using this same social media platform saw the

local barbershop's ads. But what about the advertisement currently in front of him, for flights to Italy? Is everyone else in Minneapolis also seeing this ad? He scrolls on.

If this example were real, the ad for flights to Italy would probably be the result of information linked to John and stored across multiple company databases. These data would indicate that John (1.) traveled from Minneapolis to Pisa via British Airways seated in economy class between 8-14 months ago, flagging him as having an above average likelihood to purchase economy class airfare to Italy in the future, and (2.) charged \$269.47 to his credit card at Ristorante All' Acquacotta, an amount causing a customer relationship management system to also flag John as a "big spender," as this amount is three times the average bill recorded at Ristorante All' Acquacotta in the previous year (thus meeting the system's automatic threshold for the "big spender" designation). Had John only one of these two flags, he would not have seen the personalized advertisement for flights to Italy, but he has both.

From John's vantage point he questions but is unable to determine why he is seeing this particular advertisement, but a marketer placing the ad would know. This fictitious example demonstrates the commonly experienced phenomenon of information asymmetry in personalized online advertising (Nissenbaum, 2010, p. 36-50). The disparity between who can understand the advertising process and who cannot affects fundamental aspects of our everyday experiences and wellbeing. It affects what we can see or know about the data that describes us and what we know about the other side: digital marketing and data aggregation.

Recent developments in marketing have resulted in an increasing integration between information gleaned from our online behaviors, as recorded by a range of techniques including web cookies and more persistent tracking mechanisms, with information about our otherwise offline behaviors. This has also contributed to information asymmetry, often of an unknowable



character and proportion. This itself is not necessarily a problem. However, the ways this unevenness is leveraged can be problematic. As others have noted (Turow, 2011), personalized advertising wields the power to change the way we see and experience the world and how we perceive those around us. For these reasons, investigating advertising personalization can help scholars, practitioners, policymakers, and consumers better understand the impact of increasingly personalized media experiences on contemporary life.

### **Challenges Posed by Consumer Data Collection and Advertising Personalization**

The existence of information asymmetry between marketers and consumers can limit consumer autonomy, degrade privacy (Fischer-Hübner et al., 2011), grant marketers undue influence over consumers (Calo, 2014), and jeopardize fairness in the marketplace (Culnan & Armstrong, 1999; Barocas, 2014). Alternatively, digital content personalization benefits consumers when it provides content in which individuals are measurably more likely to be interested. Yet there is also reason to believe the opaque process by which content is algorithmically-curated and selectively presented prevents online audiences from making reasonable judgments about media systems and practices.

Another source of problems for consumers is the rapid pace at which content personalization techniques change as information and communication technologies advance. For example, by the time awareness of internet tracking cookies entered the popular imaginary, marketers had moved on to newer methods of tracking consumer behavior online. Further, the use of opaque consumer data collection practices, combined with the challenge for individuals to understand an excess of methods and technologies that enable personalization, makes it unrealistic to expect most people to sufficiently understand and make informed decisions about

this process. In effect, this precludes many individuals from exerting agency among the multitude of web-based platforms and services they use (e.g., social media, e-commerce, search engines).

Naturally, as personalized digital content becomes even more prominent it is likely that this public understanding of personalized advertising practices will also grow. However, increased awareness of personalization in advertising has also been associated with increased disapproval of this practice (e.g., Turow et al., 2009, 2015; Kim et al., 2015). Therefore, progression of public knowledge on this issue could lead to serious problems for marketers who hope to rely on personalized digital advertising given its many benefits.

The opportunities posed by new abilities to collect consumer data and put these data to use for curating advertising content also create new problems for marketing firms and other technology companies. Insofar as personal data collection for ad personalization is perceived to be objectionable by consumers and regulators, firms employing these methods will incur an indirect economic challenge given the vital need to maintain trust with their customers and regulators. This is a serious challenge, as most of the web has evolved from its early origins, primarily serving universities and governments, to instead being organized around market forces and the dynamics of consumer needs and private enterprise (Greenstein, 2015). Today, many of the most popular services of what is now a commercial web rely on a revenue model that offers products and services at no monetary cost in exchange for consumers' time, attention, and personal data (Hoegg, Martignoni, Meckel, & Stanoevska-Slabeva, 2006).

Some measures of public opinion also signal that consumers increasingly disapprove of how their personal data is collected, used, rented, and sold for marketing purposes (e.g., Turow, Hennessy, & Draper, 2015). It stands to reason then that consumers may find products and

services known to employ these data collection practices to be less desirable and to be avoided when possible.

In another signal that companies may face a coming personal data collection backlash, in 2011 the World Economic Forum (WEF) launched its ongoing, multi-year initiative *Rethinking Personal Data*. To date, the WEF project has included a series of studies and subsequent reports calling for reform in how firms collect and use personal data in light of the potential negative economic impacts, most notably due to loss of consumer trust. Central to this initiative is combatting current consumer data practices that appear to be undermining the trust individuals have in firms and other organizations who collect and put to use consumer data in ways individuals would prefer they did not. The WEF initiative calls attention to the great economic possibilities linked to advances in how digital personal data can be put to use for firms, governments, and individuals, while also noting the contentious and aggressive nature with which firms, including marketers, now aggregate personal data and often with little regard for consumer preference. One of the specific prescriptions from the WEF project is to develop new ways to increase trust between consumers and organizations in regards to personal data by simply making consumers more aware of how these practices work. A recent WEF report concluded that, “An information differential exists between institutions and individuals, creating a crisis of trust that results from uses of data being inconsistent with user expectations and preferences,” offering that, “Context-aware data usage is a key element in restoring this trust” (WEF, 2014, p. 1). Similarly, another report issued the year prior, titled *The Internet Trust Bubble* (2013), included results from a global survey of 11,000 Internet users reporting that a majority believe they put their privacy at risk when going online, have mixed feelings about sharing their data with third parties (p. 10), and reported that personal data was routinely

collected about them online but without their understanding of who was collecting this data and for what purpose, all leading to reduced trust in the organizations with which people must routinely interact (p. 37).

### **Contribution of the Dissertation**

In light of these challenges posed by consumer data collection and the personalization of advertising on the internet, this dissertation examines how individuals and marketers try to make sense of one another. Marketers' goals in these interactions with consumers are somewhat straightforward. These include influencing individuals, brand building, and encouraging purchases through the use of advertising messages and other marketing techniques. When it comes to interaction with marketers through advertisements, the goals of individual consumers vary considerably by person, context, time, and many other factors. Additionally, individuals' goals for interaction with advertising messages are less focused and less consistent compared to the goals of marketers during these interactions. For example, in one instance an individual might actively seek information about a product of interest from an advertisement they happen to encounter, a message which has been selectively delivered to them by marketers based on what is known about them. In another moment, this same person might seek to avoid advertising completely, instead hoping to minimize at all costs the amount of time and attention they allot to advertising messages. Accordingly, the practice of advertising, including personalized digital advertising, can be both useful and inconvenient to consumers. To investigate the dynamics of trust, transparency, and mutual understanding between marketers and consumers within these frequent, digitally-mediated interactions, this dissertation presents three empirical studies. The overarching goal of this work is to better understand important dimensions of recent frictions

brought about by consumer data collection and personalized advertising, generally, along with some of the specific practices used to support advertising personalization today.

## **Chapter Outline**

This chapter (Chapter 1) introduced the dissertation. I presented the problem, offered a brief rationale for pursuing this work, and below I mention the specific research questions that have been explored in this work. As noted, this dissertation consists of three self-contained empirical studies, each of which is detailed in its own chapter. Though these three studies are largely independent of one another, they do follow a logical as well as chronological progression and are discussed accordingly.

In Chapter 2, I offer the context from which the dissertation proceeds. Although each of the subsequent three studies includes its own literature review, this chapter is intended to provide helpful background information for the reader by summarizing overarching literature, information, and examples relevant to the three studies of the dissertation. Chapter 2 does not present new work but rather is intentionally descriptive in nature. It also provides a foothold for the reader unfamiliar with the topic under investigation.

In the first study (Chapter 3), I focus on how marketers, systems designers, and suppliers of consumer data have imagined, classified, and offered up internet users and their corresponding consumer data for use in personalized online advertising. I explored this phenomenon not by posing questions to these different groups of people themselves but instead by placing myself in the shoes of the marketer. I examined first-hand the system for advertising personalization these individuals, together, have created. As the primary technological innovation of this system is in its affordances related to personal data rather than, for instance, the content of advertisements

themselves, I focus on the character and dynamics of the personal data now supporting advertising personalization.

Theoretically, in Study 1 I draw on and extend three key ideas from the literature related to digital privacy, consumer data, and advertising personalization. First, I leverage the notion of *digital enclosure*, as developed by Andrejevic. Second, I rely on Nissenbaum's heuristic known as *contextual integrity*. Third, I draw upon the notion of *information asymmetry*, a concept others scholars working in related areas have similarly noted. Together, these ideas supply the theoretical guidance and thrust of Study 1. Methodologically, Study 1 is a mixed-methods investigation in which I take on the position of a marketer. It combines participant observation with a short series of computational advertising tests. My participant observation revealed a new synthesis of consumer data that includes the implications of a wide range of information now readily available to marketers for selectively presenting ads online to some individuals and not others. For the computational tests, I bought online ads just as a marketer might. These tests explored the dynamics of consumer data as used in real-time auctions for online publisher inventory.

As this study provided the entry point into the subsequent two studies and my broader dissertation project, Study 1 was primarily exploratory. For Study 1, I posed the following two questions:

*What is the character of the personal data made available to marketers by third-party data providers for personalizing online advertisements?*

*How do the dynamics of real-time bidding impact the ability of marketers to personalize advertising for certain audiences over others?*

In Studies 2 and 3 I shift my focus away from examining the marketer's perspective and toward individuals, researching how ordinary people reason about advertising personalization.

In the second study (Chapter 4), I uncover how individuals think about advertising personalization. Theoretically, I focus on the role played by *mental models* in everyday life, especially during interactions with digital advertising. This focus on mental models extends work most commonly found in studies of Human-Computer Interaction (HCI), including those pioneered by Norman (1983) and since taken up for much broader application. This study complements recent research studying how individuals reason about algorithms online (Eslami et al., 2015, 2016). It also builds on a larger body of scholarship that seeks to understand the function and importance of the invisible mental models users and designers develop in response to using and designing technological systems. As mental models are theorized to influence just how successful individuals' interactions with products and systems can be, in Study 2 I extend this work from HCI into contemporary advertising studies, as few in this field have considered the role played by mental models in consumers' interactions with advertising. Methodologically, in this qualitative study I draw on six focus groups I conducted with adults. The purpose of these focus groups was to generate themes and insights that address the following three questions:

*How do people reason about personalized advertising?*

*What mental models do individuals develop in response to their experiences with personalized advertising?*

*How useful are these mental models in assisting individuals in their interactions with advertising personalization? That is, how well do they predict individuals' encounters with personalized advertising?*

In the third study (Chapter 5), I aimed to leverage the tight control and internal validity of experiments to determine what factors might be influencing the way consumers make judgments

about the costs and benefits of personalization. Study 3 draws its theoretical basis from recent empirical findings pointing towards negative effects of transparency on consumer judgments about advertising personalization. I also employ theories related to trust, both in the general form of social trust—that is trust in other people—as well as more specific feelings of trust as expressed towards advertising and other media artifacts. Further, following the work of Reeves and Nass (1996), I briefly explore the compatibility between how individuals relate to other people and how they relate to new media artifacts, including personalized ads and web platforms that deliver personalized ads. In Study 3, I explored the following related questions:

*How does transparency in personalized advertising affect individual preference for personalization, privacy concerns, trust in advertisements, and trust in the web platforms that deliver personalized advertising?*

*Similarly, how does social trust affect preferences for advertising personalization, trust in advertisements, and trust in the web platforms that present advertising?*

Finally, in Chapter 6, I briefly conclude by summarizing and synthesizing the results from the three studies reported in Chapters 3, 4, and 5. As stated earlier, these three studies are intended to be self-contained. However, across all three, I examine how today’s evolving digital marketing system organizes interaction, aiming to do so such that these investigations might compliment one another. In doing so, I provide insights into how marketers envision their audiences, how audiences envision marketing practices, and, together, how each have come to understand and use the data, information, and communication technologies that now bind them together.

Broader implications of this work and findings from the three studies are likely to be of interest to marketing managers and other practitioners who rely on personalization strategies and techniques in delivering targeted advertising, designers of technologies for personal data



collection and marketing analytics, and those policymakers in the U.S. and around the world who oversee advertising practices along with governance of privacy and consumer data. Each of these stakeholder groups might benefit from the empirical work presented in this dissertation by gaining further understanding of how the recipients of personalized advertising, that is consumers, think and act.

## **Chapter 2**

### **Context of the dissertation studies**

This chapter details background information helpful for understanding the context from which the three empirical studies in the dissertation emerged. The aim is to call attention to the conditions under which this larger research project takes place. As such, the chapter is intentionally descriptive, presenting no new work and minimal analysis of existing work. Instead, relevant examples and context are presented. I trace recent developments in technology and business practices that have contributed to some of the challenges currently faced by both marketers and consumers in light of personal data collection and efforts to personalize advertising. Most of these challenges surround efforts to customize advertisements in ways marketers find practical and highly effective, efforts consumers at times deem undesirable. As I attempt to show in this chapter, striking an appropriate balance in this regard has been challenging for both marketers and consumers. A better understanding of the context as presented in this chapter will provide the unfamiliar reader with necessary background information while also serving to situate the dissertation project in the contemporary moment.

#### **Internet Expansion**

In the U.S., the daily lives of most individuals include using internet-connected products and services. This is the result of relatively high adoption rates for broadband internet (67%) (Horriagan & Duggan, 2015) and smartphones (64%) (Smith, 2015), along with steady growth

over the past 10 years in the adoption of social media and social networking platforms, with most American adults (65%) now using at least one social media or social networking site (Perrin, 2015). In particular, mobile internet connectivity permits the constant tethering of individuals both to networked information and to other people. For instance, a 2013 study reported smartphone users had their devices with them everyday of the week and for nearly all of their waking hours, with most (79%) checking their smartphone within 15 minutes of waking up in the morning (International Data Corporation, 2013). An increasingly internet-connected populace, one gradually more and more dependent on web services for accomplishing routine tasks, recalls a similar shift in the U.S. towards daily dependence on electricity just a century prior (Nye, 1992).

This array of online platforms and services that serve to connect people and information now support diverse activities for economic, civic, and social good (Benkler, 2006; Castells, 2013; Rainie & Wellman, 2012; Zuckerman, 2013). Accompanying these activities are numerous enduring challenges, issues that far predate the web's existence. These social issues are altered and in some cases amplified by the internet's unique affordances. For instance, complications related to freedom of expression, normative notions of privacy, access to information, and the quality of news and other media, these and other problems persist, perhaps unsurprisingly, in today's online space. The digital environment is one full of opportunities and ills alike.

The challenges of online interaction have attracted the attention of academic researchers, many of whom use the internet as an object through which to investigate matters of broad social importance (Dutton, 2013). It is now commonplace for scholars in many traditional disciplines, such as sociology, psychology, and anthropology, to take up internet-related social problems. In addition to this, the web has also birthed, strengthened, and/or coalesced new sub-fields and

specific lines of inquiry, such as human-computer interaction, internet studies, digital studies, new media, computer-mediated communication, computer supported cooperative work, and the study of online communities. Linking all of these efforts are the interactions between individuals and computing technologies, tools which today almost always depend in some way on the connective capacities of the internet. Rarely do we think of computing without simultaneously conjuring the internet.

### **Digital Consumer Data & Marketing**

Entwined in many internet-related social issues is the rise of digital information emanating from and corresponding to individuals, often referred to as personal information or personal data.<sup>1</sup> Increasingly-inexpensive digital storage has supported not only the basic actions and interoperability of the web but also the broader trend towards digitally recording and sorting accounts of human behaviors (Gandy, 1993). The result is that web's default is one of remembering rather than one of forgetting. Further, ubiquitous computing supports ubiquitous records of information about individuals and their behaviors (Lyon, 2014). From routine server logs, essential for proper maintenance and security, to more elective monitoring activities, like third-party tracking cookies, digital information corresponding to individuals and their actions is increasingly abundant. This data has shown to be quite valuable, creating vast economic benefit (Deighton & Johnson, 2013, 2015), while also causing controversy (Electronic Privacy Information Center, 2016).

---

<sup>1</sup> As used in this dissertation, "personal data" refers to data (or information) emanating corresponding to people, as opposed to less person-centered forms of data (e.g., average rainfall in a given city, quarterly earnings for a business). In practice, the personal-ness of personal data proceeds along a spectrum rather than existing as a binary distinction (the city a person lives in vs. one's home street address). Further, within this realm of personal data, sensitivities also differ considerably by the corresponding person whom the data describes. For instance, one individual may consider his or her sexual orientation, religion, or current geo-location to be quite sensitive data, and seek to maintain its privacy, whereas another person may not classify these same types of data as sensitive whatsoever, sharing them freely and publically without hesitation.

This current, somewhat constant production of digital personal data endows a diversity of opportunities and obstacles (SINTEF, 2013). For opportunities, digital personal data has fast become a valuable asset across commerce, government, and personal life.<sup>2</sup> In the commercial sector, activities linked directly or indirectly to the generation, collection, storage, sale, trade, and use of personal data have spurred innovations leading to substantial economic impact and immense wealth. Most notably, Silicon Valley's more recent ventures, less silicon-driven and more data-driven, exemplify this shift towards inventing and extracting new economic value from social information (Rao & Scaruffi, 2013).

From search engines to social media and social networking sites to nearly all activities of commercial media, digital publishing, and content distribution, these domains and many of the corresponding firms that operate in this space rely heavily on the monitoring and subsequent monetization of human behavior. The resulting "consumer data"<sup>3</sup> provide key information goods, commoditized in various forms, that are necessary for many online platforms to function. As a result, personal data fuels much of the commercial web. This is achieved largely through financing from online advertising revenue and often specifically the creation, distribution, and analysis of custom internet audiences available for personalized advertising (Crain, 2013). And these activities show little sign of slowing. In terms of revenue, personal-data driven internet advertising maintains a steady historical climb. This ascent is the joint product of moving ad

---

<sup>2</sup> In addition to being an asset, personal data and specifically its proper storage and protection also exists as a liability for many organizations. This is seen in the costs and risks associated with securing potentially sensitive personal data from unauthorized access. In the US, 2014 was a particularly costly year for data security breaches with notable events for Sony Pictures, JPMorgan Chase, Apple (iCloud), Target, and Home Depot. For further information, see <http://www.networkworld.com/article/2861023/security0/worst-security-breaches-of-the-year-2014-sony-tops-the-list.html>

<sup>3</sup> Whether the individual whose data is in question is referred to as an *individual*, *user*, *person*, or *consumer* is of little consequence for the questions explored in this dissertation. This distinction is more reflective of disciplinary tradition and/or intended audience: marketing and legal scholars tend to invoke "consumer," whereas those in computing, systems design, and information studies are concerned with "users." Communications scholars often simply call human beings "people" with "audiences" perhaps a close second. I use all of these terms interchangeably in this work.

dollars from traditional broadcast and print media to buying ads on website and apps. This is evidenced, for example, in the rather dramatic 17% compound annual growth rate in online advertising revenue seen over the past 10 years, a measure that has persisted despite only 3% growth in U.S. gross domestic product during this same period (PwC US & IAB, 2016).

Despite incredibly low unit costs for any individual variety of personal data, in the aggregate elements such as internet users' keyword searches, page visits, clicks, likes, shares, purchases, and other online behaviors contribute substantial value and support substantial online enterprises.<sup>4</sup> Additionally, and increasingly difficult to decouple from online data, personal data corresponding to "offline" consumer activities augments and adds further value to information originating online, resulting in more detailed consumer profiling. Thus, copious amounts of personal data are now pursued by firms across countless touch points with consumers and users.

At the same time, commercial activities surrounding personal data creation, collection, and monetization have not come without criticism. From regulatory attention (e.g., Federal Trade Commission, 2016) and legislative attention (Interactive Advertising Bureau, 2016) to pushback from civil society and consumer protection groups (e.g. EFF, EPIC, CDT), commercial activities leveraging individuals' social data have come under considerable scrutiny in recent years.<sup>5</sup> For instance, debates over the proper custodianship of digitized information (Vaidhyathan, 2011), power imbalances linked to digital surveillance (Andrejevic, 2007), reductions in personal privacy (Solove, 2008), and related dangers of consumer profiling (Turow, 2011), each of these

---

<sup>4</sup> Values vary by data type, but at the individual unit this type of personal information might cost on the order of pennies or more likely fractions of a penny, as in the case of renting personal data from a third-party data provider to distribute a single online advertising impression. However, the volume of web activity (e.g., page loads) makes these pennies add up amounting to sizable revenue streams. In the aggregate, individual ad impressions sold for next to nothing support technology firms with multi-billion dollar annual revenues (e.g. Google, Facebook, Twitter).

<sup>5</sup> For an example of ongoing efforts on digital privacy, see the advocacy work by the Electronic Frontier Foundation (EFF), available at <http://eff.org/issues/privacy>

have emerged as pressing concerns for the future of personal data aggregation, digital marketing and advertising, and the broader commercial web.

Further, when it comes to figuring out ways to extract economic value from digital personal data, we may only be in the early days of this practice with many areas of commerce thus far only scratching the surface (Greenstein, 2014). Expansion of economic activities that leverage and monetize this class of personal/social information may amplify current tensions associated with consumer monitoring. At the same time, these activities also have potential to spur and sustain new areas of commerce. For instance, just 20 years ago few would have speculated that so many firms with multi-billion dollar valuations could be economically sustained and profitable solely by surveilling, recording, and selling information about what people do while on their computers (e.g., websites visited, search queries conducted, items “liked”). Yet, today, monetizing these digital traces of human behavior and attention now supports some of the world’s largest and most influential global firms. Despite this success of digitization and social data, it remains to be seen how exactly the interests of firms and consumers will come into alignment when it comes to the specific practices through which personal data is monetized.<sup>6</sup> To illustrate how consumer data and other items are transferred between different entities in the current digital environment, Fig. 2.1 contains a simplified representation of key components anchoring these processes.<sup>7</sup>

---

<sup>6</sup> One indicator may be the amount of consumer protection activity that emerges in the years to come advising and/or restricting how firms use information about individuals. In the U.S. this is likely to come from the Federal Trade Commission (FTC), though the White House has also been active in this area in recent years too (e.g., White House, 2012).

<sup>7</sup> Some firms function in multiple roles. For instance, Facebook is both a publisher and operates its own ad exchange. Microsoft is both an advertiser and runs its own an ad exchange, too. Google serves as a publisher, runs an ad exchange, and is also an advertiser itself. Most sizable consumer facing firms would occupy a position at the advertiser node, as these companies tend to engage in some form of advertising online.

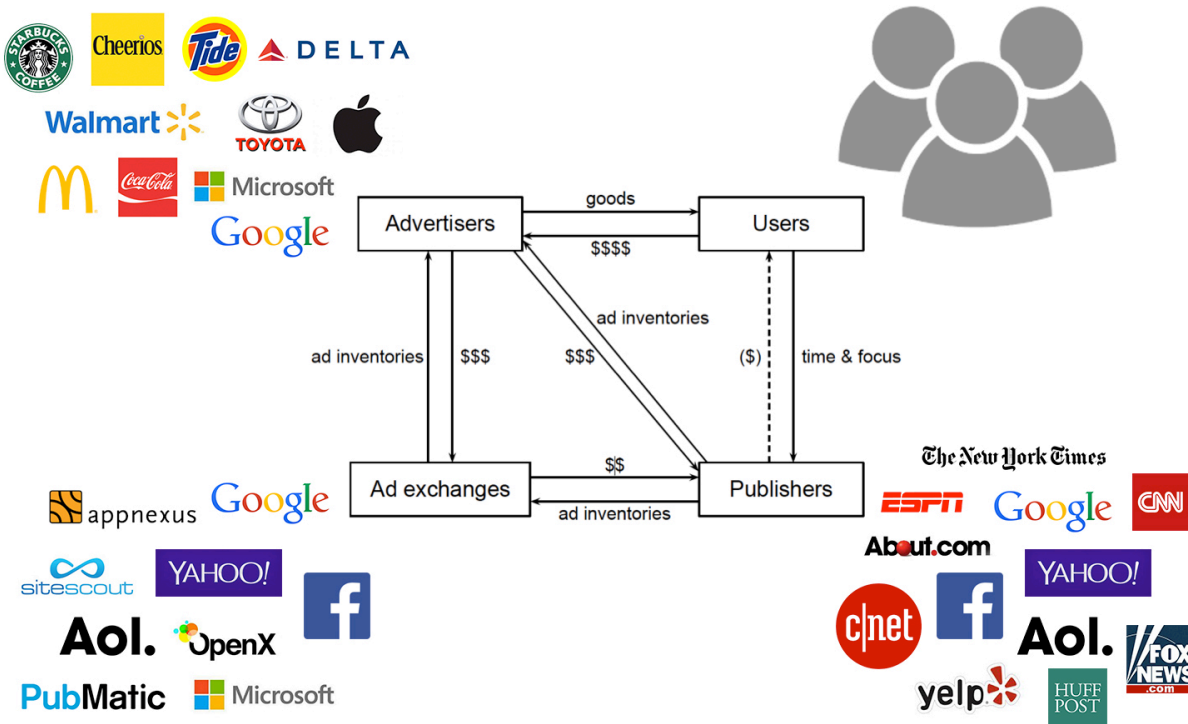


Figure 2.1. Simplified representation showing key components and flows of internet advertising processes (adapted from Yuan et al., 2012).

What is clear, as indicated by adoption and usage rates, is that individuals locate incredible value in the many free-to-use yet commercially-oriented online tools, services, and platforms—platforms that for the most part are financially dependent on the collection and monetization of users’ personal data and corresponding advertising revenues. This includes the most popular social networking and social media sites (e.g. Facebook, Twitter, Instagram, Snapchat, LinkedIn), search engines (e.g. Google, Bing, Yahoo, Baidu), multi-media and content sharing platforms (e.g. YouTube, Flickr, Tumblr), and online journalism outlets (CNN.com, HuffingtonPost.com, TheGuardian.com). Further, opportunities for firms to collect personal data are expanding, evidenced by increases in overall internet and social media use, along with the shift towards cloud-based consumer services (Anderson & Rainie, 2012). Each of these trends in adoption of internet-enabled technologies also indicates expansion of these platforms’



corresponding personal data collection activities; the volume, variety, and velocity at which personal data is generated and collected remains on the rise.

Efficiently monetizing personal data at scale, for instance, by auctioning off sponsored search engine results (simultaneously generating related clickstream and at times purchase data), has supplied the digital marketing goose that layeth golden eggs. For example, consider internet giant Google whose co-founders were credited with inventing a new “magical” source of advertising revenue in creating AdWords (Auletta, 2010). The company provides one of the most prominent examples of innovation in monetizing consumer attention online, the revenue from which has led to numerous additional advances in interactive computing and connectivity for Google along with other firms as well.

In addition to birthing tech giants, monetizing consumer information via the web has also given rise to numerous non-consumer facing firms, companies that specialize in supporting different needs within the much larger digital marketing supply chain. Many of these non-consumer facing firms specializing in consumer data (e.g. Acxiom, Datalogix, Epsilon, Experian) supply various forms of personal information about individuals to marketers, financial institutions, and governments, provided as a paid service. These firms are sometimes unaffectionately referred to as “data brokers.” Positioned between consumer facing web firms (e.g., Google, Twitter, Facebook, LinkedIn) and consumers are now a multitude of non-consumer facing firms most of which are sustained by ancillary revenues from providing marketers with access to personal data which can be used to selectively target advertisements towards consumers and put to additional uses.

For both consumer facing and non-consumer facing firms, efforts to monetize personal information involve observation and recording of behaviors in many forms. This ranges from

tracking both overt communications and activities of web users to more surreptitious forms of personal data collection, from both of which further inferences can be drawn. Messages and activity in private channels (e.g. email, instant messages), on mixed public/private platforms (e.g. posting on social media), and in public view (e.g. commenting on a blog), each of these communicative acts might generate information that feeds into a firm's larger pool of personal data, data to be later used for various audience segmentation and advertising personalization efforts.

For these more overt online actions, tracking these behaviors often occurs from firms collecting user-generated text (e.g., emails, SMS and other text/chat messages, forum comments), which is later sorted and data-mined for insights. Additionally, non-textual expressions contribute a wealth of personal data as well. Mouse clicks, touchscreen taps, upvotes, downvotes, “liking” and “favoriting” content, choosing star ratings for a particular digital (e.g. Netflix film) or physical product (e.g. product on Amazon.com), answers on multiple choice “personality tests”— these social signals along with countless others, at times, are tracked by firms, analyzed, and put to use in various ways, sometimes for audience segmentation leading to personalized advertising . Additionally, for more surreptitious behavioral monitoring, many other innovative forms of online observation are routinely deployed, the data from which often makes its way into user profiles and marketing databases. Examples include recording the “hover time” that a person's mouse cursor actively hovers over an online ad before clicking or not clicking on the ad. Similarly, the seconds or milliseconds one spends paused to look at a unit of advertising or organic content while scrolling through a social media newsfeed are recorded to derive additional insights about the viewer. On their own, metrics like mouse cursor hover time or the amount of time paused on an item while scrolling in a newsfeed are not particularly

valuable, however, when these are compared from one internet user to another (or to the average) these subtle traces can become very revealing and insightful to marketers and others.

For instance, consider the person found to spend above average time “parked” in their mobile newsfeed whenever the content currently in view is sports-related. Most likely, this person is more interested in sports compared to someone who whizzes past content categorized as sports-related, and on and on. While these behaviors may not alert themselves to consumers in the way clicking on an ad might, these subtle signals can be equally powerful indicators of personality or interest for marketers to tap into. When considering how most online actions (and inactions) can now be unobtrusively measured by firm to derive insights about the corresponding internet user, the interactive web (or “web 2.0”) emerges as an extraordinary tool for studying, segmenting, and profiling consumers.

Together, these numerous overt and not-so-overt communicative acts performed with computers, tablets, smartphones, and emerging wearable devices routinely supply a wealth of transactional social data that can be used to support contemporary marketing efforts. The resultant personal data can also be packaged, analyzed, and re-packaged, and later rented and sold as valuable products in their own right. In this way, nearly all behavior conducted in connection with a web browser, mobile app, or increasing number of devices contributing to the so-called “Internet of Things” (IoT) lends itself to automated observation and the creation of personal data and user profiles. For marketers, more recently this has presented the growing challenge of determining exactly which behaviors are worth monitoring and recording and which are not, as the volume of opportunities to aggregate these digital traces now far exceeds the capability and practicality of attempting to record all available signals.

Although the personal data-driven, online marketing system has grown in technical complexity in recent years, many of the information flows remain but with the addition of new components including a range of personal data management and analytics platforms and new ways to leverage consumer data to selectively target advertisements. Nearly a decade ago, Ashworth and Free (2006, p. 111, Fig. 1) demonstrated the exchange model in online marketing that had emerged in response to networked computing along with the relative sophistication with which marketers collect and connect consumer data across many seemingly-disparate digital systems. Since this time, increases in broadband and mobile internet adoption have completely transformed how and when consumer data is collected, thereby altering the frequency as well as intimacy with which marketers and consumers interact in this environment.

Additionally, more recently the online advertising space has also experienced an acceleration in time as well, undergoing a dramatic shift allowing nearly all of these activities to occur on-the-fly, in so-called *real-time*. The result is an arrangement of sophisticated computing systems, databases, networks, and connections that allow live auctions for both internet users' attention and online publishers' inventory—events that take place in the time it takes for a web page or mobile app to load content. These advances permit marketers to leverage audience segmentation data and compete in live auctions for the opportunity to send a single ad impression to a single individual based on known and estimated audience attributes. This represents a monumental shift away from pre-purchasing online ad space at predetermined prices and in bulk quantities. This practice of buying and selling ad space and audience attention live typically falls under the catch all term of “real-time bidding” (RTB). Zhang, Yuan, Wang (2014) depict the components of a common RTB configuration (Fig. 2.2), though arrangements vary. Nonetheless, these technological innovations support more efficient buying and selling of online

advertising space and have expanded the abilities of marketers to selectively present advertisements to specific audiences.

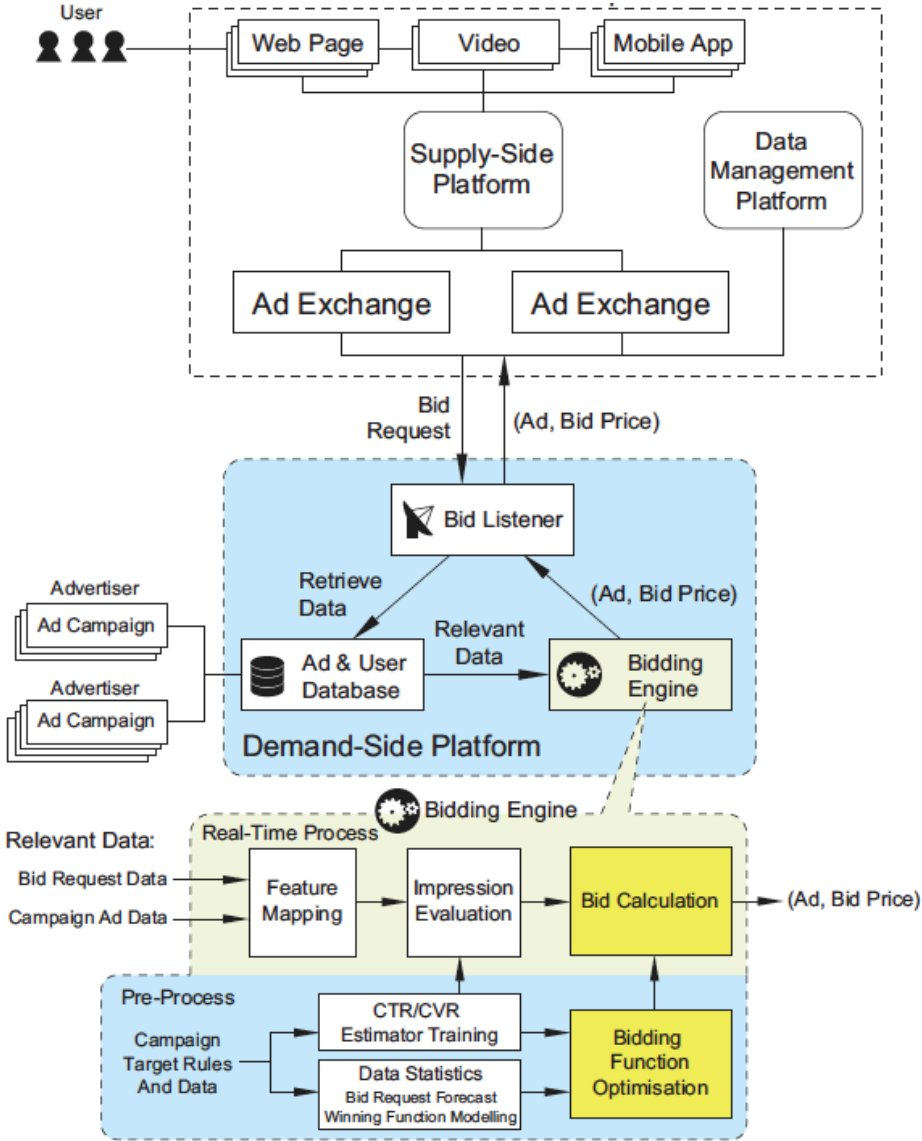


Figure 2.2. Real-time bidding process (from Zhang, Yuan, & Wang, 2014).

Incorporating live auctions at the level of individual audience members has also opened up new ways for marketers to test the effectiveness of varying their ad campaign’s parameters

(e.g., creative content, audiences targeted, web properties used), at a very granular level, and in real-time. This allows, among other things, rapid A/B testing granting immediate feedback into the performance of targeted ad campaigns. Additionally, some RTB ad-buying tools also provide automated or semi-automated selection of campaign parameters based on their performance, meaning marketers can test vast numbers of slightly different ad campaigns before these platforms automatically shift one's budget towards spending more on those campaigns that perform the best and less on poor performers. Overall, the affordances of RTB systems permit marketing managers to make strategic decisions about advertising that were not possible under the previous scheme of buying and selling online ad space at predetermined rates, volumes, and times.

### **State Surveillance vs. Commercial Surveillance**

A challenge when assessing issues linked to advertising personalization is the inseparable relationship between commercial surveillance of individuals for purposes of segmentation and targeting in the marketplace and states monitoring individuals for counter-terrorism efforts. In just the past few years this entanglement has intensified and become more apparent to the public through a number of high-profile revelations concerning state surveillance efforts, especially those leveraging popular web platforms already designed to generate detailed information and record about their users (Schneier, 2015).<sup>8</sup>

---

<sup>8</sup> Numerous disclosures about state surveillance of internet-enabled communications followed the landmark revelations made by Edward Snowden in 2013 concerning the U.S. National Security Agency's (NSA) domestic spying programs. The intertwining commercial and government surveillance was most apparent in revelations detailing the NSA's PRISM program, under which nine major internet companies (Microsoft, Skype, Google, YouTube, Yahoo, Facebook, Apple, AOL, PalTalk) were secretly providing their users' confidential communications data to the NSA.

For instance, in the case of the popular social media company Facebook, the company stores detailed profiles about its users based on activity generated on Facebook and from tracking which other sites users visit while logged in to the platform. It is not difficult to imagine why this type of detailed user data, created and typically reserved for consumer profiling and marketing purposes, would be of great interest to governments too given what it can reveal. Notably, and illustrating a unique dependency of state intelligence on commercial surveillance, most of the state surveillance practices that have come under fire in recent years would not be possible without the detailed monitoring of consumers that technology companies already engage in for purposes of content personalization including targeted online advertising.

Additionally, other state surveillance programs revealed in recent years have relied on even more aggressive efforts by directly and secretly tapping into the digital infrastructure of communications firms at various points, in some cases unbeknownst to these companies themselves. The result is a continual wiretap (or datatap) copying all communications and other personal information traveling through a company's network at a given node.<sup>9</sup>

Both this direct copy-intercept strategy and those arrangements where firms knowingly comply with requests for data have gained extreme negative attention in U.S. and around the world in recent years (Greenwald, 2014). As a result, civil liberties groups to even politicians

---

<sup>9</sup> One of the most damning examples of the U.S. government using this technique, secretly tapping a communications firm's network and copying all user traffic, was the NSA's top secret "Room 641A" located inside one of the world's busiest internet switching hubs in San Francisco. This small room was found to contain secretly-installed specialized internet routing equipment that copied web traffic including various private communications sending them to NSA servers. This secret program was unknown to internet users and apparently to most AT&T employees who worked at this site. Eventually, AT&T technician Mark Klein who worked in the building reportedly grew suspicious of the room which was said to have no doorknob and to be inaccessible even to staff working in the building. Klein eventually discovered and revealed the existence of this secret room to the public. He later served as the subject of a 2006 U.S. federal class action lawsuit filed by the EFF (*Hepting v. AT&T*). The case was eventually dismissed in 2009 when a federal judge cited an incompatibility of the class action suite with the U.S. Foreign Intelligence Surveillance Act (FISA).

have called for new privacy protections from these forms of state surveillance on digital platforms.

### **Marketing's Personal Data Problem**

While collection and use of consumer data has posed challenges for consumers and governments, it has also greatly complicated marketing management activities. For many firms this has meant shifting focus and resources towards tighter integration between IT and customer relationship management (CRM) resources. Elevating these challenges to the c-suite level, the increasingly common position of Chief Data Officer (CDO) signals this shift (Arthur, 2013). Further, within firms this change means greater emphasis on using insights gleaned from large-scale customer datasets and analytics when making strategic business decisions rather than relying on intuition alone.

Certainly, marketing has been one of the largest beneficiaries of the internet's growing ability to generate, store, and transfer various forms of digital personal data. For firms, though personalized advertising is the most prominent way of leveraging user profile data to selectively present content, the affordances of personal data go well beyond advertising, allowing for activities such as differential pricing in e-commerce,<sup>10</sup> data-driven personalized product recommendations, and custom arrangements of organic vs. sponsored content on social media, news sites, and other commercial platforms that support dynamic content.

This move towards automation and data-driven content in the broader media landscape is further evidenced by who or rather what is now responsible for much of the curation (or gatekeeping) on the most popular digital media platforms: algorithms. Models designed to

---

<sup>10</sup> To date, firms using personal data-based price discrimination in e-commerce have employed a mix of first- and third-degree price discrimination. These tactics attempt to leverage a consumer's individual reservation price. For further information see Hannak, et al. (2014), U.S. Council of Economic Advisers (2015), and Stevenson (2015).



predict who will click on a particular ad or read a public relations story are not the work of traditional “ad men” (and ad women) acting on hunches. Rather, these activities are increasingly supported by curation algorithms developed by engineers, data scientists, and statisticians who have become central to the function of online media including digital marketing.

Though the maturation of consumer data collection practices has been a boon to marketers, this shift has not been completely welcomed by consumers. Pushback from consumer includes objections linked to perceived invasions of personal privacy and notions “creepy” targeted advertising practices (Ur et al., 2012). What a marketer may view as simply more accurate ad personalization a consumer may view as getting a little too personal. Further, when customized content breaches expectations for privacy and appropriate use of personal information, this form of *over-personalization* has been known to backfire, producing the exact opposite of the intended marketing effect, causing resistance rather than persuasion (White, Zahay, Thorbjørnsen, & Shavitt, 2008; Malheiros et al., 2012).

Left on its current course, marketing’s personal data problem may be on track to undercut its numerous recent advances in audience segmentation, prediction, and customized advertising. In particular, marketers are often compelled to pursue more and more granular data about specific individuals as not pursuing these readily available data may amount to opportunity costs. At the same time, measures of US public opinion indicate that consumers regularly disapprove of certain types of data collection and how these data are used (Pew, 2014), especially when they think marketers and advertisers are on the other end (GfK, 2014). Thus, there is trust deficiency which, to date, few on the marketer side have taken steps to address. As a result, when it comes to how data is to be collected and used in the commercial setting, the relationship between marketers and consumers has been strained.

This relationship is complicated by the fact that data-driven personalized advertising is the great subsidizer of consumers' access to most digital information and online content delivery. In this arrangement, online tools and services are offered by firms to users free of charge. This includes most search engines, cloud-based email and chat clients, video sharing websites, and nearly all social media platforms. As a result, typically more tangible economic values are absent. Instead, either knowingly or unknowingly, users exchange their personal information, and more often the ability to allow advertisers to mine this data and show them advertisements, for the opportunity to use a tool, platform, or other online service. The resulting exchange often legally rests on the condition of accepting a firm's terms of service (ToS). Invariably, these ToS include stipulations that users grant rather wide permissions to firms to collect and use various information about them. Often this permission includes information gleaned while using the platform itself as well as information obtained from any third-party companies. The result of the "free" model employed by many popular online services is that users are often data-paying customers, with personal data transferred to firms in place of traditional monetary exchange.<sup>11</sup> In light of this, some have suggested users be financially compensated by firms that rely on collecting and using their personal information (e.g., Wu, 2015), though this idea has failed to gain much traction.

### **Public Opinion on Personal Data & Trust**

In recent years in the U.S., there have been abundant efforts to gauge public opinion on digital privacy and personal data use. While this work is usually not capable of providing explanations as to why people are concerned about this topic, results from probing the public

---

<sup>11</sup> Free as in no financial cost to the recipient, not in the unbounded sense of this term (e.g., *gratis* vs. *libre*). See: [https://en.wikipedia.org/wiki/Gratis\\_versus\\_libre](https://en.wikipedia.org/wiki/Gratis_versus_libre)

suggest a degree in misalignment between the efforts of marketers and some preferences of consumers when it comes to how personal data is collected and used.

Among these efforts, one of the most prolific contributions comes from a decade of related public opinion studies on this topic led by Joseph Turow at the University of Pennsylvania's Annenberg School for Communication (see Turow et al., *Americans, Marketers, and the Internet: 1999-2012*). Consistently, this work has reported that Americans say they are concerned about their privacy in the digital marketing context, while at the same time also revealing consumers are relatively uninformed and misinformed about the practices, technologies, and legal protections that accompany marketing-related personal data activities. These attempts to gauge consumers' opinions and knowledge in this area have been levied in a recurring indictment of the marketing industry over its personal data-driven collection practices and, at times, seeming disregard for what consumers at least say they prefer when it comes to data collection.

Other surveys report similar findings regarding consumers' somewhat low approval for the marketing and advertising industry when it comes to how its members collect and use personal data. When asked, a consistent majority of American adults voice concerns about these practices. For instance, among a representative sample of Americans adults, a 2014 survey conducted by the Pew Research Center reported a vast proportion of Americans (91%) say they feel they have lost control over how personal information is collected and used by companies, with most (64%) saying they believe the government should do more to regulate how advertisers collect and use their personal data (Madden, 2014). At the same time, and further illustrating the growing predicament for marketers and others who rely on user data for their business practices, this same survey found only a small proportion of people (36%) reported "appreciating" the way

web services can operate more efficiently by accessing individuals' personal data and only about half (55%) said they are willing to share personal information with companies in exchange for free online products and services. This study also reported that individuals find some types of personal data that internet companies collect more sensitive than others, lending additional support to notions of context-based information expectations and norms, or contextual integrity (Nissenbaum, 2010).

Similarly, in 2016 the annual TRUSTe U.S. Consumer Privacy Index found nearly all (92%) U.S. Internet users reported being worried about their online privacy, few (31%) saying they understand how firms share their personal information with other organizations, the majority (68%) expressing concern over not knowing how firms are using their personal information, and only around half of people (56%) saying they trust companies to do these things appropriately (TRUSTe, 2016). Numerous opinion surveys on this topic are published all the time with most reporting similar trends, revealing members of the public at least claim to be concerned about which data is collected, how, and by whom.

Viewed together, these measures of public attitude signal a degree of unease on the part of consumers regarding various forms of personal data activities by marketers. It appears Americans are not pleased with the current state of social data collection that occurs online, which many feel occurs beyond both their full knowledge or their control. While the nuances of exactly how concerned individuals are is debatable, most people report being concerned about data privacy, in some cases objecting to common collection practices by marketers and third-parties. And this trend is commonly known by marketers and policymakers alike. Though the implications of these public concerns have yet to be determined.

One dynamic that complicates attempts to gauge public opinion in this area is the difficulty in decoupling *consumer use* of a platform or tool with *informed consent* to the way it functions behind the scenes. That is, the act of using a particular online product or service is different from knowing about and approving of how it works technically or conceptually. The overarching narrative emerging in recent years from these studies of public opinions about digital privacy is that individuals are displeased with how their personal information is collected and used by marketers and other organizations including the government. At the same time, a parallel explanation and rebuttal to these concerns is that despite saying they are concerned the majority of people do very little as far as maintaining privacy best practices or other strategies to reduce personal data collection. This contradiction is summed in the title of *The New York Times* story reporting results from the 2014 Pew Research Center’s survey on digital privacy, “Americans Say They Want Privacy, but Act as if They Don’t” (Nov. 12). This argument relies on a “revealed preference” logic, claiming individuals’ decisions to use online services that rely heavily on personal data collection indicate peoples’ *true* preferences, regardless of what they say they prefer.<sup>12</sup> Applied to problems associated with consumer data collection and advertising personalization, this revealed preference counter-argument asserts that marketplace behavior, in this case electing to use certain web services and agreeing to their terms of service, is a better indicator of actual preferences than what people report in opinion surveys.

While there may be some truth to this revealed preference explanation, at the same time, the absence of viable alternatives also negates the traditional logic of revealed preference, as abstaining from those online tools and platforms that collect personal data is either impossible or

---

<sup>12</sup> Introduced by economist and Nobel Laureate Paul Samuelson (1938), under *revealed preference* to observe action is to understand preference, as opposed to say asking people what they want. For a criticism of revealed preference theory, see Wong (1978).

impractical in most cases.<sup>13</sup> Effectively, this abstention would amount to not using the internet. Further, abstaining from actively using web services also does not prevent one from having their personal information distributed online in other ways. Consider, for example, being “tagged” by name and at a specific location in a photograph uploaded by someone else to a social media platform. Further, abstaining from online activity does nothing to limit personal data collection at the many offline points of entry that also feed into the marketing data ecosystem (e.g., credit card transactions). Simply not using the internet, directly or indirectly, is not an option for most individuals.

Additionally, other studies have shown consumers do not hold marketers and advertisers in high regard when it comes to issues of trust, especially whether or not to trust firms to use and protect personal data in ways consumers deem appropriate. For example, among a representative sample of Americans, a 2014 survey by GfK reported substantial concerns by consumers over how personal data is used online, with respondents especially skeptical of how marketers and advertisers use their personal data (GfK, 2014). Respondents were asked: *How much do you trust marketers and advertisers with regard to how your personal data is handled?* 64% of Americans said they either *don't really trust them* or *don't trust them at all*, with just 25% saying they *trust them mostly* or *trust them completely*.<sup>14</sup> Moreover, the GfK survey also asked respondents to rate different kinds of organizations for how much people trust them to use personal data in acceptable ways. Of over 20 types of organizations and industries rated, the marketing and advertising industry was the *least trusted* (see survey results in Fig. 2.3).

---

<sup>13</sup> To date, alternatives to (free) online platforms that collect and monetize personal data are very limited. Often they cost users additional time and/or money while usually failing to offer the same level of benefits compared to more mainstream platforms that do collect personal data about users. For instance, consider the differences in affordances and functionality between social networking sites Facebook (<http://facebook.com>) vs. Ello (<https://ello.co>) or between search engine Google (<http://google.com>) vs. Duck Duck Go (<https://duckduckgo.com/>).

<sup>14</sup> The remaining 11% responded either “don't know” or failed to respond to this question.



## Consumers' trust is not shared equally across all organizations

Marketers and advertisers, international businesses, and social networks are the least trusted with personal data.

Hospitals and healthcare providers, online payment systems, and retailers are the most trusted:

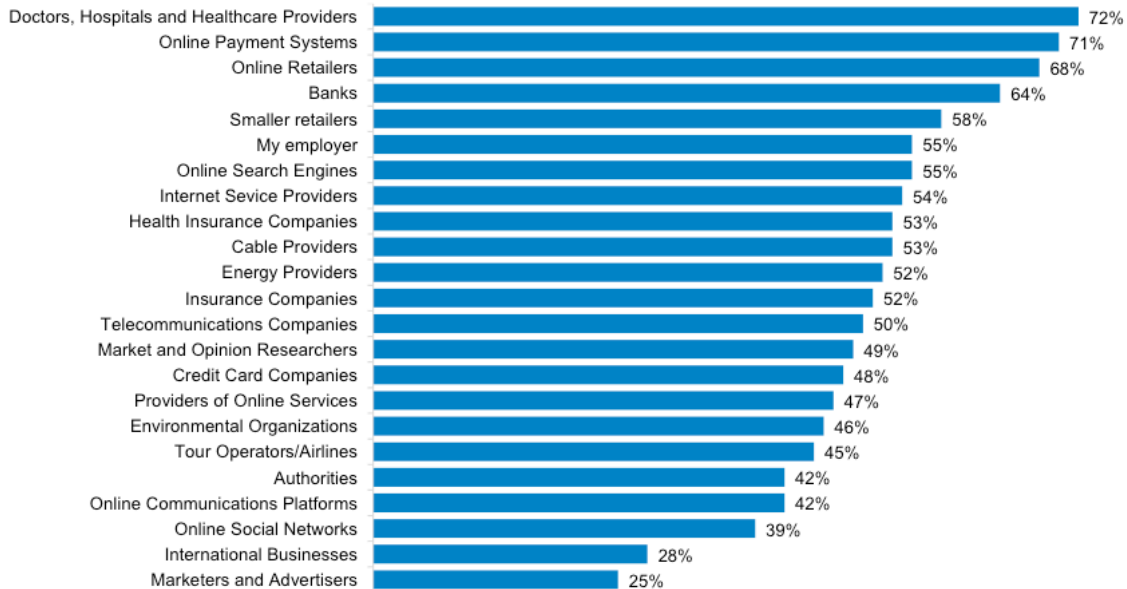


Figure 2.3. The 2014 Survey on Data Privacy and Trust (from GfK).

This relatively low trust in marketers and advertisers extends beyond the U.S. as well. In a similar global survey conducted by the World Economic Forum (2012), marketers and advertisers also ranked dead last in a rating of organizations trusted to protect personal data. Researchers asked respondents, *To what extent do you trust the following institutions to protect your personal data...* The proportion of respondents who said they trusted the institution in question was as follows: banks and financial institutions (61%), those providing health and medical services (55%), government authorities (53%), internet service providers (45%), telephone companies (44%), mobile phone operators (44%), search engine companies (40%),

shops and department stores (39%), companies that provide social networking services (37%), online marketers and advertisers (28%).<sup>15</sup> In the U.S. and around the world, people appear skeptical that marketers and advertisers will use personal information in ways consumers find acceptable.

Many of today's internet technology firms are also now indebted to a revenue model that hinges on these firms' abilities to ensure users that whatever personal data is collected will be used appropriately. Highlighting the crux of this arrangement, in Google's 2012 annual letter to its shareholders Google co-founder and current CEO of parent company Alphabet Larry Page stressed that central to this business model shared by Google and many other internet companies was maintaining two key ingredients among the firm's user base: love and trust (Page, 2012). For the latter, firms relying on a personal data exchange business model have grown in number and popularity, though from opinion surveys it appears levels of trust in how firms use personal data have not kept in step. Additionally, and as noted previously, this friction is compounded by the many revelations of wide-reaching government surveillance programs, such as those operating in the U.S. and U.K., which have relied extensively on leveraging personal data extracted from activities on popular web platforms. This has likely furthered strained levels of trust consumers feel towards internet companies and marketers operating in this heavily surveilled environment.

### **Resistance to Personal Data Collection**

Another way to consider user preferences related to consumer data collection and advertising is to simply look at individuals' efforts to resist these practices while still using

---

<sup>15</sup> Refers to percent that answered 5, 6, or 7 on 7-point Likert scale measuring trust in organizations ranging from *Distrust them a great deal* (1) to *Trust them a great deal* (7).



internet services. General strategies for individual resistance include, for instance, changing one's privacy settings on a platform to be more strict with what is shared and with whom, as the defaults tend to be more aggressive about data collection and sharing rather than less.

Additionally, in recent years a number of tools and services intended to help consumers avoid online observation and tracking have emerged. Examples include *DeleteMe* (justdelete.me), which offers web directory of around 350 personal data collectors with corresponding URLs for where to attempt to remove one's user profile and/or opt out of data collection from each corresponding firm.<sup>16</sup> Similarly, *Safe Shepherd* (safeshepherd.com) offers a paid service that uses a combination of human intervention and automated monitoring to remove individuals' personal data from websites and marketing databases. The company claims, "Safe Shepherd constantly scans the internet and private databases, looking for your personal information. When we find a company publicizing or selling your personal information, we submit an opt-out request on your behalf, which deletes your record." (SafeShepherd, 2016)

To date, one of the most centralized efforts for avoiding consumer data collection online is the proposed HTTP standard known as *Do Not Track* (DNT). This technical standard allows one's web browser to send a signal to a website whenever requesting content, requesting that the visitor not have the upcoming web session and corresponding personal data tracked. To date, the make or break weakness of the DNT standard is that compliance by websites and additional data aggregators operating in the background on websites is completely voluntary. Thus, the decision of each individual company and the decisions of additional services operating behind the scenes on each company's website, dictate whether DNT requests are honored or not.<sup>17</sup> So far the

---

<sup>16</sup> The service *DeleteMe* also ranks each firm on the technical difficulty users should expect to encounter when attempting to remove themselves from that company's records: *Easy*, *Medium*, *Hard*, and *Impossible*.

<sup>17</sup> As this is an evolving standard, its functionality is likely to change. For a more thorough explanation of Do Not Track see: <http://donottrack.us/>

incentives for compliance have been unclear and many popular web platforms (e.g., Facebook, Yahoo!, Google) simply ignore DNT requests when loading content and identifying visitors. A further blow to DNT followed a 2015 petition by a consumer watchdog group to the FCC, in which the Commission issued a response ruling offering that websites were permitted to ignore DNT requests (FCC, 2015).

Other approaches to resist consumer data collection and/or personalized online ads include various web browser extensions that proactively block ads and/or attempts by websites to utilize cookies, tags, tracking pixels, beacons, and other technologies intended to record internet users' behaviors. Examples of these extensions include Ghostery, Privacy Badger, DoNotTrackMe, DisconnectMe, NotScripts, and Track Me Not.

Some of these services have become quite popular. For instance, in 2014 the Ghostery (ghostery.com) browser extension boasted having some 40 million users (Ghostery, Press Release, 2014). Once installed, Ghostery not only blocks many of the most common tracking objects on the web but also provides the option to show users which sites are attempting to track them. Though, and perhaps ironically, the company plays a dual role in this environment by also selling data collected about its users to firms and organizations seeking to improve their website performance, a practice which some have argued represents a conflict of interest (Simonite, 2013).

Beyond these basic tools are more robust methods that can prohibit or limit the way data about web users are collected. One prominent example is the Tor web browser (torproject.org). Rather than requesting web content directly like most browsers, this privacy-enhancing technology does so by bouncing web requests through a distributed network of internet relays in effort to avoid various user tracking mechanisms employed by firms and governments.

The previous examples of efforts to evade data collection and targeted online ads represent only a portion of these techniques; new tools emerge in this space all the time.

Similar to the results from public opinion surveys on this topic, the many efforts and tools to counter online observation by marketers, data aggregators, and others signal a degree of consumer awareness and related concern over the data collection practices of firms. Several of these tools for evading data collection now report millions of regular users, pointing to their transition beyond niche use (i.e., by a small group of paranoid, privacy conscious consumers) and towards more mainstream adoption. Overall, each of these tools of resistance pose problems to the otherwise free flow of personal data on the web, though their broader impact on marketing activities remains to be seen.

### **Advertising Personalization**

Finally, to understand why marketers and others go to what might appear to be extreme lengths in efforts to observe, record, and analyze consumers, online and offline, it can be helpful to look at the effects of personalization in advertising.<sup>18</sup> Fundamentally, ads directed at audiences thought to possess certain attributes are generally more efficient in that they have a higher likelihood of influencing recipients than ads delivered en masse. This illustrates the key marketing tenant of segmentation.

Additionally, the relative ease of measuring digital advertising response rates now supports a growing body of research into the persuasiveness of personalized messaging

---

<sup>18</sup> Examining the effects of personalized advertising supports Neuman's (2010) prescription for those invested in the study of communication to, "focus on the conditions under which persuasive methods persuade and conditions under which they do not." Questions regarding the relative effectiveness of personalized messaging to persuade consumers are not addressed in this dissertation. However, the notion that personalized advertising is known to be generally more effective (more persuasive) than mass advertising is central to the studies presented in this work.

techniques (e.g., Häubl & Murray, 2003; Graepel, Candela, Borchert, & Herbrich, 2010; Tucker, 2013). Research in this area brings together marketing “science” and social psychology to locate more effective advertising strategies as they relate to using media to exert measurable influence on consumers. Research questions in this area consider, for instance, whether a message is more persuasive when appealing to a viewer based on different types of known or estimated information about that individual. And it is logical that message personalization efforts would tend to be more persuasive compared to more generalized messaging (e.g., Manchanda, Dubé, Goh, & Chintagunta, 2006; Shatnawi, & Mohamed, 2012; Farahat & Bailey, 2012; Lambrecht & Tucker, 2013).

In cases of overt forms of message personalization researchers have demonstrated, for instance, that injecting a viewer’s name into a visual advertisement simply makes it more effective (e.g. *Hey Barbara! Check out these shoes!*). This kind of personalization is shown to result in higher attention and message recall compared to non-personalized messages (Howard & Kerin, 2004; Kampe, Frith, & Frith, 2003), offering marketers plenty of reasons to pursue personalization strategies. Similarly, digitally blending an image of a viewer’s face with the face of a another person, and then using the final blended face on a model in an advertisement, has also been shown to increase favorable ad response in the experimental setting (Samat et al., 2013). This area of experimental research, referred to as “facial similarity manipulation” (DeBruine, 2002, 2004, 2005) has also shown to be effective at influencing voter behavior when merging an image of an individual’s face with that of previously unknown political candidates (Bailenson, Iyengar, Yee, & Collins, 2009). Given advances in image processing, computing, and real-time analytics, these far out examples from the experimental setting may signal future directions for personalized advertising.

Much of the work on the persuasiveness of personalized communications sheds light on the power of customizing messages towards groups of audiences and even towards specific people. These findings also call attention to some of the ethical problems faced by marketers.<sup>19</sup> These include personalized advertising techniques some consumers may find unacceptable when using personal data that that feels invasive to tap into known consumer vulnerabilities. Assessing this emerging ethical dilemma for marketers, legal scholar Calo (2014) carries advertising personalization techniques to their logical extreme. He illustrates how emerging personalization techniques in marketing may result in new forms of consumer manipulation, where firms could use real-time personal data to exploit the cognitive limitations of individuals. For instance, Calo argues marketers will have to decide whether they will exploit decision fatigue in consumers given new ways of measuring an individual's state of mind based on their recent activity. An example might be showing a certain advertisement to someone only upon gleaning that person has just made a series of complicated or emotional decisions and is less likely to resist something. Calo concludes that we may be approaching a dangerous tipping point in behavioral advertising where the ability to derive rather sophisticated insights from consumers' personal data, in real-time, may create marketing situations where consumers are manipulated rather than innocently persuaded. In the U.S., individuals have legal protections from certain forms of undue influence, where one party exerts an unfair level of influence over another party, though it remains unclear when these protections would apply in the case of highly personalized advertising techniques seeking to use various personal data to exploit consumer vulnerabilities.<sup>20</sup>

---

<sup>19</sup> Ethical considerations are of course not new to marketing. For instance, amplified by Vance Packard's (1957) book *The Hidden Persuaders*, concerns over consumer manipulation via "subliminal messaging" in advertising emerged in the US post-World War II. For a review of these concerns and impacts on advertising regulation, see Nelson (2008).

<sup>20</sup> Sometimes referred to as *undue persuasion*, see <http://legal-dictionary.thefreedictionary.com/undue+influence>

The previous examples offer just a glimpse into the type of work on advertising personalization currently under development and designed to improve message reception through leveraging in some cases detailed information about consumers. Given the direct applications of this research in the marketplace, incentives for conducting this type of applied marketing science are high. Additionally, while findings in this area sometimes become public via published research papers and reports, for private firms conducting this type of work often results go unpublished for reasons of competitive advantage.<sup>21</sup> Still, the published work in this area points toward the same overarching result: increasing message personalization, through a great number of techniques, tends to boost the effectiveness of the resulting advertising appeal.<sup>22</sup> In short, personalized communications are more effective.

---

<sup>21</sup> This is not a criticism of unpublished industry research findings but simply an effort to highlight what we do not know (that there are unknown knowns). As with research in any sector, not just in marketing, findings from studies financed by private industry may need to stay private to maintain trade secrets for competitive advantage.

<sup>22</sup> For explicit personalization, prior work indicates that when a viewer is aware of the amount of information used to tailor a message and finds this to be too much or inappropriate, this can backfire resulting in an over-personalization effect and corresponding negative response (e.g., White, Zahay, Thorbjørnsen, & Shavitt, 2008).

## Chapter 3

### Third-Party Personal Data and Personalized Advertising Up Close and Personal

This chapter presents Study 1, the first of three studies completed for the dissertation. For Study 1, I took on the role of a marketer, drawing from participant observation to examine first-hand the nature and dynamics of third-party personal data currently available to marketers on digital ad-buying platforms. This methodology was selected to provide a unique view into to the consumer data used to target ads on websites and apps towards people possessing specific attributes. Two main contributions are offered. First, based on my observations of a large number of third-party data available to marketers for ad personalization on several platforms, I synthesize some of the characteristics of audience targeting data currently in use. This serves both to illustrate and complicate the notion of *digital enclosure* advanced by Andrejevic (2007). Upon assessing the form of these data, I then assess their function. To do so, I executed a series of online advertising campaigns myself to measure the popularity of specific audience attributes from the marketer's point of view as determined by the results from these real-time ad auctions. These advertising tests suggest that some types of personal data are more popular for ad personalization models, and therefore more expensive as well. By taking on the perspective of the marketer, I assess advertising personalization up close. While the three studies in the dissertation are self-contained, my experiences conducting Study 1 informed many of questions and design decisions in Studies 2 and 3.

## INTRODUCTION

Recently, a variety of new challenges surrounding digital personal data in the new media environment have emerged. This applies to both the collectors of digital personal data, typically firms, governments, and increasingly individuals themselves (e.g., quantified self), as well as the individual consumers who supply these data in their various forms. One promising site for exploring challenges associated with personal data is today's digital marketing environment. In light of how personal data has been commoditized and perhaps fetishized by marketers while simultaneously, at times, challenged and resisted by regulators and consumer alike, digital marketing supplies a fertile yet fraught site of inquiry. Further, given the prominence of personalized advertising today on most internet-enabled platforms, from social media platforms to search engines to ecommerce portals, the link between consumer data and digital marketing has become quite central to a range of business operations supporting the commercial web.

This study brings together related ideas from communications, media studies, and marketing to explore the following questions: What is the nature of the personal data made available by third-party data providers to marketers for personalizing online advertisements? How do the dynamics of real-time bidding impact the ability of marketers to personalize advertising for certain audiences over others? Given the nature of these personal data and how they function, what are the broader implications for marketers and internet users interacting in this new *digital enclosure*?

To address these questions I temporarily assumed the position of the marketer. This participant observation method of inquiry, directly taking up the activity under investigation, has proven effective in prior work examining a range of sociocultural phenomena, such as labor (Ehrenreich 2001), immigration (McDermott, 2006), and policing (Bright, 2015), to name a few.



Similarly, this study leverages an embedded researcher approach to explore the social phenomenon of personalized online advertising in the emerging digital marketing ecosystem, specifically examining the presence and dynamics of third-party personal data made available to marketers. Taking on the perspective of the marketer was necessary due to the data in question, which exists beyond public view, yet is more or less in plain sight for marketers when using today's popular digital ad-buying platforms. Exploring this phenomenon from the perspective of the marketer provided necessary access into this world, a unique opportunity to peek across an information asymmetry.

I delivered pro bono marketing management services to a non-profit client who was seeking to drive traffic to their organization's website. This client was aware of my research interests and the primary purpose of the study, agreeing to the unique research/services arrangement due to it being mutually beneficial. I draw on two years of experience acting in this researcher/service provider role. During this time, I operated a series of online ad campaigns in which I interacted with consumer databases, interfaces, and ad-buying platforms directly to take an up-close look at digital marketing's relationship to personal data. I observed and used multiple online ad exchanges, networked computing systems that allow for the buying and selling of online ad space and internet users' attention through the practice known as *real-time bidding*. To buy and place ads in front of internet users, I used multiple industry-leading demand-side platforms—web interfaces that allow marketers to bid for and buy online ad space in live auctions, one impression at a time, targeting individuals by various audience attributes. Part one of this study assesses the nature and variety of personal data currently made available to marketers by third-party data providers across a variety of real-time ad-buying platforms. This analysis presents and discusses the character of personal data, its various origins, and ways to

leverage these data to limit or exclude different audiences when running online ad campaigns. Then, to illustrate a sample of these observed personal data in action, part two of this study reports comparative results from a longitudinal series of targeted ad-buying tests on real-time bidding ad-buying platforms. This latter portion examined the relative demand for and desirability of different audiences, as identified by personal data made available to marketers, based what marketers bid and pay to reach different groups of people online.

This work builds upon three related concepts used to explain power dynamics in today's increasingly expansive new media environment. The first is *information asymmetry*, an imbalance regarding what is visible to different parties in a system; in this study when individual internet users cannot see precisely which personal data organizations collect (Schwartz, 2004; Stefanone et al., 2015) nor how it is used (Pasquale, 2015). The second is the notion of a *digital enclosure*, an interactive realm wherein each action and transaction generates information about itself and its corresponding communicator (Andrejevic, 2007). Finally, I draw on the concept of *contextual integrity*, a privacy heuristic popularized by Nissenbaum (2010), which claims the existence and prevalence of context-relative information norms. Together, these related ideas provide the conceptual basis from which this work proceeds. Based on first-hand observations along with results from a series of ad-buying tests, I conclude that real-time bidding advertising platforms are the logical extension of the emerging digital enclosure and that these sociotechnical configurations encourage and deepen information asymmetry between marketers and internet users while upending notions of contextual integrity.

## RELATED LITERATURE

### Information Asymmetry

One product of the digital enclosure is persistent expansion of the information asymmetry between marketers and consumers. In the most basic sense, information asymmetry occurs when one party does not know what another party knows; an imbalance of who knows what in a multiparty situation. In the case of personal data asymmetry, this imbalance typically corresponds to which data one party collects and/or stores describing the attributes of another party who is unable to see, know, alter, correct, or otherwise exert control over these data. In this case the disadvantaged party is an individual unable to fully know the extent to which data describing them is held by an organization, who in this case is the advantaged party. This organization might be a marketing firm, an employer, or a government agency, for instance.

In marketing, industry practices have long been characterized by information asymmetry. For example, enabled by early computer databases containing information about individuals and households, the advent of direct mail marketing in the late 1980s and early 1990s normalized a system where attributes of consumers were surreptitiously collected and stored out of sight of the consumer, then later used to determine who received or did not receive various promotional materials in the postal mail. This practice continues today and the personal attributes used to target direct mail range from mundane (e.g., owners of SUVs) to off-putting (i.e., a promotion inadvertently addressed to “Daughter Killed in Car Crash”) (Merrick, 2014).

During the late 1990s and throughout the 2000s, the combination of large scale database systems, plummeting costs of computer storage, and widespread adoption of the Internet by U.S. households pushed the amount of consumer data collected to new levels. The technical and economic groundwork for what would become asymmetry at an entirely new scale was laid in

the early days of the web. As Crain (2014) points out, the dotcom internet bubble not only contributed to the development of a commercial web but also brought with it significant investment in new media firms. These companies were then able to parlay this momentary power to solidify longer term, wide-spread demand among marketers for purchasing online advertising.

As the web matured, so too did the advertising technology underpinning it. Much of this development helped marketers to better target online ads towards specific groups of people (e.g., sports enthusiasts, coupon clippers) and unique individuals (e.g., visitors to a specific product's URL). Pointing to a longer history in the persistent pursuit of personal data by commercial firms, Stole (2014) notes that during this period companies such as Yahoo!, Google, and Facebook discovered how personal data held significant commercial value for advertisers, prompting these and other Internet technology companies to design new protocols and systems to surreptitiously track Internet users (p. 130).

These changes, and specifically the new personal data collection infrastructure they established, allowed marketers to tailor ads towards individuals in new ways. However, individuals had no effective way to see these data, how or when they were collected or used, or know much else about the process as it unfolded. The result was the formation of information asymmetry on an unprecedented scale with marketing and technology companies on one side amassing personal data and internet users on the other side.

A disparity has emerged between what personal data is stored on the servers of various marketing and third-party data firms and what individuals are able to see about themselves along with how, when, and where these data are utilized (Crain, 2016). For instance, consider a targeted online ad shown in a web browser or mobile application. Oftentimes, this ad was delivered to a person based on some combination of verified and/or estimated personal data

originating from any number of individual transactions or behaviors. However, given the complexity of data sharing agreements by marketing firms and third-party data providers, typically it is impossible to conclusively determine the origin of the data used to personalize this ad. This unevenness points to what Dixon (2010) conceptualizes as today's "one-way mirror society," a world where individuals do not fully know what personal data about them exists, where, or how it is used. Central to this perceived information asymmetry is that the extent of the asymmetry cannot be known, as even knowing the extent of the asymmetry would require the ability to aggregate and see one's consumer data across an unknown number of organizations.

Noting the associated power relationship this imbalance erects, Andrejevic (2007) questions this personal data asymmetry, asking, "At what point does the amount of information available to advertisers constitute a form of power over consumers, especially in a context wherein consumers have very little knowledge about what information marketers have and how they are using it?" (p. 131). The benefit of looking to asymmetry in the case of personal data driven marketing is not simply to locate imbalance, as asymmetry on its own is not necessarily consequential, but rather to gauge whether this particular asymmetry results in a meaningful imbalance of power.

### **The Digital Enclosure**

Enclosure provides a way to conceptualize the current state of personal data collection now routine when using digital products and services. Andrejevic (2007) offers the notion of a *digital enclosure* as "an interactive realm wherein every transaction and interaction generates a information about itself" (p. 2). Importantly, in its use for personalized advertising, typically information describing the corresponding person who initiated the transaction is also present as

opposed to anonymous behavioral data. Andrejevic describes this enclosure as both a process as well as a physical, though largely invisible, system. It is the act of the sum of technologies monitoring and recording data describing humans and their behavior, resulting in an always-changing archive of consumer data. The digital enclosure is also the product of combining the longtime surveillance interests of firms and governments with the relatively recent abilities of enterprise computing systems, databases, and personal computing devices to generate, capture, and preserve signals linked to individual human behavior.

Notably, the digital enclosure is also flexible. Rather than beholden to any technical format, type of data (e.g., prior purchases, geolocation, online browsing history), or set of systems contributing data to the enclosure (e.g., point-of-sale, GPS, internet cookies), it is largely agnostic to the type of behavior or data it captures. This flexibility opens the door to many new additions of data types and sources all the time, with ability to capture being the lone prerequisite.

### **Contextual Integrity**

Though the affordances of the digital enclosure are indeed helpful for marketers seeking to selectively present advertisements, consumers are not always pleased with data aggregation. One explanation describing why individuals at times reject how their personal data is collected and/or put to use is the privacy heuristic known as *contextual integrity*. Popularized by Nissenbaum (2010), this somewhat common-sense theory claims individuals hold consistent expectations, linked to the social context in which the data is collected and/or used, regarding what constitutes appropriate or inappropriate collection and use of personal data.

Nissenbaum offers the framework of contextual integrity as an analytic lens describing appropriate privacy boundaries for data use. What exactly constitutes *appropriate* under this concept depends wholly on context-relative expectations, or simply norms. In this way, varying contexts or social situations (e.g., the privacy of one's home, during an interaction with a health services provider, in a retail transaction with a company), these different contexts are thought to systematically and consistently conjure different expectations from individuals for how personal data ought to be collected or not collected, used or not used. At its core, contextual integrity maintains that each context and sub-context are accompanied by predictable end-user expectations. In this way, contextual integrity is primarily as a heuristic, as Nissenbaum claims, "for determining, detecting, or recognizing when a violation [of informational norms] has occurred" (2010, p. 148).

### **Marketing & Advertising Efforts**

The marketing industry is a key stakeholder in determining how digital personal data are used and regulated. One can better understand this relationship as it relates to personal data governance by looking to the fundamental interests and functions of marketing. Among the operations of a firm, marketing is fairly porous blending into and through many other business functions (Webster, 1992). Accordingly, the term can be somewhat challenging to bound or define. In business, marketing efforts aim to encourage sales in a product or service, whether successful or unsuccessful. The American Marketing Association defines marketing as "the activity, set of institutions, and processes for creating, communicating, delivering, and exchanging offerings that have value for customers, clients, partners, and society at large" (American Marketing Association, 2013).

Similarly, Silicon Valley marketing pioneer Regis McKenna offered the following definition, “marketing is everything and everything is marketing” (McKenna, 1991). This all-encompassing picture of marketing, as not only part but the most important part of a business’s operation, illustrates an enduring concept and topic of interest stating that it is not so much what someone is selling but rather how it is sold in the hearts and minds of audiences (e.g., Packard, 1957; Williams, 1980; Cohen, 2003). Further, this distance between inherent value and perceived value, and later subsequent consumer demand, demonstrates the value added marketing through branding and persuasive advertising.

### *Models & Mixes*

To better understand its functions, one way to separate the various activities that fall under marketing’s umbrella is through various marketing models. These models, or frameworks, delineate the practice into its component parts. For instance, McCarthy’s (1964) classic “Four Ps of Marketing,” while simplistic, categorizes nearly all activities associated with an organization’s total marketing effort, including: Product, Price, Place, and Promotion. In this case, Product refers to the tangible good or intangible service being offered. Price is the cost charged to those purchasing the product. Place, sometimes also referred to as distribution, is where buyers look for and ultimately obtain the product. Finally, Promotion encompasses the many ways product information is communicated to potential customers, especially the use of advertising (Anderson & Taylor, 1995). The Four P’s remain one of the most popular ways to illustrate what marketing is and is not concerned with within a firm’s total operations.

Another common tool for understanding marketing’s pursuits is through sequential process models. One of the most popular is the three-stage Segmentation, Targeting, and



Positioning (STP) model (Dibb & Simkin, 1991). In this case, segmentation refers to the act of dividing up potential customers into specific groups, such as those with similar demographic, psychographic, or behavioral attributes. Next, once potential customer groups have been identified, targeting comprises researching and selecting which group(s) are the most attractive to market the product towards. Typically, though not always, determining the most attractive segments is based on which groups would be most profitable as potential customers given estimated demand. Finally, positioning encompasses all attempts at influencing how selected target customers perceive the product, especially efforts towards branding and encouraging purchases including advertising. From the Four Ps to STP to countless other strategic models and mixes, these tools illustrate how marketing interests and activities can often be separated, always with the hope of bringing them back together more effectively and efficiently to better influence brand perception and purchasing behavior. Indeed, van Waterschoot and Thomas Foscht (2010) argue that the concept of a marketing mix, a panoply of different activities working towards a common sales goal, is inherent to any marketing situation. Throughout its history, the state of strategic marketing has developed alongside successive schools of thought (Shaw & Jones, 2005). These ways of thinking include specification for how to parse the total marketing effort into discrete interests and how to locate the proper combination of strategic ingredients (Borden, 1964).

#### *Advertising Return on Investment*

Of all marketing activities, the fourth P, Promotion, via advertising tends to be the most conspicuous to consumers. With the majority of marketing efforts taking place beyond the view of the general public (e.g., pricing, consumer research, econometric modeling), advertising

represents the conspicuous outcropping of this much larger set of activities all designed to encourage sales.

Within advertising, a central concern of marketing departments is to limit spending on advertising and, conversely, to maximize financial return on any purchased advertising media. Return on advertising investment provides a metric to gauge correspondence between advertising expenditure and consumer spending that can be linked in some way to particular advertising effort. Similarly, the popular quip by retail pioneer John Wanamaker is rooted in this concern with return on advertising investment, in which he remarked, “Half the money I spend on advertising is wasted; the trouble is I don't know which half” (Rawson & Miner, 2005, p. 6), an attribution problem that demands substantial attention as well.

Two of the most significant advancements in improving marketer’s ability to calculate return on advertising investments have come in the ways online ad space is purchased and, subsequently, the way audience response to digital ads is tracked. For the latter, this effort has become its own topic of research for providers of digital advertising services. From early metrics such as click-through-rate (CTR) to the current more sophisticated state of cross-device tracking, where not only clicks but ad impressions are tracked and eventually triangulated with purchase data and across multiple devices belonging to the same individual, the state of advertising response measurement has changed drastically in the past decade. This has allowed new kinds of observation into consumer behavior, along with how to better influence this behavior (Schiff, 2015). In the U.S., for instance, efforts to deploy and improve cross-device tracking capabilities have been accompanied by concerns from privacy rights groups (Rotenberg, Barnes, & Gartland, 2015) and scrutiny by regulators (Federal Trade Commission, 2015).

For innovations on the ad-buying side, one of the most substantial changes arising in the early days of the web was the use of auctions to price digital advertising space sold by publishers. A popular early example of auctioning digital ad space is Google's AdWords program (Marvin, 2015). On the AdWords platform marketers place bids to buy specific keywords (e.g., "hotels") or phrases (e.g., "luxury hotels") hoping to win the auction so that their text-based ads appear in Google's search results whenever users query these words or phrases. In the case of Google's wildly successful and pioneering program, introducing the use of a second-price auction where the winner pays only the price of the second highest-bidder helped to completely shift the way digital ad space was purchased and also how attribution was assessed. Later came more dynamic "live" auctions for ad space and the solidification of publisher-agnostic RTB platforms. One of the principal features of these platforms is their reduction in advertising costs to marketers because each ad impression is bid upon individually and in real-time rather than paying in bulk. This is to provide marketers with a way to reduce ad spending by paying only the current rate based on supply and demand, bringing a new efficiency to the purchasing process and providing a means for boosting return on advertising investment. RTB platforms also afford a range of targeting and delivery options based on integration with various components such as third-party audience data providers.

### *Real-Time Bidding Platforms*

A closer look at how RTB functions illustrates this colossal shift in how ads are bought and sold online in recent years. Prior to RTB, in the past two decades personal computing has disrupted the state of advertising in its own remarkable ways. Advertisers have followed audiences as individuals have increased time spent consuming media on computers, tablets, and

smartphones. Today, online display advertising is purchased in a variety of ways. For example, advertisers still pre-purchase groups of online audiences in bulk with publishers then contractually obligated to serve a number of ad impressions to a predetermined audience of a specified composition (i.e., based on demographics like location or age). The popularity of some websites and apps, and their resulting demand from advertisers, allow their owners to sell this high-end advertising real estate at premium prices (e.g., *NYTimes.com* “above the fold” banner ad). Combining these concepts of selective audience and selective publisher inventory, in recent years more automated mechanism for buying and selling online ad space through RTB platforms have emerged. As the name suggests, RTB systems provide a way for advertisers and publishers to buy and sell online ad space, one ad impression at a time, in “real time” auctions, which takes place as webpage or app is loading. Increasingly, the integration of RTB systems with third-party data providers allows digital advertisements to be displayed to particular individuals at a particular time based on rather endless combinations of personal data (Iyer, Soberman, & Villas-Boas, 2005).

From start to finish online ad networks leveraging RTB require complex technical, business, and legal arrangements. The firms that supply the component parts of these networks must act in concert with one another, similar to a stock market with the ad network acting as a digital matchmaker, connecting on one end marketers who have advertising messages they wish to send to people and on the opposite end people capable of receiving these ads.

While specific ad network configurations vary, most often the RTB process is supported by linking up combinations of firms supplying front-end software for marketers to use directly, databases, servers, and internet networks. One way to understand all the moving parts is through the four major interests of RTB, those belonging to: internet users, marketers, publishers, and

proprietors of ad exchanges (Yuan et al., 2012). In this case, publishers are any organization with digital advertising space that they wish to sell to marketers. Again, marketers encompasses any organization wishing to buy this ad space sold by the publishers in order to distribute ads to internet users. Finally, for internet users, the largely invisible component among these players is the ad exchange itself. The ad exchange plays the crucial role of digital, real-time matchmaker and has its own array of discrete sub-parts, which are discussed in further detail below.

The common software and integrated networking components integral to an RTB system include a demand-side platform, a supply-side platform, a data management platform, an ad exchange, and an ad network. Each component is effectively connected to each other component, either directly or through other components in the system. Yuan et al. (2012) offers a comprehensive explanation for how these components function with great parsimony to achieve an operational RTB system, a system of systems so to speak.

In this study, I primarily interacted with the digital enclosure through what is known as a Demand-Side Platform (DSP). A DSP is a networked computing platform typically accessible to marketers through a web browser and DSPs work on behalf of the marketer. Often, a single DSP will be connected to multiple online ad exchanges, data management platforms (DMP), ad networks, and supply-side platforms (SSP). Importantly for this study, most DSPs allow their user, typically a marketer seeking to build and execute personalized advertising campaigns online, to leverage third-party audience databases for purposes of strategically targeting ad campaigns. The third-party data these systems supply are usually linked to specific individuals

through unique identifiers, allowing ads to be shown to certain people believed to possess specific attributes.<sup>23</sup>

While interest vary, the most prominent example of competing interests in this system are those between marketers, who wish to pay as little as possible to rent online advertising space and, opposite to marketers in the system publishers, who hope to charge as much as possible for this space. This clear difference is in addition to the diverse interests belonging to proprietors of ad exchanges and ad networks, each caught somewhere in between marketers and publishers, along with the best interests of internet users.

Also of note is that many firms provide multiple services under one roof, for instance, providing both DSP and DMP tools to marketers. The synergies from offering multiple services, given an individual firm's vantage point on industry dynamics and its own in-house data and infrastructure, make this an attractive approach yet not one without potential conflicts of interest.

Most of the individual components in an RTB system exist in multiple iterations. For instance, rather than a single DSP connected to a single SSP connected to a single ad network, and so on, more common is for multiple DSPs to be connected to multiple SSPs and multiple third-party data providers, along with various other middleware components, all linked up, unevenly, yet working together in real time across multiple ad networks and ad exchanges. In this regard, most components of an RTB system are capable of scaling up, down, over, across, and through, allowing for an unlimited number connections and combinations of each component.

It might be easy to overlook the achievement of this system of systems intended to optimize marketing efficiency and digital ad targeting while balancing various competing

---

<sup>23</sup> Or not shown to certain people, as in the case of negative (blacklist-style) online ad targeting. Some online ad campaigns target those who have *not* purchased a specific brand (e.g., Folgers) or product (e.g., Folgers Black Silk Ground Coffee), for example.

interests. This technical and organizational feat of RTB working around the clock is somewhat remarkable, as it represents an always-on coordination effort among global internet connections, networks, servers, marketers, clients, publishers, internet users, and other organizations and interests; a vast marketing machine of sorts.

The bulk of RTB activities take place behind the scenes, with millions of transactions and connections taking place each second, mostly out of sight for various users of different parts of the system. In this way, most RTB processes function primarily on a need to know basis for each other part in the system. For example, if an impending ad impression is signaled but the audience member does not meet the criteria pre-specified by a marketer for her campaign that is currently “live” on a DSP, she is typically not notified of this “impression opportunity” as it is irrelevant to her campaign. Similarly, billions of ad auctions are facilitated by an ad exchange across countless web publishers selling ad space each day, but only a small fraction of these transactions are made visible to any one publisher selling their ad space through an SSP connected to this ad exchange. Similar firm-to-firm, or component-to-component information asymmetries exist in such a system. For internet users, aside from seeing the end product—a particular advertisement displayed to them on a particular website or app—the entire RTB infrastructure and underlying processes take place entirely out of sight. This vast system of systems works in real-time to get the right ad in front of the right person at the right time.

With many parts acting together the RTB process functions sequentially as follows and the entire process typically occurring in less than 100 milliseconds (DoubleClick, 2013). First, an internet user visits a web page or launches a mobile app. Then, in the milliseconds before the page or app loads an upcoming impression is announced to all connected ad networks and their component parts. Ad exchanges query all standing bids from marketers seeking to buy ad space

to place a particular ad impression before an Internet user across each of the connected ad networks and parts. Next, information about this potential ad impression is sent through the connected networks, indicating the website or mobile app hosting the ad and (importantly for the present study) any known or estimated information about the individual who is about to view the webpage or mobile application and the ad delivered from the winning bidder. The latter is where third-party data providers enter to play a key role offering up and selling the desired audience segmentation information about the audience member who has initiated the request to load content on a website or app, that spark which began the process.

Once all information about the publisher's inventory and potential viewer of the ad are collected, the auction takes place. All preset bids matching this sent criteria are evaluated and the winner is determined. The auction winner's ad is then routed to the publisher's ad space on the corresponding website or app and the impression is delivered to the Internet user. The ad buying and selling process now concludes, at which point a host of follow-up analytics begin. These include the generation of new cookie data to be used in recording to whom the ad was serviced, capturing the relatively rare event that the viewer clicked on the ad, and a range of other less common measures that might go into effect, like the duration in seconds the viewer's mouse icon hovered over the ad, potentially valuable information especially in the case of a viewer actively hovering but not clicking on the ad.

Originating from this dynamic buying and selling process, globally billions of these micro-auctions take place every day all across the online marketing ecosystem. While not all ad impressions are sold via RTB, the proportion of ad inventory purchased programmatically has grown rapidly in recent years. For instance, estimates report around 70% of all online display ads are now bought and sold programmatically in RTB exchanges (eMarketer, 2016).



Visibility into these systems that organize personal data for advertising personalization is unequally distributed. The result is a straightforward information asymmetry. On one side, marketers who use many of today's popular online ad-buying tools are presented with an assortment of third-party personal data. This data is made available for targeting ads at internet users thought to possess certain attributes as identified usually by third-party data providers, though sometimes marketers bring their own customer data into the system. Many times, third-party audience targeting data is provided by specialty business-to-business firms with no consumer facing presence, sometimes referred to as "data brokers" (Crain, 2016). Examples of prominent data brokers include companies Acxiom, Datalogix, Epsilon, and Experian. Most internet users are not familiar with the firms that provide detailed third-party audience data to marketers. Typically, individuals have no reason or opportunity to come into contact visible with data brokers directly, as these firms' customers are marketers not individuals.

Data brokers identify, classify, and tag internet users through a multitude of tactics, such as syncing internet cookies across time and web properties, accessing individuals' account profile information, leveraging data sharing agreements with other firms, conducting consumer surveys, and integrating records from online and offline financial transactions, to name a few. As individuals go about their routine use of websites and apps, along with conducting many offline activities that now generate digital records as well, they are effectively "tagged" as having certain attributes. Identifiers associated with unique individuals or households populate databases providing the basis of user profiles containing specific attributes ranging from one's race/ethnicity, income, and family status to which brands of toothpaste and clothing they most often purchase.

On the other end, internet users are unable to see this cornucopia of personal data describing them and existing across multiple data brokers' databases, the volume of user profiles associated with them, nor the exact contents of any profile. The result is a virtual wall defining differential visibility, or information asymmetry, for marketers and internet users. This asymmetry presents accessibility issues for investigators studying the technical, social, and policy issues linked to personal data, digital marketing practices, and the dynamics of personalized advertising and RTB. Therefore, to understand the nature of these data and how they function in practice, I attempted to step through the information asymmetry by temporarily taking on the role of the marketer.

## **OVERVIEW OF RESEARCH**

Employing a hands-on, embedded approach was key to accessing and using RTB technology and beginning to understand the unique properties and affordances of current data-driven advertising platforms. For instance, to access these systems requires an up-front financial cost, sometimes referred to as a “minimum media spend,” which varies based on the DSP used. A typical minimum might be around \$1,000/year or so. In addition to financial commitment, partnering with a platform results in a commitment to bid in live auctions for online ad space, which, when successful, results in winning a proportion of the ad auctions and delivering ads to (anonymous, in this case) internet users.

To achieve this, after gaining institutional review board approval, in early 2014 I initiated a partnership with a non-profit public health organization in the U.S. This particular organization was selected so that all financial costs incurred from buying ad space on RTB platforms would in

effect be donated to this prosocial non-profit (i.e., as opposed to providing free ad buys and services to a for-profit company).

Leaders from the partner organization provided me with wide latitude to execute a series of online ad campaigns using their existing visual media. This included sidebar and banner display ads used in the organization's previous advertising efforts, which were then optimized in size for either desktop or mobile devices. The only restriction imposed by the non-profit client organization was that the ads be delivered to Americans. For this reason, web traffic for all ad campaigns was confined to that from the U.S. as determined by IP address.

Given the relative opacity of the personal data marketing ecosystem, for scholars most work addressing questions about this environment has been levied from a distance, at times favoring what might be possible in lieu of directly observing phenomena. Documented examples have been relatively uncommon. The result is a tendency to err on the side of speculation or even constructing worst case scenarios for how personal data could exist, how firms might share data with one another, or how marketers may be operating in light of abstract data possibilities (e.g., Calo, 2014).

Certainly, imagining dystopic scenarios in light of what might be possible can be a helpful exercise to stimulate critical thinking, user-centered design, and improved guidelines or regulations. And in some cases this "self-preventing philosophy" can be fruitful in preempting and curbing harms that would have otherwise emerged (Pasquale, 2015, p. 16-17). However, rooted in *what if* rather than *what is*, this dystopic-possibilities approach has its limitations. One such drawback is the continual production and reification of a discrepancy between potential problems and actual ones, or between how things may function and functionality.

Taking a different approach, for this study I relied on participant observation to gain an up-front, first-hand look at the viscera of today's digital enclosure.<sup>24</sup> In doing so, I aimed to answer the following related questions. What types of third-party personal data are currently made available to marketers through RTB ad-buying platforms? And what is the nature of these data? That is to say, in addition to observing what data are available to marketers for targeting ads online, what are the commonalities, origins, and nuances of these data?

To answer these initial questions, from 2014-2016 I explored several DSPs first-hand and documented my experiences. As the DSP is the marketer's primary interface with the other technical platforms, first- and third-party personal data, and the other players including audiences in an RTB system, using DSPs to buy ads provided both the entry point to this ecosystem as well as my primary site of inquiry. As noted, DSPs aggregate, integrate, mix, and match data from multiple data providers. This integration effort makes it possible to use a DSP to access many different firms' sources of third-party personal data. For the DSPs tested, this approach involved creating a user account and then, upon using the platform, observing its degree of integration with various data brokers, noting the types of personal data they provide for advertisement targeting, and how these data function in practice when used in online ad campaigns.<sup>25</sup>

For this study and to effectively target ad impressions for the non-profit client, initially several DSPs were evaluated. They were compared to one another weighing relative strengths and weaknesses, such as ease of use, integration with the ad exchanges and ad networks (to maximize the reach of ads), publisher inventory, customer support, and integration with a variety

---

<sup>24</sup> I refrain from calling the approach I took for this study "ethnographic," per say, as I was not immersed in an environment for an extended period but rather participated in a specific activity off and on over several years. I did this in the attempt to gain an insider's (*emic*) perspective and as required by the questions posed in this work.

<sup>25</sup> Some DSPs geared towards marketers with relatively large advertising budgets proved inaccessible due to either high minimum buy-in costs and/or interval spending restrictions (i.e., daily or monthly spending minimums exceeding my research budget).

of prominent third-party data brokers. The purpose in seeking a single DSP integrated with a variety of components on each level in the RTB system was to offer a more comprehensive first-hand assessment of the personal data currently available for targeting online ads while remaining agnostic towards any specific integrated company (e.g., data broker) as well as towards the DSP itself, which provided a window into the personal data ecosystem while functioning the same as similar DSPs.

In mid 2014, a single DSP was selected to carry out the goals of this study. At this time this particular platform was integrated with only 5 third-party data brokers. By 2016, the number had grown to around 40. This growth reflects the maturation and widespread adoption of RTB systems, including organizational and technical integration of more players and components. For marketers, middlemen, and publishers, these increased connections and data sharing agreements correspond to a win-win-win scenario. In particular, for marketers this evolution means more third-party data and audiences are accessible from a single point of entry into the RTB system. Additionally, as each data broker offers a somewhat unique blend of audience targeting data, this large number of data suppliers offered a tremendous, and usually overwhelming, volume and variety of targeting options in my observations.

I returned to this DSP regularly between 2014 and 2016 to observe which third-party audience personal data appeared in the targeting options and to execute actual ad campaigns for the non-profit client. I documented thousands of examples of personal data made readily available on this platform, on some occasion using these data to target the non-profit group's healthy living advertisements at specific internet users.

While exploring third-party personal data accessible to marketers through various DSPs, the most apparent quality of these data is their varying specificity and granularity, or just how

personal these personal data are. This varies tremendously ranging from extremely broad (e.g., *Interested in Attire, Downloaders, Holiday Shoppers*) to the oddly specific (e.g., the brand of toilet paper someone purchased *in the past 7 days* ). To make sense of these similarities and differences, in the next section I discuss these data in detail. My aim is to make clearer their nature, locating how they are derived, organized, and presented to marketers by various data brokers across different DSPs. Due to the sheer volume, variety, and granularity of these data, the findings below describe only a small fraction of the consumer data I observed available for use in online advertising personalization.

Again, the most startling characteristic of the personal data provided to marketers through third-party data brokers is the utter extent of the data, seemingly unbounded in depth, scope, and granularity. Upon exploring this environment it quickly became clear to me that any effort to comprehensively document all these data would be incomplete. Therefore, after spending time in this environment observing and documenting many examples of third-party personal data, I aimed to synthesize what I found. The next section in Part 1 describes this synthesis, some of the more salient characteristics and the nature of these data, and limited examples of along with how they are presented to marketers for personalizing online ads. Additionally, many of the categories and features identified tend to overlap rather than present discretely (i.e., the ability to use shopping data, temporal data, and Boolean modifiers, at times all together).

## **Part 1: Exploring the Nature of Third-Party Personal Data**

### **Personal Data Origins**

Given the variety of third-party personal data available, it can be difficult to know where to begin. One foothold is to recall that all personal data made available to marketers have an

origin, the original source or spark of the data. Examining this consumer data revealed some of the most common origins are past purchases for goods and services, perceived interest in and/or intent to purchase certain goods and services in the future, socioeconomic indicators and classifications, life events triggers and categories, technically-derived information about computer networks and devices, data taken from public records, media consumption, behavior on social media sites, user-supplied data from online and offline account profiles, and self-report data from online and offline surveys, among others.

### ***Products & Services Data***

#### *Past Purchases*

In terms of the volume of data observed, the most abundant as well as the most granular were data corresponding to prior purchases of products and services. As past behavior is a strong indicator of future purchase behavior, the fact that third-party data providers make this information available to marketers and in abundant detail is perhaps unsurprising. At times, past purchase data observed were accompanied by metadata indicating their origins. Other times no origin was cited.

Of note for past purchase data observed, the majority of third-party providers did not specify the origin of the data. This is not surprising, as data brokers have declined to cite their sources even upon the demands of an investigation by U.S. Congress (U.S. Senate, 2013). Often these firms cite proprietary needs to protect trade secrets, which in some cases include how and where they collect, rent, or buy information about consumers. Most providers of customer data I observed did so without indicating the particular origin, though some did indicate this such as which major credit card company was supplying which data. Nevertheless, it appears to be up to

the marketer to determine the trustworthiness or reliability of the data and how much confidence to place in its targeting abilities, likely determined by its measured performance in ad campaigns. With some providers naming their sources explicitly and others leaving this ambiguous, I was confronted by a sense of uncertainty when not able to see the origin only because some providers did specify this. When this origin metadata was included, the sources varied considerably. Past purchase data with cited sources included that from credit card transaction records provided directly from credit card companies themselves, from named brick-and-mortar retailers' point-of-sale systems (their cash registers), from named retailer's shopper rewards cards, from specific ecommerce sites as well as third-party cookie data from these sites, and taken customer self-report in online and offline surveys, among others.

Additionally, temporal indicators accompanied some but not all of these prior purchase data (e.g., *purchased during past 30 days*). These time-based signals separated certain data brokers' offerings from others based on the freshness of the monetary transaction. In some cases, the time period modifiers were well-tuned to the product (e.g., *purchased vehicle: (make/model): 5 or more years ago*). This alignment of many of the time-based modifiers with the available product types illustrates an important compatibility issue when combining data brokers' offerings; where some data can be compounded well with some modifiers but not others. Conceptually, it also illustrates the human hand in the curation of some third-party personal data offerings, as a record of an individual purchasing, for instance, organic milk in the past 30 days is likely useful in a much different way than a record of this same purchase made 5 or more years ago. Similarly, the longer period between purchasing vehicles for most consumers appears to influence, logically so, the accompanying modifiers made available for vehicle data for ad targeting (e.g., *purchased 5 or more years ago*), for example. Personal data options observed for



targeting individuals based on time since past purchase of a particular consumer good or service ranged from 7 days (e.g., grocery items) to multiple decades (e.g., cars, homes, and other major purchases).

### *Future Purchases*

Closely related to prior purchase data, another notable class of data observed was data labeled as “interested in...”, “intent to purchase...”, and similar labels indicating the specific product or good listed corresponding not to prior transaction but that there was reason to believe an individual was *likely to purchase* the item in the future.

Although less common than in the case of past purchase data, future anticipated purchase data was at times accompanied by an origin. Usually this was a reference to online browsing activity, indicating that the data broker was inferring a potential purchase based on which websites or apps a corresponding individual had visited. The fuzzy connection between web activity and future purchases is likely validated in its effectiveness on occasion, likely better than (chance). Still, it is worth noting that much intent to purchase data that did cite an origin appears to be based on the murky association between the wide variety of online content many internet users consume and the comparatively small number of purchases actually made.

Additionally, some of the future purchase intent personal data was also accompanied by a time-based modifiers, such as “likely to purchase: next 30 days” followed by various options for goods and services (e.g., motorcycle). Personal data corresponding to consumer goods and services referenced prior offline and online transactions for these items as well as individuals’ perceived interest in and/or intent to purchase items in the future. Similarly, current ownership data for certain items pointed to the present (e.g., *owns: boat* or *home owner: adjustable rate*

*mortgage*). In this way much of the personal data linked to specific or general products and services purchased or potentially purchased was also anchored to a specific temporality. When present, this time dimension modifier offers marketers an additional layer of sophistication when targeting ads based on either a person's past, present, or anticipated future consumer behavior.

### *Products & Services*

Upon examining targeting data associated with products and services, a reasonable conclusion is as follows: with minor exceptions, such as certain health procedures or illicit goods and services, if something can be purchased it is likely made available to marketers for ad targeting. Most consumer goods and services have a corresponding presence in third-party data providers' offerings.

One distinction within products and services transaction or potential transaction data is specificity of the item. Often this division is whether a specific item is captured versus a type of product or service. For instance, more general types of products available from past purchase or intent to purchase data include *tires, apparel, socks, lipstick, cookies, organic food, toaster, major appliances, sporting equipment, cellular telephone services, renter's insurance*, and so on. If it exists and can be purchased, most likely there is a corresponding data category available for targeting ads to corresponding internet users. In the case of such general product targeting, like past purchase of "apparel," the usefulness to marketers of these and similar high-level categorical purchase data is unclear. One explanation might be demonstrated effectiveness when combined into a model with other data, where the rationality of the inputs (*purchased apparel AND owns boat AND opens SMS messages from marketers*) is justified by success when applied to a

campaign. In this way what may appear in isolation as an odd personal data offering by a data broker may be a key ingredient in a highly effective ad-targeting model.

On the other end of goods and services third-party targeting data are specific items. Similar to general categories, the vastness and granularity of these data is impressive. For instance, nearly all products one can envision being sold in any variety of retail stores have corresponding data made available by data brokers for online ad targeting.

A quick thought experiment helps in grasping the extent of this consumer goods data made available for ad targeting. First, consider nearly all categories of items that can be purchased either in a brick-and-mortar store or on an ecommerce site (e.g., electronics, clothing, household cleaning supplies, frozen food, canned soup, cereal, toothpaste). This list is of course endless but provides a placeholder. Then, for each category considered, try to think of the top 5 to (in some cases) +100 brands that sell products within that product category. Then, further subdividing, for each brand consider any varieties of this item manufactured by this brand. To provide just two examples: *Purchased: Past 7 Days: Grocery: Crackers: Keebler: Club* (vs. *...Keebler: Townhouse*) or *Purchased: 3 or More Years Ago: Automobile: Luxury: BMW: 6 Series* (vs. *...BMW: 7 Series*), and so on.

Location of purchase was also a common variety of targeting data made available. In this instance rather than targeting ads based on categories or specific products or services within these categories, records of where people shop supplied a related (likely-linked) variety of personal data. For example, “shops at” data was available for a large list of retail stores, including specific grocery and retail stores (e.g., *Wal-Mart, Target, Kroger, Kohl’s, Macy’s*). Similar to product and goods data, categories of stores where individuals make purchases were also available for ad targeting (e.g. *Shops at: Big Box Stores*).

Viewed together, we can see much of the consumer decision process captured in third-party data related to products and services. From initial interest in a product and evaluating purchase decision (e.g., reading about it on a website) to making a final purchase decision (e.g., which product) and corresponding place of distribution (e.g., which retailer), many signals in this process are now captured within the digital enclosure, repackaged and made available to marketers as new digital data goods themselves.

By far, the most extensive third-party personal data observed was that linked to purchase or purchase intent for products and services. This self-referential data is logical in an environment designed to distribute advertisements encouraging purchase of *additional* products and services. Looking beyond proportionality in these personal data, a number of other categories and varieties of personal data were also observed, each made available to marketers for personalizing online advertisements.

### ***Activity Data***

A grouping of personal data available from numerous providers was activity- or event-generated data. Like products and services data, these data ranged from incredibly general to highly specific. Activity and event-based data corresponding to regular activities captured by the digital enclosure (e.g., *Leisure Activities: Nightclubs & Dancing*), along with discrete events (e.g., *Tweets about: [enter brand name]*) including events anchored in time, the data for which appear to “expire” at some point given their classifications (e.g., *Recently Single*).

### *Charitable Donations*

One example of activity-based data observed was personal data about individuals' charitable donations. For instance, data corresponding to whether an individual does or does not make financial contributions to environmental, religious, and many other types of non-profit groups was available for ad targeting. As was the ability to target those who make charitable donations to conservative or liberal political causes. While these data are financial in nature, they are distinct from products and services given the lack of connection to a particular item but rather connection to a cause or ideology.

### *Leisure Activities*

Another set of activities observed as captured by the digital enclosure and made available by data brokers were individuals' leisure activities. As with data about goods and services procured, these activity data ranged from general (e.g., *Outdoor Activities*) to highly specific (e.g., *Goes camping frequently*). Similarly, individuals who *visit casinos* were available for ad targeting, including in some cases the specific location of one's gambling activity (e.g., *Gambles in Las Vegas, Nevada*). Internet users who regularly visit movie theaters were also available for targeting ads towards along with incredibly specific combinations of this type of activity data (e.g., *Went to Movie Theater with Church Group*).

### **Network, Application, & Device Data**

One of the most concrete and arguably accurate forms of personal data made available to marketers I observed on DSPs was that emanating automatically from networks, software applications, devices, or combinations of these. Internet service providers and mobile carriers are

typically identifiable through the range of IP addresses they issue to internet users or from specific hardware (e.g., an AT&T exclusive smartphone), making the network a person is currently using trivial to detect (e.g., Time Warner Cable, Comcast, Charter, Verizon, T-Mobile, Cricket, Boost Mobile). Across DSPs and data brokers observed, the ability to target audiences based on network providers was widespread.

Similarly, software applications and the devices they run on also provide automated data to servers as they communicate back and forth, data which is made available to marketers for general (e.g., operating system) or highly specific (e.g., smartphone model) targeting. Some of these data are required for proper information exchange and optimizing how content is presented (e.g., desktop vs. mobile browser). I observed data brokers providing this automated device and software for discriminating ads based on whether someone was using a mobile or desktop computer, Wi-Fi or mobile phone data, the particular operating system in use (e.g., iOS, Android, Windows, Ubuntu), web browser (Internet Explorer, Safari, Chrome) and browser version (e.g., IE v6, IE v7, etc.), and, in the case of mobile device traffic, the exact model being used to access content (e.g., Samsung Galaxy S7 Edge, Apple iPhone 6 S Plus).

Another form of device-generated personal data offered by data brokers for ad targeting corresponded to the type of apps already installed on an internet user's device. This type of data tended to contain broad categories of apps installed (e.g., *Medical Apps*, *Trivia Game Apps*) rather than specific apps (e.g., *My Cancer Coach*, *Trivia Crack*). This app data allows marketers to target individuals not only based on the specific website or app they are currently using, but also which apps they have installed and assumedly expressed some interest in that genre or topic.

As people access web content on various networks, using different software, and on desktop and mobile devices, these automatically generated personal data are revealed to

publishers and various middleman, including data brokers who later make this technically-oriented information available on DSPs for use in targeting ads.

## **Additional Data Features & Considerations**

### *Individuals vs. Households*

For some of the data brokers examined, and for subsets of their data, I found they included metadata describing whether an attribute was linked to a single individual vs. a household. Though this was not usually specified. When present, however, this distinction might be a crucial determinant in electing to use this targeting criterion. For instance, one's living arrangement and household size directly impacts the specificity and effectiveness of using household-level data. For those living alone, whose household attributes correspond to individual attributes, this person- vs. household-level data might be identical and the distinction inconsequential. Yet, for say a family of four composed of two parents and two children all living together, or an apartment filled with co-habituating college roommates, this data distinction could be incredibly consequential for marketing efficiency. This is in addition to concerns for privacy, with postal address at times bearing different sensitivities than data associated with a unique person identifier.

This distinction between household and individual determines how addressable the corresponding targeted advertisement might be, and applies to most of the previously discussed examples and origins of these data. It also impacts just how intimate the data broker's database is with different kinds of personal data linked to either individuals or groups of people.

For marketers, personal data associated with a particular four-person household identified by say street address, IP address, or other household identifier might be considered noisy when

seeking to target only that family member who has managed to earn his household a particular label in a data broker's system (e.g., *Seeking Professional Services: Legal Counsel*). Hampering marketing efficiency, in the previous four-person household example the marketer might be wasting, on average, 75% of her ad buys which reach the wrong family members—those not currently seeking legal services.

As with many other features of these data described below, addressing an isolated individual who is thought to possess an attribute vs. addressing an individual associated with a household where she or another member of the home possesses a particular attribute, impacts many other considerations for how we might think about these data, such as how sensitive, controversial, or accurate it is. Important for considerations of addressability, in my observations more often than not data brokers did not specify whether individual targeting data were derived associated with individuals vs. households. While attributes are assumed to correspond to the individual-level given the nature of certain data (e.g., *Gender: Female*), for others, when this distinction is not made it is less clear (e.g., *Interests: Education: College Admissions*). Nevertheless, when this individual vs. household metadata was present, these indicators provided a potentially crucial distinction for the corresponding personal data.

### *Modifiers*

As noted, some personal data observed provided finer granularity and cues about accuracy through the presence of various modifier options. In some instances, modifiers were separate from corresponding data and able to be applied in addition to a targeting criterion (*Frequently Purchased: Excedrin* or *Frequently Purchased: Tylenol*). In other cases the modifier was essentially hard-linked to its corresponding data, with the two unable to be separated



(*Frequent Purchasers of Frozen Dinners*). Regardless, these enhancement data were most often used to assign either a temporal component and/or to reveal the intensity or potential accuracy of certain data. In some cases data modifiers served as a prefix to indicate the origin or nature of the targeting data (e.g., *Tweeted About: Movies*; *Reads Webpages About: Dieting & Exercise*).

Time-based modifiers corresponded to the past, present, and future behaviors and attributes of internet audiences. For instance, these modifiers looked backwards (e.g., *Over 12 months ago: Purchased: Sports Car*; *In past 7 days: Purchases made at: Trader Joes*; *Recent transaction for: Audiobooks*). They also looked to the current moment (e.g., *Owns: Boat*; *Present Home Value: >\$1,000,000*; *Current Cash Savings: <\$500*). Finally, these temporal data modifiers looked into the future (e.g., *In the Market for: Automobile: Used: Compact*; *Highly likely to purchase in the next 30 days: Cellular Phone*).

Some of the future-oriented temporal modifiers were quite ambiguous, with a lack of metadata to explain their freshness and leading to questions of about accuracy. For instance, it is unclear exactly or roughly how long a person is "in the market for" a specified item, which probably depends heavily on the item itself (e.g., automobile vs. shoes). This modifier attribute could be short-lived or may reach far into the future. It may become inaccurate when a corresponding purchase is eventually made, especially if no parallel signal indicating purchase is ever captured. Alternatively, this attribute of being in the market would over time fade away for the internet user as her interest wanes. In this way, some personal data stored by data brokers has a definite shelf life, though in my observations this duration was not rendered clear for the marketer's vantage point.

Additionally, some temporal modifiers revealed the dynamic nature of these data, in many cases constantly changing and requiring automated updating as perceived time-linked

attributes continually come and go. In this regard the freshness of certain personal data is presented by data brokers as a feature (e.g., *In past 30 days*), likely to be used or not used by marketers depending on the product being advertised, as in the case of time-linked targeting data indicating an internet user's purchase "In Past 7 Days" of a new home vs. Frosted Mini-Wheats.

Other modifiers indicated or added different levels of accuracy or intensity in the corresponding data. For example, the following modifiers instilled varying levels of confidence when attached to corresponding targeting data: *Interested in:*, *Searched for:*, *Liked on Social Media:*, *Watched:* (TV programs), *Declared:*, *Verified:*, *Likely:*, *Highly Likely:*.

From time-based modifiers along with those signaling confidence or intensity in their attached personal data, a key affordance of many data available on DSPs is the ability to combine one or more data items to produce more specific targeting criteria. Typically modifiers act as simple Boolean operations, in this case requiring the presence of two or more items simultaneously to produce a positive result—a particular internet user.

### *Boolean Logic*

Boolean logic is key to how DSPs and integrated third-party data can be leveraged in ad targeting. This applies not only in the previous case of adding a modifier to an existing type of data but for a range of other operations that dramatically alter the capabilities of third-party personal data. Most DSPs observed provided user-friendly integration of AND, OR, and NOT Boolean operators to allow combinations of personal data resulting in more sophisticated targeting. Use of the AND/OR operators in the context of targeting by data allows for complex and in theory unlimited combinations (e.g., *(Parent OR Grandparent) AND (Frequent Shopper: Wal-Mart)*).

Perhaps the most notable feature here is the ability to use negation, where a negative operator allows marketers to NOT deliver an ad to anyone thought to possess a particular attribute (e.g., *NOT: Lifestyle: Religious People*). More likely than negative targeting alone is to use exclusion statements in combination with inclusion statements to increase likelihood of reaching one's target audience (e.g., *Past Charitable Donors: International Aid Efforts AND NOT: Lifestyle: Religious People*).

Regardless the case, as endless combinations of personal data using Boolean operators is possible, the ability to *exclude* audiences based on data broker offerings is a powerful yet perhaps overlooked feature of these platforms and in critical discussions about ad targeting in general. Though the accuracy of targeting data is always in question, still the ability to systematically exclude many of those people thought to possess specific non-target attributes represents a break from prior ad-buying processes. That is to say, even though buying ads during specific radio or television programming or appearing in certain magazines or newspapers always decreases the odds that certain people who do not typically consume these media will be exposed to the advertising messages they contain, this method does not systematically exclude these non-target audiences from receiving this advertising.

Yet the affordances of Boolean logic when using DSPs to execute online ad campaigns, specifically via the exclusionary NOT operator, signal a marked shift in the ability to exclude unwanted audiences when conducting advertising. In some applications this exclusionary capability might raise concerns over fairness (e.g., *NOT Income: <\$500,000*) or even social discrimination (e.g., *NOT Ethnicity: African American*). At the same time, this targeting affordance and the efficiencies it offers could lead to decreased costs associated with executing online campaigns ultimately improving a marketer's return on advertising investment.

## **PART 2: Exploring the Dynamics of Third-Party Personal Data**

In addition to assessing the nature of third-party personal data and exploring the resulting information asymmetry, I further examined how these data function when put to use by marketers. Noting the character of data provides insights into what is captured in the digital enclosure and how it can be put to use by marketers to personalize online advertisements. Examining the functionality of these data adds an additional layer of understanding regarding their broader implications in the digital marketing environment. For instance, if utilizing these personal data lead to more democratic outcomes, we might ascribe a net benefit to their underlying systems and the activities of marketers. Alternatively, if these data appear to reinforce existing disparities, this could bring its own set of new challenges.

One way to explore the personal data available on RTB platforms for ad targeting is to look at how popular certain data are compared to others. In this case, popularity provides a direct proxy for which personal data and corresponding people or audiences marketers target ads towards more than others. As the financial costs of advertising largely determine who gets to send promotional messages to whom, when, and where, in addition to what these costs reveal about who is targeted, the pricing dynamics of the emerging advertising ecosystem warrant further attention. Therefore, it stands to reason if some consumer segments are easier to reach than others, due to being in less demand and/or less expensive at auction, this may have both positive and negative implications linked to consumer vulnerability and the type of advertising in question such as political ads, ads for products and services, or prosocial public service ads. Accordingly, I examined the market-based audience costs on RTB advertising systems and considered the following: How do supply and demand forces in real-time bidding ad networks impact audience pricing at the level of individual traits?

To explore the bidding component of the RTB phenomenon, I conducted a series of tests assessing differential pricing for online advertisements when targeted towards single-trait audiences identified by third-party data providers. Audiences were segmented by one of the following: *gender, level of education, income, U.S. political affiliation, or relationship status*. From the existing partnership with a non-profit organization, their healthy living ads supplied the promotional media and ad campaigns. I purchased online display ad space for delivering the healthy behavior ads on 30 prominent U.S.-oriented websites. These publishers (websites) were selected due to their immense popularity based on comScore (comscore.com) traffic rankings. The websites included a variety of the most popular websites for news, entertainment, e-commerce, general information, and other highly prominent websites based on web traffic originating in the U.S. (by total monthly visitors). Ad space was bid upon using third-party audience data from multiple providers. Bids were placed from the single DSP which accessed multiple industry-leading ad exchanges and ad networks.

Of importance to the present study with respect to ad impression bidding and pricing is how the auction component functions in RTB, specifically how winning bids are determined and charged to the marketer (see diagram in Ch. 2, Fig. 2.2, Real-Time Bidding Process).<sup>26</sup> As noted, typically, RTB systems use so-called “second price” auctions to determine the cost charged to a winning marketer and, conversely, what is paid to the publisher of the ad (Edelman, Ostrovsky, Schwarz, 2007). Under this scheme the highest bidder pays only the price of the second highest bidder, or sometimes a fraction of one cent more, rather than paying their actual maximum

---

<sup>26</sup> Different RTB systems use slightly different flavors of second price auctions and not all platforms specify to users which is used, though the logic is unchanged in that the winner only pays the same or slightly more than the second highest bidder not the winner’s original bid amount. Whether a platform uses a standard Vickery auction scheme or Generalized Second Price option is inconsequential to the present study. More importantly, the integration of second price auctions in RTB systems facilitates the design of this study, which relies on a fixed maximum bid amount used across all campaigns reported, allowing for comparison of win/loss ratios within categories (demographics).

amount bid. The subtle brilliance of the second price auction is its ability to instill confidence in buyers who are assured they are never paying more than their closest competitor is willing to pay. For example, a maximum bid level of \$20.00 (per 1,000 ad impressions) would win ad impressions at a wide range of auction prices almost always paying below this \$20.00 bid amount (divided by 1,000 in this example) and a different price for each impression. The amounts paid depend almost entirely on the other marketers' maximum bids.<sup>27</sup> This competition is the central conceptual thread knitting together the components of ad exchanges via real-time bidding. Additionally, to provide more robust assessment of the dynamics of these data, accounting for any seasonal variation or timing anomalies, identical ad-buying test campaigns were executed at three time points during April, July, and December of 2014. Additional ad campaigns were tested and executed for the non-profit client. For a straightforward example, I present results from these three months in 2014. Results were merged across the three time points to provide a longitudinal view of pricing dynamics.

Additionally, by using an assortment of ad networks and ad exchanges, resultant auction data was not specific to any particular company's network (e.g., Google, Facebook, Yahoo!, OpenX) but instead reflective of the much broader network of ad networks and exchanges and, therefore, characteristic of the RTB space more generally.

All bids were limited to U.S. web traffic and all auctions operated under a generalized second-price scheme as previously described. In total, for all tests during the three time periods I

---

<sup>27</sup> There are a number of technical details beyond the scope of this chapter that prevent RTB ad-buying from being purely auction driven, including mechanisms to prevent all ad impressions from being sold at the current best-available second price. As there are times when doing so would favor one party in the system too heavily over another (e.g., a marketer buying ads vs. a publisher selling ad space). For instance, "price floors," degree of "session depth," and "waterfalling" each influence how bids are evaluated in different RTB scenarios. For further explanation, see <http://adexchanger.com/data-driven-thinking/real-time-bids/> and <http://adexchanger.com/the-seller/the-programmatic-waterfall-mystery/>

competed in >200,000 RTB ad auctions. The bids, which were all placed at the same consistent dollar amount, were successful on ~45,000 occasions for an overall RTB win rate of 22.5%.

Using a consistent bid amount, win rates can be compared using Pearson chi-squared goodness of fit tests. Expected values for the proportion of successful bids for each trait were computed using an assumption of equality within that category. Essentially, this test imposes the assumption that categorical outcomes will be similar; that is, that the distribution of outcomes among a category (i.e., in this case Democrats vs. Republicans) will be equal even though the number of observations for each likely differ.

Below, I briefly summarize the results from this series of ad-buying tests. Results are compared only within each group (e.g., gender) to locate differential demand and pricing for these categories, as comparison among levels in different categories (e.g., female vs. Republican) will always result in statistical differences. Instead, I was interested in whether within the categories tested, whether certain attributes were in greater demand by other marketers, which would be apparent by the relative difficulty and corresponding price of winning ads targeted at each attribute within a category (e.g., Democrats vs. Republicans).

## **Results**

For ads targeted by gender, the proportion of bids won differed significantly between females and males,  $\chi^2 (1) = 41.43, p < .001$ , with males more difficult to win at auction than females. Ads targeted by education level also differed significantly,  $\chi^2 (2) = 2787.29, p < .001$ , with those believed to hold a high school diploma or college degree in higher demand than those estimated to have an advanced degree. Targeting by income differed significantly,  $\chi^2 (4) = 742.67, p < .001$ , with the lowest income individuals (<\$25K) in less demand and easier to win

than middle (\$25K-\$60K, \$60K-\$75K, \$75K-100K) and upper (\$100K+) income consumers. Democrats were significantly more difficult to win than Republicans,  $\chi^2 (1) = 225.72, p < .001$ . Finally, the proportion of successful bids differed significantly when targeting by relationship status,  $\chi^2 (4) = 4755.31, p < .001$ , with those labeled single or married substantially more difficult to win than those thought to be divorced, engaged, or widowed. All differences observed were significant and highly unlikely due to chance.

This test over three-time points offers just one characterization of today's online attention marketplace. The main observation from these tests is that certain audiences, as identified and segmented by individual traits provided by third-party personal data suppliers, were shown to consistently be more costly and therefore difficult to reach with targeted online advertisements than others. This is inherent to the real-time bidding system, where the bid process inflates the costs of certain audiences over others as a direct product of different bids placed by marketers, illustrating supply and demand pricing dynamics of RTB platforms. Across these ad campaigns, differential audience pricing was located within each of the categories targeted with the healthy living ads, including: gender, income, level of education, political affiliation, and relationship status.

## **DISCUSSION**

Based on observations of an array of personal data offerings made available to marketers for personalize ads, from a multitude of third-party data providers on different ad-buying platforms, I found compelling evidence for what scholars sometime speculate about but less often provide tangible evidence towards, the existence of a grand digital enclosure. Andrejevic (2007) describes, "...when we go online, we generate increasingly detailed forms of



transactional information that become secondary information commodities: information that may eventually be sold to third parties or used by marketers for targeted advertising campaigns” (p. 2). Part one of this study located and synthesized just some of these information commodities in the context of personalized advertising. The digital enclosure appears to be vast, alive, and well. Few have contested this is the case, yet first hand examinations into the nature of these consumer data have been largely absent. I have aimed to present and synthesize some of these data in efforts to provide a clearer picture of the information asymmetry that many believe to characterize the interactions between online systems and consumers today.

Additionally, the way in which these data are available in plain view and ready access to marketers, yet typically unknown to many consumers, further erects and exacerbates information asymmetry. Many of the data located in my observations do not originate online, that is they come from many “offline” sources, such as in-store purchases, tracked by credit card transactions, shopper reward cards, and sophisticated couponing (e.g., those mailed to households with unique identifiers, later matched to households when redeemed). This availability of consumer data from a multitude of non-online sources call into question the notion of contextual integrity. Understanding online ads as being delivered in the context of online behavior, it would seem a great many of the forms of consumer data I observed might violate current notions of contextual integrity. That is to say, consumers may not expect that data corresponding to their activities such as the type of job they have, the number of children they have, and which charitable or political causes to which they make financial donations would later be used to selectively show or not show them a particular advertisement on the internet. Yet, it appears this is the case.

In part two of this study, I conducted a series of tests in which I bought advertisements targeted at specific attributes in a selected group of examples. The results from these tests, repeated at three time points over different months during the same calendar year, illustrated the way in which attributes become more or less expensive to target based on supply-and-demand and also how much other marketers are currently paying when bidding to show ads to these audiences. Certainly, this simplistic example only suggests differential pricing and does not necessarily indicate that the attributes that were more difficult to win in my testing are always more costly and difficult to target.

A further caveat when interpreting the outcomes from part two of this study is that these real-world ad buying tests are unable to account for audience size within each category tested, which is also likely to have an impact. A hypothetical example helps to illustrate this limitation. Consider the category of gender: If we knew there were only half as many females as males using the internet, then we would expect ad auctions for female audience members to occur roughly half as often as those for male audiences. As marketers set bids for audiences ahead of the auctions themselves (prior to an audience member loading a web page or app), these bids by multiple marketers are essentially sitting and waiting for the same targets, in this example for female internet users to visit certain web pages or apps. Thus, given equivalent demand, the nature of the second-price auction would effectively drive up the price for females given the limited supply. In this case, if there really were half as many female as male internet users, marketers would expect to pay a premium for them compared to males; that is, if they are attempting to target females and males at the same frequency, which may not be the case. Despite this possible effect of audience size, or perhaps an effect of over/under supply for a particular trait given its relative demand from marketers, the outcome of differential pricing

remains just that: differential. Some attributes become more or less costly than others. This difference is the result of both marketers willing to pay more for certain audiences and the relative number of internet users possessing this attribute (e.g., male, Democrat, single, etc.). The amount of variance in pricing that could be attributed to audience size is not accounted for by the method used to examine relative audience values in the current study. However, despite underlying explanations for *why* it is more costly to reach people with some attributes compared others, in these tests targeting certain audiences within a given category (e.g., relationship status: married vs. divorced) was more expensive than others due to supply-and-demand affordances of a real-time bidding.

Overall, policymakers interested in understanding the degree to which online environments offer a level playing field might take note of these differences. In the aggregate, systematic differences in individual attribute pricing may affect the degree to which more broadly targeted ad campaigns send advertising messages to certain groups and not others. Consider the case of setting a maximum bid to conserve your marketing budget, which then has the secondary effect of limiting the exposure of your advertisements to those who are less expensive due to being in lower demand by other marketers. Or, similarly, consider the context of online political advertising, where the ability to target more easily young or old, rich or poor, liberals or conservatives, could favor one political candidate over another in the aggregate. At the same time, a similar advantage emerges for targeting ads that promote, for instance, predatory products and services or, conversely, advertisements that promote healthy behaviors, as in the case of the ad campaigns used in the current study. If more advertisers are in competition for a particular audience, thus willing to set higher bid prices for that segment, then consumers possessing this trait (e.g., *household income > \$100K*) become more difficult (expensive) to win

at auction when attempting to deliver targeted promotional messages. The opposite holds true when considering less desirable consumer segments who by being in lower demand are therefore easier to win at auction under RTB supply and demand pricing conditions. Additionally, though not likely to be the case in my empirical tests, if the unevenness of these effects were to meet the criteria for disparate impact, then the proprietors of these systems may incur legal liabilities related to providing means for prohibited forms of social discrimination. Due to a range of legal, commercial, and technical reasons, locating potential harms embedded in these technical arrangements, like social discrimination, has not been easy (Stevenson, 2014). Importantly, each of these previous considerations correspond to understudied dynamics of proprietary user data, algorithms, and systems that operate beyond the view of the audience members whose lives these processes impact.

## **CONCLUSION**

Overall, I find ample evidence for the presence of the digital enclosure. While every transaction and interaction may not generate information that ends up in a data broker's repository located in some far off data center, it seems a great proportion do. Records of these transactions, along with derived and inferred consumer attributes based on these records, are exchanged in a marketplace for consumer data and internet user attention. Initially, it was a perceived information asymmetry linked to advertising personalization systems that prompted my up close inquiry reported in this study. In attempting to further understand and explain this phenomenon both to myself and others, I attempted to step through the one-way mirror that often characterizes the interactions between marketers and consumer on the web. The result is a documented information asymmetry, supporting, to some degree, the criticisms levied by

scholars and advocates in recent years regarding the digital enclosure and the degree to which consumer data persists outside the view of those individuals it describes.

In locating worthy research problems related to marketing, Shugan (2003) calls for academic researchers to engage in studies of interest beyond the academy; that we might aim to appeal to a variety of external audiences in addition to scholarly interests. He identifies 15 stakeholder groups to consider who influence or are influenced by marketing related research, including: policy makers, marketing practitioners, members of the news media, and consumers, among others. By taking on the vantage point of the marketer and analyzing digital consumer data and advertising personalization from the inside, so to speak, I hope this study might engage several of these audiences who otherwise may not have adequate information from which to form opinions nor be in conversation with one another about this topic given the opacity with which these systems operate.

## Chapter 4

### **It acts like it knows me: How consumers think about personalized online advertising**

In this chapter I present a study drawing from a series of focus groups I conducted with adults to examine how individuals reason about advertising personalization. I set out to explore the mental models people rely on for interacting with online advertising in their daily lives, in particular advertising they believe has been personalized for them. Prior work examining consumers' perceptions of advertising has mainly focused on what people think about the act of advertising, gauging public opinions about advertising in general or advertising on specific forms of media. The result is that we know a great deal about what people think about advertising. Instead, we might gain additional insights by exploring not only what people think but also how they think about various advertising practices. Given the prominence of personalized advertising today, primarily delivered through websites and apps, this study seeks to understand how individuals reason about online ad personalization by exploring consumers' mental models of this process. In doing so this article builds on established work investigating mental models in human-computer interaction, applying and extending this line of inquiry into contemporary advertising studies. Based on observations from a half dozen focus groups with adults in the U.S. and U.K., it appears consumers may have fairly coherent and even consistent mental models of how online advertising personalization is achieved. Similarly, the common parts these models lack may also reveal important insights about how consumers may not be thinking about ad

personalization, absences that could be equally important in understanding how people reason about this everyday online experience.

## INTRODUCTION

Advertising researchers have advocated for increased public knowledge of advertising practices to enhance consumers' understanding of how persuasive mechanisms work. For instance, recently investigators have examined the viability of various advertising education interventions such as those coming as early as elementary school (e.g., Nelson, 2016). These efforts highlight a conviction few in the field would contest: consumer knowledge of advertising mechanisms is important in furthering public understanding of advertising, and that overall this is a worthy goal. Further, this conviction crosscuts the various stakeholders and interests in the field. Practitioners, academics, and policymakers all stand to benefit from greater public understanding and engagement with advertising, though each in different ways. Therefore, it follows that insights into how ordinary people think about advertising practices should benefit each of these stakeholder groups in informing them about how best to interact with their various constituents, whether they be consumers, students, or citizens generally.

Most research on consumer perceptions of the now well-established construct advertising-in-general have focused on *what* consumers think about advertising, but less commonly on *how* they think about advertising. This distinction can be subtle, yet important particularly as these two lines of inquiry, along with associated behavioral implications, are not the same. This void is likely more pronounced when it comes to more recent developments in the state of the art, such as the rise of digital advertising messages that can be tailored towards individual viewers using consumer data and in fairly sophisticated ways.

This study draws on a series of focus groups with U.S. and U.K. adult consumers to explore how individuals think about one process among today's much larger contemporary advertising practice—personalized online advertising. In doing so, this article builds on previous work theorizing the impact of mental models. Prior work on mental models has primarily been relegated to studies of human-computer interaction, a line of inquiry the current study extends into contemporary advertising studies in which inquiry into consumers' mental models of advertising process itself have been largely absent. The rationale for this union of scholarship, between what we know about people's invisible mental models and consumers' perceptions of advertising, is twofold.

First, and perhaps obvious, an increasing proportion of consumers' interactions with advertising are inseparable from their interactions with computers. Advertising and digital advertising are now largely synonymous; marketing strategy has become digital strategy. Moreover, the marriage of advertising and networked computing extends to “offline” marketing channels too, as evidenced in promotion of brand's social media accounts and campaign hashtags now prominent in television commercials, magazine ads, radio spots, and roadside billboards (i.e., the text “Follow us on Instagram @Starbucks” appearing at the bottom of a printed newspaper advertisement).

Second, advertising researchers have routinely cited the need for more cross-disciplinary approaches to advance the field (e.g., Rotted & Taylor, 2009). While this appeal for cross-disciplinarity is far from unique to advertising studies, given the increasing embeddedness of advertising in consumers' online experiences, the field is particularly well suited for this kind of cross-pollination drawing from human-computer interaction. Connecting these areas of scholarship, advertising's ongoing inquiry into consumer perceptions and human-computer



interaction's emphasis on the mental models of systems employed by "users," reflects both the need as advertising researchers to go where the field has taken many of us—to studies of computer-mediated advertising. This type of cross-disciplinarity might result in new insights by leveraging existing bodies of knowledge from other domains instead of struggling to reinvent them.

## **RELATED LITERATURE**

### **Public Perceptions of Advertising**

Given the attention advertising research has given to studying advertising itself, we may know more about how the public feels about this practice than most other components of contemporary society. Similar inquiry has taken place across the globe as well including the U.K. (O'Donohoe, 1995), across Europe (Petrovici & Marinov, 2007), in Asia (Tsang, Ho, Liang, 2004), and elsewhere (Yu, 2011).

This line of inquiry tends to examine either perceptions of advertising generally or perceptions of advertising when delivered across specific channels. The later include studying public attitudes about advertisements on the radio (Sayre, 1939), television (Derbaix & Pecheux, 2003), and the internet (Schlosser, Shavitt, & Kanfer, 1999), among other media. Further, as society-level attitudes are not static and thought to evolve over time, the constant need for updating this understanding of public attitude accompanies new generations of people and emerging media technologies.

Examinations of public perception of advertising typically serve a range of interests, too. As consumers' perceptions of advertising are thought to impact subsequent attitudes toward specific products and services advertised (Mittal, 1994), practitioners benefit from understanding

how ads may be pre-judged and how ads are received irrespective of the ad content. Other motivations are to inform policymaking and to more efficiently and effectively regulate the advertising industry itself.

With some exceptions (e.g., Shavitt, Lowrey, & Haefner, 1998), public sentiment towards advertising tends to be more negative than positive in many parts of the world, with consumers wary of the practice of advertising itself. The picture is muddled, however, as in addition to attitudes being more negative than positive they also tend to be highly mixed. For example, consumers express both an appreciation for the utility of advertising, which provides a means of obtaining information about products and services of interest, while simultaneously voicing a distaste for the perceived moral or cultural degradation also often associated with contemporary advertising (Pollay & Mittal, 1993). Consumers have been known to assign negative personifications to advertising in general, for instance, envisioning the practice as a *con-man*, *seducer*, or *evil therapist* (Coulter, Zaltman, & Coulter, 2001). Analyses attempting to understand public distaste for advertising have looked to demographic (Alwitt & Prabhaker, 1994) and psychographic (Dutta-Bergman, 2006) explanations, though these potential explanatory factors are not found in a consistent pattern. In addition to common negative sentiments expressed toward advertising more specifically, prior work has shown that overall consumers tend to be distrustful of advertising as well (Boush, Friestad, & Rose, 1994; Soh, Reid, & King, 2007; Soh, Reid, & King, 2009). Complicating this understanding, public perceptions of advertising are multi-dimensional comprised of attitudes toward the institution of advertising along with toward the specific instruments used by advertisers, for instance, when using certain media or persuasive techniques (e.g., Muehling, 1987; Andrews, 1989). This multi-dimensionality, combined with a conflicting perception that advertising is at times very useful

and at times harmful or dangerous, provides a rich albeit slippery means through which to further examine the contemporary everyday experience of advertising.

In the online context, negative attitudes toward online advertising appear even more pronounced than those toward advertising that is delivered on older media channels (i.e., newspapers, radio, television). The many ways consumers seek to avoid online advertising also indicates a degree of displeasure felt towards internet advertising that outpaces avoidance of advertising delivered on other forms of media (e.g., Seyedghorban, Tahernejad, & Matanda 2015). At the same time, perceptions of advertising delivered on mobile devices and within video games have produced mixed reactions, with consumers turned off by unsolicited ads delivered on mobile phones, such as those received via text message (Peters, Amato, & Hollenbeck, 2007), though largely apathetic rather than angry towards in-game advertisements (Lorenzon & Russell, 2012). Consumers also express even greater privacy concerns over targeted advertising on mobile devices than that on non-mobile devices, in some cases advocating for tighter restrictions on advertising delivered on mobile phones (Okazaki & Hirose, 2009). Finally, personalized online advertising can conjure a unique form of resistance by consumers due to its perceived privacy violations mixed with the perceived relevance personalization affords (e.g., Gigya, 2015), thus tapping similar yet distinct factors from those underlying attitudes toward advertising messages delivered en masse.

### **Mental Models**

One strategy for understanding how individuals think about a multi-part system like the one supporting personalized online advertising is to explore the mental models people develop in response to using this system. A mental model is a representation of a system or process inside a

person's head explaining how it works in the so-called real world, or simply an abstraction of something real. Importantly, mental models do not necessarily represent how something actually works, only how a person believes it works (McDaniel, 2003). This distinction is paramount when considering and probing the impact of these abstractions. Similarly, a mental model need not be very detailed, complex, nor correct in order to be highly effective, as its purpose is to reduce complexity and thereby allow individuals to better comprehend and take advantage of how a process or system functions. In this regard mental models can be best understood as tools for reduction that intentionally suppress details rather than seeking to catalogue each and every one.

Early efforts to understand end-users' mental models in human-computer interaction is seen in the work of Norman (1983), who describes the mental model as a naturally evolving understanding reflecting one's beliefs about a system, acquired through observation, instruction, and/or inferences, and which enables its user to predict the operation of the system (p. 7, 12-13). Prediction is often a key feature of a mental model, upheld by some as the single criterion of an effective mental model. Norman also stresses that mental models are always incomplete, unstable, and unscientific and exhibit porous boundaries (1983, p. 8). Generally, mental models are thought to play important roles in the ways humans interact with the world. This is despite being relatively difficult to articulate or document in most cases.

### ***The Role of Mental Models***

The role of mental models in problem-solving tasks is now a well-studied activity. For instance, instructional techniques that rely on explaining and depicting how an unfamiliar device works using visual diagrams have been shown to facilitate device operation much more

efficiently than techniques that rely solely on repeated practice (Kieras & Bovair, 1984; Fein, Olson, & Olson, 1993). In the case of using diagrams to influence mental model construction, simplified visual representations, such as those of the human circulatory system (Butcher, 2006) have been shown to be more effective at learning than more complex representations, with diagrams in general improving a person's ability to connect and comprehend complex concepts (Fiore, Cuevas, & Oser, 2003). While much of the work on knowledge- and skills-transfer investigating mental model development has focused on structural or spatial mental models, such as describing how a device works or how electricity gets from a power plant to one's toaster, others have demonstrated the vital role played by mental models in temporal reasoning as well (Schaeken, Johnson-Laird, & d'Ydewalle, 1996).

Mental models tend to differ among people considerably as well. Prior research on mental models has revealed domain experts tend to possess more informative and effective mental models than non-experts. For example, experienced pilots were found to possess more effective mental models of invisible flight control procedures compared to the models possessed by novice pilots, which also lead to superior performance in a flight simulator compared to those with less understanding of these invisible processes (Bellenkes, Wickens, & Kramer, 1997). Similarly, scientists tend to exhibit higher-order mental models compared to non-scientists, a difference shown to directly influence applied reasoning and approaches to problem solving about complexity and/or invisible processes (Jacobson, 2000). Similar links between mental models and differing task performance between experts and non-experts has been shown in other domains (Acton, Johnson, & Goldsmith, 1994; Hsu, 2006; Al-Diban, 2008).

That experts would exhibit more information-rich and arguably more technically-accurate mental models when compared to non-experts is of course unsurprising. Similarly, it is logical

that an expert in a given process or system would then be able to invoke these more effective mental models and perform better on measured tasks compared to non-experts. For researchers, locating these distinctions between the mental models and subsequent performance levels of experts and novices can further explain the role played by these abstractions in learning about and understanding invisible processes and systems. It also sheds light on some of the ways in which the structural, ordinal, and/or temporal organization of a person's ideas can impact decision-making (Hester et al., 2012; Mumford et al., 2012; Barrett et al., 2013).

### ***Models in Human-Computer Interaction***

In studies of human-computer interaction, there is enduring interest in analyzing the role of mental models and assessing how end users' understanding of computing systems (e.g., software, hardware architecture) are formed and how mental models function in everyday use. This includes, for example, investigating the role of ordinary people's mental models of information retrieval systems (Borgman, 1986), how hypertext works (Marchionini & Schneiderman, 1988; Gray, 1990), locating information online (Brandt & Uden, 2003; Zhang, 2008; Dinet & Kitajima, 2011), and how the internet works (Thatcher & Greyling, 1998).

For developers of human-computer interaction systems, mental models have also been shown to have great influence on the strategies of computer programmers in software development (Storey, Fracchia, & Müller, 1999), highlighting the role of mental models on both ends of a system. That is, in its initial design and later during its use. In general, the line of research into the mental models of system developers ultimately aims to improve usability for end users. In the current study the ends user is simply an internet user who is a de facto consumer of personalized online advertisements.

Human-computer interaction researchers and designers have gone to great lengths to understand mental models. This is because it is generally accepted in this domain that ill-conceived or unnecessarily complex mental models developed by end users may inhibit system use and, further, that optimizing users' mental models minimizes barriers to the user experience. This notion of mental model optimization for end users involves encouraging models that are minimally complex, or only as complex as absolutely necessary to facilitate system prediction and use. This simple model strategy favors reduction whenever additional details in one's mental model do not directly enhance the use of the corresponding system. Principally, this upholds the underlying goal of mental models—to reduce complexity. Importantly for this concept as it relates to this study, all individuals who interact with a system will develop, and over time refine, their mental models of that system. This process is automatic and unavoidable. This is true both for the designers of a system and the users of that system, though the current study focuses on end users, or consumers of online ads.

Finally, another type of mental model, distinct from that of the designer(s) and end user, is the mental model that designers/engineers *anticipate* users will develop to aid in using a tool or system. This is mental model once removed, so to speak, and might be best understood as the mental model developers project onto system users, projections that have been shown to have substantial implications for designers themselves and the artifacts that result from their efforts (Agre, 1995; Bardini & Horvath, 1995).

Key to each of these types of models is the idea that end-users' mental models need not be the same as those of system experts, which logically tend to better reflect how a system functions, nor the same as the models experts project onto end users. And this idea of a model's accuracy differs from a model's effectiveness. For instance, investigation into users' mental

models has shown that a “wrong” mental model can be just as effective in helping someone operate a system as a “correct” mental model (e.g., Kempton, 1986).

### ***Model Accuracy***

Finally, the notion of a mental model’s correctness highlights a crucial point for any investigation examining end users’ mental models. The accuracy of one’s mental model of anything is largely irrelevant. For instance, an end-user might have a much simpler mental model of system than its designer simply because that is what is most effective for each person. Yet, each of these individuals’ mental models would be wrong in very important ways and neither can be fairly described in terms of their “accuracy” but rather only their effectiveness. Thus, attempting to evaluate a person’s mental model on the basis of how similar it is to some perfect, ideal model misses the point. As the statistician George Box (1979) offered, "All models are wrong but some are useful." Accordingly, mental model effectiveness often hinges on relative utility including ability to predict, not accuracy, which is largely immeasurable and, in “situated action” often irrelevant (Suchman, 1987; 2007). Thus, a mental model of a system that strays very far from how that system actually functions (e.g., technically, physically, organizationally) can be quite useful, regardless of how it goes about eliminating complexity.

Extending these theories and concerns into advertising flows somewhat naturally, given the growing synergy between digital advertising and consumers interaction with personal computing devices. In this case, the end user is simply the consumer of an advertising message, whereas those behind the curtain in human-computer interaction, the “designer,” could be any number of players including the marketing practitioner or a brand needing to advertise its products or services. Both of these “designers” of advertising interaction may benefit from



considering the unseen mental models consumers develop in response to advertising technologies and practices.

## **METHODS**

### **Study Design**

To explore how consumers think about personalized online advertising, six one-hour focus groups were conducted with U.S. and U.K. adult participants in 2015 and 2016. Four of the group interviews were held in Chicago and two in London. I moderated all focus groups myself. Additionally, for two of the groups, a paid research assistant was also present who took notes and occasionally interjected follow-up questions of participants. The focus groups ranged in size from five to eight people and there were a total of 38 participants across the six groups. The University of Michigan Institutional Review Board approved this study and all participants signed informed consent forms prior to participating. At the conclusion of each session individuals were paid \$60-\$70 in cash (depending on the location) in exchange for their participation.

The decision to collect data through focus group interviews (e.g., as opposed to one-on-one interviews) was motivated by the technical nature of personalized online advertising and the relatively opaque processes that combine to support this everyday practice, such as third-party consumer data collection, cookie matching, and real-time bidding operations. Therefore, by allowing participants to build conversations off of one another's previous stated experiences and ideas, rather than in the isolation of a single interviewer-interviewee setting, participants' comments helped others acclimate to the topic more quickly rather than relying solely on interviewer prompts. Footholds for entering the conversation at times emerged organically. In

this way, the group discussions allowed participants to conjure, compare, and contrast their own experiences with others out loud, something not afforded by isolated one-on-one interviews. Compared to a series of individual in-depth interviews piloted for this study, the focus group atmosphere elicited a richer discussion and a wider set of ideas due to interactivity. At the same time, the usual shortcomings of focus groups limited what could be learned during these sessions, primarily that not all participants are comfortable expressing divergent opinions amongst a group of complete strangers and some individuals are less articulate in general when made to speak in front of a group compared to one-on-one conversation.

The focus groups consisted of several activities including a word association exercise where participants generated lists of terms they associate with personalized online advertising, a drawing exercise where participants sketched a picture depicting how online ad personalization works, and a larger oral discussion between participants about advertising personalization. There was no attempt to analyze group dynamics or the nuances of participant interactions. Participants were encouraged to discuss their ideas and opinions with one another and not just the moderator, which they often did. However, the particular group dynamics as they relate to the questions posed to participants were outside the interests of this study and these participant interactions were not examined. In this way, the sessions served as group interviews to study individual beliefs, but not conversation dynamics as they relate to personalized advertising.

Additionally, in pilot one-on-one interviews conducted for this study, it became clear that participants needed to be acutely aware of the scope of the topic prior to discussion. As people's experiences with advertising tend to occur across many channels (e.g., billboards, print, television, internet), along with the increasing convergence of integrated marketing communications on these channels, focusing group members on personalized online advertising

required a brief introduction and set of examples aimed to ensure participants were on the same page about the topic. To do so, at the beginning of the focus group participants were told the topic of the discussion was “personalized online advertising,” which was equated with “targeted online advertising,” as this concept is also commonly referred to. These two terms (*personalized* and *targeted*) were used interchangeably by participants, though effort was made to steer the conversation towards “personalization” in online advertising rather “targeting,” as the latter has been shown to have negative connotation. Participants were informed that the discussion was only concerned with their experiences with ads on website and apps that they believe had been personalized in some way for them.

Most groups followed the same basic sequence: a moderator introduction of the topic including presentation of a few visual examples of personalized online advertisements, a word association exercise, a drawing exercise, and a broader discussion. At the conclusion of these activities a secondary follow-up discussion proceeded about how consumer data is used. The results from this discussion will appear in a separate research report.

The rationale for expanding the focus groups beyond oral discussion, adding the word association and drawing tasks, stems from prior work which demonstrated that administering supplementary tasks before, during, or at the conclusion of focus groups can enhance primary data collection efforts expanding what can be learned from the verbal component of a session (e.g., Cresswell, 2002; Yoder & Lopez, 2013; Fahey, Verstraten, & Meyers, 2014). Overall, all participant activities and question ordering was intended to draw out individuals’ ideas by providing minimal moderator explanations especially early on in the discussions.

Professional transcriptionists in Chicago and London transcribed the focus groups verbatim. Transcriptions of the six focus groups served as the primary source of data along with

lists from the word association exercise and the set of drawings created by participants. First, prior to analyzing the transcription text, video and audio recordings from all focus groups were reviewed in full during long undisturbed periods. As the focus groups were held across two years time, this provided a refreshment step prior to analyzing the transcripts, drawings, and other notes. This review also helped to put the different groups' ideas into conversation with one another. Then, all transcribed text was manually unitized separating individual ideas that were spoken consecutively during longer responses from participants (Morgan, 1997, p.120). Once unitized, a first pass of coding proceeded. Rather than imposing predefined themes, an emergent coding scheme was used with labels, categories, and themes growing directly out of manifest content (Creswell, 2007; Berkowitz, 1997). Themes were then revised, combined, and condensed. This was followed by a second round of reviewing and in some cases re-coding the unitized text under the revised coding scheme. During this second and final round of coding, repeated themes were noted to distinguish the most salient ideas expressed by participants. Following final coding of the transcription text, each of the audio-video recordings were reviewed in full once again in effort to confirm, reject, and otherwise update the themes generated based on unitized text. Manual unitization of transcripts, all coding, and theme generation was performed using the NVivo (v.11) software platform (qsrinternational.com).

In addition to analyzing transcripts, all word association lists were later analyzed by manually grouping individuals' lists as one of the following: negative, neutral, conflicted, and other. This cursory analysis provided insights into participants' initial reactions prior to the more substantive group discussion.

For the participant drawings, these were analyzed and coded as well. An initial pass of analysis simply included displaying all 38 drawings laid out next to one another to scan for

visual similarities and differences. After spending extended time with the drawings laid out next to one another on multiple occasions, a closer analysis was performed of the various symbols, connections, and text annotations. Similarities and differences in structure, symbols/items, activities, players, and interests, were noted. Additionally, key absences in the drawings of different components that support personalized online advertising were noted.

Overall, themes and conclusions were generated through an iterative data analysis process (Bogdan & Biklin, 1998), taking into account the video recordings and transcribed dialogue of the focus groups along with the word associations and drawings generated by participations. As is common in exploratory qualitative studies, an emergent and interpretive framework was used in this analysis, arriving at relevant understandings, explanations, or theories rather than attempting to test a priori hypotheses (Lofland, Snow, Anderson, & Lofland, 2005).

## **Participants**

Local market research firms in London and Chicago recruited participants from the general adult population. In both cities, the one-hour sessions were held at professional focus group research facilities. Using market research firms to recruit and provide space for conducting focus groups has been successful in similar studies that incorporate focus groups in multiple locations (e.g., Humphreys & Wilken, 2014). For recruitment, potential participants had to be age 18 or older and fluent in conversational English. Additionally, potential participants were eliminated if they reported ever working in the media industry (including marketing, advertising, journalism, or market research). Finally, potential recruits who reported participating in any kind of focus group in the previous two years were also eliminated in effort to avoid any potential

“professional” research subjects. While no strict demographic quotas were used, the individual groups were filled by balancing gender, age, education, and income among successful recruits in efforts to make groups as diverse as possible. The resulting focus groups were relatively diverse along these criteria (e.g., sessions were balanced among gender, age, income, etc.).

Among the participants, 18 were female and 20 were male. The number of participants in each of the following ranges was as follows, age 18-25: 5, 25-34: 11, 35-44: 8, 45-54: 9, 55-74: 5. The income for participants (UK converted to USD) was <\$25K: 6, \$25-50K: 15, \$50-75K: 9, \$75-100K: 6, >\$100K: 2. Participants’ highest level of education completed (UK converted to US equivalent) was high school diploma: 10, some college: 7, associate’s degree or vocational training: 5, four-year college degree: 11, post-graduate degree: 5. The participants’ race/ethnicity was as follows, White: 21, Black: 8, Hispanic/Latino: 3, Asian: 3, Other race/ethnicity: 3.

Participants were not aware of the topic of the session until arriving at the focus group facility. The participant “no-show” rate was relatively low with 38 of the 40 (95%) recruited participants attending. One participant arrived seven minutes late and was allowed to participate. Participants were provided with light refreshments during the group interviews and were not permitted to leave the room. The execution of the groups was without disruption and, likely due to the relatively non-sensitive nature of the topic, participants were fairly talkative and conversational. No participants exited the focus groups prematurely. Though one participant expressed feeling moderately distressed immediately afterwards after learning about types of consumer data collection described by other participants that she found to be invasive. All focus groups were audio and video recorded. For several of the groups, one or more participants stayed after their session was over to ask questions, voluntarily electing to continue discussing the topic. These post-session, ad-hoc conversations were not recorded and are not included in this analysis.

## **Educative Authenticity**

For the last five minutes of each focus group, participants were asked to complete a brief, four-question feedback form asking them to rate the quality of the session, how to improve the session, and their comfort level. Participants were encouraged not to identify themselves on their forms. All but one participant reported they felt completely comfortable during the session. On the quality of the group, all participants rated the group as being either “excellent” or “very good” (response options: *excellent, very good, good, fair, poor*).

The feedback form also included the open-ended question, “Do you feel you learned anything during the focus group? If so, would you tell us one or two things you learned about that were not familiar to you prior to the focus group?” This provided a way to try to gauge so-called “educative authenticity” (Means Coleman, 2000, p. 278), where research participants are exposed to ideas or perspectives different from their own and learn something new as a result of participating in academic research. Lincoln and Guba (2000) describe educative authenticity as a secondary research goal for academic researchers where participants gain a “raised level of awareness” (p. 180) about the research topic, thereby benefiting from their participation beyond any monetary incentives paid. For this question on the feedback form, nearly all participants reported multiple items they learned during the focus group. Most responses related to gaining an improved understanding of how personal data is collected and how online advertisements are personalized. The focus groups appear to have been educational for the majority of participants.

## **DATA & ANALYSIS**

Following a word association exercise and drawing exercise, participants engaged in a discussion based on a set of questions. This discussion was moderator-led using a “funnel-based”

focus group interview. This technique begins with less structured discussion and proceeds to more specific probing of focus group participants (Morgan, 1997, p. 41). The list of moderator entry-point questions appears in online Appendix A. This group discussion occupied the bulk of the focus group time and broader themes emerging from the focus group discussions, including mental models observed to be in use by participants, are described later in the Findings section below. Additionally, as participants also engaged in a brief word association task and created a drawing, these data are first detailed before presenting broader findings.

### **Word Associations Task**

Following a brief introduction of the topic, participants were first asked to think about their recent encounters with online advertising on websites and apps and to specifically think about experiences where they may have received advertisements that were personalized or targeted for them online. They were then instructed to write down five words that came to mind when they think about “personalized online advertising.” After a moment for reflection and time to write several words, participants were asked to share a few of the words they wrote down and why. The words participants generated appear in Appendix B.

### ***Negative Sentiment***

The most common sentiment expressed in word associations in response to the phrase “personalized online advertising” was negative, captured in various valances. For instance, one participant with negative associations wrote: *fraud, hidden persuasion, seduction, dishonesty, deletion*. Two others listing primarily negative words included the lists: *presumptuous, predatory, sloppy algorithm, targeted and stalking, too much frequency, everywhere, annoying,*



*competitive*, respectively. Additionally, two others wrote: *annoying, interruption, distraction, irrelevant, sales pitch, time waster* and *repetitive, annoying, money grab, invasive, disingenuous, smart*, respectively.

This generally negative attitude towards personalized (or targeted) online advertising was unsurprising as skepticism and negative attitudes toward advertising in general are well documented in the literature (Pollay & Mittal, 1993; Coulter, Zaltman, & Coulter, 2001; Alwitt & Prabhaker, 1994; Dutta-Bergman, 2006; Boush, Friestad, & Rose, 1994; Soh, Reid, & King, 2007; Soh, Reid, & King, 2009). Though notably, most of the completely negative lists have little to do with the “personalized” component and appear to reveal negative feelings towards either advertising in general (*seduction, sales pitch*) or perhaps towards online advertising generally (*annoying, interruption*) but not necessarily personalization or targeting. From this blurring we might conclude, logically, that participants were not thinking about personalized online advertising as a substantively different activity than either advertising in general or perhaps (non-personalized) online or digital advertising. It is logical that these concentric advertising processes would stimulate similar associations with broader parent concepts including reactions to advertising in general.

### ***Neutral Sentiment***

Less common than negative sentiment words were participant lists containing neutral and in some cases positive associations with personalized online advertising. For example, a participant with a neutral word association list wrote: *sports, interests, economics, family, politics*. Another participant included: *pics, words, wording, innovation*.

Again, given negative sentiment typically associated with advertising in general, a somewhat global distaste for online “pop-up” ads, and reported privacy concerns by many in the U.S. (Madden, 2014) and U.K. (TRUSTe, 2015), the lack of overtly positive words in participants’ lists reflects prior attitudinal measures on this topic. For most consumers the notion of personalized (or targeted) online advertising does not conjure blissful feelings; this was the case for most participants across all focus groups.

### ***Conflicted Reactions***

While simplistic, this exercise captured one of the key tensions repeatedly expressed by participants in later dialogue and discussed in further detail later: participants’ frustration at both “aggressive” consumer data collection practices and at online ads they perceived as targeted that are irrelevant to their interests. This sentiment is somewhat contradictory, which many have described as the “personalization-privacy paradox” (Awad & Krishnan, 2006), the benefits for consumers of personalized content appear in contradiction with the information required for personalization and the privacy invasions some associate with relinquishing this information.

This conflicted position of the participants on the topic of personalized (or targeted) online advertising is seen in individual participant’s words that, in the context of online advertising, included positive and negative sentiment words in the same list. For instance, two participants wrote: *inundated, useful, helpful, scary, indifferent* and *privacy, selling, interest, smart, scary*, respectively. Two other participants with somewhat conflicting reactions listed: *annoying, in the way, not needed, too many, convenient* and *inspired, economic, convenient, persistent, annoying*, respectively.

Though the word association lists were overall more negative than positive, two important dynamics illustrate the heterogeneous sentiment many participants expressed towards personalized online advertising. First, not all lists were composed of all negative words and not all lists were negative; yet some lists were completely negative and others completely positive. This alone signals these different consumers' diverse associations and reactions to personalized online advertising. Second are the individual participants who themselves, within a short five word list, included both clearly positive and clearly negative sentiment words (*helpful* and *scary*). This points to a different diversity of thought, or conflictedness, at the individual level, signaling that, consciously or unconsciously, some participants were considering the tradeoffs of personalized advertising as they described the process. Further, even many of the mostly negative or mostly positive sentiment word lists included at least one contradicting or indifferent word, revealing confliction among most of the participants' initial reactions to the concept of personalized online advertising. Overall, the responses to this task point to a more mixed or nuanced perception of personalized advertising than purely positive or purely negative associations with this concept.

### **Drawing Task**

Also early on in the focus groups, following the word association task and a brief warm-up discussion, participants were asked to draw a picture depicting "how personalized online advertising works" on a single sheet of paper. Positioning this exercise prior to more in-depth discussions about how participants believed this process works was intended to minimize the degree to which participants' comments influenced one another's drawings. They were given five minutes to create their drawing and told they should be prepared to discuss it aloud with the

group afterwards. Not all participants elected to show their drawings to the others, but most participants described them aloud even if not revealing their sketches visually.

Previous work aiming to solicit and examine individuals' mental models has also relied on simply asking participants to make drawings describing how a given process and/or system works. For instance, to investigate how individuals think about and navigate hypertext in online environments, Gray (1990) asked participants to draw pictures of how hypertext works and later made conclusions from these drawings. Further, this draw-me-a-picture method has been used in a number of studies to investigate how individuals think about systems and processes which they cannot see, such as how electricity is generated and gets into an appliance in one's home (Devine-Wright & Devine-Wright, 2009). More closely related to the current study, this drawing technique has been used to study individuals' understandings of how systems inherent to personalized online advertising work, including computers in general (Denham, 1993), databases (Kerr, 1990), and the internet (Thatcher & Greyling 1998; Zhang, 2008; Dinet & Kitajima, 2011). Notably, this rendering technique is imperfect. Ad-hoc drawings of a system or process may differ from how the creators of these drawings actually think of and approach that same system or process. Hence, it remains a lossy method of extracting information as individuals may not be readily aware of the mental models they use. The challenge of accurately depicting and documenting individuals' mental models is well established (e.g., Norman, 1983). Yet, and as prior studies have shown, this method is promising for gaining a glimpse into the "pictures inside people's heads" so to speak. Nevertheless, this activity was used to gain insights into consumers' mental models and how they think about how online content personalization is achieved.

The research goal of this sketching activity was to get a glimpse of participants' mental models describing how personalized online advertising works, albeit a fuzzy or impartial

snapshot. Thereby illustrating to some degree how consumers might be thinking about this highly technical and somewhat opaque process of online ad personalization. Participants were given five minutes to create their drawings. They were told in advance they would need to discuss their completed drawings aloud with the group and assured they were not being evaluated on their artistic merits, but to do their best in drawing how advertising personalization works. Still, when discussing their drawings a common opening statement by participants was to remark about how “bad” or “wrong” their drawing was before proceeding to discuss it with the group.

When analyzing drawings, similarities and differences in drawing structure, the distinct symbols/items included (referred to here as “nodes”), activities, and players were identified. Additionally notably absences, or simply what was not present in the sketches that perhaps could have been given the many ways in which online advertising personalization occurs, were also considered. The following section provides an overview of these commonalities. For example purposes, four participants’ drawings appear in Fig. 4.1 below. The complete set of all the participants drawings appears in Appendix C.

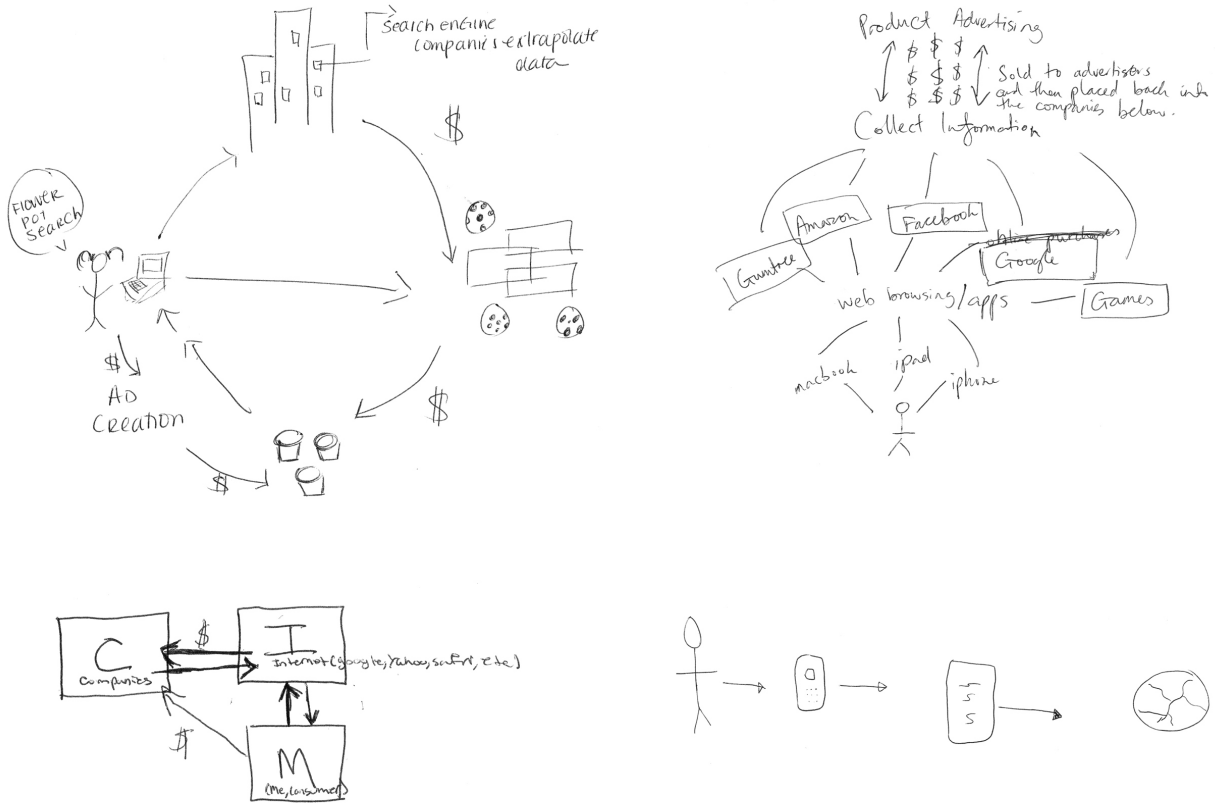


Figure 4.1. Example drawings of personalized online advertising created by participants.

### Structure

The majority of participants provided a drawing resembling a common flowchart-style or network diagram. These basic drawings were typically composed of distinct items (or *nodes*) contained by boxes or circles and connected to one another by lines (or *edges*). Among these common node-edge diagrams, three styles or subtypes were observed each similar yet different in important ways. These included a basic linear model, a linear cycle model, and an interconnected model.

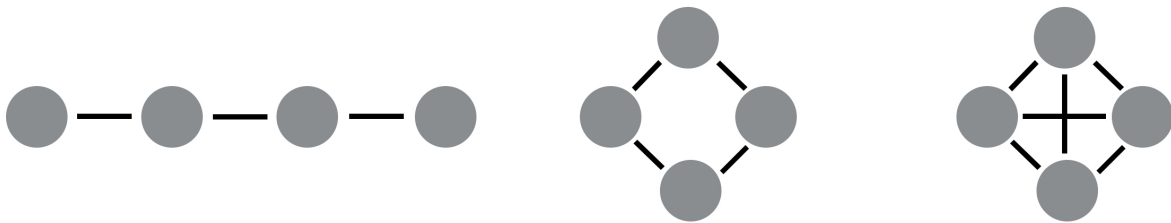


Figure 4.2. Three styles of network diagrams drawn by participants.

**Basic Linear Model.** In this case, the representations depicted personalized online advertising through a very rudimentary linear process. This style of drawing used lines or arrows to show progression from an origin point to future nodes in the network to typically the final node, the personalized ad itself.

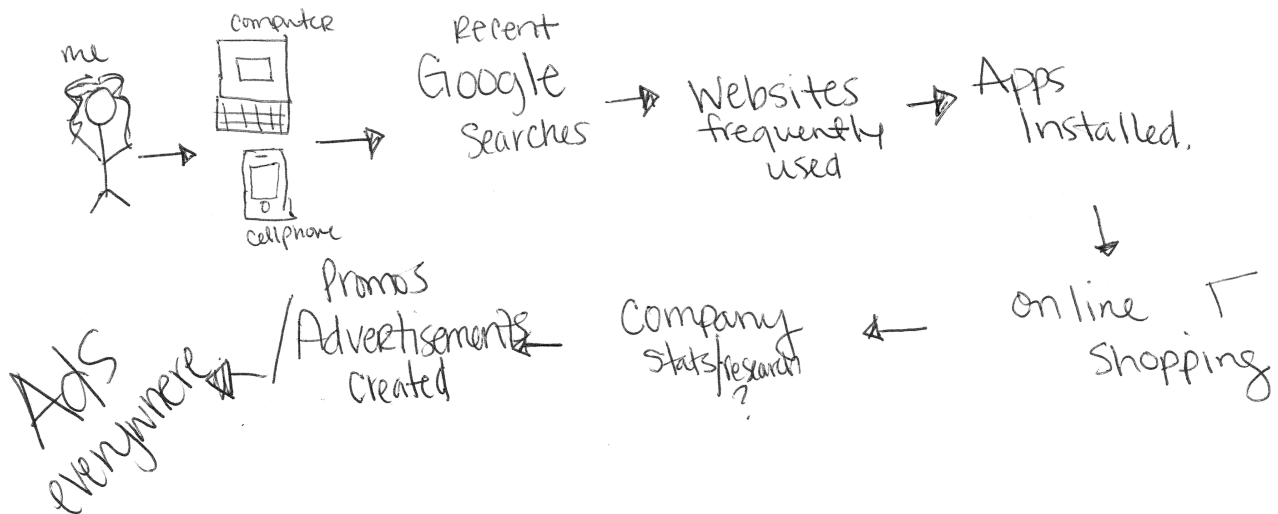


Figure 4.3. Example of a basic linear model drawn by a participant.

**Linear Cycle Model.** In addition to the most basic linear process diagram, some participants drew cyclical process models, resembling more iterative or repetitive processes. In these cases the models still tended to have a beginning and end, but with additional lines and arrows

indicating the process returned to its original starting point. This notion of a linear cycle illustrates how some participants considered the ongoing nature of personalization processes through a distinct set of events in sequence, yet cycling back on itself as the process repeats.

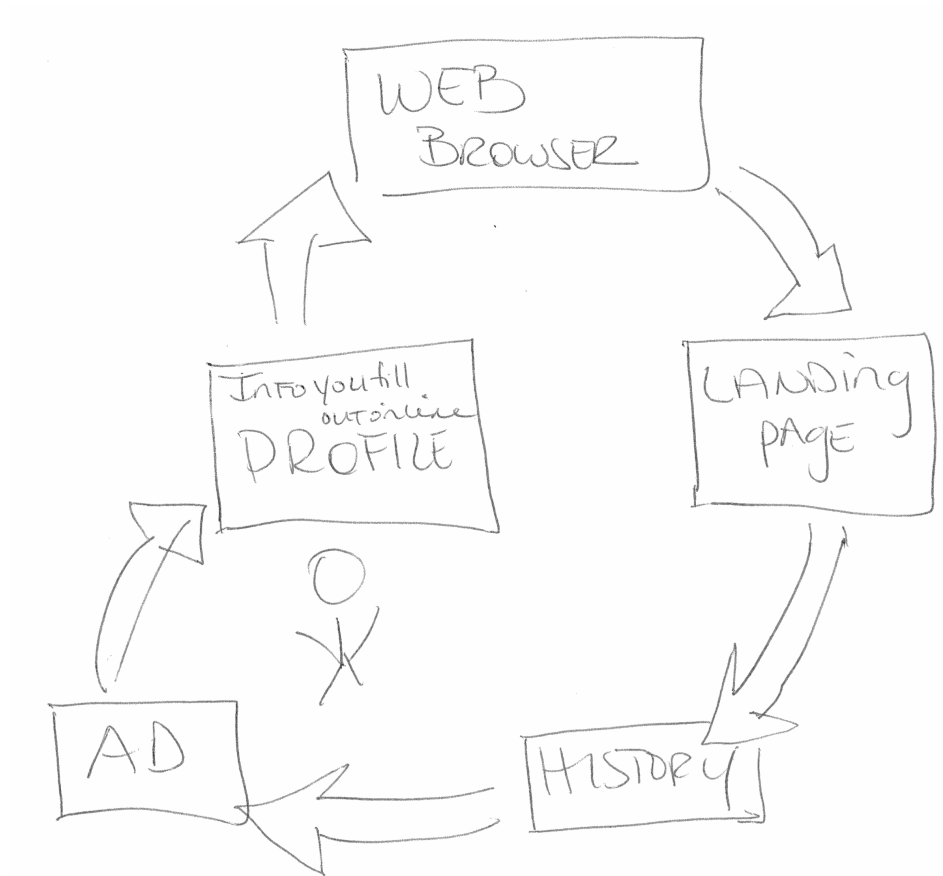


Figure 4.4. Example of a linear cycle model drawn by a participant.

**Interconnected Model.** The third type of network diagram depicted an interconnected model.

As it sounds, the interconnected model contains many nodes connected to multiple other nodes, with a web of connections appearing in place of a single sequence of one node to the next.

Interconnected model drawings corresponded to some of the most elaborate depictions of the ad



personalization process. For instance, some participant drawings with interconnected nodes in effect described what marketers refer to as “cross-device tracking” where a single consumer is identified across multiple devices, such as her laptop, tablet, and smartphone, in order to collect more accurate consumer data for ad personalization (Díaz-Morales, 2015).

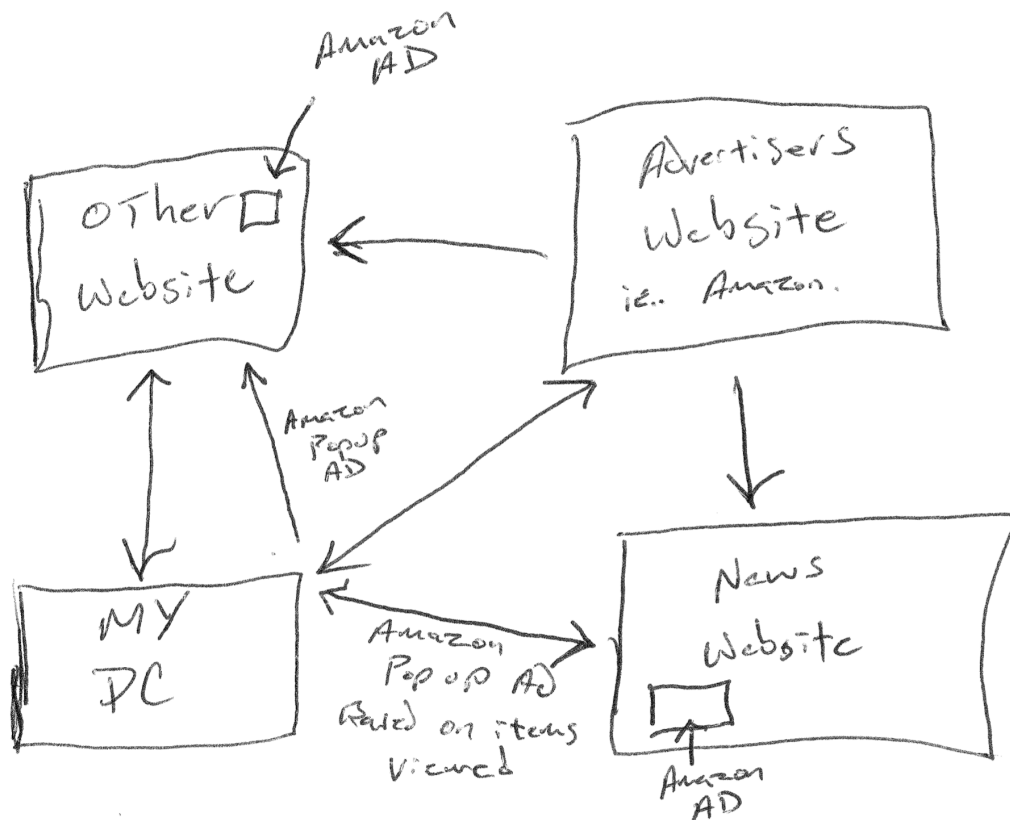


Figure 4.5. Example of an interconnected model drawn by a participant.

### Nodes

With the majority of participants across all focus groups sketching and labeling something resembling a network diagram, with the labeled components often connected by lines. Additionally, many of the nodes connected by lines were also similarly labeled, pointing towards

further consistency among participants' mental models. Among various types of nodes technical artifacts were common in the drawings. Most often individuals drew one or more personal computing devices as the point of origin in their drawing, especially for those sketching a linear process model. These nodes were labeled *computer, laptop, PC, tablet, phone, smartphone*, which provided a common origin for participants. Other technical nodes appearing included those labeled as *search engine, database, website, web browser, network, server, script, app, cookies*, and *search/browsing history*, among others.

Participants often drew lines connecting the nodes in their sketches, including lines with and without arrowheads. Arrowed lines signal how some participants perceived a directional flow of activities, from one to the next, shaping the ad personalization process. Some direction was one-way, with arrows appearing on one end and still others connected items using two-way arrows, indicating processes flowing in multiple directions, often back-and-forth between nodes. This two-way flow indicating more cyclical or iterative processes, was less common with most connecting nodes revealing a basic ordering of events.

### ***Players***

Among nodes appearing in participants' models, many corresponded to specific people or organizations that play a role in the process of generating and personalizing ads. Key players commonly appearing in drawings included the participants themselves, companies advertising a product or service, internet technology firms including providers of search engines and social media platforms, companies specializing in (online) personal data collection, and generic conceptions identified, for instance, simply as *marketers* or *advertisers*.

Among the commercial players included, it was common for participants to label individual nodes as corresponding to specific firms. A small group of named companies appeared repeatedly in participants' drawings suggesting the dominance of these firms in participants' thinking about personalized online advertising and the way some consumer facing firms with substantial market share stand in as synonymous with their category in the system (i.e., "Facebook" = social media, "Amazon" = ecommerce). Companies appearing repeatedly in participants' drawings included Amazon, Apple (iPhone, iPad, etc.), Facebook, Google, Twitter, and Yahoo. Additionally, though unnamed, several additional players were included in individuals' drawings. These included nodes labeled as *advertising gurus*, *marketing companies*, *media minions*, and *retail outlets*, for example.

### ***Activities***

The various symbols, artifacts, nodes, connections, and other items appearing on participants' drawings typically described a set of discrete activities leading up to the delivery of a personalized advertisement. These events were most often linear and sequential in their depiction. In addition to the popular decision to draw an end point of the personalized ad delivery, the most common activity to depict was some type of deliberate action on the part of an internet user that set the process in motion. Among these, most included the act of using a search engine, which typically served as the origin of the entire process. For instance, participants depicted and labeled activities and nodes in their drawings as "search products online," "what I search," "flower pot search," "recent Google searches," "computer/app searching," and so on. This reliance on consumer initiated activities, such as using a search engine, to impact the ad

personalization process demonstrates the active role individuals may feel they have in instigating this type of tailored advertising.

### ***Absences***

While trying to accurately discern what is *not present* is perhaps a fool's errand, two related activities (or nodes) were conspicuously absent from participants' drawings. The first is any mention of actions or data resulting from activities taking place offline, events that marketers routinely use to personalized online advertisements for consumers. This offline-originating data includes, for instance, that generated by purchases in brick and mortar stores and captured by point-of-sale systems, rewards cards, and credit card companies, all of which are commonly used in segmenting consumers and personalizing online advertisements. Instead, participants commonly referenced their online behaviors and the presences of "databases" such as those depicted as connected with lines from a laptop or smartphone in their drawing. Yet participants appear to be thinking these databases and the data they contain are relegated to consumers' online activities. While uncommon, a few exceptions to online activity data included single references to a bank account, library books, age, and gender.

The second notable absence is more process-oriented: most participants drew process diagrams beginning with a consumer taking a deliberate and online action (e.g., performing an internet search), which later determined the content of the tailored ad this person received. This was almost universal across the drawings. However, this neglects a substantial proportion of online advertisements personalized for consumers. As most personalization (or targeting) is not the product of retargeting, which makes up only a fraction of all attempts to personalize online ads using consumer data. On this absence, notably participants' drawings commonly revealed

some familiarity with social media platforms and that they play a role in the process of online ad personalization. However, it appears the role of social media platforms, for instance, was not perceived by participants as providing the personalization and delivery of *uninitiated* tailored ads as these platforms tend to do based on conclusions made about users from other online behavior (e.g., likes, comments, sharing posts) as well as pairings with offline data. As noted, while accuracy of an individual's mental model is neither the a requirement nor a goal to be pursued, these two absences highlight ways in which consumers may not be thinking about online personalization processes, in light of how these processes sometimes work. (These two absences are further discussed below in the notion of an *Online Only World* and a *Self-Catalyst Mental Model*.)

## FINDINGS

A set of themes, metaphors, personifications, and broader mental models emerged from participants' responses. These findings are discussed in detail below. I first discuss six personas emerging from how participants described personalized online advertising practices. Then, I discuss some additional themes observed also emerging from ways individuals discussed ad personalization. Finally, I discuss two mental models participants appear to be relying on when reasoning about online ad personalization—the Self-Catalyst Model and the Eye of Providence Model. Based on the shortcomings of these observed models, I offer a third prescriptive mental model that may provide a more useful way for people to think about online ad personalization. I refer to this proposed model as the Broken Clock Model.

## Personifications Identified

When describing online ad personalization, in some instances participants personified targeted online ads directly: “*They want to get you. They want to grab you*” (P26). Similarly, other times participants described different personalized ad processes as exerting a human-like agency (e.g., utilizing deception), invoking metaphors linked to a number of human behaviors (e.g., getting “hustled” by someone). This personification and use of metaphor revealed a number of insights for how participants reasoned about online ad personalization.

The link between metaphor and mental model development is well established; with individuals at times relying on metaphors to connect knowledge within larger mental models (van der Veer et al., 1990). Other times a metaphor provides the necessary foundation on which a mental model is established (Rupietta, 1990). Further, the selection of appropriate metaphors is thought to be crucial in developing effective mental models of how things work (Carroll & Thomas, 1982), especially because “people develop new cognitive structures by using metaphors to cognitive structures they have already learned” (p. 109). Which metaphors someone invokes in response to a system drastically affects their success when trying to take advantage of it (p. 110). For these reasons, in this study I take seriously the personas and metaphorical understandings held by participants to better understand how they reason about online ad personalization.

Reflecting overall negative attitudes toward advertising in general, nearly all of the ways participants personified and deployed metaphorical language about ad personalization were tied to unwanted behaviors (e.g., oppression, harassment, trickery). Here, I present six personas identified based on how participants described their thinking about online ad personalization.

## Dictator

Some participants described online ad personalization and accompanying data collection practices as though consumers are captives living under the rule of an oppressive dictator:

*“We’re being tracked. We might as well put a thing around our necks and be done with it”* (P5).

Participants described how they feel unable to break free: *“There’s no privacy nowadays anyway, so you do things, like go on Google, type someone’s name and address in, and [a photo of their house] comes up. You can’t escape. So worrying about a bit of privacy? [...] You can’t let it bother you because you can’t escape it. That’s how it is”* (P10). And participants felt they had no choice but to live under this rule of this regime: *“Do you really have a choice? If you want to use Facebook, for example, can you say, ‘I don’t want them to collect that information?’* (P34). Feeling they had little autonomy also influenced some people’s preferences for personalization: *“Given that [marketers are] going to take that information anyway, I would prefer the personalized version [of online ads] because it relates more to me”* (P23).

Participants discussed personalization and consumer data collection as instilling conditions under which they were unhappy but had little recourse. Some considered ways to escape this oppressive rule including the drastic option of disconnecting entirely from digital communications and internet-connected devices: *“You can always disconnect yourself from the internet and get a normal phone [...]. You haven’t got to be part of it. [...] if you’re not happy with the internet or your phone, your iPhone, get a basic phone and just don’t be on the internet”* (P10). Although not using the internet at all is unviable for most people today, leaving this escape route an unlikely option for consumers who discussed data collection and ad personalization with terms describing conditions of an oppressive regime.

## Hustler

I found some participants feel they are being cheated out of money they should be entitled to based on the way marketers and others profit from consumers' attention and actions online: *"I think, from [the marketer's] point of view, it's actually smart. You know, it's a smart hustle. It's a hustle. Yeah, that's what it is. The whole system is a hustle"* (P20). Participants who thought data collection and advertising personalization processes were hustling money away from them, some articulated feeling left out of the financial transactions in this system: *"I feel like it makes money off of your interests without your permission. [...] I'm not getting any money off of it. Because they are making money off of [me] I guess a part of me feels like it's a business operation where I'm involved and I'm not a part. Not a member who has a voice"* (P9).

Similarly, another participant voiced feeling he got "played" by personalized ads, which he found unnecessary: *"I'd rather them not pester me with ads coming up. [...] If I want to pick something and buy something, I'll go and look for it myself. I don't really need [marketers] to keep playing me as I'm looking [online], as I'm maybe just reading up on something"* (P6).

Similarly, some participants exhibited awareness of ways marketers earn money from clicks on personalized ads and selling customer data, leading them to feel cheated out of something that was rightfully theirs: *"But [the data is] about you, so shouldn't you have ownership? I don't want [marketers] to have it [...] so I would say the ownership is mine because it's about me"* (P5). Others had similar expectations but were more forgiving of marketers. Still they felt they should receive some money too: *"I want a cut out of it. [...] Because, at the end of the day, we're giving [marketers] information and they're pocketing. They're selling on our information. [...] Give me a percentage of what I buy online [...] 10% or something. It [should] come back to me if they want my data"* (P6). Though many of these



practices result in free online services, it appears some consumers feel they are caught up in a hustle, an exchange of money, personal information, and time spent viewing personalized ads.

### **Presumptuous Stranger**

One tension articulated by participants was feeling as though personalized online ads presumed they “knew” the participant when in fact they did not *really* know participants. When inferred customer attributes and interests were wrong, and ads were based on this, some participants were offended.

This *you don't know me* sentiment was mainly based on experiences where participants thought ads had been targeted at them and were deemed irrelevant: *“I find it a bit ridiculous because I always get adverts that aren't anything to do with me, but it's my tablet. It's like diet pills. I never diet. [...] it's totally irrelevant when they think it is relevant”* (P38). For others, the practice of retargeting simply contributed to feelings of ads “acting” like they knew them: *“I go into my Yahoo [account] every day, but I also go online and look at sales. And so on the days that I just may look in my Yahoo, Macy's [ads] or something may pop up on the side that I may have looked at yesterday or two days ago. It acts like it knows me.”* (P33)

Other participants were put off by the audacity of tailored ads to suggest any products whatsoever: *“If I want to go shopping for something, I know how the Internet works. I know where Amazon is. I know where eBay is. [...] I don't need a ‘Hey! You probably would like this’ ad. It's presumptuous”* (P12).

Generally, participants expressed dissatisfaction with the inability of marketers or personalized ads to know them to the degree they desired. For some this included wishing marketers knew them less: *“I don't want [marketers] to know me that well”* (P9). Yet others

were disappointed with the superficial nature of the connection: *“It might be targeted but there’s no personal relationship there between us.”* (P1).

## **Stalker**

Participants commonly described feeling “stalked” all over the internet by personalized ads. Some described feeling smothered by tailored ads; just needing a break, but unable to get away from ads that steadfastly followed them wherever they went: *“I can be at home. I can be on my cell phone. I could be at work, on a trip or something, and it’s following me everywhere. It never gives me a break”* (P33).

Participants also portrayed ad personalization as unable to distinguish between genuine affection and being uninterested: *“If I look something up on Amazon. [then] I go to a different website. That same product that I looked up, the next couple days, is popping up in an ad. And I just took a look at it. That’s it. Just took a look at it. And all of a sudden it’s stalking you for the next two days”* (P37). Others saw personalized ads as unable to take a hint, not realizing when participants were no longer interested: *“it’s happens for days and days. I wouldn’t mind if I searched today for shoes, tomorrow [saw] another [ad]. That’s fine, you know, but it’s continuous, over and over. I looked it up like two months ago. Like, get over it!”* (P2) and *“I looked up [online information about grad] school like twice, and it wasn’t even for me. So, relax”* (P2). Participants saw their months-old web queries for goods and services or ‘just taking a look’ at a product page as insufficient grounds for the marketer to continuing pursuing a relationship. Participants described feeling harassed by personalized ads: *“I’m trying to do something and you are in my way and your harassing is annoying”* (P33). Overall, some

participants expressed feeling like they were victims of stalking, generally when receiving repeated ads for items they were not interested in.

### **On Again Off Again Lover**

Many participants described enjoying the relevant advertisements afforded by online personalization at times and being distraught by personalization at other times; they seem to like personalized ads, on occasion or for a while, and then wish them to go away on other occasions: “sometimes when [personalized ads are] asking me to buy something, it frustrates me and annoys me because it’s [violating] my privacy. But other times, I’m a musician, so sometimes [personalized ads] will say ‘here’s some production techniques about writing music’ [...] I quite like that because I guess it’s easier to access information that way.” (P34). This on again off again way of thinking, enjoying the free benefits of personalization sometimes and at other times being turned off by this same functionality was repeatedly voiced by participants in different ways: “maybe you were checking [online] for a bank. And here’s some bank ads or some credit cards ads. It’s like, ‘God, leave me alone.’ Just because I looked up something. [...] [I] can go both ways because I’m always looking for a deal. [...] If it’s a deal it’s a deal.” (P26). These conflicted perceptions led some participants to view the practice as simply “tolerable,” acknowledging their inconsistent feelings towards online ad personalization: “It’s like a kind of love-hate relationship. We’ll tolerate it. It’s tolerable” (P18).

### **Eavesdropper**

At times, participants’ thoughts around having their every behaviors “stalked” during the ad personalization process bordered on paranoia. In multiple focus groups different participants

mentioned they believed their smartphones listen to their ambient environment, monitor the content of any conversations they might have, and later show them personalized ads based on this eavesdropping. As one participant, who felt advertising personalization technology was eavesdropping on his life, put it: *“You will be talking in a conversation with somebody, and I think your phone could, you know... And then you will look at it and [ads on your phone] will be like what you were talking about or something you were watching on TV. Maybe your phone could hear it or something. It might pick up from where they could hear your phone”* (P18). “To make sure I understand correctly, the phone could be listening to your conversation or to your television to personalize the ads you see?” (moderator). *“Yeah. Yes. While it’s not even on”* (P18).

While this way of reasoning about how online ads are personalized may sound like conspiracy theory at first glance, it may stem from first hand experience where participants consciously connect ambient audio to the tailored ads they receive. In 2016 the U.S. Federal Trade Commission issued a warning to mobile app developers using the *SilverPush* audio beacons framework, which turns a smartphone’s microphone on when not in use to pick up background audio (FTC, 2016). *SilverPush* uses this background audio to create a hidden list containing the television programs that a smartphone owner watches, or at least the programs one’s phone has been in the audible presence of, and this data is then sold to marketers for advertising purposes. Thus, the eavesdropper conceptualization of ad personalization may stem from these actual participants’ experiences, real or perceived, of personalized advertising received on their mobile phones

## Themes Identified

In addition to personas identified, a number of additional themes and metaphors emerged from participants' descriptions and drawings of online ad personalization. These broader themes are described below.

### Personification of Personalized Advertising

As revealed in the personas described in the previous section, one overarching theme emerging from discussions was the tendency expressed by respondents to discuss personalized advertising behaving in human-like ways. The resulting personifications, many of which took on negative characters (e.g., hustler, stalker, dictator) provided participants with a way to relate their understandings of ad personalization practices that they feel exist but often cannot see themselves, such as the collection of personal information by *eavesdropping* marketers.

Notably, these personifications often went beyond simply being negative. Participants described ad personalization acting in cunning ways, not only exhibiting unfavorable behavior but relying on intentionally deceptive maneuvers behind the scenes, as though marketers were scheming to harm the participants in some way. While not the case across the board, that some participants perceived personalized advertising as full of *intentionally* harmful processes illustrated a rather surprising degree of negativity and lack of trust some participants associated with personalized marketing and advertising techniques. Along these lines, one participant thought about advertising personalization as though "he" was a type of scheming little man living inside her computer: "*You are looking up something and all of a sudden, the computer, whoever is back in there, that little man, he's thinking, 'Oh, I'm going to invade her privacy and put this on there and put that on there and put that on there...' And then what if someone else*

*used your computer? And now they are getting your ads for the things that you looked up” (P26).*

In this example, prior experience somehow led the participant to envision a “little man” who intentionally thinks of ways to harm her, including how to invade her privacy by tracking her online so “he” can later deliver a revealing and embarrassing personalized ad to someone else with whom she shares her computer.

### **Online-Only World**

One way of thinking about ad personalization emerging from participant comments and drawings is the Online-Only World. In this way of thinking the online world is everything; that is to say, there are few if any activities or corresponding consumer data that play an active role in the ad personalization process that do not also originate online. Thus, the online world is the only environment under consideration given this understanding. For example, in the participants’ drawings, in addition to including the use of search engines other activities appearing in the drawings were mainly confined to visiting particular websites, using and installing apps, generating and having one’s web browsing history analyzed, liking and sharing content on social media, making purchases on ecommerce sites, sending or receiving emails, providing online account profile information, and other online activities. However, there was little to no mention by participants in their drawings nor in the group discussions of offline-originating activities or corresponding consumer data that influence how ads get personalized online. This includes, for instance, credit card data, data from customer loyalty cards, or demographic and socioeconomic data, all of which are commonly used to personalized online ads. Similarly, the Online-Only World way of thinking about ad personalization is consistent with the environment in which the ad itself is received. To include mechanisms, players, or data contributing to online ad

personalization that are not overtly linked to using the web might introduce a rupture, or a violation to the Online-Only World that most participants appear to be living in when reasoning about how online ads become personalized.

### **Unprotected Protected Classes**

When reasoning about how ad personalization does (and does not) work, participants imposed erroneous legal restrictions regarding what types of consumer data are permitted to be used to personalize ads online. This resulted in some people operating as though unprotected classes of data were actually protected, with marketers prohibited from using them to personalize ads online: *“They’re not able to collect really personal data, are they? [...] it’s illegal, therefore, the stuff they’re going to store and know is going to be somewhat generic, albeit personalized, but it’s not going to be how much you earn per year”* (P3). Similarly, though consumer financial information including income is incorporated in online advertising personalization schemes, other participants invoked imagined consumer protections regarding the use of financial information: *“for advertising, the type of information they’re going to want to advertise something to you is not necessarily that personal. It’s not like they’re looking at your specific bank details. They’re just looking at the types of things you’re buying”* (P16). Going along with these erroneous restrictions, some participants who invoked protections imagined a comprehensively regulated system, while also expressing doubts: *“I would say [marketers] must be policed heavily, [...] and I would imagine they are, but that may be a very naïve thought”* (P3). Though some participants expressed the exact opposite: *“I don’t believe [marketers are] policed heavily. I think it’s a free-for-all at the moment”* (P6).

These unprotected protected classes of data revealed how some consumers think about ad personalization under an inflated sense of consumer protection that, in ways described by participations, does not exist in the U.S. nor in the U.K. Still, this erroneous assumption is not surprising, as previous studies have found that consumers tend to dramatically overestimate legal restrictions on personal data use in commerce (e.g., Hoofnagle & King, 2008). Similarly, given the number of companies most consumers interact with online, understanding one's individual protections on a firm by firm basis has been shown to be highly impractical (McDonald & Cranor, 2008). In the related context of social media curation algorithms, individuals have been found to invoke self-serving understandings of these mechanisms, however, self-defeating understandings have also been observed (Eslami et al., 2015; Eslami et al., 2016). Though, once again, consumers' misunderstandings of how processes actually work do not necessarily render them ineffective (e.g., Kempton, 1986).

### **Transparency & Unseen Chefs**

In describing how online ad personalization works, participants often discussed their inability to know or see the process. Some articulated the role played by transparency and trust in response to the black box nature of ad personalization. One participant likened trusting marketers to use customer data to personalize ads in ethical ways as being the similar to the type of trust relationship one has with an unseen chef when dining in a restaurant when the kitchen is out of sight for the customers and, importantly to his understanding and conjuring this metaphor, of unknowable cleanliness: *"It's like trusting a chef to cook you a meal. [...] You can't always (think) 'Oh, what if that information... ?' or 'What if I get poisoned when I go to a restaurant?' It's that kind of thing. You don't know how dirty the kitchen is. It's the same"* (P28).



The unseen chef metaphor revealed another way of reasoning about online ad personalization. That personalization occurs in a black box, with many activities taking place out of sight. This means the consumer must trust that those operations occurring inside this black box (or restaurant kitchen) occur in ways the consumer would approve of if they could see inside. This association with a lack of transparency characterized many participants' experiences with personalized online advertising and was commonly associated with a feeling of disempowerment expressed in relation to data aggregators. These were common sentiments voiced by participants across the focus groups.

### **Responsibility & Taxi Drivers**

Similarly, trust was also invoked in comparing marketers appropriate use of data when personalizing ads to trusting a taxi driver: *“if someone is driving you somewhere, what if his driving is bad and he crashes? I’m just comparing it [to personalized advertising] really”* (P28). In both cases, trusting the unseen chef and the unknown taxi driver, the notion of the service provider relationship brings up a potentially fruitful metaphor not only for how consumers might on occasion think more creatively about this arrangement, but also the associated set of expectations for acceptable behavior in any successful transaction with an unseen or unfamiliar service provider. Key to these service relationships is often trust, which may weigh heavily in how some consumers are reasoning about the broader online ad personalization process and their relationships with marketers and other collectors of personal data.

## **Mental Models Identified**

Personas and themes identified from the focus group interviews and drawing exercises point to common ways participants expressed thoughts and feelings towards ad personalization. Related to these outcomes, and providing further insights into how individuals think about personalized advertising, overall participants appear to be relying on two mental models to facilitate their expectations and predictions when encountering online ads they believe have been personalized for them. These are the Self-Catalyst Model and the Eye of Providence Model. These models are commensurate rather than competitive with one another, although contradictory mental models are not uncommon. However, these models appear to be used, at times, in concert with one another while employing one is no requirement for possessing the other and some participants comments indicated they relied on only one not both. The Self-Catalyst Model and Eye of Providence Model are described below.

### **Self-Catalyst Mental Model**

The first overarching mental model emerging from both the drawings and understandings voiced by participants is what I will call the Self-Catalyst Model. First, many of the drawings depicted online advertising personalization as a process that *participants themselves set in motion* through discrete and deliberate action. Most often, this action included searching for a product on a website or app. Then, as participants responded to questions and dialogued with one another, often they recalled experiences where they received ads online they inferred had been personalized based on a discrete act they themselves initiated.

The notion of the consumer providing a catalyst, and one they are aware of for that matter, which directly sets the process of ad personalization in motion is consistent with the most

common end point in participants' drawings—not only a personalized ad but one that was the specific result of a subset of advertising personalization known as “retargeting” or sometimes “remarketing.” Retargeted ads are simply those generated for consumers based directly on their specific prior actions, most often linked directly and overtly to a specific webpage(s) one has visited or terms entered into a search engine (ReTargeter, n.d.). For retargeting, the domination concept appearing in many of the drawings and when participants described their experiences with personalized advertising, the link between a specific catalyzing event on the part of the consumer and the specific content of the resulting personalized ad is often strong and direct. For example, visiting the URL for a specific blouse on an ecommerce website and later seeing a personalized ad for that exact blouse on another website's ad space. Thinking stemming from the use of this Self-Catalyst Model of ad personalization was very common among focus group participants and appears to dominate these individuals think about this process.

This self-catalyst way of thinking is opposed to, for instance, having one's interests and personality assessed by an advertiser, social media platform, or algorithm over any longer period of time with these attributes assigned to a user profile gradually and later used to selectively present ads to viewers. Then, regularly and casually, seeing personalized ads across websites and apps based on this consumer profile data. Or, alternatively, seeing personalized online ads based on ones socioeconomic status or based on data from offline activities that advertisers use for ad personalization (e.g., job title, automobile, grocery purchases). Instead, as drawn by participants, the Self-Catalyst Model suggest advertisers leverage more discrete consumer behaviors to tailor ads. This way of thinking conceives as personalization as directly reactionary to online behavior.

This Self-Catalyst Model of reasoning about ad personalization is also notable as it both represents, and somewhat accurately, a common way online advertising personalization is

achieved while at the same time differs considerably from how a substantial portion of online ads are personalized. That is to say, many online ads exhibiting some form of selective targeting towards specific audiences are not the product of retargeting based, such as that based on URLs visited or search engine queries. While these activities do stimulate personalized ads, on occasion, they represent only a fraction of personalized advertising as oftentimes ads are selectively presented, or personalized, for viewers based on ad campaigns targeted towards individuals tagged as possessing a specific consumer attribute. For instance, this includes ads based on income or interests expressed over time, rather than retargeting and personalization based on a live, catalyzing action on the part of the consumer.

Yet, it is completely logical that consumers would most readily reason about advertising personalization using the most visible way this process manifests. That is, under retargeting savvy consumers connect the invisible dots between their previous recent behavior and the current ad in front of them. Still, the popularity of the Self-Catalyst Model in participants' reasoning suggests consumers may be thinking very narrowly about online advertising personalization given the scope of retargeting among the broader targeting of ads online using a number of additional data and approaches.

### **Eye of Providence Mental Model**

The second mental model participants commonly appeared to be using to reason about how online ad personalization works is the Eye of Providence Model. In their study of "folk theories" of how the Facebook algorithm works, Eslami et al. (2016) found participants operating on the assumption that the popular social media platform was not only all-powerful but also all-knowing, aware of every possible detail and capable of using this information to its own

advantage and/or that of its users. The authors term this folk theory the Eye of Providence Theory after the concept in many religious traditions of God's all-seeing eye constantly watching over everyone and at all times.

Similarly, focus group participants in the current study appear to also rely on an Eye of Providence mental model when reasoning about how online ad personalization functions. Repeatedly, in describing their experiences with online advertising and their specific expectations for ad personalization, individuals projected an all-seeing omniscience onto a singular system both capable of and responsible for knowing all possible details of individuals' lives, even including invisible desires and preferences—a monotheistic vision of an all-knowing system.

For instance, participants commonly voiced being annoyed by ads they (rightly) perceived as personalized for them under retargeting schemes when they had already purchased a product yet continued to receive personalized ads for this same item. Repeatedly, participants were dismayed that ad personalization systems failed to take note that they had purchased an item online or in a store and therefore were no longer in the market for this product. In this way, participants appeared to rely on an Eye of Providence mental model in expecting that some singular, all-knowing system not only existed but was able to keep track all of their purchases made online or offline. In expecting this singular ad personalization system to somehow stay apprised of all of one's purchases, as not to continue delivering ads for products that were no longer being considered for purchase, is somewhat remarkable when considering what this would require. For instance, this would necessitate a system that monitored the variety of places individuals can make purchases along with those made with cash and/or in the absence of

purchase tracking mechanisms such as loyalty cards. Accordingly, the Eye of Providence Model observed in use by many participants predicts an omniscient system tracking all purchases.

In the case of participants expressing disappointment when the Eye of Providence Model often led them to view ad personalization as failing by continuing to show ads for items they had already purchased, this way of reasoning about personalized advertising is of course ambitious. Yet one can imagine a time in the not so distant future when improvements in tracking technology keep retargeted ads in step with purchases more effectively. In this way, in part, the Eye of Providence Model may become more useful if ad personalization becomes more intelligent. Though the assumption that personalized advertising will only become more and more accurate in the future is a contested notion (e.g., Searls, 2012, p. 21-42).

Further, participants expectations when invoking the Eye of Providence Model extended to technological capabilities that were ambitious beyond simply tracking all of one's purchases. Participants also expected online advertising personalization technologies to know and act on participants' unseen desires (e.g., knowing whether a particular internet search query was related to their own interests or conducted for someone else such as a relative of theirs). In this way, individuals projected capabilities onto ad personalization platforms that might allow them to act on invisible and largely unknowable user preferences—further suggesting consumers at times ascribe a God-like omniscience to this system, where the all-seeing eye penetrates beyond what can be learned through data analysis alone. Participants also described being annoyed when their variable, day-to-day levels of interest in specific products were not accompanied by corresponding response in the ads they were shown. This is despite these interests levels being invisible and unexpressed, information only an all-knowing being or system might possess and be able to act upon.

Additionally, participants invoked the Eye of Providence Model in expecting that ad personalization systems were capable of determining whether items they had previously read about online or purchased were for themselves or for other people. For instance, a participant described dissatisfaction that his shopper rewards card might give him physical or digital coupons related to purchased he had made for his mother not himself. Participants reasoned with these expectations of omniscience and were then disappointed that the all-seeing eye had failed to properly account for nuanced differences.

Ascribing omniscience or at least a form of high intelligence to ad personalization systems, another participant expressed being disappointed when she began seeing ads on her smartphone that were clearly targeted towards children after her niece installed a children's game on the participant's smartphone: "*All my advertisements after that were, literally, little kid things for like a month. It was terrible. I had to delete that app*" (P2). This line of reasoning expects advertising technology to operate at an all-knowing level, for instance, capable of distinguishing between which software has been installed for oneself versus others and to personalize ads accordingly discerning this level of detail among individual desire and preferences, which are often unseen desire and preferences.

Overall, employing the Eye of Providence Model results in user expectations for a type of *perfect personalization*, where the advertisements one receives are perfectly tailored towards one's unique and dynamic interests, even those unexpressed or unseen, while simultaneously knowing and honoring one's expectations perfectly, for instance, for which consumer data will be collected under which circumstances and how this data is used or not used to selectively present advertising content. While the goal of perfect personalization is noble, it is more an ideal to aspire towards. In this way, invoking the Eye of Providence Model and expectations for

perfect personalization can be more damaging than useful. Further, rather than experiencing perfect personalization, the most notable outcome expressed by participants employing the Eye of Providence Model was disappointment. As mental models are useful insofar as they are good at predicting the operation of a system, the way in which Eye of Providence thinking was often observed alongside experiences when advertising personalization failed to live up to this omniscience suggests this is not a very effective mental model in terms of its usefulness. The Eye of Providence Model often led participants astray. In this regard, though it appears quite popular it is not very useful.

### **A Proposition: The Broken Clock Mental Model**

Based on the mental models of advertising personalization participants appear to be relying on, perhaps more useful than the narrow Self-Catalyst Model and somewhat the inverse of the Eye of Providence Model, is what I will describe as the Broken Clock Model. This Broken Clock Model of advertising personalization is completely prescriptive, as it does not come from participant descriptions but rather is proposed as a potential solution and replacement for the mental models observed in use by participants, models which appear to often leave individuals disappointed when reasoning about and interacting with advertising online.

The key deficiency common to the Self-Catalyst Model and Eye of Providence Model, as observed in how participants use these models, is the severely limited ability of these constructions to live up to the primary function of mental models: prediction. Both of these models are problematic for those relying on them due to the infrequency with which they help consumers predict how ad personalization will function. Instead, both are highly likely to disappoint people due to their weak predictive power. In expressing these models, through



drawings and spoken dialogue, participants appear to be routinely disappointed by the lack of parity between outcomes these models lead them to predict (e.g., anticipating one's unseen desires) and their lived experiences with online advertising. *A more effective model would better predict these experiences while leaving its users with less disappointment.*

Of the two, the Self-Catalyst Model appears to be more useful, or perhaps simply less harmful, than the Eye of Providence model. The former is correct, occasionally, and when it does accurately predict the outcome of personalized advertising, as in the case of seeing an ad for a particular swimsuit after earlier in the day searching for this style of swimsuit using a search engine, it predicts this outcome very well and its users are not disappointed. At the same time, the applicability of the Self-Catalyst Model is very narrow, corresponding to just a portion of the many mechanisms by which online advertisements become personalized. Therefore, it will often disappoint its user if this is the expectation for how all personalized advertising works not just occasions of retargeting. Additionally, and as discussed previously, the Eye of Providence Model is also often ineffective in accurately predicting outcomes, typically leaving its users with mismatched expectations between what is possible and how marketers actually collect and cross reference consumer data for use in advertising personalization. As there is no such singular all-knowing omniscient system supporting ad personalization, the Eye of Providence Model elevates expectations of this system to a God-like level only promising to disappoint its users when failing to live up to the expectations and predictions it invokes.

An alternative to these two largely unsatisfactory mental models some individuals appear to construct and rely would be to use a mental model with lower, more realistic expectations for how well various processes within the system function. Though perhaps bleak at first glance, a model defined by low expectations could end up being quite useful if it could predict what

appear to be somewhat negative experiences with ad personalization, including disapproval with some of the ways consumer data is collected along with dissatisfaction when ads that are deemed by consumers to be personalized are also deemed inaccurate based on not matching individuals' expressed or unseen interests. In this regard, a low expectations mental model of ad personalization, where the model predicts ads to be personalized yet only on occasion to an individual's liking, may be more effective than the observed Self-Catalyst Model or Eye of Providence Model employed by participants in this study.

Therefore, a proposed mental model for reasoning about online advertising personalization is the Broken Clock Model. As it goes, broken clocks are seldom right, correct exactly two times each day. Accordingly, their owners and any others who happen to use them come to develop low expectations for their performance, as these clocks accurately predict the time on the rare occasion, though usually fail to do so. For example, the owner of a broken clock may happen to glance up at the clock at the exact moment when the time it shows is indeed accurate. Therefore, the clock does not always fail. In fact, it sometimes works just as one expects a clock to work by indicating the correct time. When used, the Broken Clock Model would lead its users to expect personalized ads to usually, most often, miss the mark. This means not being matched exactly or even closely towards one's interests. In this way the Broken Clock Model might be very accurate at predicting outcomes.

Similarly, this proposed model might also be useful in predicting not only the end product of advertising personalization—the ads themselves—but also its user's experiences with consumer data collection practices necessary to support personalization, activities focus group participants also voiced repeated disappointment with based on their understandings of how consumer data collection either does or show work. For instance, the Broken Clock Model would

lead its user to predict website and apps to be aggressive in their collection and use of consumer data, at times tracking behaviors that individuals would prefer were not tracked. Similarly, in this sense, employing this model would lead individuals to expect and predict mobile apps, for instance, to require access to many types of data from one's smartphone such as location, contacts, camera, microphone, searches, and other data even when it is unclear why the app would need many of these items to function properly. Invoking the Broken Clock Model for these data collection activities linked to ad personalization further aids users in expecting a system that does not always work as one might wish, while sometimes doing so.

Finally, and importantly, the Broken Clock Model occasionally succeeds in predicting ad personalization not only for the previously described negative experiences but also on those occasions when individuals are well-pleased with personalized ads—similar to those times when one happens to glance up at the broken clock on the wall when the clock “works” in displaying the current time, in seeing a well-targeted ad that one not only enjoys but clicks on to make a purchase. In this regard, the model is one of low expectations, but not incapable of predicting positive experiences in addition to broken ones.

Of course, the proposed Broken Clock Model of advertising personalization is not without drawbacks. The most notable is that it is rather bleak. It assumes a broken system in the sense of a system that will typically fail to deliver advertising content that a person deems to be personalized appropriately towards their expressed or unseen interests and/or failing to honor expectations for personal data use. In this way, the Broken Clock Model it is quite the opposite of the Eye of Providence Model discussed previously, as the former assumes the system typically does a poor job in “knowing” the individual to whom it delivers ads. In general, this expectation of brokenness is intended to curb disappointment resulting from other models that assume higher

levels of ability, such as the all-knowing Eye of Providence Model, and associated expectations for perfect personalization.

## **DISCUSSION**

### **Contextual Integrity, Online-Only, & Retargeting**

The Self-Catalyst Model and Eye of Providence Model highlight how consumers may be reasoning about online advertising personalization by projecting their own sense of what constitutes appropriate online information use by marketers when tailoring ads. This line of reasoning observed in many participants' drawings and discussion points to what Nissenbaum (2010) identifies as "contextual integrity," a heuristic for determining whether context-relative information norms have been violated. In the case of personalized advertising, consumers experience this phenomenon in the online context. When speaking about the data used to personalize advertising, they spoke from the perspective of an Online-Only World and even then almost exclusively considered their more overt online behaviors (e.g., typing in terms to a search engine) as affecting the customization of the ads they encounter on the web. Therefore, and affirming the central thrust of contextual integrity, these individuals expect online advertising personalization and its underlying information practices to be confined to what can be learned from online behaviors (i.e., as opposed to offline behaviors/data), especially over behaviors like visiting a particular product's URL. This might explain why marketers' use of customer data corresponding to, for instance, a consumer's web browsing history (e.g., a specific product URL) to personalize ads was both the most salient example among focus group participants and generally the most permissible, given the similar context of the original behavior and where the subsequent personalized ad is viewed.

The other reason why retargeting and the Self-Catalyst Model was so widely popular among participants is obvious but it should not go without mention. Retargeting and similar personalization processes (e.g., sponsored search results) connect actions stored in consumers' short-term memory, such as visiting a particular product URL, with personalized advertisements. This is in contrast with, for instance, highly personalized online ads based on psychographic inferences made by a social media platform according to a consumer's behavior on the platform. Instead, retargeting, though arguably only a small proportion of personalized online advertising, is the most perceptible and therefore salient way of reasoning about this practice.

Additionally, contextual integrity might also explain, in part, why participants failed to describe or depict online ad personalization based on offline behaviors and corresponding data, for instance, such as their transactions in brick-and-mortar retailers. Further, this offline/online connection, though apparently less salient for participants, may lead to violations of contextual expectations for appropriate information use, such as delivering online ads for Cheerios based on previous cereal purchases at one's grocery store. For consumers to imagine online ad personalization incorporating activities and data not originating online (e.g., occupation, income, whether they are a parent or not) might result in dissonance (Festinger, 1962) given that there are few ways for people to learn about the use of offline data in online ad targeting. As a result, the use of consumer data from sources not originating from online behaviors (e.g., web searchers, account profile information) is likely to strain individuals' expectations for which personal information will be used, how, when, and by whom with many people's conceptions of ad personalization understandably confined to data emanating from these online behaviors.

## **Limitations and Future Research**

The ways individuals think they think about something, and how they describe this, may be incorrect, especially when compared to how people behave in the moment. This potential distinction highlights the need for examining more situated action (Suchman, 1987; 2007), such as in-situ lab experiments or unobtrusive observation of consumers (Webb et al., 1999). Norman (1983) highlights the difficulty in discovering and representing individuals' mental models, warning that you can not simply go up to the person and ask them to discover their mental model of something (p. 11). He further cautions how attempts seeking to characterize a person's way or reasoning about something may yield erroneous information because people can state they believe one thing about a system when in practice they deploy a completely different mental model when actually interacting with this very system.

Similarly, in this study I have attempted to draw out, analyze, and make conclusions based on the mental models and other perceptions individuals hold about personalized online advertising. But this approach is lossy at best. At worst, it is misrepresentative. As it is impossible to extract a person's actual mental model of how something works. Mental models are abstract and, by definition, immaterial. Nevertheless, attempts to gauge and further understand the unseen conceptual models people possesses about online personalization may provide information directly useful to a range of advertising stakeholders. This includes the practitioners and policymakers who influence the development and regulation of different technologies and marketing practices, especially, in this case, those that support consumer data collection and personalized online advertising. Further, improving how advertising stakeholders perceive consumers may benefit consumers indirectly through improved system design and regulations.

Another limitation of this work is the inability to completely separate attitudes toward ad personalization and those toward related practices. Clearly distinguishing between participants' attitudes toward personalized online advertising, those toward online advertising more generally (such as any promotional messages delivered across a digital channel), and those toward the institution of advertising itself is challenging, likely contributing some unknown degree of noise in the current study. To address this challenge, questions posed to focus group participants and follow-up probes were worded to continually steer participants towards thinking specifically about "personalized online ads," but broader feelings about online ads and advertising in general were articulated at times when discussing personalization (i.e., annoyance with pop-up ads). The second overlap is between the consumer surveillance by marketing firms and citizen surveillance conducted by governments. The notion of "big brother" came up in multiple focus groups along with some examples of state surveillance in discussions of participants describing how they believed their behaviors are constantly monitored by the government (e.g., via state CCTV systems). Further, ways of reasoning about government surveillance practices are increasingly difficult to separate from customer data collection by marketing firms. And this entanglement has only been compounded by the many revelations in recent years of state surveillance tapping the customer data of internet, social media, and telecommunications companies (*The Guardian*, 2013).

Additionally, of immeasurable influence in this work is my positionality and grasp of personalized online advertising processes, specifically the impact of my own mental models and how I reason about ad personalization myself. This bears influence on the study design, stimuli and questions posed to participants including follow-up probes, and overall dynamics of the open-ended focus group discussions. Efforts taken to minimize these biases include neutral

question wording and care given to the ordering of focus group activities and questions to minimize moderator influence. Yet, bias in interpretive work is unavoidable. As Packer (2010) argues, investigators need not and further cannot completely eliminate their own subjectivity in designing and executing qualitative inquiry (p. 79). Likewise, the aim of this exploratory study was not objectivity but rather to provide a rich set of descriptions about consumers' ideas and experiences to tap intimate knowledge of otherwise unexamined beliefs (Lindlof & Taylor, 2011, p. 45). The hope is that any conclusions presented will stimulate other researchers and practitioners to explore related questions and/or triangulate these conclusions using approaches more suited toward testing the themes I have identified (i.e., using surveys or experiments).

Finally, findings from focus group interviews do not lend themselves to generalizability and no attempt to do so has been made for the broader populations in the U.S. or U.K. Additionally, regarding participants sampled, this study relies on focus group interviews with residents of global media hubs, London and Chicago, whose experiences with advertising could differ widely compared to those in more rural areas, even in these same countries let alone in other parts of the world. While this (potential for) variance is unknown, there is no cause to believe individuals living in dense urban environments reason in substantively different ways about personalized online advertising systems than others. Though this remains an empirical question and beyond the scope of this work. As for groups conducted in the U.S. versus those in the U.K., no substantive differences were observed between participants in these different countries. The most notable distinction was participants' use of different national chains when discussing experiences with brands (e.g., Walgreen's vs. Sainsbury's) though of no noticeable consequence for how participants described related experiences. And even then a small set of media technology companies (e.g., Google, Facebook, Twitter) entered into and dominated



responses from both sides of the pond, highlighting the multinational influence of these firms. Future work might seek to leverage comparison among urban vs. rural, along socioeconomic difference and other demographics, or between countries generally especially as either similarities or differences cross-culturally could prove insightful.

## **CONCLUSION**

This study examined consumers' ways of thinking about how online advertising personalization works and how they reason about this process. Based on a set of focus groups with adults in the U.S. and U.K., it appears how consumers think about ad personalization is at times consistent. Common themes emerged in how they described and depicted this process. Participants voiced resistance to personalized advertising practices, often discussed alongside broader distaste towards marketers' attempts to persuade them in any form of advertising. At the same time, and somewhat ironically, many participants voicing these negative reactions also expressed disappointment that the online ads they perceived to be targeted for them were often not accurate enough or not personalized enough. For instance, participants often voiced that marketers did not respect their privacy, "stalking" participants across websites and apps, while also voicing disappointment that ads they received were not specific enough and clearly did not understand their interests very well.

In response to the practice of personalized online advertising, in an initial word association exercise focus group participants expressed quite varied word lists with some providing altogether negative associations, others more neutral, and many others demonstrating conflicted feelings about the process, such as describing it as simultaneously "useful" and "scary." While participants' overall discussions about advertising personalization were also far

more negative than positive, participants showed diversity both in attitudes about this process and how they reason about ad personalization. Seeking to explore how these same consumers think about this practice, when asked to draw a picture depicting how online advertising personalization works they offered relatively consistent diagrams both in structure and the processes included. Many participants appear to possess a related mental model where through deliberate online behaviors they themselves provide the necessary catalyst for online advertising personalization, even though this Self-Catalyst Model corresponds to only a fraction of ad personalization online. Consumers' experiences may be more similar than different, impacting these corresponding mental models and ways of reasoning about personalization.

Participants thought about ad personalization often relying on personification and metaphor. Some went so far as to consider that marketers were inflicting not only harm on consumers through data tracking and personalization but doing so intentionally, as in the case of one participant thinking about ad personalization processes as a scheming little man inside of her computer invading her personal privacy. Other understandings of ad personalization relied on metaphor and seem to assume more neutral intentions on the part of marketers, as in the case of the unseen chef with whom consumers must entrust their digital data behind the scenes of ad personalization, similar to the way consumers must trust that unseen chefs prepare their food in an unseen clean kitchen when dining out.

If consumers reason about online advertising from a negative starting point, viewing this process as opposed to consumers' interests, influence of this attitude likely shapes how they think about the system itself—either as one that serves consumers' interests first versus that of marketers, or vice versa. Given certain economic realities, consumers have little ability to influence directly how this system works. Accordingly, operating with a mental model that

predicts advertising personalization as serving the interests of marketers first, and not organizing interaction in ways consumers prefer, may actually be a more useful way of reasoning about this process, at least for the time being. This may sound bleak but it may also be the most useful way of approaching and predicting the function of the current imperfect system that supports online ad personalization.

Overall, the themes, personas, and metaphors identified correspond to how participants described online ad personalization. The value in any mental model, regardless of how close it resembles that of an expert, is how useful it is. Participants in the series of focus groups revealed two distinct, yet commensurate, mental models that appeared to be in widespread use among these individuals. These are what I have termed the Self-Catalyst Model and, borrowing from the related phenomenon described by Eslami et al. (2016), the Eye of Providence Model. Additionally, given the degree to which both of these mental models appear to disappoint their users, I have proposed the alternative Broken Clock Model of online advertising personalization as a way to mitigate current expectations in this context. Advertising practitioners and researchers might take note of the ways participants in this study appear to be thinking about advertising personalization, including the use of a number of personifications of this process (e.g., hustler, stalker, presumptuous stranger) along with reasoning that they themselves provide the catalyst in this process governed by all-knowing system both of which lead to missed expectations for consumers.

## Chapter 5

### **The trouble with transparency: Experimental evidence for the negative effects of transparency on trust and preference for personalized advertising**

In this chapter I present four experiments I conducted to examine the relationships between transparency, trust, and consumer preference for personalized advertising. Experiments one and two varied participants' levels of awareness regarding how ads are personalized online by altering transparency in this process, while experiments three and four varied participants' levels of social trust. Similar or identical outcome measures were used across the four experiments to examine the possible effects of transparency and social trust on concepts of interest such as preference for advertising personalization along with individual trust in ads, websites, and apps. Viewed together, results from all four experiments point towards a quagmire for marketers. Across the board, greater transparency about some of the ways online ads are personalized for individuals appears to diminish their support for this practice, while at the same time increased social trust appears to do nothing to improve how ad personalization is perceived nor the degree to which consumers trust online advertising and the websites and apps that deliver personalized content including advertisements. Marketing practitioners and policymakers should take note as findings point towards possible consumer backlash, including reduced trust in marketing and digital environments generally, when some of the practices currently supporting online advertising personalization are made more transparent.

## INTRODUCTION

Increasingly, marketers have championed the practice of using detailed information about individuals to create advertising content that has been personalized, to varying degrees, and delivered to consumers through web browsers and, more recently, in mobile apps as well (Sherbin, 2012). This effort parallels the development of the commercial internet (Greenstein, 2015), routinely spurred forward by the need to store information about its users. Examples include the invention of the “persistent client state object,” or internet cookie, to facilitate online shopping carts and the completion of multi-page web forms (Schwartz, 2001). It also follows advances in information technology and associated synergies with consumer data collection efforts resulting in a new nexus of business interests linking developers of commercial websites and apps, third-party data providers, and marketing services and other media companies.

This move towards personalization in online advertising comes as no surprise, too, given that audience segmentation and the use of personalized messaging is highly effective. The use of personalized appeals to consumers in advertising, both when viewers are aware and not aware of the personalization, tends to be more persuasive compared to general appeals (e.g., Kampe, Frith, & Frith, 2003; Lambrecht & Tucker, 2013; Shatnawi & Mohamed, 2012). Hence, marketers and others with shared business interests including most of the web’s most popular platforms and services providers (i.e., for social media, email, search, photo/video sharing) go to tremendous effort and expense to personalize advertisements for consumers delivered on these websites and apps.

More broadly, personalized content on web platforms comes in many forms, including online advertisements but also extending to discounts, product recommendations, e-commerce prices, and assorted organic content on social media platforms. Each of these examples can be

selectively presented and tailored towards individuals using combinations of verified and estimated consumer data. Thus, a considerable amount of the content appearing online is now personalized, to some degree, for the consumer. Complicating this matter, following the path already entrenched by social media platforms (e.g., Facebook.com) and recently enabled by advances in third-party native advertising platforms (e.g., Sharethrough.com), more traditional websites and apps have also begun to integrate and intersperse personalized organic and editorial content with personalized advertising. As a result, not only is the line between editorial, organic, and marketing content increasingly fuzzy online, so too is distinguishing between which content has been selectively presented, or personalized, for an individual viewer versus that delivered en masse. For internet users this distinction between which content is and is not personalized for them is increasingly impossible to make. At least, this distinction is difficult to make with much confidence.

In advertising, personalization is further justified by the increased efficiencies for both marketers and individual audiences. Firms benefit from cost savings by reducing advertising expenditure wasted on audience members who fall outside the target audience, while at the same time offering better return on investment whenever personalized messages, discounts, and/or prices increase favorable response from this targeted group (Goldfarb, 2014; Evans, 2009; Lin, Ke, & Whinston, 2012). Of course, market segmentation and tailoring marketing communications towards anticipated audience attributes has long been used to achieve better results in advertising (Plummer, 1974; Wedel & Kamakura, 1998) and, similarly, to extract greater consumer surplus using various forms of price discrimination (Ekelund, 1970; Varian, 1989; Elmaghraby & Keskinocak, 2003).<sup>28</sup> However, the maturation of high-speed computer

---

<sup>28</sup> The study emphasizes the marketing phenomena of *advertising personalization* (sometimes called targeted advertising), rather than, for instance, focusing on audience segmentation, with segmentation being a broader

networks and online advertising technology combined with expanded offerings to marketers from data brokers have been a windfall for many marketing efforts centered on data-driven digital ad personalization (Deighton & Johnson, 2013, 2015).

For consumers, personalization of content (organic, editorial, or promotional) provides a means of filtering a cacophony of digital content pushed, pulled, and streamed across the websites and apps. The aim is to deliver more relevant information to consumers. For promotional content, personalized advertising messages are generally upheld as providing more relevant product information compared to mass delivery or even context-based online advertising.

To achieve online content personalization, marketers rely on opaque consumer data collection practices. This lack of transparency in how online ads are personalized is justified in the name of competition, with consumer data and proprietary personalization algorithms constituting trade secrets. As a result, internet users are usually unaware of both the degree to which the online content they see differs from what others see and, equally important, unaware as to which information about them has been used to perform the personalization. A general lack of transparency characterizes the process from the consumer's point of view. As daily use of interactive media continues to increase for many populations, largely in the form of time spent using smartphones and tablets, automated content personalization practices including highly targeted advertising techniques may pose new challenges for both marketers and consumers.

---

activity under which ad personalization exists. The emphasis on personalization rather than segmentation reflects recent advances in marketing, particularly how consumer data at the individual level is used to determine online marketing content in a way that exceeds traditional notions of simply dividing audiences into groups based on a shared attribute. Today, many forms of personalized advertising are not based on broad group identity segmentation (e.g., one's age, income, location) but combinations of these and actions linked to specific people (e.g., purchasing a specific item online or in a store). Of course, advertising personalization and market segmentation are two sides of the same coin, but this minor distinction and the move by marketers towards advertising based on discrete behavior motivates the focus on personalization.

Exploring these challenges, this study reports results from four experiments examining consumer response to the online advertising personalization. The following research questions are considered: Under what conditions do consumers prefer and trust personalized advertising and when do consumers trust the websites and apps that deliver these messages? For instance, does being more aware of how online ads and other content are personalized for individuals using consumer data affect how they feel about this practice? If so, then what underlying social factors might be contributing to these perceptions? For instance, does feeling more or less trusting towards other people increase or decrease support for advertising personalization? And how does this “real world” trust, that is social trust in other people, affect other forms of individual trust in the online environment, such as trust in online advertising and trust in the web platforms that deliver personalized online ads? Overall, the individual influences of these two factors—transparency in personalized advertising and social trust—are examined.

The first two experiments reported, Experiments 1 and 2, test the effects of transparency in personalized online advertising assessing if heightened awareness of these practices affects consumer trust in personalized advertising, trust in commercial websites and apps, online privacy concern, personal data control self-efficacy, opposition to consumer data collection, and stated and revealed preference for personalized advertising messages. The second two experiments reported, Experiments 3 and 4, investigated the influence of social trust (that is, trust in other people and faith in humanity in general) similar outcome measures previously tested in Experiments 1 and 2. Examining the impact of social trust on these outcomes may supply insights into how this broader form of trust might impact more specific and actionable forms of consumer trust, such as trust in advertising and trust in commercial web platforms. While much is known of consumer privacy preferences (e.g., Rainie & Duggan, 2015), far less is understand



about individual preferences for the growing number of activities that depend on these privacy-arousing data collection activities, such as the personalization of online content, including advertising.

## RELATED LITERATURE

### Trust

In marketing research consumer trust is understood to be a key component necessary for creating value, through the exchange of goods and services, and for maintaining consumer loyalty (Sirdeshmukh, Singh, & Sabol, 2002). More broadly, trust typically refers to a directional, multi-dimensional construct functioning in-between two or more parties and rooted in expectations held by the *trustor*, one who trusts, for what constitutes acceptable interaction, behavior, or attitude by the *trustee*, one who is trusted. Levels of trust between parties do not always match and can differ considerably, in addition to being variable over time. Negative considerations often accompany trust as well, such as consideration of risk, harm, embarrassment, and/or costs.

Trust is often thought of either as a process or as a discrete quantity, or both. While many have attempted to define trust, there is no agreed upon definition of trust in the literature. For instance, Khodyakov (2007) opts for a process definition, offering that trust is “a process of constant imaginative anticipation of the reliability of the other party’s actions based on 1) the reputation of the partner and the actor, 2) the evaluation of current circumstances of action, 3) assumptions about the partner’s actions, and 4) the belief in the honesty and morality of the other side” (p. 126). Similarly, Rousseau, Sitkin, Burt, Camerer (1998) conducted a cross-disciplinary review of scholarship on trust, synthesizing the most common uses of the term in academic

inquiry and arriving at their own definition of trust, “a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another” (p. 395). Consideration of vulnerability on the part of the trustor and reputation on the part of the trustee are especially pertinent to the topic of online advertising given the potential sensitivities surrounding some forms of personal data used by marketers to selectively target ads towards consumers, as well as how a firm’s reputation impacts consumer adoption of its products or services.

Another common distinction in the literature on trust is to treat interpersonal trust, the trust from one person to another person (Simpson, 2007), differently than trust individuals have towards an organization, the trust from one person to an organization (Tan & Thoen, 2000). Further, both of these forms of trust have been examined and treated differently from trust between two or more organizations (Burchell & Wilkinson, 1997).<sup>29</sup>

### **Trust in Online Environments**

In online environments, trust is assumed to be a necessary precondition for consumer adoption and sustained use of digital products and services (Mutz, 2005; Beldad, de Jong, & Steehouder, 2010). However, this precondition is challenged by recent online marketing practices, where consumers are often unaware of the actors they are being called upon to trust, at times with quite sensitive consumer data, when adopting and using web-based technologies such as social media platforms, search engines, cloud-based email services, and most commercial websites and apps, most of which surreptitiously collect consumer data in the background during regular use. Still, consumer trust has deep roots in nearly all areas of commerce and remains a

---

<sup>29</sup> There is scant investigation into the fourth and final combination of these directional person/organization trust distinctions—organizations trust towards persons—which for that reason is not included in this list.

key concept in understanding business transactions, with theoretical and empirical work continually finding trust to be a key facilitator of consumer- and business-to-business interactions (Macaulay, 1963; Fox, 1974; Husted, 1989; Burchell & Wilkinson, 1997; Kim & Kim, 2005; Weisberg, Te'eni, & Arman, 2011).

Additionally, emphasis on trust as a major driver and facilitator of transactions extends to the type of non-traditional consumer-to-business relationships that characterize today's popular business model used by many web platforms—free use in exchange for personal data. Under this business model, consumers (or “users”) supply their personal information to firms, at times knowingly and often unknowingly, in exchange for using no-cost online products and services, such as social media sites, cloud based email, search engines, and, more recently, operating systems (e.g., Windows 10). While this “free” relationship complicates traditional knowledge of how consumer-to-business transactions function, it is likely that the role of trust is not diminished in these arrangements only different. This should be the case, as regardless of specific context, trust is typically considered key for a healthy and functioning society (Schneier, 2012). That is, trust should be important in understanding online consumer behavior, even when money is not technically exchanged, because trust is simply important to people in general.

Beyond online environments, general social trust is thought to be highly impactful in the way individuals go about their daily lives. General social trust is often understood as the trust an individual feels toward their fellow humans or simply faith in humanity. The scholarly literature is full of efforts to understand this type of broad trust, for which there is an unsurprising degree of overlap with interpersonal trust—the variety of trust felt toward specific groups or even specific people (Rotter, 1971; Rotter, 1980; Hinde & Groebel, 1991; Zaheer, McEvily, & Perrone, 1998).

Previous work in this area has also found various conceptions of trust to be crucial factors in determining, for instance, willingness of consumers to conduct financial transactions with firms in both online and offline (brick and mortar) settings (Hoffman, Novak, & Peralta, 1999; Chen & Dhillon, 2003; Petrovic et al., 2003; Paakki, 2008; Beldad, De Jong, & Steehouder, 2010). Similarly, beyond financial transactions, forms of trust have also been shown to play vital roles in interactions between users and online systems more broadly (e.g., Li, Valacich, & Hess, 2014), pointing to the possible wide application of trust in facilitating user preference for other digital decision making.

### **Trust and The Media Equation**

In a large body of work by Reeves and Nass (1996), the researchers discovered a strong connection between the way people treat one another and how they act towards computers and various forms of media and media technologies. For instance, people routinely act politely towards inanimate media artifacts, including computers, and also ascribe thinking linked to human stereotypes (e.g., based on gender) to computers and other media technologies just as they would in “real life” social situations. The researchers call this connection “the media equation” in reference to their overall theory, which posits that “media = real life.” This equation represents their conclusion that there is little distinction in how people relate to other people and how they relate to media and media technologies including computers.

Relying extensively on experimental methods, Reeves and Nass attribute the inability of humans to react differently to non-human media and technology to the slow pace of human biological evolution. In particular, after repeatedly failing to locate differences in how humans behave towards artifacts and towards other people, Reeves and Nass arrived at the conclusion

that the human brain has failed to evolve at the same pace as media technologies (1996, p. 12-13). Consequently, they claim, this makes people inept at treating media differently than they treat fellow human beings.

If this is the case, that is, if the social rules that apply in human interaction transfer rather directly to how individuals treat computers and media, and if these dynamics are universal such that “the media equation applies to everyone” (p. 252) and does not just apply to rare occasions (p. 253), then, consistently, there should be little to no difference in how people respond to others versus how they respond to media technologies. In the case of trust, under the media equation the way humans express trust towards people should exhibit correspondence in trust towards media. That is to say, social trust should equal media trust, or to borrow the notation from the media equation,  $\text{media (trust)} = \text{real life (trust)}$ .

### **Third-Party Trust Online**

Further, there are shortcomings to be noted regarding how trust has been studied and reported in the literature in light of the way firms, particularly non-consumer facing firms, often operate out of the view of consumers in many digital environments. Most work theorizing the role of trust in web-based, consumer-to-firm interactions has not accounted for the range of actors that accompany the bulk of online transactions today. Instead, in the online context, theorizing the role of trust has primarily been limited to examinations where the trustor is aware of both the transaction taking place (e.g., how and when personal data is collected) and the trustee on the other end (Beldad, de Jong, & Steehouder, 2010). This focus on known (or first-party) entities may seem obvious or even necessary. But it is problematic for understanding how trust functions today given the range of third-party players that facilitate online interactions and

influence content. Simplistic understandings of online trust, that only take into account what are primarily transparent transactions between known trustor-trustee relationships, do not allow for understanding the role of integral third-party actors and their impacts on user trust and desire for content personalization. This is mostly because consumer awareness of the range of firms involved (e.g., third-party data brokers) in personal data transactions varies considerably. This makes investigating the influence of third-parties (parties who are often largely unobservable) on individual trust challenging. Further, considering only activities and trustees that consumers are readily aware of may profoundly limit current conceptions of online trust. The numerous information asymmetries that characterize the use of the web today have now complicated traditional vectors of trust online. Further, how digital marketing technologies and practices often function without the user (trustor) fully aware of the interacting firm's (trustee's) presence, nor the nature of many underlying transactions linked to users' personal data, has not been considered in assessing trust online. That is to say, consumers transact with a multitude of firms often without ever knowing about these transactions nor the firms who conduct them. Yet these transactions are just as real and tangible as their first-party counterparts and constitute relationships (e.g., social, legal, ethical) between consumers and firms despite how aware or unaware consumers are of these activities.

Prior work examining trust online has been largely limited to, for example, examining internet users' trust towards websites or apps used for e-commerce (Chen & Dhillon, 2003), news (Bucy, 2003), or social networking sites (Valenzuela, Park, & Kee, 2009). This type of online trust corresponds to trust by an individual towards a known entity and typically known activities. A trustee might be the proprietor of a website or app, the website or app itself (platform or digital object), or the type of content a website or app distributes (e.g., online news,

product reviews, posts on social media). As a result, current understandings of trust online are mainly built on analyses of trust relationships between parties who are cognizant of one another and almost always specifically where the trustor (user/consumer) is cognizant of the trustee (firm). While this is not surprising, this straightforward approach to the study of trust in online environments is now overly simplistic given the current landscape and technical architecture of most commercial platforms, especially those that support marketing content personalization. Almost without exception, third parties operate in the background of most online services and/or push and pull personal data acting in concert with platforms. As a result, most work on trust in the online setting fails to account for the influence of a number of non-consumer facing actors that now accompany most online interactions, particularly those involving online behavioral monitoring and data sharing for marketing purposes, as well as individuals' varying levels of awareness of these additional unseen trustees.

Additionally, the role of context is commonly cited as a key explanatory mechanism in determining appropriate flows of personal information, manifest in the heuristic for privacy violation referred to as "contextual integrity" (Nissenbaum, 2010). The distinction between original context and unintended- or out-of-context personal data use has also been linked to Internet users' privacy expectations (Martin, 2012, 2015). It is anticipated that individuals primarily associate targeted advertising on websites and apps with the website or app that is displaying the ad, commonly referred to in online advertising practices as *contextual targeting*. Therefore, use of information about one's family members, income, relationship status, and purchases in physical stores, for instance, to selectively present advertisements online are likely perceived as out-of-context uses of this consumer information. As these data are not typically thought to be essential for interaction with most websites and apps. However, despite the role

played by context in how consumers react to uses of information that describe them, it stands to reason that trust—being the social and economic lubricant it is—may alleviate some of these concerns commonly associated with out-of-context use of personal data, as is common in personalized online advertising.

### **Transparency in Advertising Personalization**

Findings describing the effects of transparency in online advertising are scarce. However, recent experimental work by Kim et al. (2015) suggests consumer preference for how personal data is used to tailor online ads is nuanced. Researchers found that, depending on the source of the consumer data used (first- vs. third-party), providing greater transparency in the ad personalization process backfired when using third-party consumer data. This resulted in individuals reacting less favorably towards advertisement that were known to have been personalized when consumers were aware of how the ad personalization was achieved (e.g., first- vs. third-party data). Kim et al. also found that individual preferences for personalization varied depending on whether consumers believed online ads had been personalized for them based on their own user-provided information (i.e., information users entered on their account profiles, such as age or interests) vs. information that a website claimed had been inferred about them (i.e., inferred age or inferred interests). Researchers found nuanced shifts in participants' perceptions of online ads when informing them that they were seeing (an identical) online ad to buy artwork either because of their *stated interest* in oil paintings vs. an example stating the website had *inferred* they were interested in oil paintings. Revealing that an ad was based on inferred interests proved less persuasive than ads said to be generated from participants' stated interests.



Similar findings have been observed in the U.S. in public opinion surveys, however, these studies lack the degree of control and internal validity afforded by experimental methods (Mutz, 2011). For instance, in a survey of Americans researchers found people to be more opposed to online advertising personalization than favor it (Turow et al., 2009). Interestingly, in the majority of cases, researchers also found that informing survey respondents about some of the techniques used by marketers to personalize online ads including personal data collection across websites and apps lowered respondents approval of personalized advertising.

Additionally, outside the advertising context and in regards to having one's behavior predicted generally, Ybarra et al. (2010) found that individuals exhibited an aversion to being predicted especially when they thought they were going to be predicted by a competitive rather than a friendly entity. As attitudes toward advertising are generally negative and consumers skeptical of marketing messages overall, marketing activities that aim to predict consumers' interest may be perceived as coming from a competitive rather than cooperative source. This might further explain some of the reason why, at times, participants express distrust and negative attitudes towards advertising personalization, reporting they prefer not to have ads personalized based on information about them. Other problems attributable to the lack of transparency in personalization may be linked to confusion, as consumers appear to express conflicting attitudes and responses in describing preferences whether they wish to have ads personalized or not. Accordingly, individuals often describe personalized advertising in ways that signal conflicted or unsettled preferences for this practice, for instance, as "useful," "smart," "scary," and "creepy," all at the same time (Ur et al., 2012).<sup>30</sup> Further, given their common link to consumer data,

---

<sup>30</sup> I also observed these conflicting descriptions of personalized advertising when running focus groups on this same topic, as discussed in Chapter 4.

preferences for ad personalization may be related to privacy preferences, which have been shown to be incredibly malleable (Acquisti, John, & Loewenstein, 2013).

## **OVERVIEW OF RESEARCH**

Four experiments were conducted to examine the relationships between transparency, trust, and consumer preference for personalized advertising. Experiments 1 and 2 varied participants' levels of awareness regarding how ads are personalized online, or simply altered transparency of this ad personalization process, while Experiments 3 and 4 varied participants' levels of social trust. Similar outcome measures were used across the four experiments to examine the possible effects of transparency and social trust on concepts of interest, including preference for advertising personalization and trust in advertisements, websites, and apps.

In the two transparency experiments, Experiments 1 and 2, participants were asked to rate how acceptable they found different types of consumer data marketers routinely use to selectively target, or personalize, advertisements across websites and apps. In the first transparency experiment, using a within-groups repeated-measures design, all participants were asked to evaluate the practice of advertising personalization, generally doing so before and after engaging with examples of consumer information used by marketers to personalize ads online which provided the manipulation. In the second transparency experiment, using a two-group between-subjects design, this time half the participants engaged with examples of consumer information used by marketers to personalize ads online while those in a control group did not. This served to examine the effect of ad personalization transparency on several outcomes beyond those investigated in the first transparency experiment, including individual trust in websites and apps as well as both stated and revealed consumer preference for personalized advertising.

Participants in Experiment 1 came from an undergraduate participant pool (N = 412) while those in Experiment 2 were recruited on MTurk (N = 1,558).

In the two trust experiments, Experiments 3 and 4, with each using a three-group between-subjects design, participants' levels of social trust were manipulated. Both experiments used the identical social trust manipulation. In this manipulation, participants in one group were made to feel more trusting towards other people, those in a second group less trusting towards other people, and a third control group whose participants were not manipulated. This served to examine the effects of social trust on several outcomes of interest previously explored in the transparency experiments, including stated and revealed preference for personalized advertising, online privacy concern, and trust in websites and apps. Participants in Experiment 3 (N = 1,181) and Experiment 4 (N = 883) were both recruited on MTurk.

For comparative purposes, in parts these experiments employ identical or very similar outcome measures. Some measures correspond to previously validated scales, such as those for online privacy concern (Buchanan et al., 2007) and trust in online firms (Bhattacharjee, 2002). Additionally, as some of the novel measures developed for this study appeared to be effective in tapping constructs of interest (though others less so), an ancillary contribution of this study is the creation of multiple new measures related to advertising personalization; that is, novel scales appearing effective in the reported experiments offer an initial step to assess their validity, though further testing is needed to gauge their effectiveness.

## EXPERIMENTS 1 & 2

### **The Effect of Transparency on Consumer Attitudes Toward Advertising Personalization**

Given the general lack of transparency for internet users into which of their consumer data is used to personalize the online advertisements they see, combined with relatively low levels of awareness on the part of internet users for how marketers use these data to selectively present ads on many popular websites and apps today, in Experiments 1 and 2 the following research question is posed:

*How does increasing awareness about ad personalization effect preferences for personalized advertising and perceptions of web platforms that deliver personalized advertising?*

Accordingly, Experiment 1 tests the following four hypotheses:

- H<sub>1</sub>: Increased transparency in online advertising personalization practices will decrease stated preference for online advertising personalization.
- H<sub>2</sub>: Increased transparency in online advertising personalization practices will increase opposition to consumer data collection.
- H<sub>3</sub>: Increased transparency in online advertising personalization practices will decrease consumer data control self-efficacy.
- H<sub>4</sub>: Increased transparency in online advertising personalization practices will increase online privacy concern.

Then, using a more robust experimental design to replicate selected results from Experiment 1 (H<sub>1</sub>) and further assess the effects of increased transparency on not only stated preference but also revealed preference for ad personalization, in addition to trust, Experiment 2 tests the following three hypotheses:

H<sub>1</sub>: Increased transparency in online advertising personalization practices will decrease stated preference for online advertising personalization. (repeated in Experiment 1)

H<sub>5</sub>: Increased transparency in online advertising personalization practices will decrease revealed preference for online advertising personalization.

H<sub>6</sub>: Increased transparency in online advertising personalization practices will decrease trust in commercial websites and apps.

### *Manipulation of Transparency in Advertising Personalization*

In Experiments 1 and 2, the manipulations involved increasing participants' awareness about the use of consumer data to personalize online advertisements, thereby introducing greater transparency in how this process functions. Importantly, to ground these experiments in contemporary marketing practices, for their respective (similar) manipulations it was important to present participants with types of personal data *currently used* to personalize online advertisements, as opposed to fictitious examples. This decision was made to maximize the likelihood that any effects of the manipulations—of increasing transparency in how online advertising personalization works—would be directly linked to and consistent with current marketing practices. This is noteworthy because in Experiments 1 and 2 it would be far simpler to achieve significant outcome effects by simply using much more sensationalist manipulations employing examples of personal data that are likely to be the most inflammatory to participants. Instead, to increase the external validity of the designs, the manipulations for online ad transparency in Experiments 1 and 2 reflect current targeted advertising practices. In Experiment 1, participants were asked to rate 75 examples of personal data all of which came directly from current ad-buying web platforms located in a prior observational study (as reported in chapter 3 on real-time bidding ad-buying platforms). These varieties of consumer information presented to participants in Experiment 1 are readily available to marketers for delivering personalized ads

across websites and apps. Experiment 2 uses a similar manipulation but instead, for efficiency, reduces the number of examples of consumer data participants rate from 75 to only 15, along with using a between-subjects design. Further, consumers are typically unaware that many of these kinds of personal data are used to selectively present advertisements online and, when aware of some of these examples, tend to think narrowly about data corresponding to prior websites visited and search engine queries (as previously discussed in chapter 4, based on findings from focus groups). Therefore, asking participants to rate these data provides a way to measure the effects of *increased* awareness of online advertising personalization practices, as it is reasonable to think this awareness can be easily increased due being relatively low to begin with for most participants. Additionally, due to its between-subjects design, Experiment 2 included a manipulation check the results of which also lend support to the similar manipulation used in Experiment 1. These manipulations are presented in detail in their respective experiments below.

## **Experiment 1**

### **Participants**

A survey experiment was administered online during the fall of 2013 and spring of 2014 to students enrolled in a communications course at a large Midwestern university ( $N = 412$ ). Participants received partial course credit in exchange for their participation. Among participants, 72% were female and the average age was 19 ( $SD = .85$ ). The sample consisted primarily of participants identifying as Caucasian (79%), Asian American (11%), and Bi-racial/Multi-racial (5%). Participants also identified as Catholic (27%), Christian (27%), Jewish (26%), and no religion/other (20%).

## Design and Procedures

A within-subjects repeated measures survey experiment was administered to anonymous participants through a web browser. The experiment examined the effects of increased transparency in online ad personalization on stated preference for advertising personalization, online privacy concern, data control self-efficacy, and opposition to consumer data collection. Survey questions measuring these outcome variables were asked twice: once before participants completed a manipulation survey, which asked them to rate the acceptability of various types of consumer data for use in online advertising, and once again after taking the survey.

The consumer data rating exercise supplying the manipulation asked participants to rate 75 different examples of personal data. As discussed previously, the examples of consumer data rated by participants in this manipulation survey were taken from current ad-buying interfaces. All participants were asked a set of questions related to the outcome measures both before and after rating the 75 types of consumer data. Additionally, prior to rating the examples of consumer data, participants were asked first shown a brief prompt to explain the ratings exercise and also to underscore that participants were being asked to rate examples of data currently used by marketers to personalize online ads (e.g., that the examples were real and not made up).

Participants were first prompted:

*For each type of information, indicate how acceptable it is to you for marketers to use this information to personalize the online ads you see.*

*Note: These types of information are currently used by marketers to determine who sees which ads online.*

Participants were then repeatedly asked the same question appearing below while varying the example of consumer data in each question. All 75 examples of consumer data rated by participants appear in Appendix D. Participants were asked, for instance:

You see an online ad based on: *grocery products you've purchased*

Response options for all consumer data acceptability questions were: *totally unacceptable, unacceptable, slightly unacceptable, neutral, slightly acceptable, acceptable, perfectly acceptable*. The 75 short questions were presented in randomized order and across five pages of 15 examples each during in the survey experiment. Additionally, due to the large number of questions, at the top of each of the five pages of questions respondents were shown the identical prompt from earlier to remind them they were to choose how acceptable or unacceptable they found it to be for marketers to use each type of data to personalized the ads they see online and, importantly, that these examples they were rating were *currently used* by marketers to selectively present online advertising. Responses to these 75 items supplied the manipulation rather than questions of primary interest to the experiment.

Scores on the pre- and post-survey questions measuring stated preference for ad personalization, online privacy concern, opposition to consumer data collection, and consumer data control self-efficacy were compared using paired samples t-test to indicate whether the manipulation—rating the 75 examples of consumer data for acceptability in ad personalization—had a corresponding effect on these measures of interests. Scales for all outcomes measures are described in detail in Appendix E.



## Results

### *Descriptive Statistics*

As expected, there was strong correlation between individuals' responses to identical pre-test/post-test questions asked before and after rating the 75 examples of consumer data. Scores for stated preference for personalized advertising before ( $M = 4.66$ ,  $SD = 1.18$ ) and after ( $M = 4.37$ ,  $SD = 1.17$ ) were correlated,  $r(407) = .64$ ,  $p < .001$ . Levels of online privacy concern were also correlated before ( $M = 5.59$ ,  $SD = 1.08$ ) and after ( $M = 5.59$ ,  $SD = 1.15$ ) rating the consumer data examples,  $r(406) = .65$ ,  $p < .001$ . Likewise, consumer data control self-efficacy was correlated before ( $M = 3.49$ ,  $SD = 1.46$ ) and after ( $M = 3.36$ ,  $SD = 1.37$ ) the manipulation,  $r(409) = .45$ ,  $p < .001$ , as was opposition to consumer data collection before ( $M = 5.19$ ,  $SD = 1.16$ ) and after ( $M = 5.38$ ,  $SD = 1.11$ ) rating the examples of consumer data,  $r(407) = .56$ ,  $p < .001$ . This correlation between pre-test/post-test measures suggests internal validity, increasing confidence in conclusions based on subsequent tests for experimental effects from the paired samples t-test when using a within-subjects repeated measures experimental design (Warner, 2013, p. 908).

For responses to the survey where participants rated the 75 types of consumer data as acceptable or unacceptable for use in advertising personalization, which supplied the experimental manipulation, participants were consistent in their ratings. That is, individual responses were internally consistent across the 75 items ( $\alpha = .97$ ). Taken as a 75-item scale of sorts, measuring acceptability of all the examples rated, individual participant responses were highly reliable. Though as Cortina (1993) cautions, Cronbach's alpha will be inflated given a large enough number of items, as was likely the case with 75 questions. Regardless, while participants' approval ratings for the 75 data examples were internally consistent, still, the average rating for the 75 items combined from each participant did vary ( $M = 3.94$ ,  $SD = .87$ ).

As expected, participants did not find the consumer data examples to be overwhelmingly acceptable. At first glance, the average combined rating for all 75-items of 3.94 (7-point scale) hovering around the midpoint may appear rather forgiving towards marketers' use of these varieties of consumer data to personalize online ads. However, closer inspection reveals that among the 75-items participants rated, 21 of these 75 examples of consumer data had a median rating of 3 or less corresponding to either *slightly unacceptable*, *unacceptable*, or *totally unacceptable* for these 21 of the items rated. This disapproval for some examples of consumer data—even among a college student experimental sample most likely to be on average *more accepting* than the general population—illustrates the sometimes contentious nature of using these types of data to personalize online advertisements. Again, as all examples rated by participants came from current online ad-buying platforms, meaning these consumer data options are currently available in one form or another for ad targeting, practitioners and policymakers might take note of the disapproval voiced by participants, in some cases quite strongly (e.g., *totally unacceptable*), of what amount to everyday uses of consumer data in online advertising.

### *Experimental Effects*

Despite strong correlation between all pre-test/post-test measures, examining the experimental effects of increased participant awareness of consumer data use in ad personalization reveals effects on stated preference for personalized advertising and opposition to consumer data collection, confirming H<sub>1</sub> and H<sub>2</sub>. First, for personalized advertising stated preference, after rating the acceptability of the 75 types of consumer data participants were less likely to say they wanted the online ads they see to be personalized for them,  $t(408) = -5.81, p < .001, d = -.41$ . Increased awareness of consumer data use in personalized online advertising

appears to diminish support for this practice. Second, increasing awareness of advertising personalization also caused participants to assert greater opposition to consumer data collection,  $t(408) = 3.36, p < .001, d = .23$ . Similar to causing lower support for personalization, injecting transparency in this process also caused participants to be more resistant to the underlying data collection activities necessary for some ad personalization efforts to function. Both of these findings suggest that informing consumers about some of the ways consumer data is used by marketers to personalize online ads, and thereby injecting greater transparency in the ad personalization process, is detrimental to how consumers feel about these personalization practices. Simply stated, alerting individuals to some of the ways ads are personalized causes them to disapprove of online ad personalization at greater rates.

When examining if participants felt more or less empowered to control how their personal information is used by marketers after seeing how this data is used, differences between pre-test/post-test scores only approached but did not reach significant differences:  $t(410) = -1.73, p = .08, d = -.12$ . Therefore, Experiment 1 failed to reject the null hypothesis for  $H_3$ . Despite inconclusive results, participants' responses for this measure did change in the negative direction, indicating, as anticipated, they may have felt *less* in control of their data rather than more after rating the 75 types of consumer data. Additional testing is needed to determine whether ad personalization transparency, and specifically greater openness about consumer data use in ad targeting, effects consumer data control self-efficacy. Additionally, as no established scales for consumer data control self-efficacy could be located, this represents an opportunity for scale development to measure this construct that may be of additional interest to researchers studying related problems in marketing, interface design, privacy, and public policy issues.

Finally, despite being less inclined to receive personalized advertising online and more opposed to consumer data collection upon rating the 75 types of consumer data, this manipulation had no measurable effect on participants' online privacy concerns,  $t(407) = 0, p = 1.00$ . Thus, Experiment 1 failed to reject the null hypothesis for  $H_4$ . Though, contrary to the other dependent measures, which were likely less established constructs for participants prior to entering the experiment than privacy concern, the absence of an effect on online privacy concern could in part be attributable to the unique characteristics of the sample. Unlike the other three experiments reported in this study, the sample for Experiment 1 was composed entirely of college undergraduate students. Additional testing on a more diverse sample may yield an effect of transparency on online privacy concern. This remains as an empirical question requiring additional study. In addition to sample characteristics that may nuance effects on privacy concern, Experiment 1 also relied on a single-item measure for online privacy concern, which may also have suffered from measurement insensitivity. For these reasons, when designing Experiment 3 examining the effects of social trust (reported below) a validated 16-item scale tapping online privacy concern, developed by Buchanan et al. (2007), was used instead of this single-item measure used in Experiment 1.

## **Experiment 2**

### **Participants**

A survey experiment was administered online during the spring of 2016 to anonymous participants recruited through MTurk ( $N = 1,558$ ). Participants were paid the federal minimum wage rate at the time of their participation (\$7.25/hour) prorated for the estimated time to taken to complete the survey. Among participants, 47% were female. Participants were diverse in their

age: 18-24 (11%), 25-29 (22%), 30-34 (21%), 35-44 (23%), 45-54 (12%), 55-64 (8%), 65-74 (2%), 75-84 (<1%). Participants were relatively well-educated as highest level of education achieved was: less than high school diploma (1%), high school diploma or GED (12%), technical, trade or vocational training beyond high school (4%), some college or Associate's degree (33%), bachelor's degree (35%), post-graduate training or professional degree (14%). Participant annual income was distributed as follows: <\$25K (24%), \$25-50K (30%), \$50-75K (23%), \$75-100K (11%), \$100-150K (9%), >\$150K (3%). The sample consisted primarily of participants identifying as Caucasian (79%), in addition to Black/African American (6%), Asian/Asian American (5%), Hispanic/Latino (4%), and Bi-racial/Multi-racial (2%).<sup>31</sup>

Additionally, the MTurk platform allows requesters to only show tasks to MTurk participants based on their country of residence, using taxpayer information from the MTurk participant's account profile. This allows requesters to specify that MTurk respondents come only from the U.S., for instance. For all participants recruited through MTurk, this US-only specification was selected. Additionally, the Qualtrics online survey platform has its own built-in location metadata tool that determines a survey respondent's country of origin based on their IP address. Despite specifying US-only respondents on MTurk, the platform allowed a small proportion of participants to complete the survey from an IP address originating in a country outside the US according to Qualtrics. This discrepancy could be due to a number of unknowable factors including fraudulent MTurk accounts, legitimate MTurk account holders from the US who happened to be traveling outside the US at the time of completing the task, or the use of a VPN or other IP address proxy while completing the task. Nonetheless, to provide a more robust screening for participants only from the US, a two-tiered check was used. This involved first limiting the survey to US-only MTurk accounts and later, after data collection, removing any

---

<sup>31</sup> Where these numbers do not add to 100% this is due to the rounding error.

responses originating from an IP address outside the U.S. In this way, nearly all responses are likely to be from U.S. residents. This process was applied to all three samples recruited on MTurk for Experiments 2, 3, and 4.

## **Design and Procedures**

A two-group, between-subjects survey experiment was administered to anonymous participants through a web browser. This experiment examined the effects of increased transparency in online ad personalization on stated preference for advertising personalization, revealed preference for advertising personalization, and trust in commercial website and apps.

For those in the experimental group, survey questions measuring these outcome variables were asked following a manipulation, which asked this half of the participants to rate the acceptability of various types of consumer data for use in online advertising. The other half of the participants, those in the control group, were directed straight to the questions for the dependent variable measures.

Similar to the manipulation used in Experiment 1, in Experiment 2 half the participants rated examples of consumer data as acceptable/unacceptable for marketers to use to personalize online ads. This supplied the manipulation for Experiment 2. Additionally, and differing from Experiment 1, the list of examples of consumer data rated by participants was shortened from 75 items to 15 items. These 15 items were manually curated from the original longer list attempting to create a diverse yet much shorter set of examples of consumer data for participants to rate. Additionally, an exploratory factor analysis of participant ratings for the original 75 data items, though not reported in detail here, revealed these items to load intuitively among coherent categories of consumer data. For instance, financial-related examples among the 75 items

presented to participants (e.g., credit card type, loan balances, income) clustered on a single factor fairly well, as did many other examples of data within the 75 items, which revealed categories or factors present in this long list (e.g., demographics, online activity, media consumption). The presence of intuitive factors and corresponding redundancy among participant responses to the 75-items provided further support for the creation of the shorter, largely representative 15-item manipulation.

Also similar to Experiment 1, in Experiment 2 all examples of consumer data rated by participants were taken from current ad-buying interfaces. This decision connected the manipulation and its effects with today's marketing practices in efforts to increase the external validity of the experimental design and any potential findings.

For those in the experimental group, this short 15-question survey asked participants how acceptable they found various types of personal data to be when used by marketers to personalize the advertisements they see on websites and apps.

Those in the experimental group completing the manipulation survey were first prompted as follows:

For each type of personal information listed below, indicate how acceptable or unacceptable it is to you for marketers to use this information to personalize the online advertisements you see.

0 = This type of information is *not at all acceptable* for marketers to use to personalize the online advertisements I see. *I do not want this.*

10 = This information is *completely acceptable* for marketers to use to personalize the online ads I see. *I want this.*

Then, participants in this experimental group were asked to individually rate how acceptable or unacceptable they found it to be for marketers to use different categories of personal information to selectively present ads to them online. The 15 categories of consumer information used by

marketers to personalize online advertisements were presented to participants in randomized order. For each consumer data example, participants were asked the following question, for instance:

How acceptable or unacceptable is it to you for marketers to use the following information to personalize the online advertisements you see? *Your Precise Geographic Location (GPS latitude and longitude)*

The complete wording for each of the 15 questions, along with ratings for each of the consumer information categories are reported in Table 1 and category averages are compared in Fig. 5.1. For each question participants used an 11-point slider tool to indicate their individual acceptability rating for each type of consumer information, positioning the slider between 0 (*not at all acceptable*) and 10 (*completely acceptable*), inclusive.

Additionally, completing this survey not only supplied the manipulation for Experiment 2 but also provided an economical way to collect nearly 800 responses describing how consumers feel about having various types of data used to personalize ads online. Though designed as the experiment's manipulation, this survey's responses were not without value in their own right. Results from the manipulation survey itself offer insights about the degree to which individuals feel differently about different types of personal information used by marketers to selectively present online ads. Therefore, able to be viewed as an independent, embedded study, responses from the manipulation survey are reported in the n independent descriptive statistics below.

Outcome measures in Experiment 2 were stated preference for online advertising personalization, revealed preference for online advertising personalization, and trust in commercial websites and apps. The between-subjects design also facilitated a manipulation check to test whether those in the experimental group were more aware of online advertising personalization after rating the 15 types of consumer data at the beginning of their experiment



compared to those in the control group who did not take this manipulation survey at the beginning. This 5-item scale was developed to assess the internal validity of the transparency manipulation and questions were asked to respondents in the experimental and control groups. Individual questions examined the degree to which participants believed online ads were personalized for viewers, asking about both the frequency of ad personalization and the proportion of ads online that are personalized for individuals. It was expected that those in the experimental group, who should have an increased awareness of ad personalization practices after rating the 15 examples of consumer data used to target ads, would correspondingly believe the degree to which ads were personalized online was greater than those in the control group.

Scores on the outcome measures of stated and revealed preference for advertising personalization and trust in commercial websites and apps were all compared between those in the experimental group and those in the control group. Independent samples t-tests were used to assess whether the manipulation—rating the 15 examples of consumer data—had a corresponding effect on these outcome measures. Further details describing the manipulation check and all scales used for the outcome measures appear in Appendix E.

## **Results**

### *Descriptive Statistics*

Participants in the experimental group rated the 15 examples of consumer information at the start of the experiment and those in the control group did not. For the half of the participants who rated these examples, on average they rated some categories of consumer information as more acceptable for marketers to use to personalize online advertisements and deemed other varieties less acceptable. Participants were reliable in their ratings of the consumer data. That is,

individual responses were internally consistent across the 15 items ( $\alpha = .92$ ). Taken as a 15-item scale measuring overall acceptability of all 15 examples rated, individual participant responses were highly reliable. On the 11-point scale with 0 as *not at all acceptable* and 10 as *completely acceptable* the overall average for the 15 items was relatively low ( $M = 3.22$ ,  $SD = 2.08$ ).

Further, a closer inspection of these ratings reveals a median rating of *4 or less* for 12 of the 15 examples of consumer data rated by participants. This places the majority of examples rated by participants towards the bottom third of the 11-point scale towards the end of *not at all acceptable*. Thus, strong opposition to the examples of consumer data rated was revealed.

Additionally, disapproval expressed by participants in Experiment 2 appears much greater than the disapproval expressed by the college student sample in Experiment 1 who rated the larger (75-item) list of consumer data examples. Nonetheless, participants in Experiment 2 appear highly unfavorable towards having these examples of consumer data used to selectively present the ads they see online.

**Table 1**

Responses to manipulation survey completed by experimental group in Experiment 2.

---

*How acceptable or unacceptable is it to you for marketers to use the following information to personalize the online advertisements you see? Your...*

	<i>M</i>	<i>SD</i>
<i>...Family Members' Personal Information</i>	0.87	1.97
<i>...Credit Card Transactions</i>	1.20	2.20
<i>...Current Financial Debt (loans, credit card balances, etc.)</i>	1.22	2.16
<i>...Financial Net Worth</i>	1.75	2.56
<i>...Annual Income</i>	2.11	2.72
<i>...Place of Employment</i>	2.38	2.88
<i>...Precise Geographic Location (GPS latitude and longitude)</i>	3.70	3.42
<i>...Relationship Status (single, in relationship, engaged, married, divorced, widowed)</i>	3.84	3.31
<i>...Vacation Locations</i>	3.84	3.30
<i>...In-Store Purchase History</i>	3.91	3.40
<i>...Social Media Activity</i>	4.03	3.32
<i>...Online Purchase History</i>	4.08	3.33
<i>...Ethnicity</i>	4.26	3.44
<i>...Automobile (make and model)</i>	4.56	3.21
<i>...Gender</i>	6.60	3.10

---

*Notes.* Items were asked in randomized order. 0 = not at all acceptable, 10 = completely acceptable. N = 774.

Of the 15 categories rated by participants, broad types of financial information, such as one’s income or debt, appear to be thought of as less acceptable for marketers to use for personalizing online advertisements compared to other kinds of information, such as one’s gender or activity on social media platforms. Additionally, and unsurprisingly, information about one’s family members was rated as highly unacceptable for marketers to use in ad personalization. These differences in consumer approval are similar to what others have found when exploring individuals’ differential willingness to disclose different types of personal information (e.g., Milne et al., 2015).

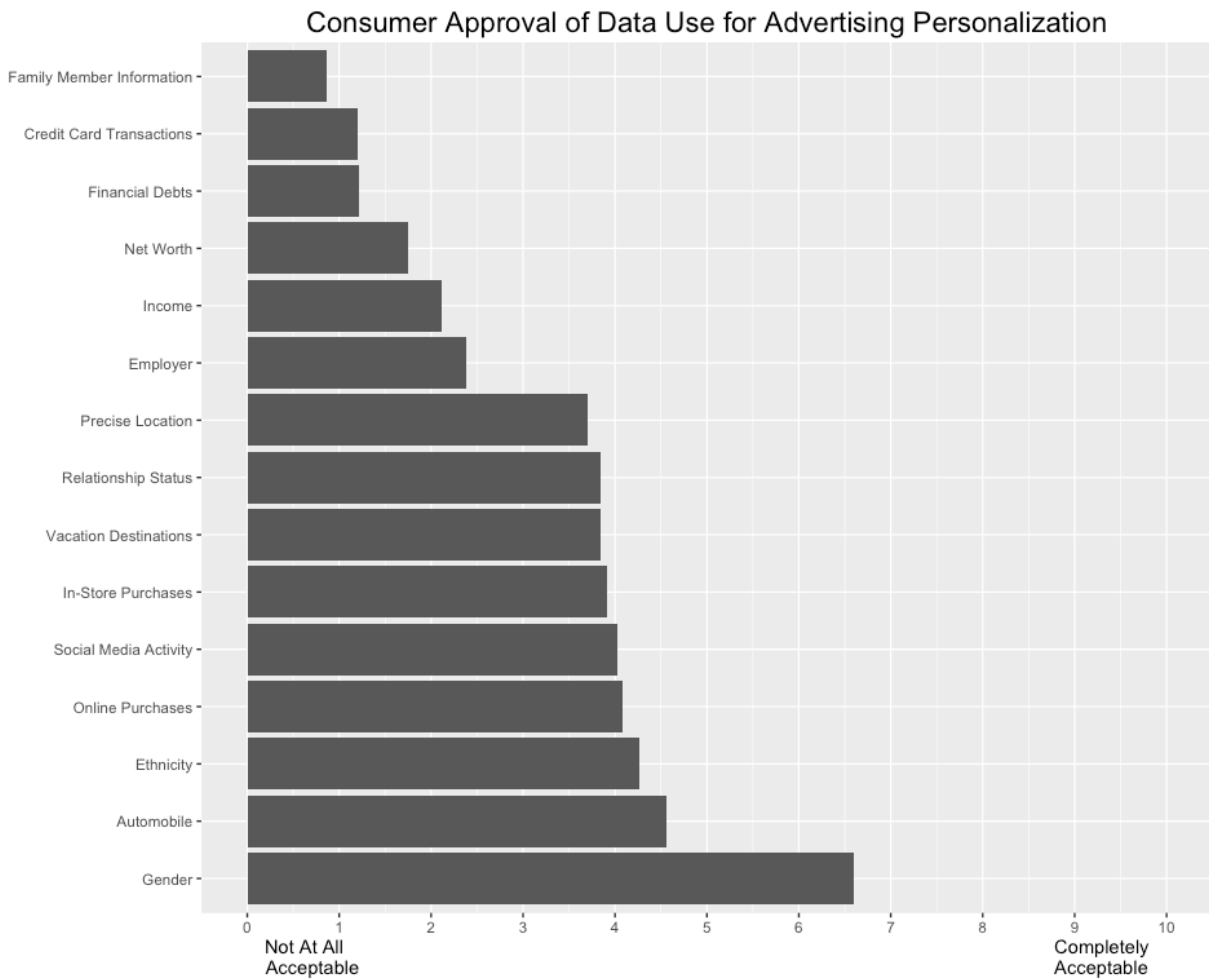


Figure 5.1. Average approval ratings of consumer data use in Experiment 2 (N=774).

### *Experimental Effects*

After the measures of stated preference for ad personalization and website-app trust were conducted, but prior to the revealed preference measure at the very end of the survey, a manipulation check was employed to assess the degree to which rating the consumer data examples had been effective in increasing awareness of online advertising personalization among those in this experimental group. The manipulation was found to be effective. As anticipated, those rating the examples of consumer data scored higher on the awareness of advertising personalization scale ( $M = 6.01$ ,  $SD = 1.58$ ) than those in the control group ( $M = 5.59$ ,  $SD = 1.09$ ),  $t(1548) = 7.34$ ,  $p < .001$ ,  $d = .26$ . This difference is shown in Fig. 5.2. These results from the manipulation survey reported for Experiment 2 demonstrate how consumers find certain types of personal information more or less acceptable, and likely more or less sensitive for using in marketing.

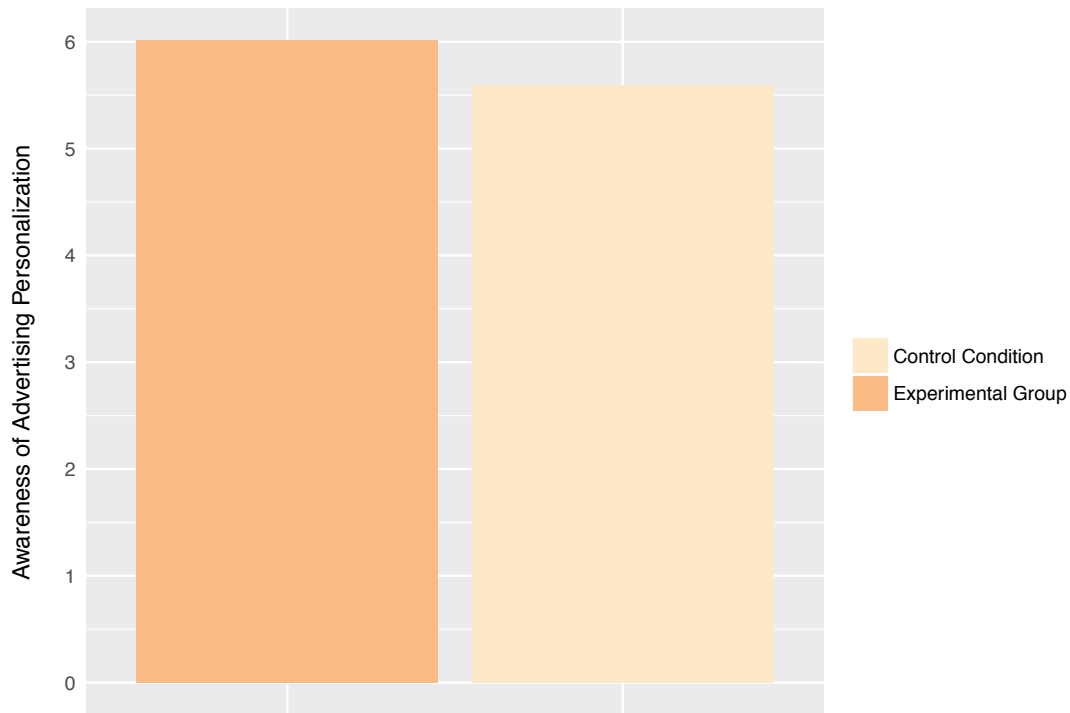


Figure 5.2. Manipulation check for Experiment 2. The effects of completing the manipulation survey on awareness of online advertising personalization. Group differences were significant,  $t(1548) = 7.34, p < .001, d = .26$ , indicating the manipulation was effective.

Examining the experimental effects of increased awareness of consumer data use in ad personalization reveals a negative effect on each of the outcome measures. Similar to Experiment 1 but this time using a between-groups rather than within-groups experimental design, in Experiment 2 participants who rated examples of consumer data used in ad targeting said they preferred to receive personalized advertising less ( $M = 3.25, SD = 1.58$ ) than those in the control group ( $M = 3.48, SD = 1.59$ ),  $t(1556) = -2.89, p < .01, d = -.10$ , once again confirming  $H_1$ . Transparency in this process appears to diminish stated preference for online advertising personalization.

Additionally, Experiment 2 examined the effects of transparency in ad personalization on revealed preference for personalized advertising. As described, this was assessed using a novel

measure where at the end of the survey participants were given the option of revealing three ads for books they believed to be personalized for them and whether they clicked on these faux-personalized ads. Participants who rated the examples of consumer data used in ad personalization were less likely to choose to see the ads they believed had been generated uniquely for them and to click on these “personalized” ads ( $M = .26, SD = .32$ ) compared to those in the control group ( $M = .29, SD = .33$ ),  $t(1555) = -2.12, p = .03, d = -.08$ , confirming H<sub>5</sub>. Thus, in addition to diminishing *stated* preference for ad personalization, increasing transparency in this process also negatively impacted *revealed* preference for ad personalization. A heightened awareness of ad personalization practices caused participants to decline the opportunity to see personalized advertising and reduced the rate at which participants clicked on the ads they believed had been generated uniquely for them.

Finally, the effects of increased awareness of ad personalization practices on participants’ trust in commercial websites and apps were examined. Participants who rated the examples of consumer data expressed lower levels of trust in websites and apps ( $M = 2.58, SD = .79$ ) compared to those in the control group ( $M = 2.70, SD = .74$ ),  $t(775) = -2.18, p = .03, d = -.11$ , confirming H<sub>6</sub>. Individuals who rated the 15 examples of consumer data for use in online ad personalized were less trusting of commercial websites and apps.<sup>32</sup>

---

<sup>32</sup> As respondents were paid by the minute and due to the sufficiently large sample in Experiment 2, the scale for website-app trust was only asked of half the participants in the control group and half the participants in the experimental group. The results for this measure indicate there was still sufficient power to detect a difference ( $p = .03$ ) in website-app trust. This is despite making only about half as many comparisons ( $N = 775$ ) for this measure as was possible for the other dependent measures ( $N = 1,555$ ). Overall, this budget-based decision proved inconsequential for the results.

### *Discussion of Experiments 1 & 2*

In Experiments 1 and 2, the effect of increased transparency in advertising personalization was examined. In Experiment 1, awareness of some of the ways marketers personalize online advertisements lead participants to be more opposed to consumer data collection. Similarly, in Experiments 1 and 2, providing transparency into how consumer data is used to personalize ads online diminished participants' stated preference for personalized advertising. In Experiment 2, increasing awareness of ad personalization practices also caused participants to feel less trusting towards commercial websites and apps and exhibit lower levels of revealed preference for personalized advertising.

Viewed together, Experiments 1 and 2 both found that transparency in how online ads are personalized diminished preference for advertising personalization. Again, this finding is similar to results reported in recent work by Kim et. al (2015) previously mentioned, who found informing people about how ads had been personalized caused individuals to select non-personalized ads when given the choice between the two. Additionally, and extending the work of these researchers, Experiment 2 also found increased awareness of how ads are personalized negatively impacted trust in commercial websites and apps. This finding suggests that when consumers are aware of some of the ways ads are personalized online they become less trusting of the websites and apps that often deliver these messages. This finding is important as, over time and as they are used more commonly, practices in use by marketers gradually become known to consumers.

Consider, for instance, the way many consumers today are likely aware that providing detailed contact information “for a chance to win a vacation” will likely result in unsolicited offers. This assumed knowledge would not have been the case when marketers begin using this



prize entry bait-and-switch technique. However, over time, consumers became more cognizant of this practice, awareness that grows out of media coverage, personal experiences (e.g., being called upon entering one's information for such a sweepstakes), and socialization with others. A similar example is seen in the more recent case of the online advertising practice referred to as re-targeting, where visiting a webpage results in seeing ads for the product or service originally viewed but later on and across other websites. This practice, as with bait-and-switch prize entry, has entered popular understanding over time. Similarly, if over time consumers become more aware of other ways advertisements are personalized online, this heightened awareness would appear to pose a problem for marketers and the websites and apps they use to generate and deliver personalized advertising. As the results from Experiment 2 suggest consumer awareness of these practices directly reduces trust individuals express towards web platforms.

This is also an important finding because, generally speaking, information transparency in the online environment tends to be taken for granted as a net positive. And few would argue that it is good nor permissible to impede consumer education when it comes to how marketers attempt to persuade consumers. Consider, for instance, consumer backlash against marketing in the 1950s-1960s linked with overreactions to so-called "subliminal advertising," a practice thought by some of the public to be in use extensively by marketers at the time (Packard, 1957; Nelson, 2008). Granting consumers access to and accurate information about how they are being persuaded is now viewed as imperative under advertising education best practices and general marketing ethics. Nevertheless, in the digital era and in a time of advertising automation and personalization beyond anything imaginable in the 1950s and 1960s, simply revealing to consumers some of the ways consumer data is used to personalize the online ads they see appears

to backfire. As people are less likely to prefer personalized ads and less likely to trust websites and apps upon learning how these media artifacts operate under the hood.

Further, as Experiments 1 and 2 demonstrated negative impacts of increased transparency in advertising personalization on both stated and revealed preference for personalized advertising while at the same time diminishing the trust consumers feel towards the media and media technologies that deliver personalized advertising, marketing policymakers and practitioners should especially take note given the push for greater transparency in these areas. Findings in Experiments 1 and 2 may encourage marketers to reconsider best practices and how techniques typically unseen to consumers might be better communicated to them. It could be that simply educating consumers about how these practices work would cause people to perceive these activities as more favorable not less. If this is the case, negative reactions to these practices may simply be a response by consumers who are confronted with information that contradicts their current understandings, the effect of surprise rather than genuine opposition. That is to say the less amazed consumers are when seeing how advertising personalization works under the hood, and perhaps the more aware they are made of the benefits of personalization (e.g., encountering irrelevant content less frequently), the more they may simply find this practice favorable. Further, improving transparency in these commercial processes—activities that usually leverage opaque algorithms dependent on detailed consumer data as inputs—is now viewed as being squarely in the interest of the public given seemingly-ubiquitous personal data (e.g., The White House, May 2016).

## **EXPERIMENTS 3 & 4**

### **The Social Trust Digital Disconnect?**

Experiments 1 and 2 demonstrated negative impacts of increased transparency in advertising personalization on both stated and revealed preference for personalized advertising while at the same time diminishing the trust consumers feel towards the media and media technologies that deliver personalized advertising. These findings further motivate investigation into the potential underlying factors leading consumers to be generally resistant to ad personalization when more aware of how this process works. Based on work demonstrating the importance of trust in facilitating online social and commercial interactions (Beldad, de Jong, & Steehouder, 2010) along with the general understanding that social trust in particular is key to successful human interaction in networked society (Petrovic et al., 2003), one reason for this resistance might be that consumers simply do not trust marketers to act in individuals' best interests whenever using varieties of detailed personal information to personalize online advertising. If this is the case, then increasing trust on the part of consumers might alleviate apprehension on the part of consumers to allow marketers to use granular consumer data to selectively present advertising. Additionally, as Reeves and Nass (1996) posited under their Media Equation that "media = real life," humans react to media and media technologies just as they react to other people. It follows then that altering how trusting consumers feel towards other people should correspondingly affect how trusting they feel towards media, including commercial websites, apps, and advertisements. Similarly, if this link between trust "in real life" connects to digital actors and artifacts, then it stands to reason that increasing social trust may also alleviate reservations about personalized online advertising causing individuals.

Additionally, if trust in other people has a direct spillover effect into how individuals feel towards digital actors and artifacts, then increasing or decreasing social trust should correspondingly influence online privacy concerns as the connection between privacy and trust is well established (Luo, 2002).

Given each of these possibilities, in Experiments 3 and 4 the following research questions were explored:

*How does increasing and decreasing consumers' levels of social trust affect feelings toward personalized advertising practices and those web platforms that deliver personalized advertising?*

Accordingly, Experiment 3 tested the following 10 hypotheses:<sup>33</sup>

H<sub>7a</sub>: Increased social trust will increase stated preference for online content personalization.

H<sub>7b</sub>: Decreased social trust will decrease stated preference for online content personalization.

H<sub>8a</sub>: Increased social trust will increase revealed preference for surreptitiously-personalized online content.

H<sub>8b</sub>: Decreased social trust will decrease revealed preference for surreptitiously-personalized online content.

H<sub>9a</sub>: Increased social trust will increase revealed preference for overtly-personalized online content.

H<sub>9b</sub>: Decreased social trust will decrease revealed preference for overtly-personalized online content.

H<sub>10a</sub>: Increased social trust will increase trust in commercial websites and apps.

H<sub>10b</sub>: Decreased social trust will decrease trust in commercial websites and apps.

H<sub>11a</sub>: Increased social trust will decrease online privacy concerns.

H<sub>11b</sub>: Decreased social trust will increase online privacy concerns.

---

<sup>33</sup> Because increasing vs. decreasing levels of social trust may produce incongruent effects on the outcome measures, all hypotheses were tested in both directions. This resulted in the three-group experimental design.

Then, as a follow-up experiment to Experiment 3, Experiment 4 also tested the following eight hypotheses:

- H<sub>10a</sub>: Increased social trust will increase trust in commercial websites and apps. (repeated in Experiment 3)
- H<sub>10b</sub>: Decreased social trust will decrease trust in commercial websites and apps. (repeated in Experiment 3)
- H<sub>12a</sub>: Increased social trust will increase stated preference for online advertising personalization. (variation of H<sub>7a</sub>)
- H<sub>12b</sub>: Decreased social trust will decrease stated preference for online advertising personalization. (variation of H<sub>7b</sub>)
- H<sub>13a</sub>: Increased social trust will increase revealed preference for online advertising personalization. (variation of H<sub>8a</sub> and H<sub>9a</sub>)
- H<sub>13b</sub>: Decreased social trust will decrease revealed preference for online advertising personalization. (variation of H<sub>8b</sub> and H<sub>9b</sub>)
- H<sub>14a</sub>: Increased social trust will increase trustworthiness of example advertisements.
- H<sub>14b</sub>: Increased social trust will increase trustworthiness of example websites/apps.

### ***Manipulations in Experiments 3 and 4***

Experiments 3 and 4 were both three-group between-subjects experiments and used the identical social trust manipulation. In both cases, participants were randomly assigned to either a positive (increased) social trust condition, negative (deceased) social trust condition, or a control group. Within their respective experiments, the three groups were equivalent in size.

Additionally, a manipulation check appeared near the end of the experiments to assess the effectiveness of attempting to make participants feel more or less trusting towards other people.

Both the manipulation and the manipulation check came from a related study by Mutz (2005), which found experimental evidence for the effect of social trust on stated intent to use e-

commerce, though only among individuals who had never before used e-commerce. Wording in the manipulation and manipulation check was minimally modified for added clarity and to account for inconsistency in the original vignettes.

### *Positive Social Trust Cue*

For the positive social trust manipulation, participants read the following four short vignettes describing a selected portion of a report by Reader's Digest, modified slightly for consistency in the manipulation. The vignettes were presented across four pages (four paragraphs below, each shown on a new page). Respondents were instructed that reading the vignettes carefully was necessary to answer the remaining questions in the survey.

*In a recent study done by Reader's Digest magazine, its employees purposely dropped thousands of wallets each containing \$50 in cash and an identification card with a name and phone number, so that the finder would have no trouble returning the wallet—presuming the finder wanted to return it. The wallets were left on sidewalks and benches, in front of office buildings, discount stores, and churches, in parking lots and in restaurants. Then observers watched and waited to see what would happen.*

*Reader's Digest repeated this test in big cities and small towns across the United States, and then all over the world, to see how many people would keep the wallet with the money, and how many would do the right thing and try to return it. Interestingly, the overwhelming majority of people who found the wallets tried to return them to the owner. Surprised by the honesty of so many people, Reader's Digest interviewed many of them. Some who handed back the wallets cited their religious beliefs as what compelled them to act. Others pointed to their upbringing and the emphasis their parents put on honesty.*

*Mary, a little girl in a pink floral dress, found a wallet on a bench in a Seattle park. She ran to her father, Yong Cha, who immediately handed it back to her. "You must take this to someone who can help find the owner," he said. The nine-year-old took her dad's hand and they walked to the park's office. "Honesty is the most important thing a child can learn," Cha said.*

*Time and again, around the U.S. and all over the world, even those who possibly could use an extra \$50 often turned it in. Consider Dirk Engel, who works as a restaurant waiter in Keokuk, Iowa. After handing in the wallet, he said, "I put in long hours and I know how hard people work to earn that much money."*

Then, immediately after reading the positive trust text vignettes in the survey experiment, on the next page of the survey participants were asked to respond to a set of reinforcement questions intended to further increase levels of social trust. As these questions served only to strengthen the positive social trust manipulation, participants responses were discarded.

Participants in this group were asked the following questions:

*Do you think positive news stories like the Reader's Digest experiment get the kind of news attention they deserve, or does the news tend to overemphasize the bad things that people sometimes do to each other?*

Response options were:

*Good things that people do get the attention they deserve in the news.  
The news puts too much emphasis on the bad things people sometimes do.*

Then, participants in this group were asked:

*Why is it that people often go out of their way to help a total stranger?*

*For each of the explanations below, indicate whether you think it is part of why people act this way by selecting Yes or No.*

*People often go out of their way to help a total stranger because of.. [followed by]*

*...parental emphasis on moral values in children's upbringing.  
...a cultural practice of treating others the way they would want to be treated themselves.  
...feeling a connection with fellow humans, even complete strangers.  
...religious beliefs that emphasize an ultimate reward or punishment for actions while on earth.  
...schools' emphasis on codes of good conduct.  
...human nature.*

Response options were: *Yes, No.*

### *Negative Social Trust Cue*

For the negative social trust manipulation, participants in this group read the following four short vignettes describing a selected portion from the same report by Reader's Digest,

modified slightly for consistency in the manipulation. The vignettes were presented across four pages (four paragraphs below, each shown on a new page.) and respondents were instructed that reading the vignettes carefully was necessary to answer remaining questions in the survey.

*In a recent study done by Reader's Digest magazine, its employees purposely dropped thousands of wallets each containing \$50 in cash and an identification card with a name and phone number, so that the finder would have no trouble returning the wallet—presuming the finder wanted to return it. The wallets were left on sidewalks and benches, in front of office buildings, discount stores, and churches, in parking lots and in restaurants. Then observers watched and waited to see what would happen.*

*Reader's Digest repeated this test in big cities and small towns across the United States, and then all over the world, to see how many people would keep the wallet with the money, and how many would do the right thing and try to return it. Disappointingly, a large number of the wallets simply vanished, with no effort ever made by their finders to return them to the rightful owner listed on the identification card.*

*Even when the wallets were found by people who didn't appear to need the money at all, many kept them anyway. For example, in an upscale resort town, a well-dressed woman in stiletto heels was walking hand in hand with her daughter. The woman stooped over to grab the wallet. With the young girl looking on silently, the mother removed the cash and placed the money in her pocket. She then sat the empty wallet back on the ground and they continued walking.*

*And then there was the man who pulled his luxury car up to the entrance of a palace in London, who jumped out and snatched the wallet. Back in the car, he picked through the wallet carefully, removing only the cash before driving through the palace gates never to be heard from again. On another occasion, at least two apparently devout Christians who kept the wallets made the sign of the cross after picking them up and noticing the cash. The money, they must have decided, was heaven-sent, despite the identification card of the owner. Overall, the Reader's Digest study confirmed that many people of all sorts of backgrounds can be untrustworthy when they think no one is watching.*

As with the positive social trust manipulation, immediately after reading the negative trust text vignettes, on the next page of the survey participants were asked to respond to a set of reinforcement questions intended to further reduce levels of social trust. As these questions



served only to strengthen the negative social trust manipulation, participants responses were discarded. Participants in this negative social trust group were asked the following questions:

*Do you think news stories like this one from the Reader's Digest experiment, which reveal how people often behave badly when they think nobody is watching, teach people a valuable lesson?*

Response options were:

*Stories revealing people behaving badly teach people a valuable lesson.*

*Stories revealing people behaving badly do NOT teach people a valuable lesson.*

Then, participants in this group were asked:

*Why do many people tend to be untrustworthy when they think nobody is watching?*

*For each of the explanations below, indicate whether you think it is part of why people act this way by selecting Yes or No.*

*People tend to be untrustworthy because... [followed by:]*

*...of not enough emphasis on moral values in the school*

*...of a cultural emphasis on wealth and consumption*

*...people benefit financially from being dishonest*

*...of absentee parents due to many single-parent families*

*...of not enough emphasis on religion in the family*

*...of human nature*

Response options were: *Yes, No.*

### *Control Condition*

Experiments 3 and 4 also each included a third control group. Participants randomly assigned to this condition did not read the short vignettes nor answer the corresponding short set of reinforcement questions. Instead, following the consent page in the survey, participants in the control group proceeded directly to the first set of questions measuring the dependent variables.

Additionally, pilot testing revealed that participants read these manipulation vignettes and answered the list of questions afterward so quickly that there was no need to add an additional

vignette for the control conditions, as differential survey fatigue was a non-issue. While some advocate for strict parity in this design the case against parity is equally compelling. As any text read by the control condition would have had its own effect on participants, which could not be accounted for in this design and further justifying use of an unaffected control group in the purist sense. Results from the manipulation checks in both Experiments 3 and 4 provide further, post-hoc support for this experimental design decision.

### **Experiment 3**

#### **Participants**

A survey experiment was administered online during the spring of 2016 to anonymous participants recruited through MTurk (N = 1,181). Participants were paid the federal minimum wage rate at the time of their participation (\$7.25/hour) prorated for the estimated time to taken to complete the survey. Among participants, 45% were female. Participants were diverse in their age: 18-24 (14%), 25-29 (20%), 30-34 (19%), 35-44 (23%), 45-54 (12%), 55-64 (9%), 65-74 (3%). Participants were relatively well-educated as highest level of education achieved was: less than high school diploma (<1%), high school diploma or GED (12%), technical, trade or vocational training beyond high school (4%), some college or Associate's degree (33%), bachelor's degree (36%), post-graduate training or professional degree (15%). Participant annual income was distributed as follows: <\$25K (26%), \$25-50K (30%), \$50-75K (22%), \$75-100K (11%), \$100-150K (8%), >\$150K (4%). The sample consisted primarily of participants identifying as Caucasian (78%), in addition to Black/African American (7%), Asian/Asian American (6%), Hispanic/Latino (5%), and Bi-racial/Multi-racial (3%).<sup>34</sup>

---

<sup>34</sup> Where these numbers do not add to 100% this is due to the rounding error.

## Design and Procedures

A three-group between-subjects survey experiment was designed to examine the causal influence of social trust on stated preference for online content personalization, revealed preference for surreptitiously-personalized online content, revealed preference for overtly-personalized online content, trust in commercial websites and apps, and online privacy concern. As previously described, participants were randomly assigned to equally sized groups corresponding to either positive social trust, negative social trust condition, or a control group.

For revealed preference for content personalization, individuals were assessed in their choices of which short videos they chose to watch given a series of selections between two videos, one of which was personalized based on information about the participant. In the first half of the video selections presented to respondents, the personalized video was not pointed out. In the second half of participants' video selections, the personalized video options were explicitly indicated to them. This additional split in revealed preference for content personalization was designed to assess whether social trust influenced revealed preference for personalization *differently* when participants were and were not aware that one of the options had been recommended for participants based on information about them, as related work has demonstrated that being aware of content personalization can influence its desirability (Kim et al., 2015). This additional nuance to the videos options questions resulted in two separate measures of revealed preference for online content personalization, one for surreptitious personalization and one for overt personalization.

Scores on these outcome measures were compared using one-way ANOVA to examine whether increasing or decreasing social trust in participants had a substantial effect on stated and revealed preference for content personalization, trust in websites/apps, and online privacy

concern. Additionally, a manipulation check measuring levels of social trust was used to assess the effectiveness of the manipulation. Scales for all outcomes measures are described in detail in Appendix E.

## Results

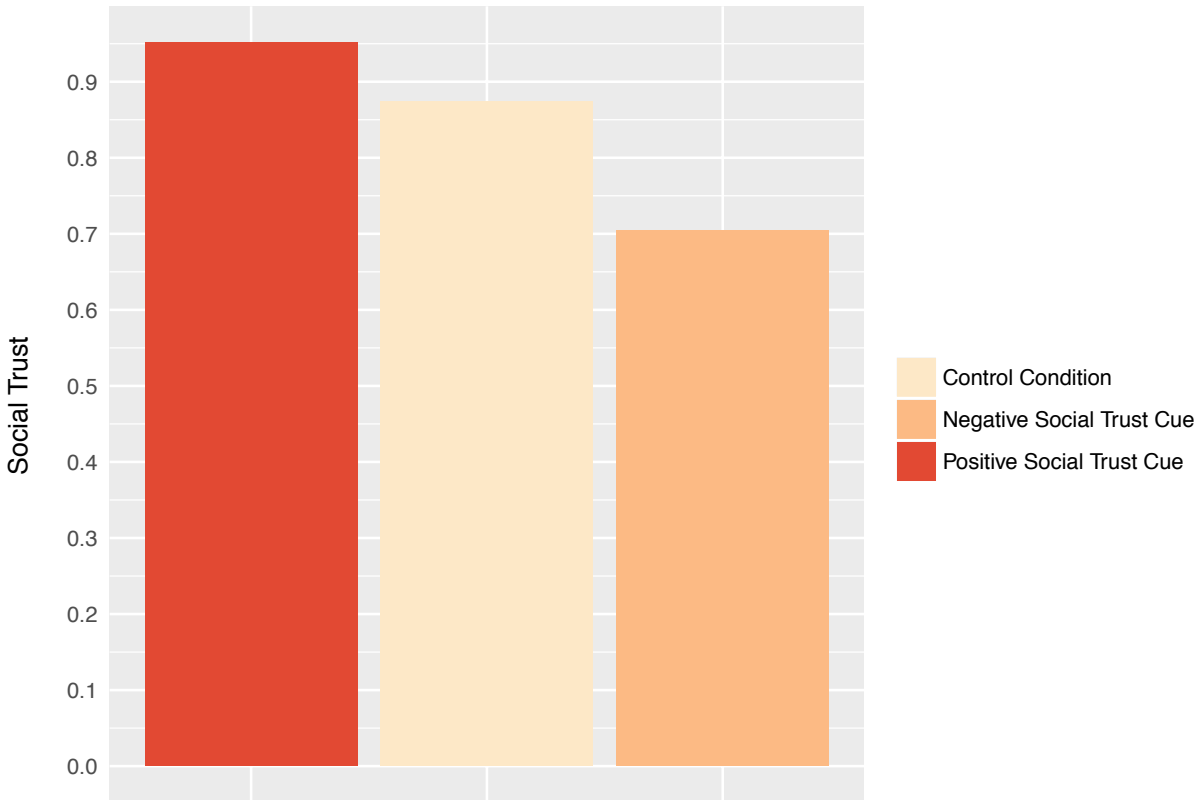
A manipulation check was employed to assess the effectiveness of the social trust cues at the beginning of the experiment. The manipulation was found to be effective across the three groups. Those in the positive social trust condition exhibited elevated levels of social trust ( $M = .95$ ,  $SD = .42$ ) compared to those in the control group ( $M = .87$ ,  $SD = .43$ ). Those in the negative social trust condition expressed decreased levels of social trust ( $M = .70$ ,  $SD = .42$ ) compared to the control group. These observed differences in levels of social trust between the three groups were highly significant,  $F(2, 1178) = 34.98$ ,  $p < .001$ ,  $\eta^2 = .06$ , attesting to internal validity of the experimental treatments.

Additionally, and further underscoring the effectiveness of the manipulation in Experiment 3, the effect on participants levels of social trust was twice that observed by Mutz (2005) in the study for which she developed this social trust manipulation and corresponding manipulation check measuring increased/decreased levels of social trust.<sup>35</sup> Notably, Mutz still found a significant effect of social trust on consumers' stated intent to use ecommerce. Though this effect was only detectable among a subgroup, those who had never made an online purchase. Nonetheless, for Experiment 3, for participants in the positive and negative experimental groups

---

<sup>35</sup> For instance, effects detected in the manipulation check reported by Mutz (2005) were sufficiently present,  $F(2, 814) = 12.63$ ,  $p < .001$ ,  $\eta^2 = .03$ , yet apparently less effective than those observed here in Experiment 3. There is likely a ceiling effect on the degree to which social trust can be influenced using this technique and one should proceed with caution when comparing effect sizes for manipulations. Interestingly, the sample in Mutz's study was a probability sample of the U.S. population recruited by Knowledge Networks (Menlo Park, CA) while I recruited participants in Experiment 3 from MTurk. Determining whether MTurk recruits are more easily influenced compared to a general population sample, as these results suggest, would require further investigation.

levels of social trust were momentarily increased or decreased, respectively. These differences are shown in Fig. 5.3 below.



*Figure 5.3.* Manipulation check for Experiment 3. The effects of the social trust cues on levels of social trust in participants. Group differences were significant,  $F(2, 1178) = 34.98, p < .001, \eta^2 = .06$ , indicating the manipulation was effective.

Upon determining the social trust manipulations were effective, assessing the effects on the dependent variables revealed somewhat surprising results for Experiment 3. Across the board, altering how trusting participants felt towards other people had no significant effects on how trusting they felt towards commercial websites and apps, online privacy concern, nor revealed preference for surreptitiously- and overtly-personalized content.

Despite the conceptual link of trust between trust directed at other people and trust directed at websites and apps, no corresponding effect was observed in either direction. Altering how trusting participants feel towards other people had no causal effect on how trusting they found commercial websites and apps to be. Comparing the group mean differences using ANOVA, differences in trust towards websites-apps between individuals feeling increased social trust ( $M = 2.88, SD = .78$ ), the control group ( $M = 2.97, SD = .77$ ), and those feeling decreased social trust ( $M = 2.91, SD = .70$ ), were not significant,  $F(2, 1178) = 1.32, p > .05$ , thus failing to confirm  $H_{10a}$  and  $H_{10b}$ .

Similarly, levels of social trust had no significant effect on individuals' reported online privacy concerns when comparing the positive trust ( $M = 2.95, SD = .90$ ), control ( $M = 2.90, SD = .90$ ), and negative trust groups ( $M = 2.97, SD = .89$ ),  $F(2, 1178) = 0.78, p > .05$ , failing to confirm  $H_{11a}$  and  $H_{11b}$ . Notably, this apparent lack of effect of social trust on online privacy concern, and when using a validated 16-item scale (Buchanan et al., 2007) for this construct, is highly counterintuitive given both commonsense expectations and empirical associations between privacy and trust (e.g., Luo, 2002). Additional work examining the relationship between social trust and privacy concern is warranted.

Neither of the revealed preference for content personalization measures were significantly different among the social trust conditions either. Group differences for revealed preference for surreptitiously-personalized content among the positive social trust ( $M = .86, SD = .19$ ), control ( $M = .87, SD = .19$ ), and negative social trust conditions ( $M = .86, SD = .18$ ) were not significantly different,  $F(2, 1176) = 0.32, p > .05$ , failing to confirm  $H_{8a}$  and  $H_{8b}$ . Nor were differences between positive ( $M = .90, SD = .18$ ), control ( $M = .91, SD = .15$ ), and negative ( $M =$

.90,  $SD = .17$ ) conditions for the measure used to assess revealed preference for overtly-personalized content,  $F(2, 1176) = 0.62, p > .05$  failing to confirm  $H_{9a}$  and  $H_{9b}$ .

Besides the manipulation demonstrating clear effects on how trusting participants felt towards other people, as identified in the manipulation check, the only outcome on which social trust had a significant effect was the measure of stated preference for online content personalization and this effect was in the direction *opposite* of that hypothesized. This measured effect is likely to be spurious given it did not occur in step with the three differing levels of trust, as the control condition ( $M = 2.73, SD = .95$ ) and negative social trust group ( $M = 2.73, SD = .88$ ) reported identical mean scores on this measure, which were both higher than those in the positive social trust condition ( $M = 2.58, SD = .86$ ) for this stated preference for online content personalization measure.  $F(2, 1178) = 3.64, p = .03, \eta^2 = .01$ . In addition to the effect detected most likely being spurious, the extremely small effect size suggests this relationship, if real, is unlikely to be of substantial interest. Thus,  $H_{7a}$  and  $H_{7b}$  were not confirmed, exhibiting significant differences counter to those hypothesized but with mixed differences out of step with a complete opposite effect with control and negative trust conditions scoring higher than those in the positive social trust condition. Given this result, the possible counterintuitive effect of social trust on stated preference for personalization warrants further examination, which occurred in follow-up Experiment 4 reported below.

Overall, Experiment 3 failed to locate the hypothesized effects on stated and revealed preference for online content personalization. Perhaps more interestingly, the experiment also failed to detect an effect of social trust on two concepts that have much greater logical support for being linked to social trust. That is trust in websites and apps and online privacy concern. Social trust appeared to have no effect on these two outcomes.

## Experiment 4

### Participants

A survey experiment was administered online during the spring of 2016 to anonymous participants recruited through MTurk (N = 883). Participants were paid the federal minimum wage rate at the time of their participation (\$7.25/hour) prorated for the estimated time to taken to complete the survey. Among participants, 52% were female. Participants were diverse in their age: 18-24 (15%), 25-29 (23%), 30-34 (20%), 35-44 (21%), 45-54 (12%), 55-64 (8%), 65-74 (1%), 75-84 (<1%), over 85 (<1%). Participants were relatively well-educated as highest level of education achieved was: less than high school diploma (1%), high school diploma or GED (11%), technical, trade or vocational training beyond high school (4%), some college or Associate's degree (36%), bachelor's degree (36%), post-graduate training or professional degree (13%). Participant annual income was distributed as follows: <\$25K (25%), \$25-50K (31%), \$50-75K (21%), \$75-100K (11%), \$100-150K (8%), >\$150K (4%). The sample consisted primarily of participants identifying as Caucasian (78%), in addition to Black/African American (5%), Asian/Asian American (7%), Hispanic/Latino (5%), and Bi-racial/Multi-racial (2%).<sup>36</sup>

### Design and Procedures

A three-group between-subjects survey experiment was designed to examine the causal influence of social trust on stated and revealed preference for online advertising personalization, trustworthiness of simulated advertisements, trustworthiness of simulated websites/apps, and trust in commercial websites and apps. As previously described, participants were randomly

---

<sup>36</sup> Where these numbers do not add to 100% this is due to the rounding error.



assigned to equally sized groups corresponding to either positive social trust, negative social trust condition, or a control group.

Scores on these outcome measures were compared using one-way ANOVA to examine whether increasing or decreasing social trust in participants had a substantial effect on preference for online advertising personalization, trustworthiness of simulated advertisements, trustworthiness of simulated websites/apps, and trust in commercial websites and apps. Additionally, a manipulation check measuring levels of social trust was used to assess the effectiveness of the manipulation.

Experiment 3 failed to locate a significant effect of increased or decreased social trust on the outcomes as measured. Therefore, to further probe whether social trust might impact these outcomes, Experiment 4 was designed to be very similar to Experiment 3 while using modified measures for many of the outcomes previously examined. Experiment 4 used the identical social trust manipulation and manipulation check used in Experiment 3, resulting in the same three-group between-subjects design with participants randomly assigned to conditions of positive social trust, negative social trust, and a control condition. However, for the dependent variables, this time rather than measuring preference for content personalization broadly, a measure that also included advertising, instead stated and revealed preference for personalized advertising was assessed specifically.

Additionally, six simulated online advertisements and websites/apps containing these ads were used to measure trustworthiness in actual ads and actual websites/apps, rather than solely relying on traditional question-answer scales that require an additional level of abstraction. Instead, participants were presented with six different online display ads across six different websites/apps and asked to evaluate how trustworthy they found these example ads and example

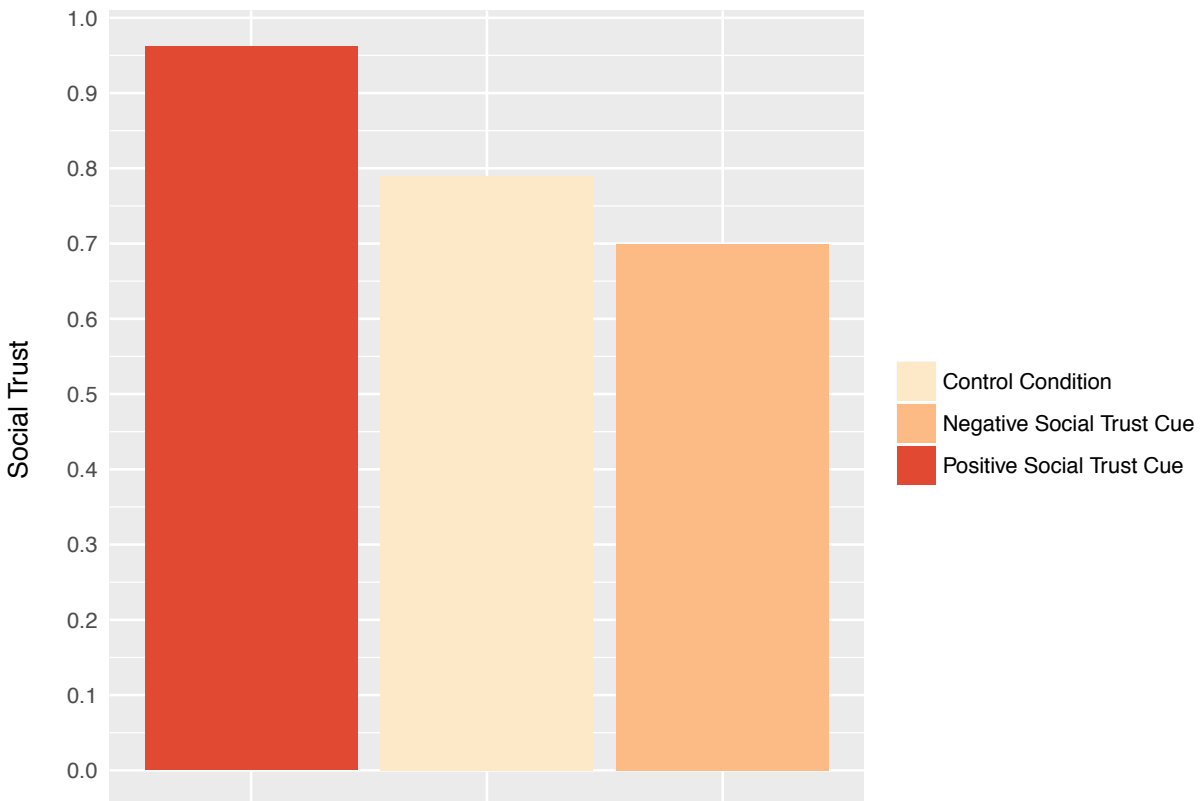
websites/apps to be. This more situated measure was conducted in addition to the validated 7-item scale used previously in Experiment 3 which also measures trust in commercial websites and apps. As Experiment 3 failed to locate causal influence of social trust even on this scale measuring trust in websites and apps, the additional example ads and websites/apps were added in efforts to further reject or confirm the link between social trust and trust in digital media artifacts (i.e., and further investigate nuances of The Media Equation put forth by Reeves & Nass, 1996). In this case examining trust in example ads and trust in example websites and apps presented to participants in their relation to trust consumers feel towards other people.

Finally, this time to measure revealed preference in personalized advertising, the identical measure previously used in the transparency Experiment 2 was implemented. This measure was placed at the very end of the survey experiment asking respondents whether they wished to see ads for three books that had been selected just for them (personalized) and, then, whether participants who elected to see these faux-personalized ads then clicked on them or not.

## **Results**

A manipulation check was employed to assess the effectiveness of the social trust cues at the beginning of Experiment 4. The manipulation was found to be effective across the three groups. Those in the positive social trust condition exhibited elevated levels of social trust ( $M = .96$ ,  $SD = .40$ ) compared to those in the control group ( $M = .79$ ,  $SD = .44$ ). Those in the negative social trust condition expressed decreased levels of social trust ( $M = .70$ ,  $SD = .42$ ) compared to the control group. These observed differences in levels of social trust between the three groups were highly significant,  $F(2, 880) = 30.28$ ,  $p < .001$ ,  $\eta^2 = .06$ , signifying a degree of internal validity for the experimental treatments. Similar to Experiment 3, in Experiment 4 the effect size

on participants' levels of social trust was double that reported in the original study (Mutz, 2005) for which this manipulation and manipulation check were developed. In Experiment 4, for participants in the positive and negative experimental groups, momentarily, levels of social trust were successfully increased or decreased, respectively. These differences are shown in Fig. 5.4.



*Figure 5.4* Manipulation check for Experiment 4. The effects of the social trust cues on levels of social trust in participants. Group differences were significant,  $F(2, 880) = 30.28, p < .001, \eta^2 = .06$ , indicating the manipulation was effective.

After checking that the social trust manipulations were effective, assessing the effects on the dependent variables in Experiment 4 once again revealed somewhat surprising results, though less surprising given results of Experiment 3. Manipulating how trusting participants felt

towards other people had no significant effects on how trusting they felt towards commercial websites and apps, how trustworthy they found example ads and example websites and apps, and participants' stated and revealed preference for personalized online advertising.

Again, despite both measures commonly rooted in trust, no effect was observed in either direction when examining the impact of social trust on participants' trust in commercial websites and apps. Comparing the group mean differences using ANOVA, differences were not significant between individuals with increased levels of social trust ( $M = 2.84$ ,  $SD = .68$ ), those in the control group ( $M = 2.92$ ,  $SD = .72$ ), and those feeling decreased social trust ( $M = 2.81$ ,  $SD = .61$ ), when assessing trust in commercial websites and apps,  $F(2, 586) = 1.24$ ,  $p > .05$ , thus failing to confirm  $H_{10a}$  and  $H_{10b}$ .

Altering social trust also had no corresponding effect on participants' stated preference for personalized online advertising online, as seen when comparing the positive trust ( $M = 3.26$ ,  $SD = 1.60$ ), control ( $M = 3.48$ ,  $SD = 1.66$ ), and negative trust groups ( $M = 3.35$ ,  $SD = 1.60$ ),  $F(2, 761) = 0.46$ ,  $p > .05$ , failing to confirm  $H_{12a}$  and  $H_{12b}$ .

The revealed preference for personalized advertising measure—identical to that used in Experiment 2 which did locate an effect of transparency on this measure of revealed preference—was not significantly different among the social trust conditions in Experiment 4. Group mean differences for this measure of revealed preference for advertising personalization among the positive social trust ( $M = .29$ ,  $SD = .32$ ), control condition ( $M = .26$ ,  $SD = .30$ ), and negative social trust conditions ( $M = .31$ ,  $SD = .34$ ) were not significantly different,  $F(2, 880) = 1.83$ ,  $p > .05$ , failing to confirm  $H_{13a}$  and  $H_{13b}$ .

Finally, after failing to locate an effect of social trust in Experiment 3 on hypothesized relationships of interests (e.g., privacy concern, trust in web platforms), for Experiment 4 two

additional measures we developed to assess more embedded (applied) measures of trust in ads and trust in websites and apps displaying ads. For these, example media were shown to participants who were asked to rate how trustworthy they found ads embedded on websites/apps along with the trustworthiness of the websites/apps themselves. Despite the logical link to trust, there was no measurable effect of altered social trust on how trustworthy participants found these example ads and websites/apps. Those in the increased social trust group ( $M = 2.61, SD = .73$ ), control condition ( $M = 2.62, SD = .77$ ), and decreased social trust condition ( $M = 2.59, SD = .74$ ) did not differ significantly in how trustworthy they rated the example advertisements framed by website or apps,  $F(2, 880) = 0.20, p > .05$ . Nor did those in the positive ( $M = 3.23, SD = .74$ ), control ( $M = 3.11, SD = .75$ ), or negative trust groups ( $M = 3.17, SD = .77$ ) rate the example websites and apps as more or less trustworthy based on these groupings and respective manipulations,  $F(2, 880) = 1.75, p > .05$ . Thus, Experiment 4 did not confirm  $H_{14a}$  and  $H_{14b}$  despite using more situated measures as they relate to the context of personalized online advertising. It appears feeling more or less trusting towards other people has little to no effect on the degree to which participants felt ads and websites/apps they viewed were trustworthy or not.

Besides the manipulation demonstrating clear effects on how trusting participants felt towards other people, as identified in the manipulation check, none of the additional outcome variables exhibited a significant difference in step with the three social trust levels when analyzing group means using ANOVA.

Overall, Experiment 4 failed to locate the hypothesized effects of social trust on concepts presented, many of which were previously tested in Experiment 3 using different measures. Repeatedly finding no effects of social trust on a range of outcomes, some of which would seem to overlap conceptually with trust towards other people (e.g., trustworthiness of simulated ads,

websites, and apps) is once again surprising. Thus, Experiment 4 further suggests that people may not feel towards these media technologies as they do towards other people in the way previous work has suggested (e.g., Reeves & Nass, 1996) or that these mappings between “media” and “the real world” are more nuanced than previously suggested. Implications of Experiments 3 and 4 are further discussed below.

### ***Discussion of Experiments 3 & 4***

After locating a negative effect of increased awareness of ad personalization practices in Experiments 1 and 2—the apparent effect of transparency in advertising personalization—on both stated and revealed preference for advertising personalization, opposition to consumer data collection, and trust in commercial websites and apps, Experiments 3 and 4 explored whether social trust might explain some of the underlying reasons for locating negative effects of transparency in the first two experiments. If seeing how ads are sometimes personalized has a corresponding negative effect, it stands to reason that consumers may be reacting negatively towards marketing practices because they do not trust marketers and/or do not trust the tools and technologies used to collect consumer data and put this to use in advertising personalization on the web. This potential lack of trust, similarly, could negatively impact consumers’ desire for personalization if individuals are thinking about advertising and web platforms in ways they are thinking about people.

However, despite Experiments 3 and 4 both being effective at increasing and decreasing measurable levels of social trust in respondents, no corresponding effect related to stated or revealed preference for personalization was seen in either experiment. Further, neither Experiment 3 nor Experiment 4 detected a measurable effect of social trust levels on how

participants felt towards personalized content. This was despite attempts in Experiment 4 where constructs from Experiment 3 were maintained (e.g., stated and revealed preference for personalization) while varying the measures used to assess these concepts. This repetition was done in an effort to continue examining possible effects of social trust using alternative measurements in case Experiment 3 had failed to locate effects due to measurement insensitivity.

Still, even when deploying alternative measures for both stated and revealed preference for personalization in Experiment 4, there was no measurable impact of increased/decreased social trust on participants' preferences for ad personalization. Somewhat ironically, using multiple measures across repeated experiments appears to have triangulated a type of null effect of social trust on these outcomes of interest, rather than doing so in favor of the stated hypotheses. There two experiments found no causal influence of trust towards other people on preference for online content personalization, including the personalization of online advertising.

Further illustrating this disconnect between how trusting individuals feel towards other people and the trust individuals feel towards those artifacts these people create (e.g., online ads, websites, and apps), in both Experiments 3 and 4, when looking at group means (e.g., via ANOVA) there was no measurable impact of increased/decreased social trust on participants' levels of trust in commercial websites and apps nor on how trustworthy participants deemed examples of online advertisements and the example websites/apps displaying these ads.

Across the board, as seen in Experiments 3 and 4, there were no substantial effects of social trust on individuals' desire for personalized content nor the degree to which they found advertisements, websites, and apps to be trustworthy. The lack of measurable influence of social trust on outcomes of interest goes against the original hypotheses, which predicted that participants' levels of trust in people would spill over not only into their trust in advertisements

and the web platforms these ads are displayed on, but also into individuals' stated and revealed desire to have these web platforms personalized the ads and other content they deliver to consumers. However, this does not appear to be the case as measured in Experiments 3 and 4.

This observed lack of connection between how people feel about other people and how they feel about the artifacts people produce, goes against an understanding of behavior that conceives the two as tightly coupled. In their pioneering work on how people treat computers and media objects, Reeves and Nass (1996) theorized that people not only project human-like characteristics onto media technologies but also treat them very similar to how they treat other people. For example, individuals were found to act in accordance with human social norms when interaction with computers, acting polite towards computers in the way we might act politely towards other people as not to offend them. Based on "a great deal of evidence," Reeves and Nass concluded, "we have found that individuals' interaction with computers, television, and new media are *fundamentally social and natural*, just like interactions in real life." (p. 4-5, emphasis original). Under this understanding, one's feelings toward media artifacts, which would include digital advertisements along with websites and apps displaying them, are thought to exist just as they do *in real life*.

But do feelings such as trust toward *real life* individuals extend to feelings about *real life* media artifacts? Based on these results, it appears this is not always the case. This is demonstrated in Experiments 3 and 4, where individuals were clearly and measurably moved to feel more or less trusting towards other people yet they displayed no measurable and corresponding change in how much they trusted different media artifacts examined, including digital ads, websites, and apps. Reeves and Nass go on to argue that rules for how individuals interact with one another, "from the world of interpersonal interaction, and from studies about



how people interact with the real world ... apply equally well to media” (p. 5). While people may treat computers politely, and in many cases respond to them *socially and naturally*, as Reeves and Nass argue, it could be a mistake to assume this is always the case. Though the null results from Experiments 3 and 4 only suggest that no effect was observed and do not necessarily refute The Media Equation.

Of course, this complication and incompatibility of a complete offline/online unity, at least in the case of social trust and trust in artifacts, remains both preliminary and inconclusive. Though a range of experimental best practices were used in Experiments 3 and 4, along with repeating the social trust experiment using modified measures of these same constructs, further in-depth investigation of many other social concepts and media contexts is needed to arrive at any broader conclusions regarding the ability or inability for emotional or psychological constructs to map neatly from people to things. It may be that sometimes “people treat computers, television, and new media like real people and places” (Reeves & Nass, 1996).

Further, while Reeves and Nass posit their theory as universally applicable claiming it applies automatically and with nearly all forms of media (p. 252-253), they also momentarily concede that people are “quite capable of thinking their way around it” (p. 7), treating media and technologies not as they treat other people. The researchers claim doing so requires substantial cognitive effort and is difficult to sustain, and that when people break free from The Media Equation they typically reverting back to relying on automatic processes that support this idea, *media = real life*.

It may be the case that social trust does not affect how individuals feel towards websites, apps, and personalized advertising because, for this specific media technology, individuals’ preexisting attitudes toward advertising are too well established to be altered by current levels of

social trust. If this is the case, perhaps Reeves and Nass were correct in their totalizing theory under the caveat that sometimes people expend the resources to think their way around The Media Equation. It may be the case that people, generally suspicious of advertising-related activities, are simply expending the extra effort to treat personalized advertising differently than how they feel towards other people.

Additionally, several measures were developed specifically for these experiments as there were no existing scales available, such as multiple measures of stated and revealed preference for online advertising personalization and the awareness of personalized advertising scale. In particular, some of the measures developed rely on multiple moving parts, so to speak, as in the case of the revealed preference for personalization measures in Experiment 3 where participants selected between videos they had no reason to believe had been personalized for them, but effectively included a personalized option based on real participant data including demographic information participants entered at previous points in the survey, the U.S. state from which they were completing the survey (derived on-the-fly by a participant's IP address), and detecting participants' computer operation systems using HTTP request data. Similarly, in Experiments 2 and 4, a revealed preference for personalized advertising measure was developed and "hidden" at the end of the survey. This measurement observed whether participants clicked a button to see 3 book recommendations they believed had been personalized for them based on their previous survey responses. And then, additionally, whether or not they clicked on the "personalized" ads for these books upon being informed this would take them to books' corresponding Amazon.com pages (which it did). This question effectively measured click-through-rates for personalized ads as a revealed preference. All of these measures and all scales used, both validated measures from

previous studies as well as those developed for the current experiments, are described in detail in Appendix E.

## GENERAL DISCUSSION

As consumers' levels of awareness are generally low and variable, failing to incorporate the multitude of additional actors operating below the surface for most web transactions severely limits our ability to understand trust in today's online context. Yet this is essential for understanding why consumers trust or prefer online personalization, where numerous third-parties complicate established notions of consumer-firm trust and challenge traditional understandings of consumer-firm relationships.

Many of these consumer-to-firm trust relationships are difficult to pinpoint, with those in a third-party trustee position often obscured. These relationships are also multi-faceted. For example, an individual may have high trust towards the definitions that appear on Merriam-Webster.com (content) and also feel great trust towards this website/company (publisher), while at the very same time feel very low levels of trust towards the numerous data tracking companies (e.g., data brokers, marketing firms, social media companies) that Merriam-Webster permits to operate under the hood. In this example trust is complicated by third-parties who monitor and track the behavior of site visitors through placement of third-party cookies on a visitor's computing device and use of other, more persistent tracking mechanisms.<sup>37</sup> Therefore, what it

---

<sup>37</sup> A number of browser extensions allow end users to get a glimpse of these third-party partner sites' activities. Popular examples include Lightbeam for Firefox (<https://www.mozilla.org/en-US/lightbeam/>) and Ghostery for Chrome, Firefox, Safari, and Opera, along with different iOS and Android app-based versions on mobile devices (<https://www.ghostery.com/en/>). I conducted a quick test using Lightbeam (v.1.2.1), visiting Merriam-Webster.com, which revealed 61 third-party consumer tracking companies operating on this site's homepage. Of these 61 companies were a handful of well-known firms (Facebook.com, DoubleClick/Google, Yahoo). However, most companies detected were not those commonly known to most people (e.g., adexcite.com, adnxs.com, spotexchange.com, openx.net, mathtag.com, change.com, powerlinks.com, adtech.de, ru4.com, bidswitch.net).

means to trust Merriam-Webster.com (or similarly Facebook, Google, Twitter, etc.) is far less straightforward than our current understandings of online trust are capable of explaining.

As noted, for these types of convoluted, multi-tiered trust relationships existing between consumers and various first- and third-party firms, generally, consumers are relatively unaware of the volume or nature of the transactions that occur between themselves and these firms, nor the presence of most of these firms with whom they transact. This lack of awareness is exacerbated by the frequency of these transactions, which typically occur each time a webpage or mobile app loads new content. In some cases, monitored behavior transactions take place when content has already loaded on webpage, such as recording the length of time internet users hover their mouse cursor over an ad on a desktop. Additionally, in the many cases when there is no active clicking in agreement to terms of service, many of these transactions are still governed by formal and legal agreements covered under all-encompassing privacy policies effectively granted by consent by use. In the E.U. this consent is typically governed by an initial click on websites to accept its respective personal data collection scheme. Though in the U.S. this is not the case, where most often the act of accessing content on a particular website is viewed as consenting to its subsequent data collection practices.

On the other hand, in the case of today's online marketing system this trustor/trustee asymmetry is not limited to consumers. For instance, while a particular data broker would technically know about its transactions with internet users (by nature of recording them), under the current ad exchange model this data broker may not be aware of the particular advertiser(s) to which it rented a consumer's information. In this case, the monetary transaction between first- and third-party firms is obscured by the system itself. As a result, and perpetuating asymmetries within this system, a data broker is aware of some but not all of the parties it transacts with in

this example. Similarly, while Internet users are often aware of a portion of their interactions with NYtimes.com, OverStock.com, or Facebook.com for instance,<sup>38</sup> most consumers are oblivious to their many interactions with numerous other firms whom these consumer facing sites partner with and permit to monitor website visitors (or app users) for marketing purposes.

Whether or not an unknown entity in these type of web transactions (e.g., data broker) is a trustee or not becomes more of a semantic difference than a substantive one, as relationships exist despite varying degrees of awareness. In the current study, examining trust relationships between consumers and entities they may or may not be fully aware of is intentional, with the proposed studies leveraging this varying degree of awareness and trust on the part of the user.

When it comes to actually using computing systems, pioneering studies in human-computer interaction drawing from situated action have taken for granted the role of trust in the adoption and use of these interactive systems (e.g., Suchman, 1987). Situated action approaches have not fully considered the complexities surrounding trust, proceeding as though an individual who decides to use a system has already decided to trust this system. Yet the two are different. Individuals adopt and use systems they do not fully understand nor necessarily trust all the time. This point is key to the current investigation of trust as it functions in online marketing systems.

Further, the dichotomy of in-context vs. out-of-context has been used to explain how individuals evaluate whether information about them has been used appropriately or not. Nissenbaum (2010) conceptualizes this phenomenon through the notion of *contextual integrity*; a heuristic for evaluating if an instance or ongoing practice of personal information use aligns with the expectations (i.e., for privacy, redistribution, granularity) of the individual who the

---

<sup>38</sup> Even with these types of consumer facing groups (not including the non-consumer facing firms they partner with), users are largely unaware of many of their interactions with firms that occur surreptitiously. For instance, when logged in to Facebook, the company and its partners track Facebook users' activities across non-Facebook websites. This means Facebook users are frequently and somewhat continually transacting personal data back to Facebook when using the web, even when not actively using the Facebook website or app.

information describes, or whether this use of their information violates what are theorized as context-relative information norms.

For instance, most people would affirm the contextual integrity of an interaction with their doctor when she mentions one's prior or chronic medical conditions or even family medical history based on information in one's medical chart. Similarly, a bank teller might discuss one's balances from various loans, investments, or savings accounts. As the bank employee's knowledge of this type of personal information is not only necessary for the bank to function but is expected by consumers given the context. Though necessity is not always a requirement, as in the examples of the doctor's office and bank, for maintaining information norms. For instance, a person's roommate would likely be intimately familiar with his drug store purchases due to sharing a medicine cabinet, despite not needing to know this information for the apartment to function. In this way, context provides a powerful means of setting information expectations regardless of information requirements. These and many other information expectations are fairly obvious. However, far less clear is the question, what constitutes appropriate information use in marketing.

### *Future Work*

Building on these results, future work might further investigate the effects of transparency and trust on preference for advertising personalization in many additional ways. For example, two key questions left unanswered by these studies related to the nuances of different specific forms of trust and also what might be learned from conducting this type of experimental testing but using more naturalistic or embedded study designs to examine situated practices. First, what is the effect of more specific forms of trust, such as trust in commercial website and

apps, on consumer preference for personalization? From a recent related study surveying American adults (Stevenson & Pasek, 2015), we know these constructs are moderately correlated. However, the causal role played by trust in commercial entities (versus generalized social trust) in developing preference for personalization remains unknown. Understanding these concepts, which are likely to have more overlap than with social trust, could be especially helpful for marketers and systems designers. A simple modification of the third experiment could shed light on this question, manipulating trust in firms to examine its effects personalization. Another question stimulated by this work but remaining unaddressed relates to more situated or embedded effects of transparency in advertising personalization transparency on consumer reception to ads themselves. Are consumers more or less likely to be persuaded, or perhaps persuaded differently, when they are more or less aware of how the ad in front of them has been personalized? Though this genre of applied persuasion research is well studied in marketing and advertising, to date there is little published work documenting the specific effects of transparency in the use of consumer data for generating ads online. As more and more consumers consume media through online channels, and as many of the ads distributed on these channels are selectively presented to viewers based on information about them, further understanding of the dynamics of personalization should benefit marketers and consumers in reaching their respective goals.

## **CONCLUSION**

Together, the empirical studies I have reported examine the relationships between trust, transparency, and preference for personalized content including advertising, in addition to related factors of secondary interest, privacy concern, data control self-efficacy, approval of various

consumer data categories for use in personalized advertising. The aim has been to execute these individual studies and experiments successively and so they are in conversation with one another, with each study motivating the next. How factors such as awareness of how media we use function under the hood and social trust affect more specific forms of trust such as trust in websites and apps and trust in online ads, along with preference for personalization of these ads and other content, are important inasmuch as we know that the ways individuals think about advertising—both its content and form—directly influence consumer response to attempts at advertising persuasion. This alone should motivate marketers to try to further understand not just how consumers feel about advertising personalization but how these positions impact reactance to marketing messages. This is in addition to the implications for consumers and policymakers, who should take note of the relatively low levels of approval and trust in ads, websites, and apps when consumers are made more aware of how these processes go about personalizing content.

As demonstrated in Experiments 1 and 2, arousing consumer awareness about how consumer data is used to personalize online ads, or simply increasing transparency in this process, appears to diminish several constructs most marketers would logically uphold as worthy of pursuing in the digital era: trust in commercial websites and apps along with both stated and revealed preference for personalized advertising. As marketing messages are routinely personalized, often using person-specific consumer data, establishing trust and consumer preference for these practices seems vital to successful interactions and relationships between marketers, web platforms, and consumers. However, as the studies here have illustrated, being up front about precisely which consumer data is collected and how it is used for targeting online ad campaigns appears to diminish consumer trust and support for this ubiquitous practice. This should be troubling for marketers, as trust is often considered a key ingredient in facilitating



successful interactions. And no marketer or web platform intends for its audiences or user base to disapprove of its standard practices. Yet, in the case of online advertising personalization, when consumers are more aware of how consumer data is used they disapprove of related activities including ad personalization.

One remedy for the trouble with transparency might be for marketers to pursue novel ways of establishing better trust with consumers. So that consumers, upon finding less reason to be concerned about how data describing them is used to personalize ads and more reason to trust marketers in this regard, may increase in desire to receive content selectively presented to them based on what marketers and other entities know about them. For measures to increase social trust, regrettably, as reported in Experiments 3 and 4, increasing social trust appears ineffective at improving consumers' trust in websites, apps, and advertisements and how trustworthy they find actual ads and websites/apps, along with having apparently no bearing on individuals' stated or revealed preferences for ad personalization. This is unfortunate given the apparent reactions by consumers to increased transparency in online ad personalized, as reported in Experiments 1 and 2. This suggests that marketers should be on the look out for alternative ways to improve both consumer trust and preference for advertising personalization. As improving trust in people in general appears to have no corresponding uplift on how consumers feel about the technical artifacts (ads, websites, apps) these same people create.

Finally, the concepts examined in these studies, especially those linked to consumer preference, are anything but static. Rather, they are likely to change in some way over time. This likely change over time may work for or against the interests of marketers and web platforms that use detailed consumer data for personalized advertising. First, working against these interests, over time, naturally more consumers will become cognizant of (consumer data/ how ad

personalization functions). If this is the case, this could signal trouble for marketers and proprietors of online platforms that deliver personalized advertising. Second, and contrary to the previous problem posed by time, it would be a mistake to interpolate the negative effects of transparency very far out into the future. Preferences, and especially those as nuanced as preference for consumer data privacy and personalized (vs. non-personalized) advertising on the internet, are likely to change over time. Thus, in this instance, as more consumers naturally become more aware of different consumer data practices and the ways advertising personalization is achieved online, these same people may come to accept or even prefer these practices *more*. This remains a possibility, especially as these practices become not only more widely known in the popular imaginary but simply normative. This history of media technologies suggests consumers can be highly resistant to adoption only to reverse their preferences after some time. The same may be the case for automated advertising personalization that relies on rather detailed consumer data. Therefore, there is reason to believe that time may simply alleviate problems for marketers associated with greater consumer awareness of ad personalization practices rather than expand them. In the meantime, and in the absence of a crystal ball, it would serve marketers to take consumer preferences, both stated and revealed, seriously. For now, for marketers to ignore the advertising and personal data use preferences of consumers may be to the peril of both groups.

## Chapter 6

### Conclusion

Nearly 40 years ago, commenting on what he observed to be a loss of human agency in response to the growing complexity of technology, Langdon Winner (1977) concluded,

“...members of the technological society actually know less and less about the fundamental structures and processes sustaining them. The gap between the realities of the world and the pictures individuals have of that world grows ever greater. For this reason, the possibility of directing technological systems toward clearly perceived, consciously chosen, widely shared aims becomes an increasingly dubious matter” (p. 295-296).

I share Winner's concern that as technologies of everyday life grow more complex we are less able to fully grasp how these important parts of our lives function. However, while the problem of directing complex sociotechnical systems towards mutually beneficial outcomes might seem daunting, complexity alone does not necessarily prevent people from understanding how technologies and processes work. At least, it should not.

The findings presented in this dissertation point towards another possibility. One where those involved in building and maintaining a system of rather astounding sophistication, capable of delivering precise communications to people possessing very specific attributes by parsing and matching vast quantities of data and in real time, have also done a very poor job at showing individuals the impressive way this system works. In some cases, this is intentional obfuscation, as when system designers deploy user interfaces and policies that exacerbate information

asymmetry with the consumer. More than anything, this business strategy is likely caused by a fear of the consumer. Fear that allowing individuals to peek inside today's advertising machinery will result in backlash rather than awe. I suspect awe may actually be possible. However, because the marketer's default has been secrecy for so long, the targets recoil. As observed across the three studies presented in this dissertation, whenever individuals, myself included, gain the occasional glimpse into some of the ways advertising personalization works, they are often surprised, displeased, and even shocked. Were this system more open and transparent to begin with, I suspect people might instead come to appreciate the tremendous lengths marketers and others go to in efforts to try to get our attention and sway our thinking.

### **Summary of Dissertation**

The three studies in this dissertation examined the many efforts on the part of marketers to selectively present digital advertisements to individuals.

In the first study (Chapter 3), I took on the role of the marketer to gain an up close look at the character and dynamics of consumer data as it functions in advertising personalization. Examining real-time bidding interfaces, I found the volume, variety, and granularity of third-party consumer data available to marketers to be nothing short of dizzying. Quite simply, if you can imagine a type of consumer data, it likely exists, in some cases with stunning specificity. In this study, I synthesized just a snippet of the variety of consumer data available for advertising personalization that I observed while using real-time bidding ad-buying platforms. Further inquiry might explore how these data are used by marketers and how they are not. From my vantage point, I primarily assessed what is possible but had no way to determine what is most common via this research design. I also reported results from a series of tests which found that,

in a number of examples, personalized advertising pushes and pulls on some audience attributes more than others. These results also warrant further inquiry, as some have pointed to the potential for social discrimination when members of protected classes experience content personalization or other algorithmically-determined outcomes.

In the second study (Chapter 4), I spoke with consumers in focus groups to explore how people imagine and reason about advertising that they believe has been personalized for them. Findings from this study demonstrate the generally negative associations with personalized advertising. I further explored the possible mental models that people rely on when interacting with online advertising today. I interrogated some of the shortcomings of these models, which seem to be at times quite ineffective in that they leave people disappointed. If our mental models are indeed the most useful when they provide a basis for action that allow us to achieve our goals with a technological system, then it appears consumers' mental models of personalized advertising might not be very good. This might explain the widespread negative opinions about many consumer data collection practices. Teaching web users different mental models demonstrating how advertising personalization works may prove to mitigate some of the problems faced by consumers and marketers alike.

Finally, in the third study (Chapter 5), I reported results from four experiments that examine the interplay and effects of transparency and trust on how consumers feel about advertising personalization. Across the board, greater transparency about some of the ways online ads are personalized for individuals appears to diminish their support for this practice. As mentioned earlier, this may simply be due to the backlash of being confronted with information one finds surprising rather than truly objectionable. Or, these objections may be enduring. Further study might examine whether the effects of transparency are momentary or whether

awareness over time about how ad personalization functions leads to more sustained opposition to these practices. Marketing practitioners and policymakers should take note, as findings point towards a consumer backlash, including reduced trust in marketing and digital environments generally.

In light of this, what is one to make of the present state of advertising personalization? the ways in which marketers envision consumers? and how consumers perceive these attempts to persuade them? Across all three studies I repeatedly found the notion of information asymmetry relevant and compelling. The view from the consumer's perspective is so limited that it appears any glimpse into how advertising personalization functions wakens perceptions of the asymmetry itself. One possible solution might be for marketers to take the lead in chipping away at the information asymmetry they have themselves created. Efforts to educate the public about matters deemed to be of current cultural importance (e.g., healthy eating, anti-smoking, exercise) are common. Thus, marketers might take a page from the public health playbook and seek to educate consumers about the digital processes in which consumers and data describing them are routinely involved. In doing so, it stands to reason that "the gap between the realities of the world and the pictures individuals have of that world" might be reduced.

## Appendix A

### Focus group moderator question guide

#### (Chapter 4)

*Note.* Given the open-ended nature of the focus group discussions, not all questions below were asked of all participants in each focus group.

*Please take a minute or so to write down five words that come to mind when you think about personalized online advertising.*

*Try to think back about a time or two when you noticed an ad online, this could be on a website, in a mobile app, anywhere on the internet where you noticed advertising. What happened? Where was the ad shown? What else comes to mind?*

*Do you think this ad you saw was generated uniquely for you? That it had been personalized for you in some way?*

*For the advertisements you see on websites and apps, would you say you expect these ads to be personalized for you? Why or why not?*

*How would you say it feels to you to have advertisements personalized for you, based on information about you?*

*Can anyone recall a time when you saw an online ad and then either clicked on the ad to get more information or used it to make a purchase?*

*In the past, would you say you've found personalized online ads to be useful? Either ones you've clicked on or used to help make a purchase or just the ones you have viewed but didn't click on?*

*Generally, would you say you want the online ads you see to be personalized just for you? Or, would you say you would prefer to see non-personalized ads online? (similar to the type of ads you see when watching live television, which are typically the same for everyone watching at that time and not personalized for you)*

*Do you enjoy seeing online ads that are relevant to your interests, background, habits, etc.?*

*Do you believe advertisers have your best interests in mind when it comes to creating ads personalized for you individually? Why or why not?*

*Was there anything that came up during today's discussion that was surprising to you?*

*Is there anything from our discussion you would like myself or someone else to try to clarify?*

*Is there anything else that comes to mind that you would like to say today?*



## Appendix B

### Participant word associations for personalized online advertising

#### (Chapter 4)

<i>abundant</i>	<i>economic(s)*</i>	<i>millennials</i>	<i>scroll</i>
<i>age</i>	<i>everywhere</i>	<i>money</i>	<i>search</i>
<i>aggressive</i>	<i>eye catching</i>	<i>money grab</i>	<i>seduction</i>
<i>aggravating</i>	<i>family</i>	<i>needs</i>	<i>selling</i>
<i>algorithm</i>	<i>fraud</i>	<i>not ideal</i>	<i>simple</i>
<i>annoying**</i>	<i>free</i>	<i>not needed</i>	<i>sloppy</i>
<i>appealing</i>	<i>frustration</i>	<i>nuisance</i>	<i>smart*</i>
<i>assuming</i>	<i>garbage</i>	<i>overkill</i>	<i>spend</i>
<i>avoid</i>	<i>helpful*</i>	<i>persistent</i>	<i>sporadic</i>
<i>basic</i>	<i>hidden persuasion</i>	<i>personal*</i>	<i>sports</i>
<i>business</i>	<i>ignoring</i>	<i>pics</i>	<i>stalking*</i>
<i>busy</i>	<i>in the way</i>	<i>politics</i>	<i>target age group</i>
<i>car</i>	<i>indifferent</i>	<i>predatory</i>	<i>targeted</i>
<i>competitive</i>	<i>information</i>	<i>presumptive</i>	<i>time waster</i>
<i>confusion</i>	<i>innovation</i>	<i>privacy*</i>	<i>too many</i>
<i>constant</i>	<i>inspired</i>	<i>profile</i>	<i>too much frequency</i>
<i>consumer</i>	<i>interest(s)*</i>	<i>profiling</i>	<i>truth</i>
<i>convenient*</i>	<i>interference</i>	<i>quizzical</i>	<i>understand</i>
<i>cool</i>	<i>interruption</i>	<i>random</i>	<i>unnecessary</i>
<i>curious</i>	<i>intruding</i>	<i>repetitive*</i>	<i>unreal</i>
<i>deletion</i>	<i>inundated</i>	<i>resource</i>	<i>useful*</i>
<i>dishonesty</i>	<i>invasive</i>	<i>sale</i>	<i>user specific</i>
<i>disingenuous</i>	<i>inventive</i>	<i>sales pitch</i>	<i>word/wording</i>
<i>distraction</i>	<i>irrelevant/not relevant</i>	<i>scary</i>	

\* Term appeared in 2-3 participants' lists.

\*\* Term appeared in 7 participants' lists.

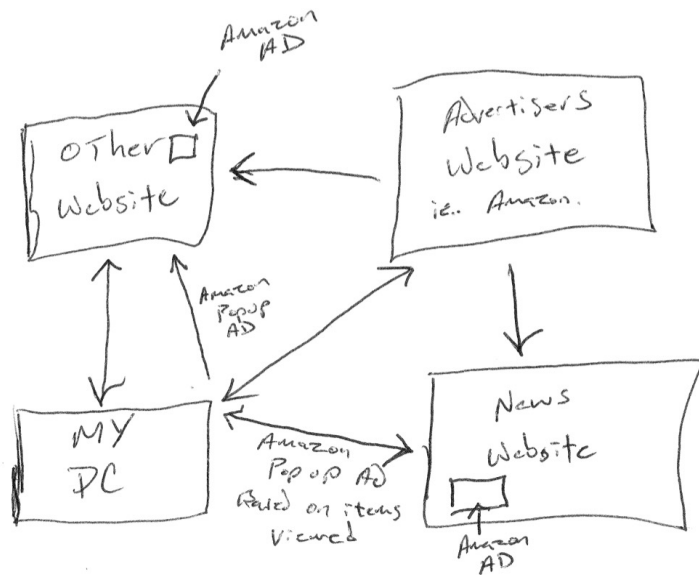
Some participants wrote less than 5 terms.

## Appendix C

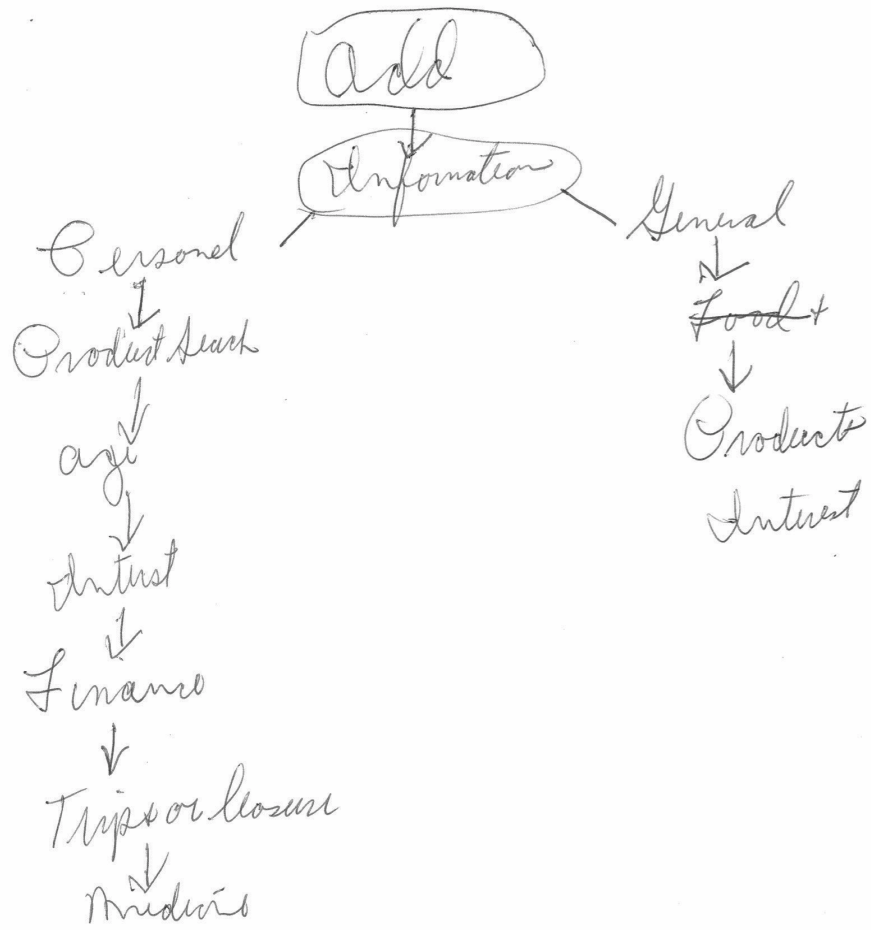
### Participant drawings of personalized online advertising

#### (Chapter 4)

*Note.* Some participants wrote their real names in various places on their drawings, typically when labeling items (i.e., *Fred's data*, *Mildred's Facebook account*). Where present these identifying marks have been removed resulting in blank spaces (white rectangles) in a few of the drawings below.



Leaving cookies on my PC to identify what I have searched



Step 1

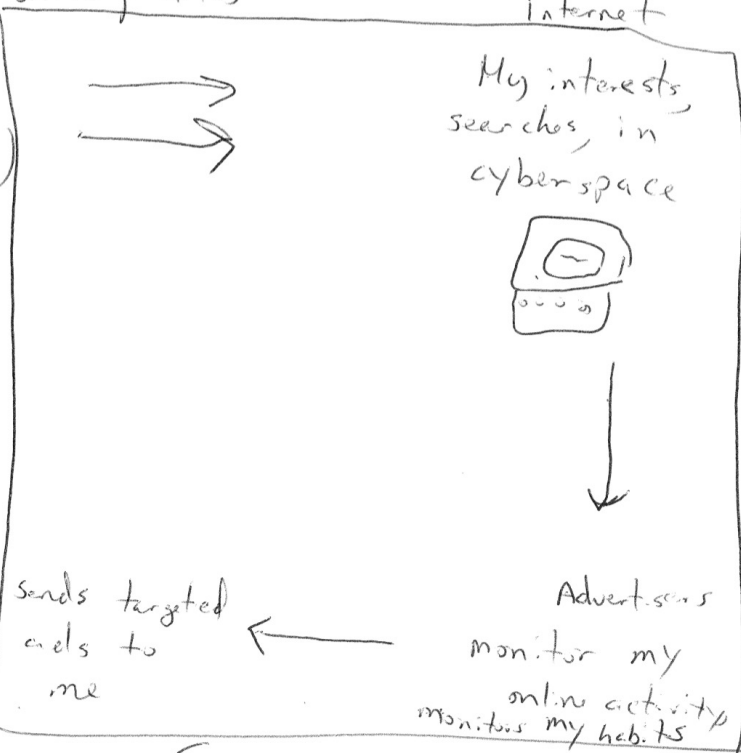
Me setup accounts  
online profiles



Repents



Step 4

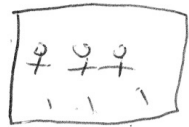


Step 2

music, movies  
news, friends



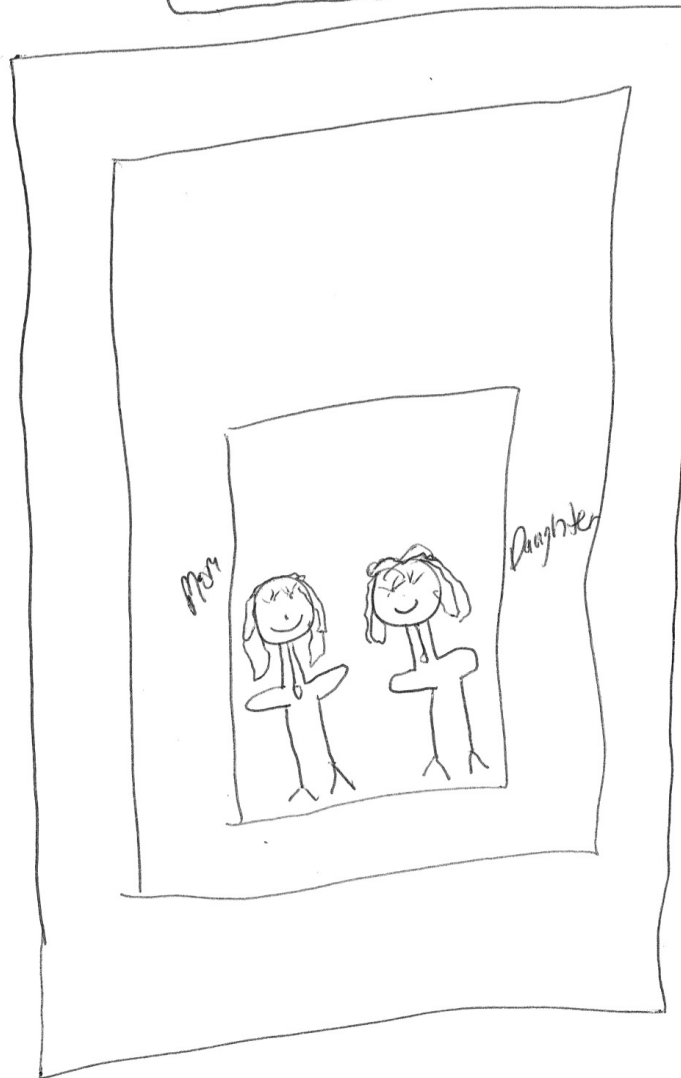
Advertisers  
monitor my  
online activity,  
monitor my habits

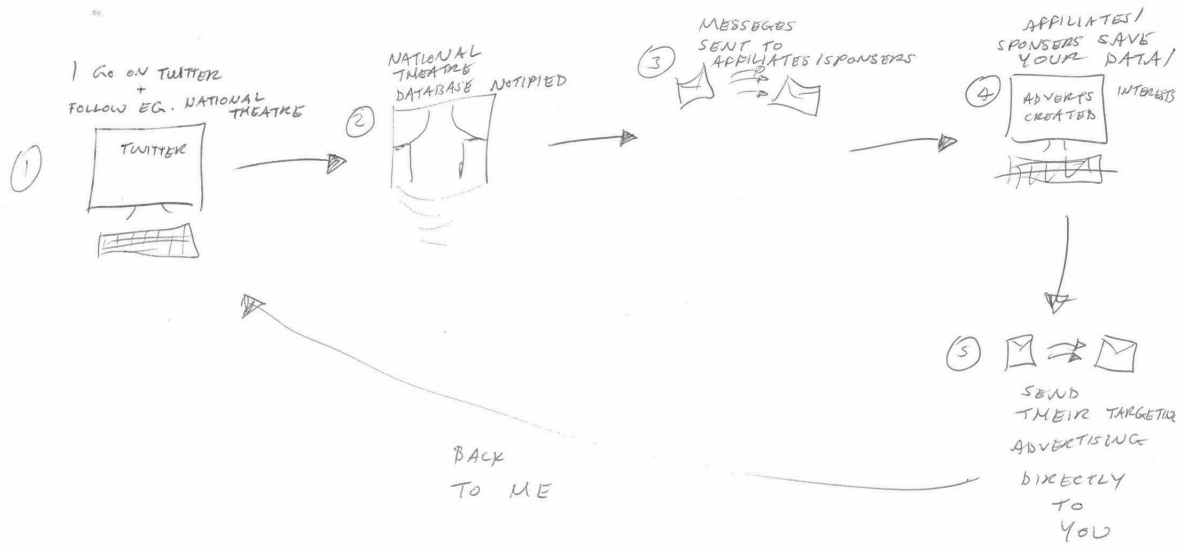


Step 3



Yahoo	NY
Young+Older	U+Habe Kaspa

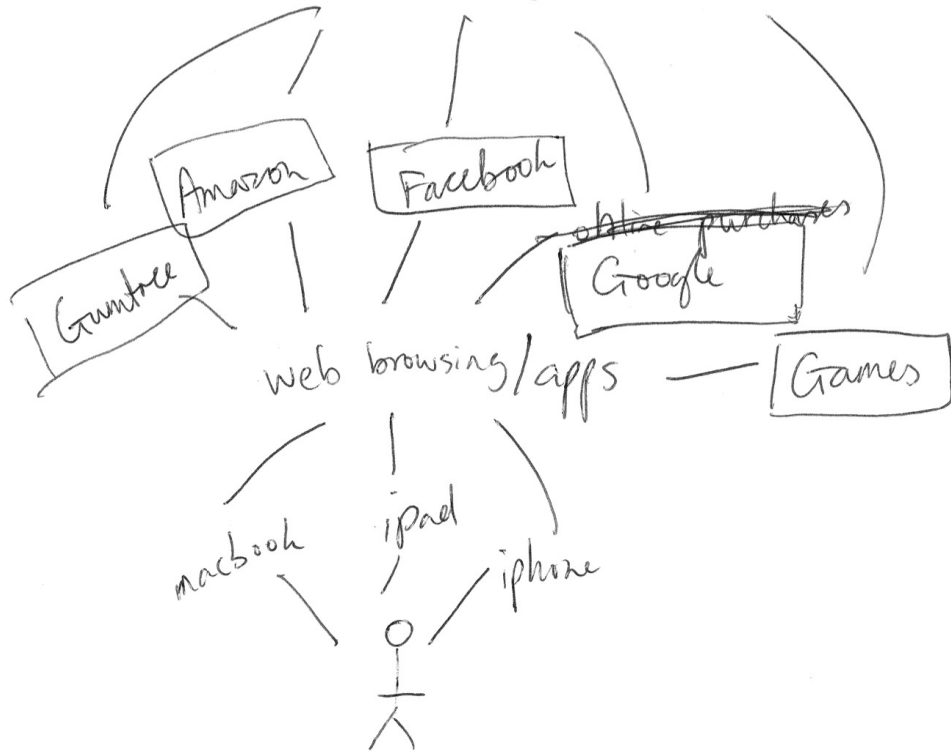




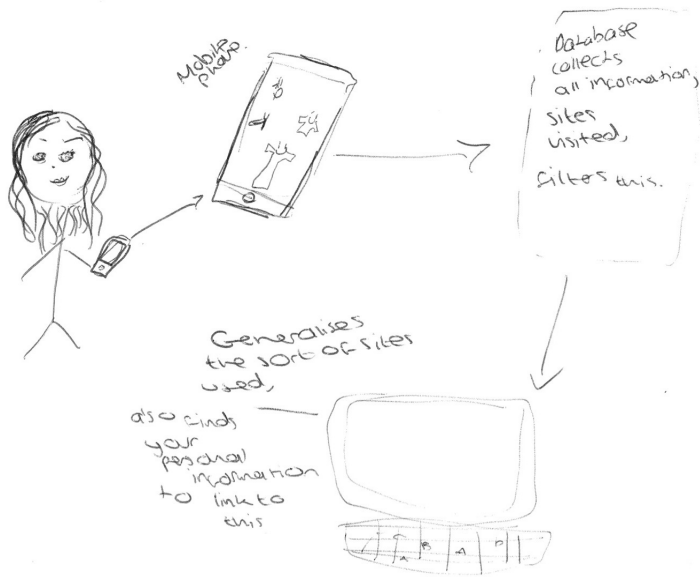
# Product Advertising

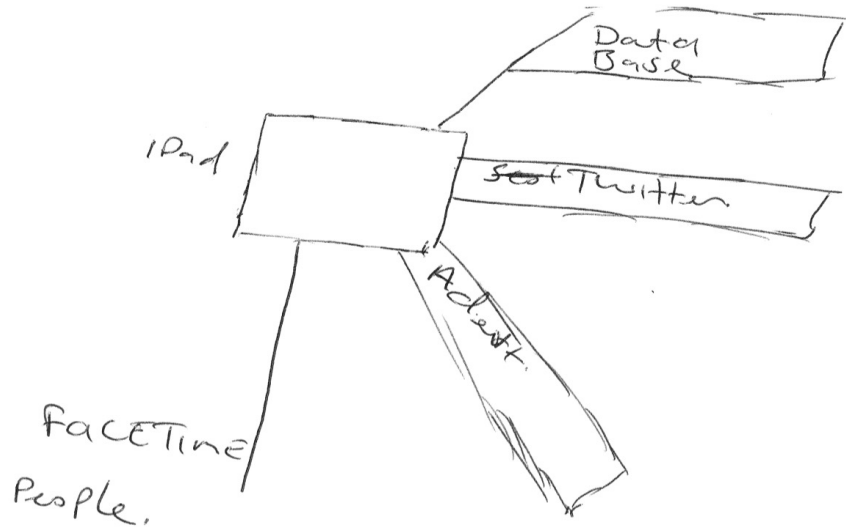
↑ \$\$\$ ↑  
↓ \$\$\$ ↓  
Sold to advertisers  
and then placed back into  
the companies below.

# Collect Information



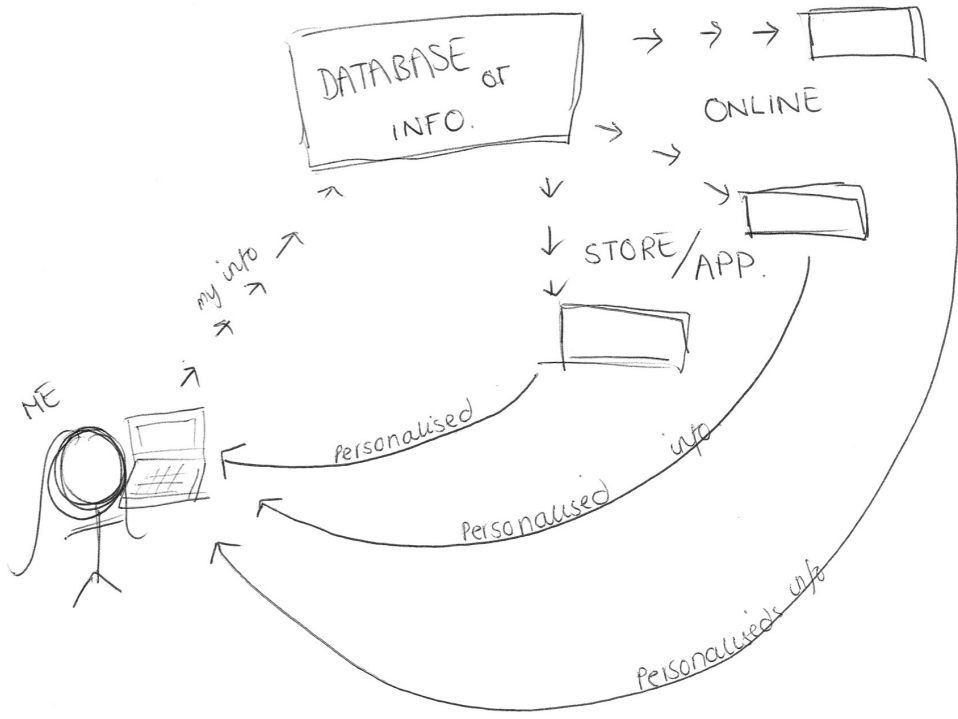




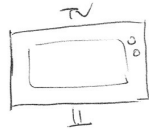


The Web





~~MARKETING~~



SHOPPING.



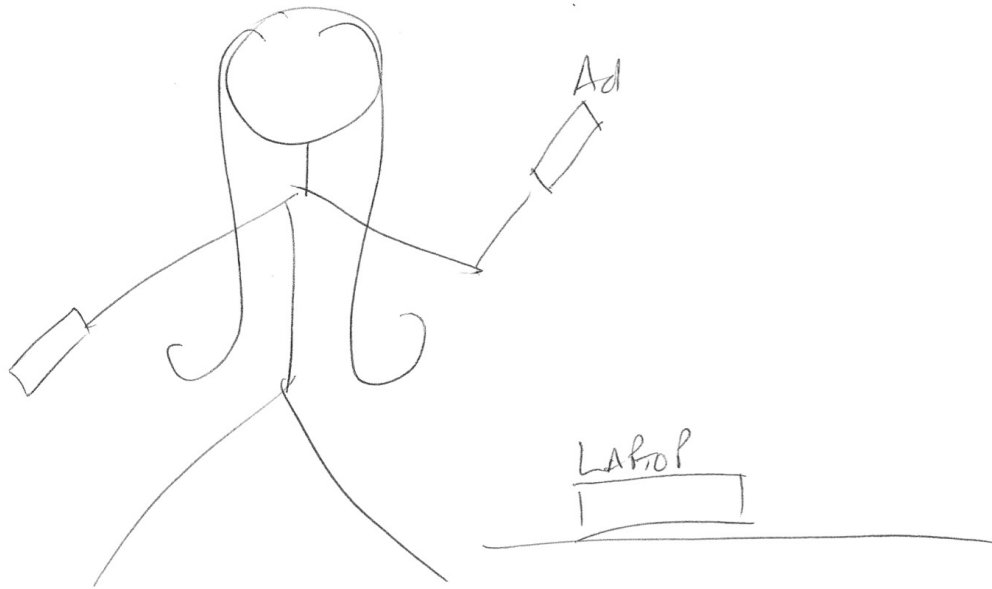
SOCIALISING



Home  
owners

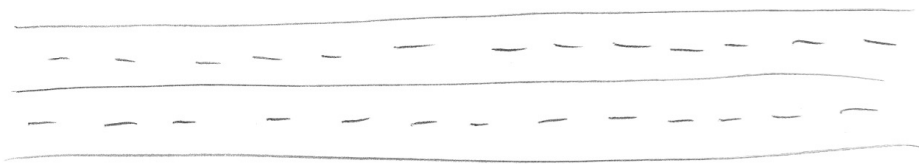
Shops

Shopping  
Centres



Billboards

TV  
ADVERTISING



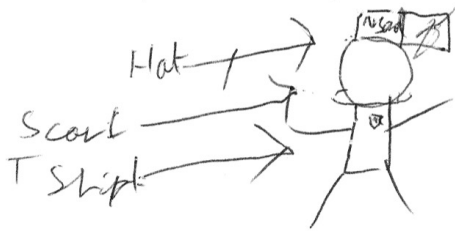
Home

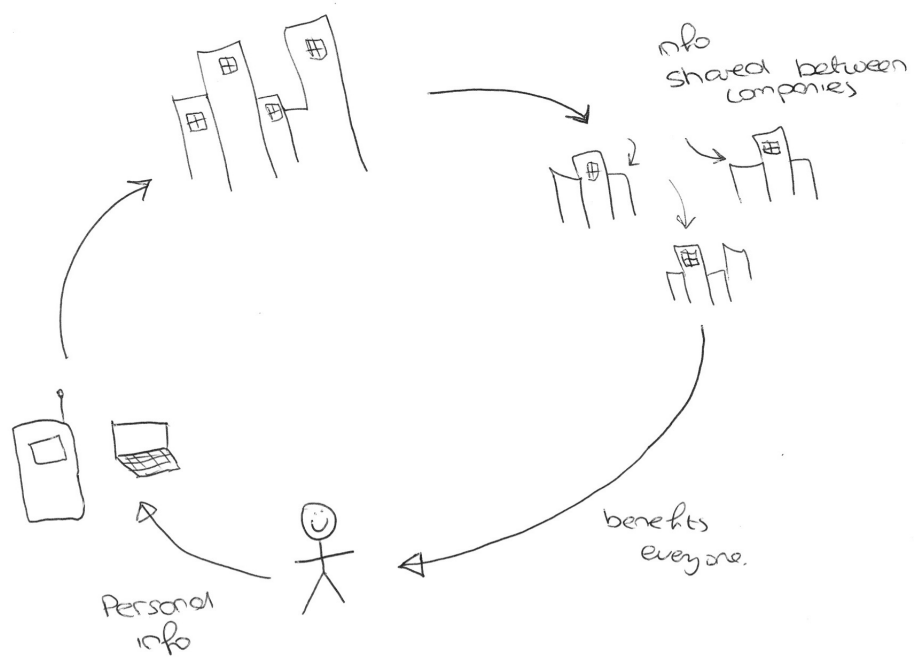
Clothes

Mugs

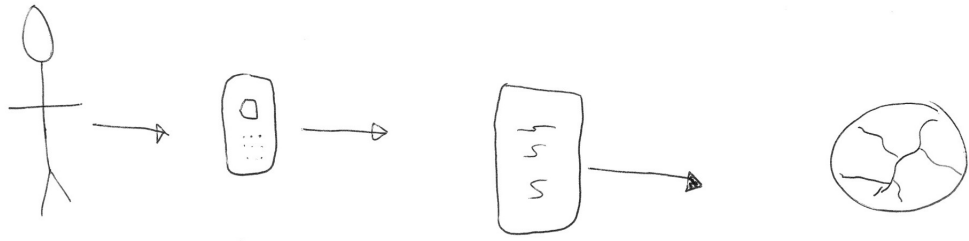
Kids

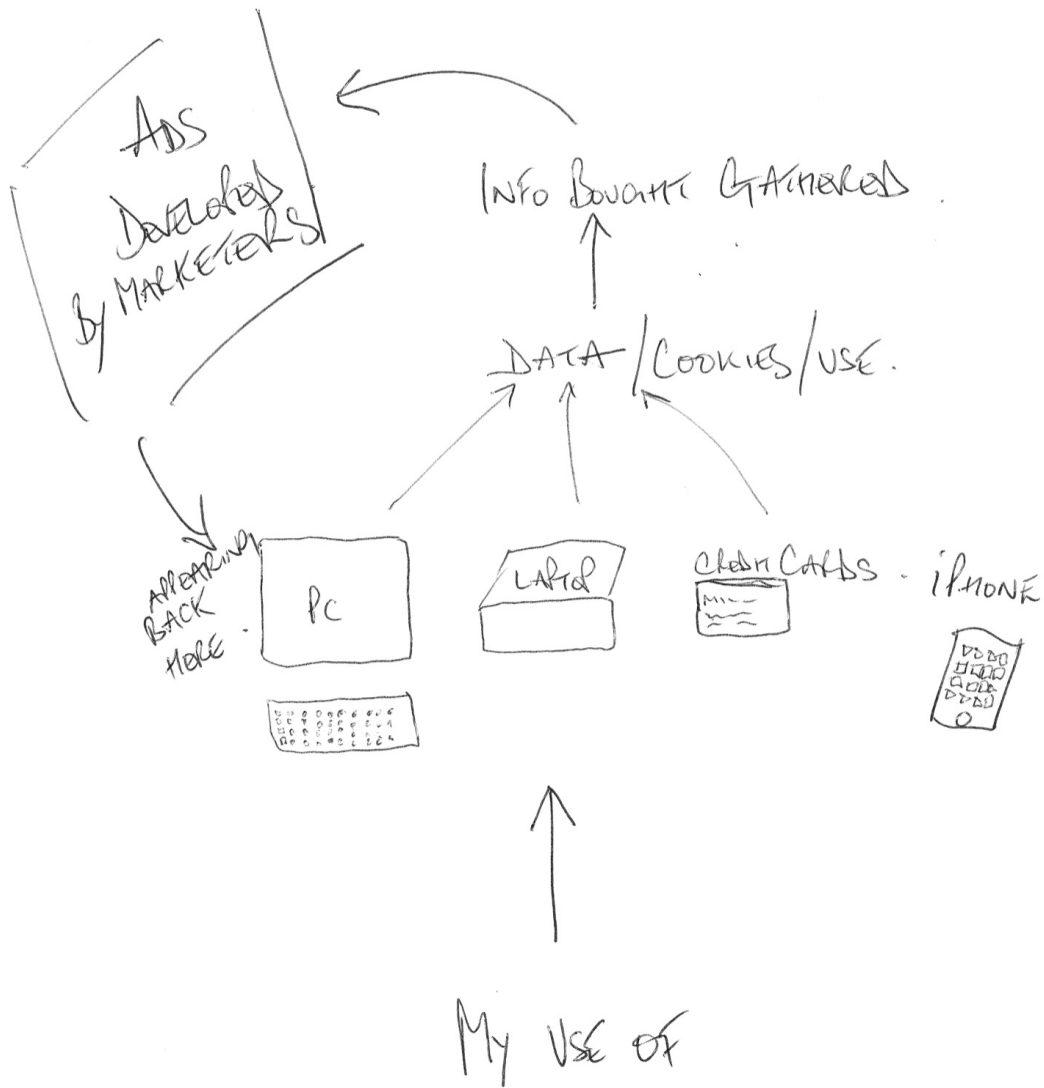
Arsenal Shop online

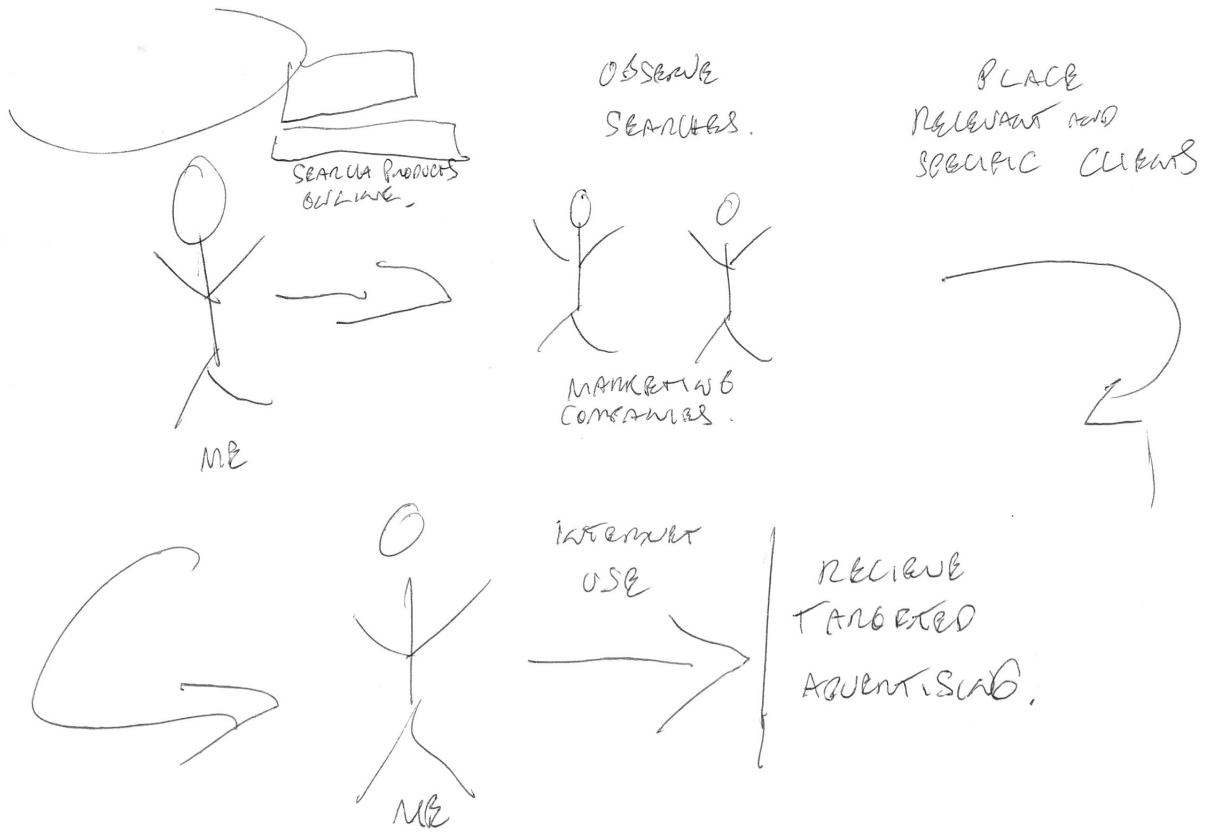


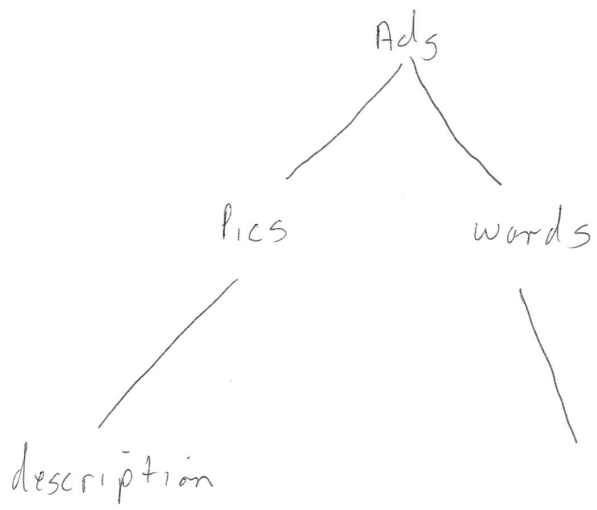


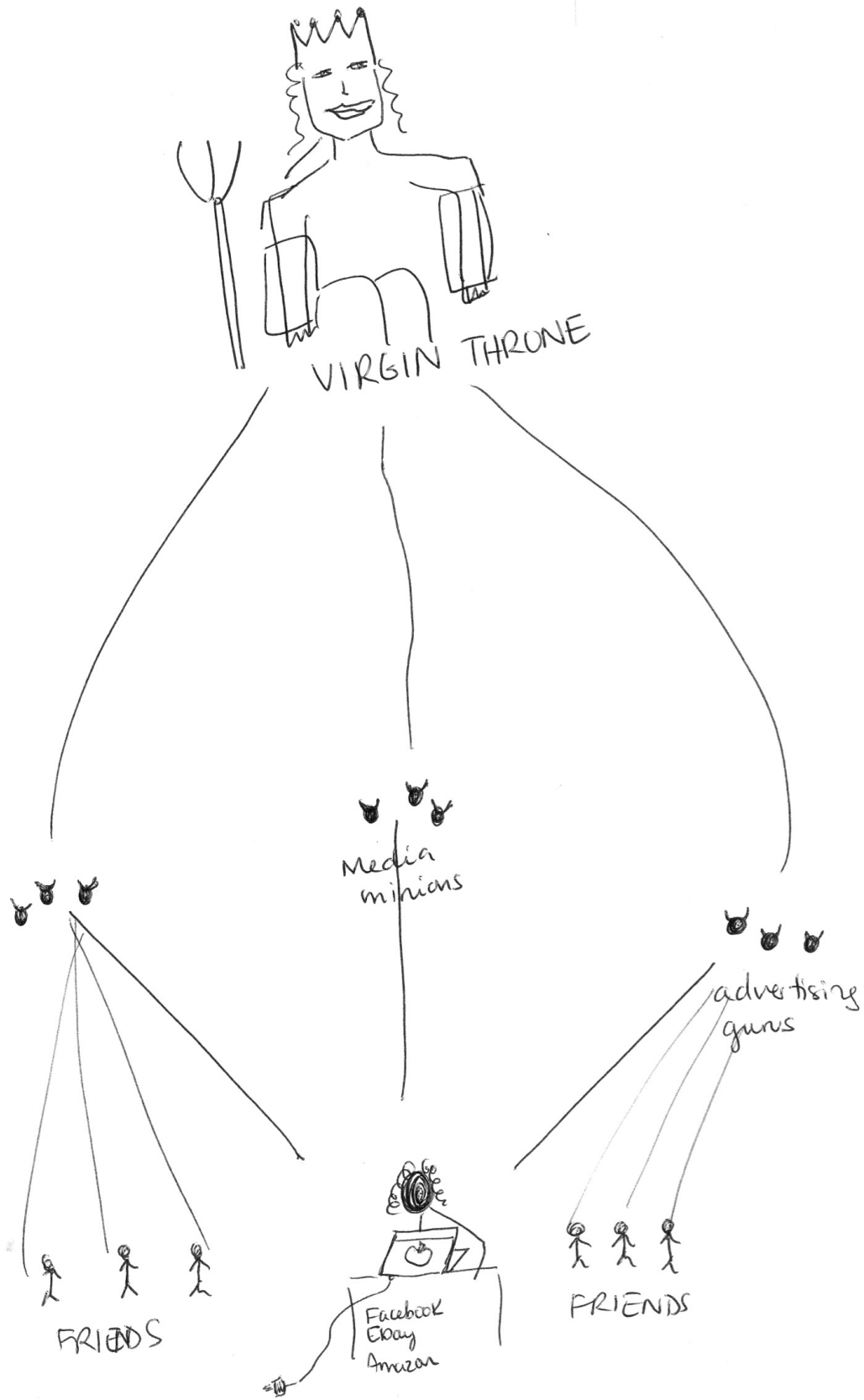


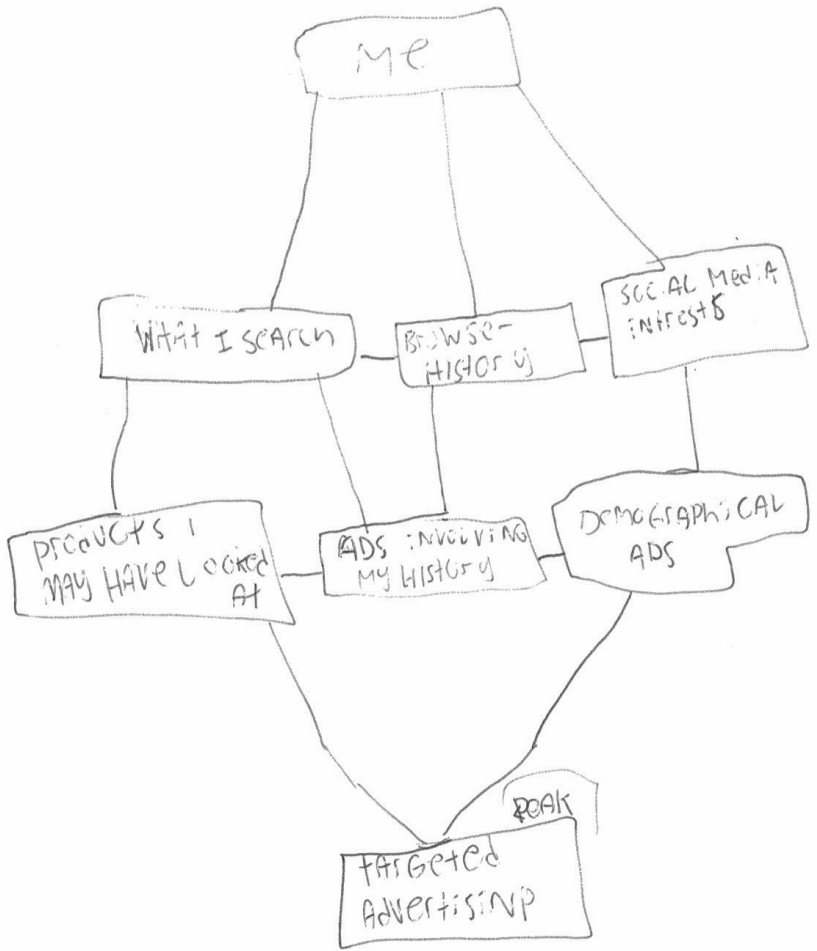


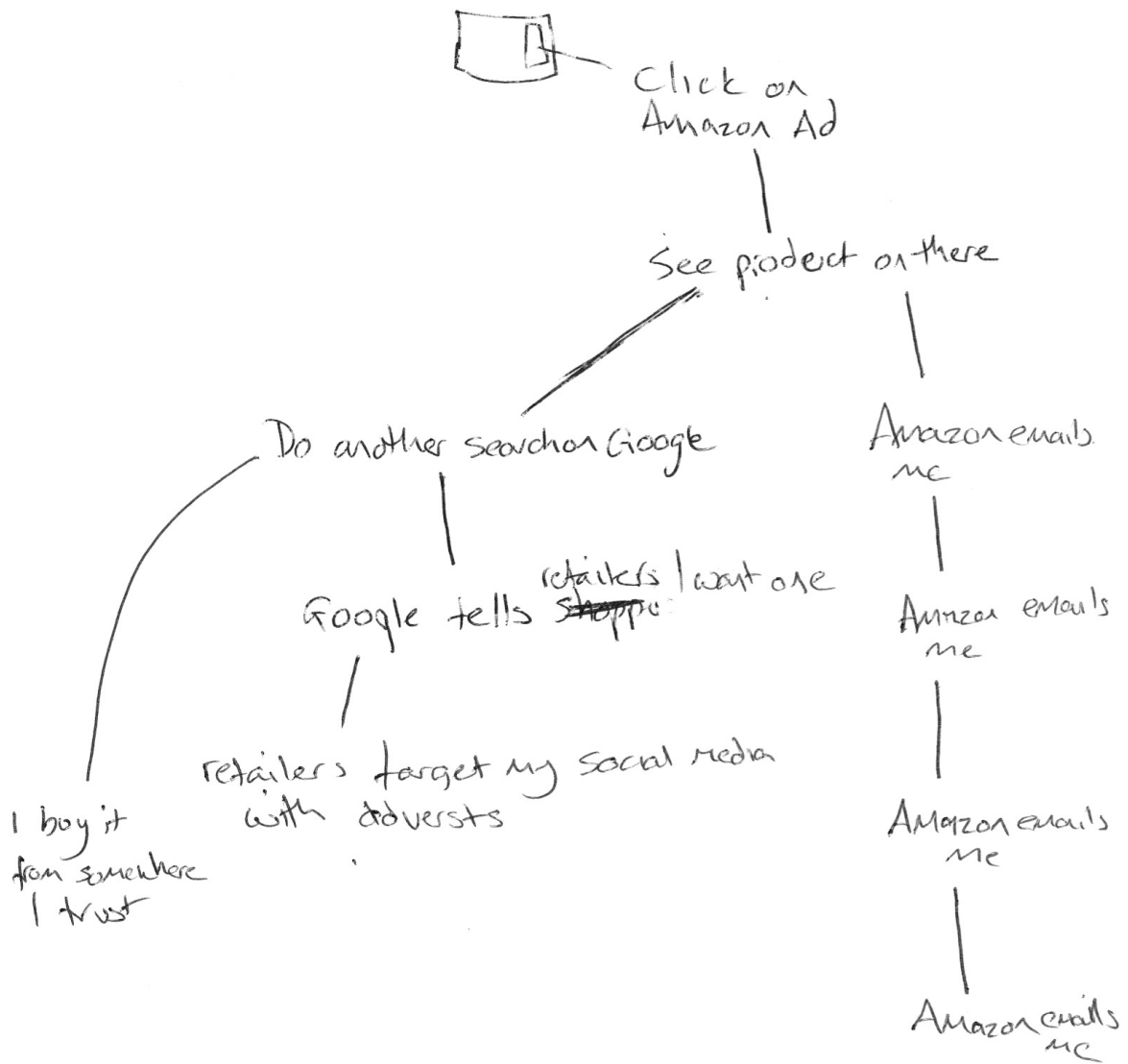




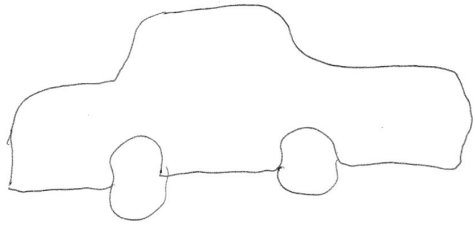








Cars 9 US



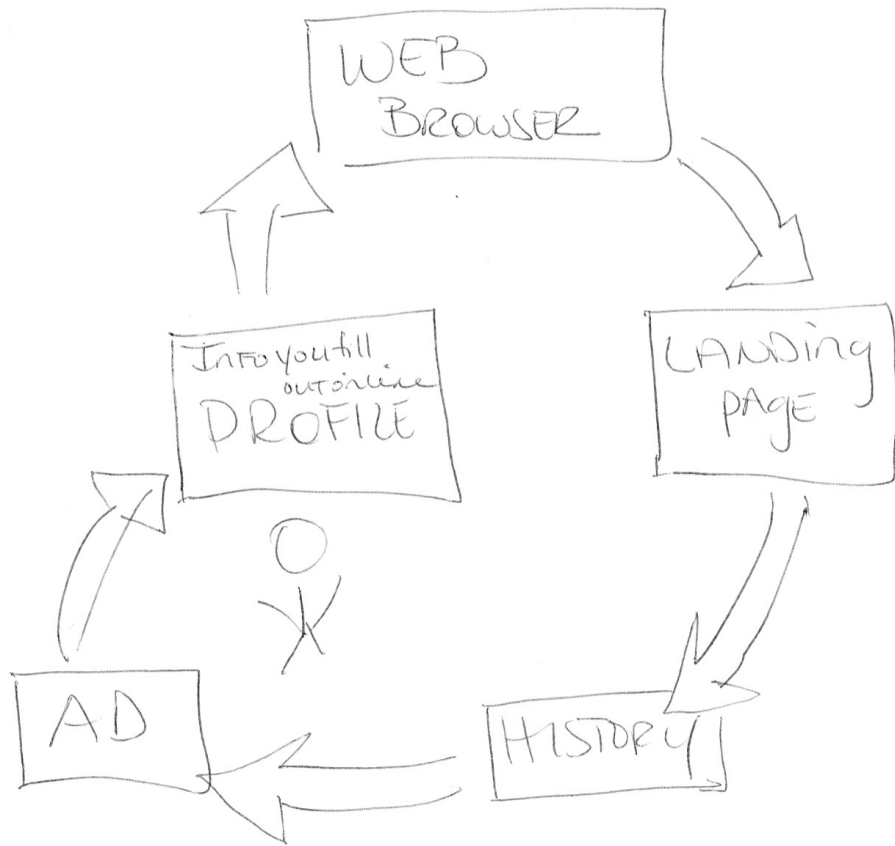
History  
Maintained 9/3/16

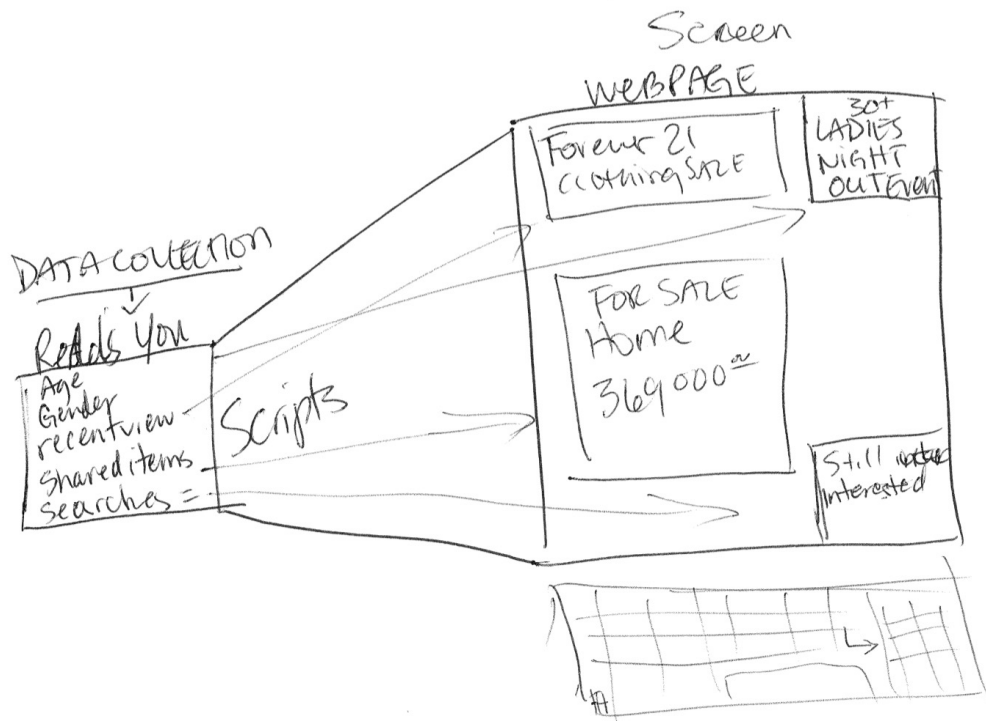
11  
11  
11  
11

Milage

11  
11  
11







google search for item



Direct website search

~~AD ON WEBSITE~~

ADS ON DIRECT WEBSITE

Go to next website  
SIMILAR ADS  
SEE ~~THE SAME~~  
NEW AD AND ON WINDOW  
AND ON

Searching



Target what you are searching



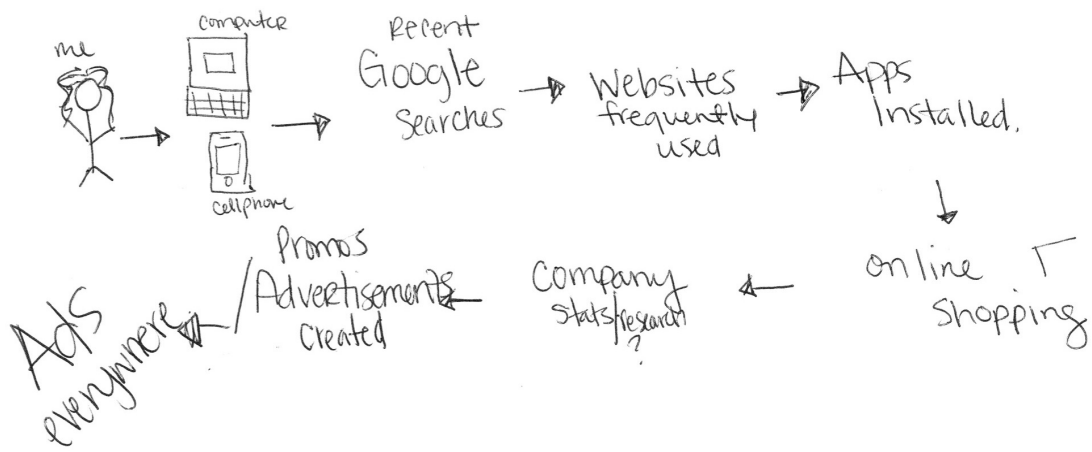
Intice you to buy

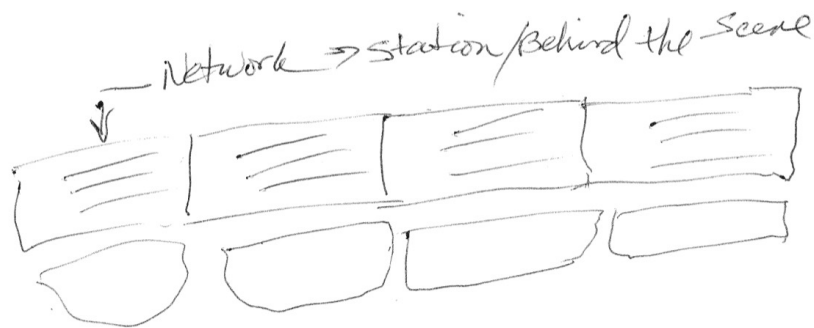
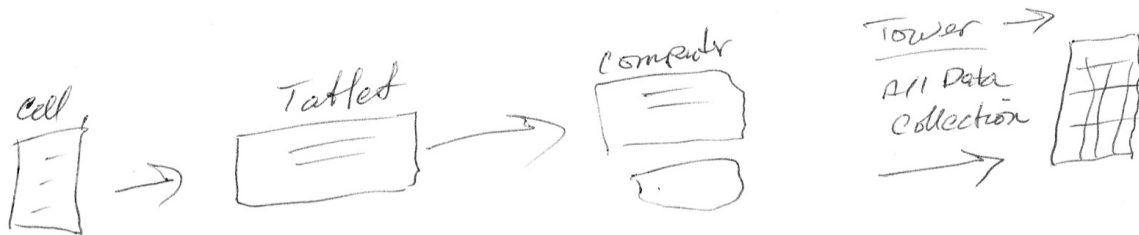


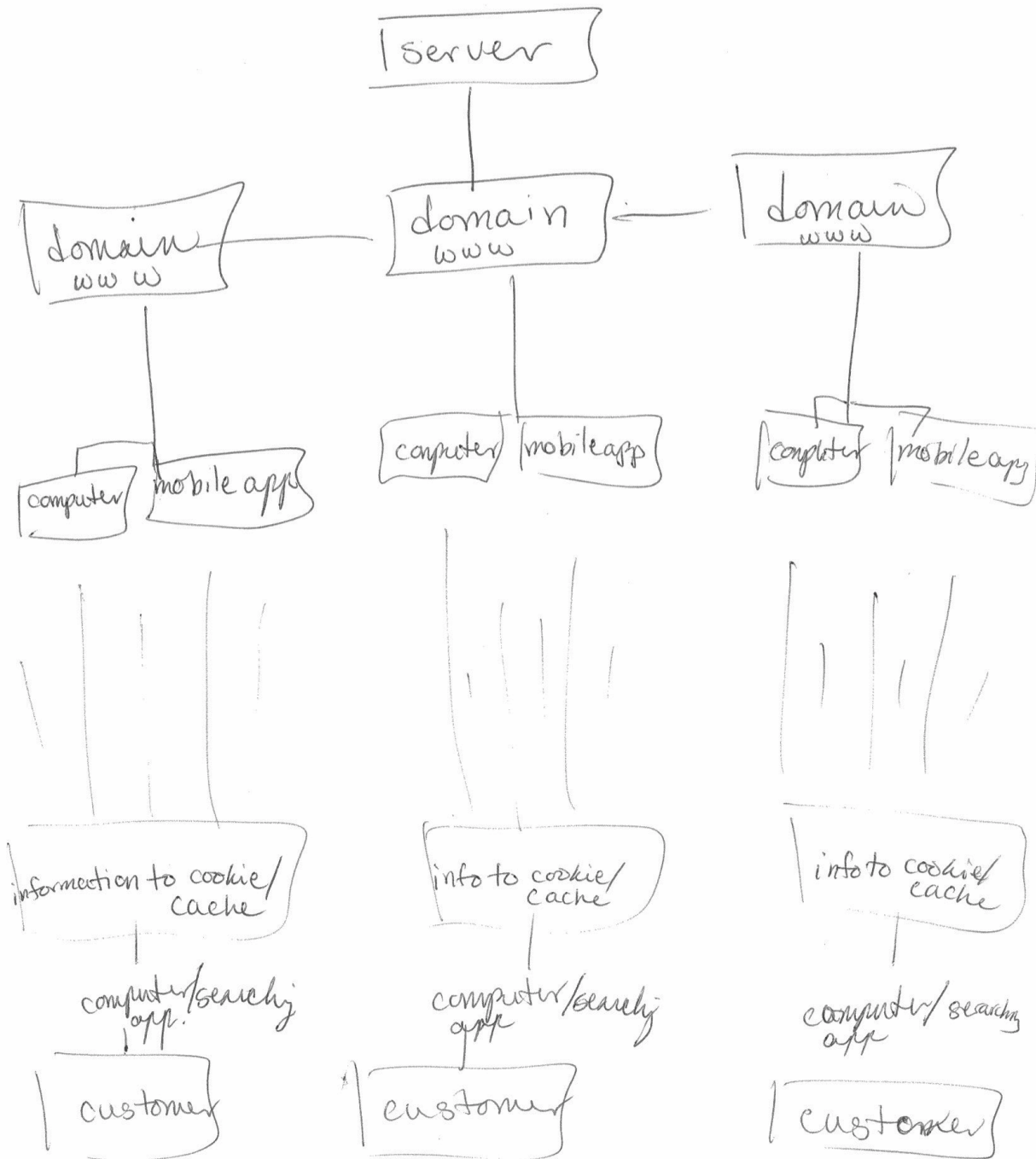
Pop up All the time

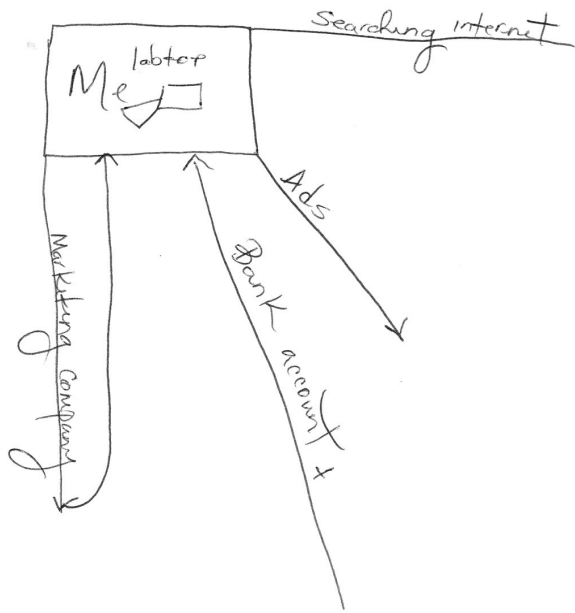


Never stops

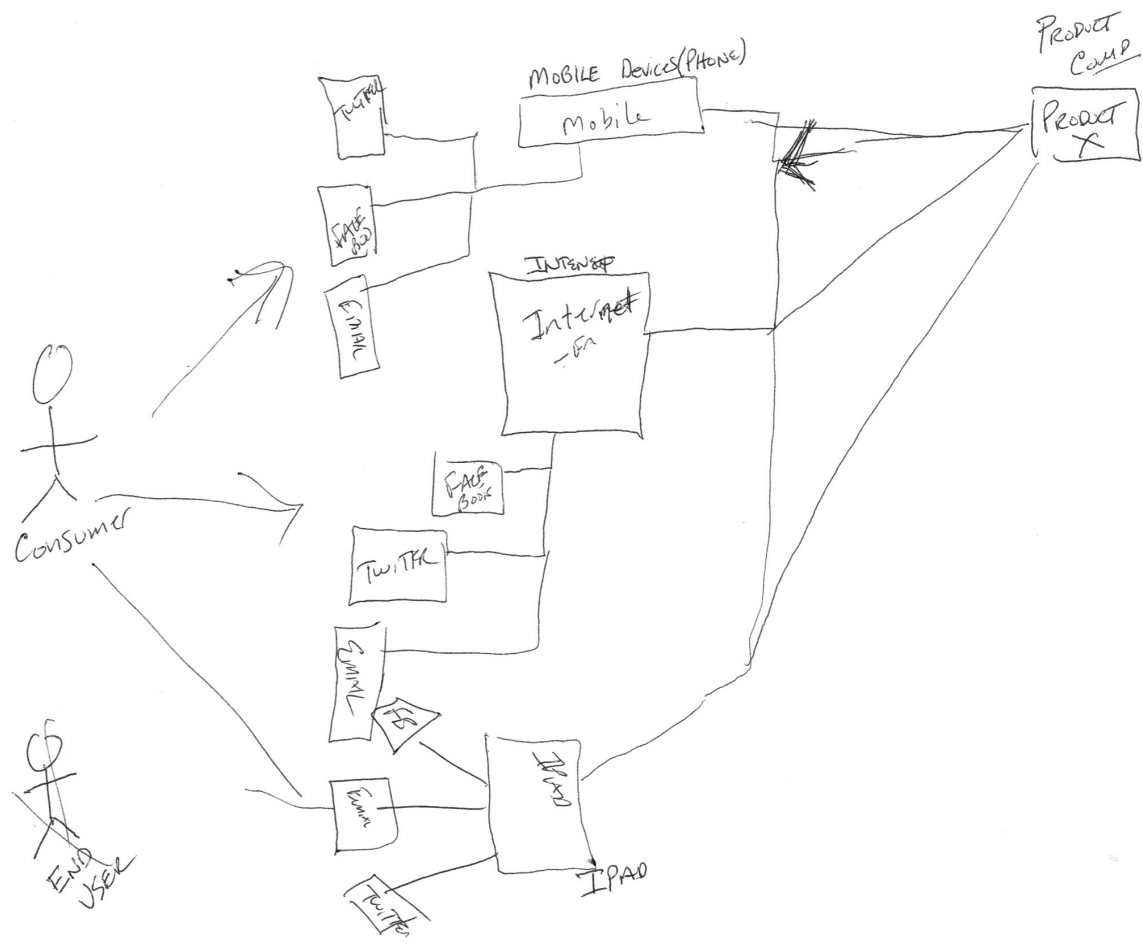


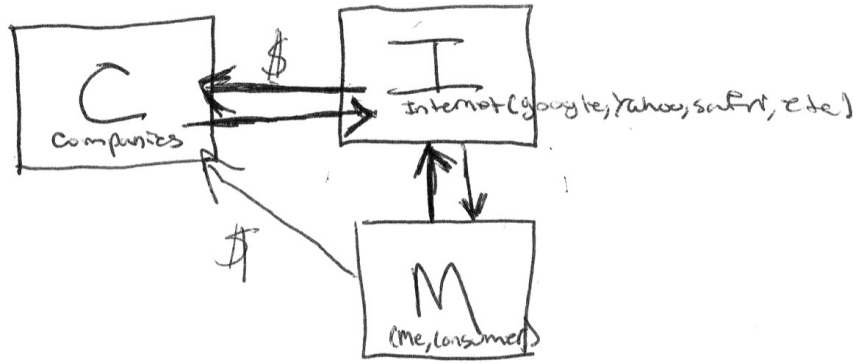


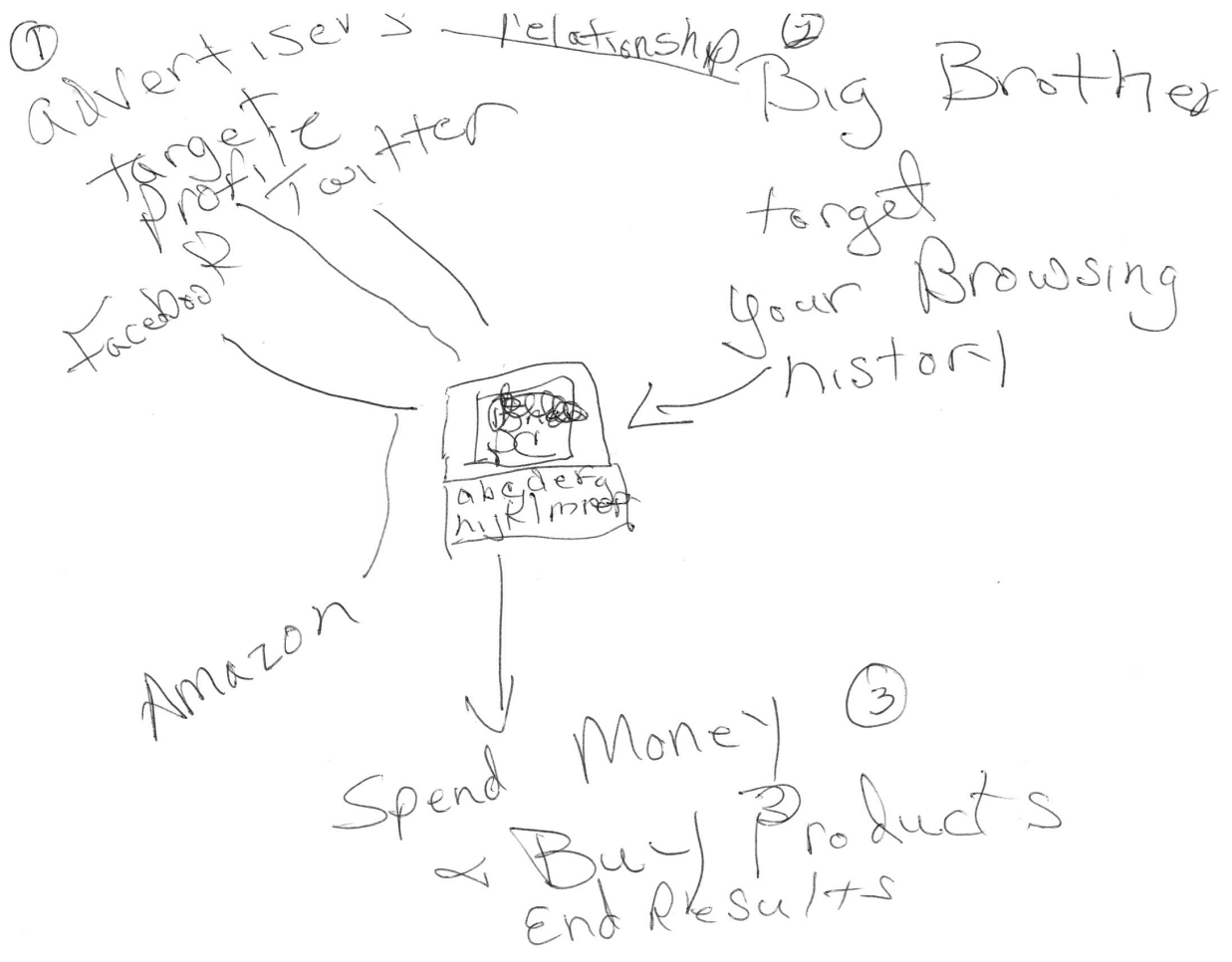


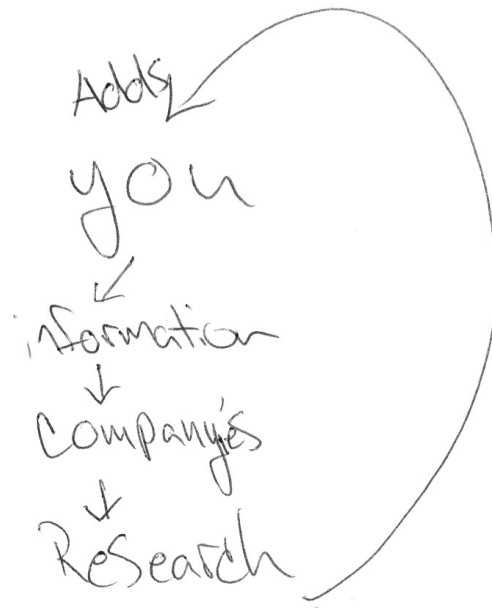


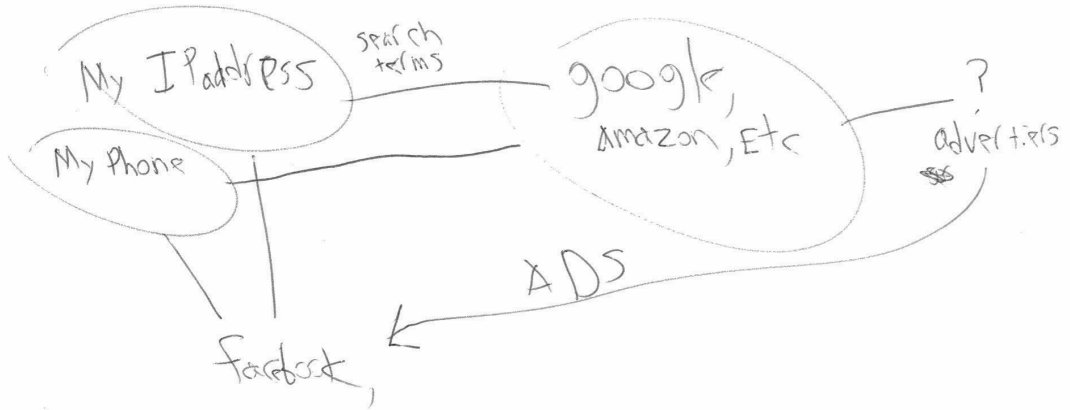




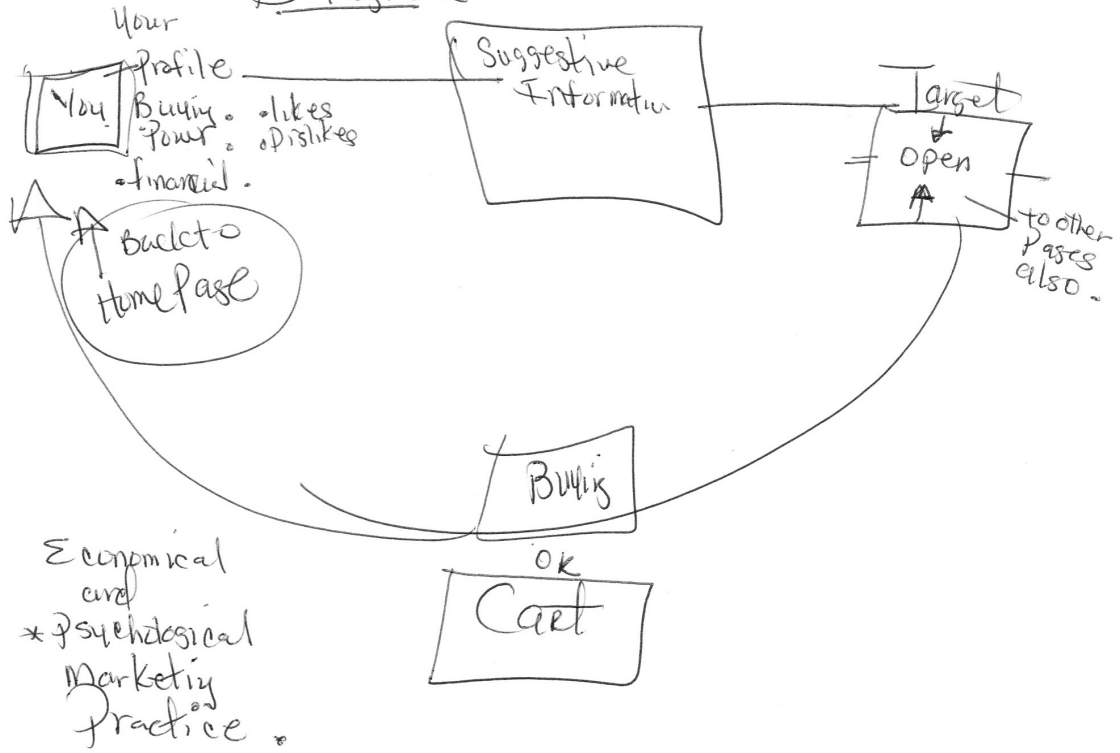








Diagram

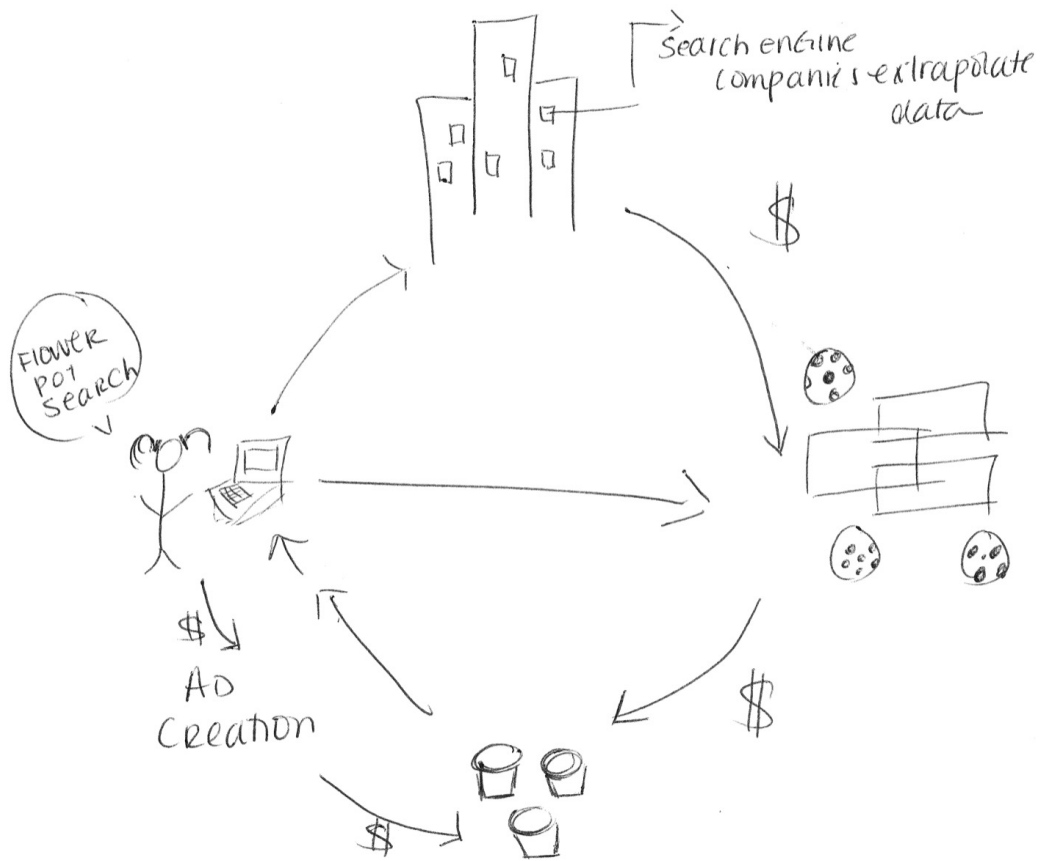


Computer  
Screen

Ad

Triggered by:

- Need for sponsors to finance their website
- Past purchases
- Visits to other websites
- Vendors who want me to buy





## Appendix D

### Manipulation for experiment 1

#### (Chapter 5)

In Experiment 1, all participants completed the following 75 question manipulation survey in a web browser. These 75 questions were presented in randomized order to participants. Differences between measurements taken before and after completing this survey are reported in the results section of Experiment 1. For the manipulation participants were initially prompted, and then again after completing every 15 questions, with the following message:

*For each type of information, indicate how acceptable it is to you for marketers to use this information to personalize the online ads you see.*

*Note: These types of information are currently used by marketers to determine who sees which ads online.*

Then, all questions appeared in the following format:

*You see an online ad based on:* [example of consumer data]

Response options for each question were: *totally unacceptable, unacceptable, slightly unacceptable, neutral, slightly acceptable, acceptable, perfectly acceptable.*

All 75 examples of consumer data rated by participants in Experiment 1 for acceptable use in online advertising personalization appear below.

*You see an online ad based on:*

*whether you use online dating services or not*

*your gender*

*your web browser (e.g. Mozilla Firefox, Internet Explorer, Google Chrome, Safari)*

*the temperature of the city in which you are currently located \**

*the age of your computer*

*the age(s) of the people in your immediate family*

*whether you have been to a casino recently or not*

*the type(s) of news you consume (e.g. political news, international news, sports news, entertainment news)*

*health care products you have purchased*

*the type of music you listen to*

*the age of your mobile device*

*the charitable organizations to which you have made a donation*

*the combined value of any financial securities you own (e.g. stocks, bonds)*

*current street traffic conditions for the city in which you are located*

*the content of the website you are currently viewing*

*bank(s) with which you have an account(s)*

*your health insurance provider*

*the type(s) of movies you watch*

*grocery products you have purchased*

*your political leaning (e.g. conservative, liberal, independent)*

*beauty products you have purchased*

*the type of pet(s) you own*

*personal care or hygiene products you have purchased*

*countries to which you are planning on traveling in the near future*

*your computer's operating system (e.g. Windows, Mac OS X, Ubuntu)*

*your sexual orientation*

*your activity on other mobile apps*

*your hobbies or recreational interests*

*the type(s) of magazine(s) to which you subscribe (print or digital editions)*

*your zip code*

*the make and model of your automobile (e.g. Honda Accord, Ford Focus, Toyota Prius, BMW 3 Series)*

*whether you are a parent or not*

*your ethnicity or race*

*your age*

*the type of home in which you live (e.g. apartment, condo, single-family home)*

*your level of education (e.g. high school diploma, some college, college degree, advanced degree)*

*your telephone number's area code*

*words or phrases you have searched for online (e.g. your searches on Google.com, Bing.com, Yahoo.com)*

*your relationship status (e.g. single, engaged, married, divorced, widowed)*

*countries to which you have recently traveled*

*your precise geographic location (latitude and longitude)*

*which credit card(s) you have (e.g. Visa, Visa Signature, MasterCard, MasterCard Platinum)*

*whether you frequently eat at fast food restaurants or not*

*whether you or your partner are pregnant (expecting a child) or not*

*your place of employment*

*whether you use coupons in a physical store (e.g. grocery store) or not*

*the quantity of online connections (e.g. friends, followers, colleagues) on your social networking or social media accounts*

*which internet service provider (ISP) you are using*

*today's U.S. stock market performance (e.g. NASDAQ Composite, Dow Jones Industrial Average, S&P 500, etc.) \**

*content you post via your social networking or social media accounts*

*the type of mobile phone you own (e.g. basic phone, feature phone, smartphone)*

*the language(s) you speak*

*the type of video game(s) you play*

*items you have purchased online (e.g. from a website)*

*the type(s) of television programs you watch*

*the type of device you are using to access the internet (e.g. personal computer, tablet, smartphone)*

*the weather conditions for the city in which you are currently located (e.g. sunny, overcast, rain, snow) \**

*the city, state, country in which you live*

*today's average Air Quality Index (AQI) across the ten largest US cities \**

*whether you rent or own your place of residence*

*your activity on other websites*

*the type of restaurant(s) at which you most frequently eat*

*the specific retail store(s) at which you shop*

*the amount of content you share online*

*your household income*

*the balance of any loans you have (e.g. auto loan, student loan, personal loan)*

*the current calendar season (e.g. spring, summer, autumn, winter) \**

*the video game console(s) you use*

*household products you have purchased*

*items you have purchased offline (e.g. in a physical store)*

*the content of the mobile app you are currently using*

*financial investment company(s) with which you have an account(s)*

*the number of people in your immediate family*

*who you are connected to online (e.g. friends, followers, colleagues) on your social networking or social media accounts*

*your mobile device's operating system (e.g. iOS, Android, Windows Phone, Blackberry)*

\* These five items do not correspond to personal data but simply data more generally and were included to measure response to online ad customization based on data not specific to the viewer per say.

## Appendix E

### Dependent measures for all experiments

#### (Chapter 5)

***Social Trust.*** A measure of social trust provided the manipulation check in Experiment 3, described below. Sometimes referred to as generalized social trust, this measure consists of seven items including questions taken directly from the social trust scale of the standard General Social Survey and World Values Survey. This 7-item scale was used as the manipulation check in a related study by Mutz (2005), which also supplies the manipulation used in Experiment 3. Participants were asked the following five binary response questions: *Would you say that most of the time people try to be helpful, or that they are mostly just looking out for themselves?*, *Do you think most people would try to take advantage of you if they got a chance, or would they try to be fair?*, *Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?*, *Which statement comes closer to your views, even if neither is exactly right? (Human nature is basically bad, and you can't be too careful in your dealings with people.) versus (Human nature is basically good, and people can be trusted), Which statement comes closer to your views, even if neither is exactly right? (People benefit in the long run if they are honest and fair in their dealings with others) versus (Many people get ahead by being dishonest and unfair in their dealings with others.)* Participants were also asked to respond to the following two items: *Most people in this society are trustworthy* and *When they face temptations, people are not very honest*, with response options: *strongly agree, somewhat agree, somewhat*

*disagree, strongly disagree.* All seven responses were coded so that a higher score pointed towards higher trust response, which were then averaged resulting in the Social Trust measure.

***Website-App Trust.*** Participants were assessed on how much they trust the commercial websites and apps with which they interact. Importantly, this measure taps trust in commercial firms' (as opposed to all organizations') websites and apps. It is a slight modification of the validated *Individual Trust in Online Firms* scale previously developed by Bhattacharjee (2002) to better match the construct of interest to the current studies. Bhattacharjee's trust in online firms measure differs in two notable ways from that used in the studies reported here. First, the original scale was developed using specific internet firms rather than online firms in general as is used in the current studies. Second, to more directly tap trust in the digital platforms themselves on which consumers encounter personalized online advertisements (the Twitter App), rather than trust in these platforms' parent organizations (e.g., Twitter) questions were modified asking respondents to "consider the commercial websites and apps that you use..." at the beginning of each question. In this way, website-app trust as used in these studies taps individual trust in commercial web platforms generally, rather than trust in a specific internet company or in a specific website/app. Additionally, for each of the questions, item-specific response options were used in place of Bhattacharjee's generic Likert response options. (As demonstrated by Pasek & Krosnick, 2010, item-specific response options tend to be more effective than generic Likert scales.) Participants were first prompted: *For the questions below, consider the commercial websites and apps that you use, which includes those owned and operated by a company.* Then, they were asked seven questions. Each of these questions began with the phrase: *For the commercial websites and apps that you use...* followed by each of the following taken directly from the original scale validated by Bhattacharjee: *...how fair are they in the way they use*

*information about you? (not at all fair, a little fair, somewhat fair, very fair, extremely fair); ...how often do they act in your best interests? (never act in my best interests, rarely act in my best interests, sometimes act in my best interests, often act in my best interests, always act in my best interests); ...how fair are their Terms of Service (ToS) agreements? (not at all fair, a little fair, somewhat fair, very fair, extremely fair); ...to what extent are they receptive to your wishes? (not at all receptive to my wishes, a little receptive to my wishes, somewhat receptive to my wishes, very receptive to my wishes, extremely receptive to my wishes); ...how fair are they in the way they interact with you? (not at all fair, a little fair, somewhat fair, very fair, extremely fair); ...how often do they try to address your concerns? (never try to address my concerns, rarely try to address my concerns, sometimes try to address my concerns, often try to address my concerns, always try to address my concerns); ...how trustworthy are they? (not at all trustworthy, a little trustworthy, somewhat trustworthy, very trustworthy, extremely trustworthy).* The seven responses were coded so that a higher score pointed towards a higher trust response, which were then averaged resulting in the Website-App Trust measure.

***Personalized Advertising Awareness.*** A measure of heightened awareness of advertising personalization practices supplied the manipulation check for Experiment 2 described below. A 5-item scale was developed to assess awareness of personalized online advertising linked in response to confronting some participants with information about personalization and the use of consumer data to selectively present ads. Participants were asked to respond to the following three items: *On the Internet, different people receive different advertisements when accessing the same website/app at the same time; Advertisements appearing on websites/apps are personalized for each viewer; Advertisements appearing on websites/apps are randomly generated for each viewer.* Response options for these three items were: *never, very rarely, rarely, sometimes, most*



*of the time, almost always, always*. The third item was reverse coded, so that all three questions tapped increased awareness of advertising personalization. Participants were also asked to respond to two related estimation questions: *In your estimation, what percent of advertisements appearing on websites/apps have been personalized for the viewer?; In your estimation, what percent of advertisements appearing on websites/apps are randomly generated for the viewer?* Response options for these two questions were: *0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%*. The latter question was reverse coded, so that both estimation questions tapped heightened awareness of advertising personalization, using proportion of ads personalized as proxy for heightened awareness about personalization. Responses from all five questions were averaged resulting in the measure of Personalized Advertising Awareness.

***Stated Preference for Personalized Advertising.*** Preference for personalized advertising was assessed using multiple measures capturing participants' stated and revealed preferences for receiving personalized (vs. non-personalized) online advertising. Stated preference was assessed using a 3-item scale, which asked participants the following questions: Indicate how much you agree or disagree with the following statement: I would prefer to see advertisements on websites and apps that have been personalized for me using information about me, rather than random or generic ads. (strongly disagree, disagree, somewhat disagree, somewhat agree, agree, strongly agree); To you individually, how acceptable do you find this practice of personalized online advertising, where marketers use information about you to determine which ads you receive? (totally unacceptable, unacceptable, slightly unacceptable, slightly acceptable, acceptable, perfectly acceptable); In general, do you favor OR oppose having online advertisements personalized for you using information about you? (strongly favor, favor, somewhat favor,

somewhat oppose, oppose, strongly oppose). The last item was reverse coded, so that all items pointed towards stronger preference for personalized advertising. Responses to the three items were averaged to generate the Stated Preference for Personalized Advertising measure.<sup>39</sup>

***Revealed Preference for Personalized Advertising.*** In addition to asking participants whether they preferred personalized advertising online, a measure of revealed preference was developed. As what participants think they prefer and actually prefer may differ. When used this measure was placed at the very end of the experiments immediately following demographic questions. This positioning was intended to maximize arousal in participants regarding their personal information. First, participants were prompted with the statement: *Based on your answers to some of the previous questions in this survey, we've selected 3 books for you that you might be interested in.* This was followed by the question: *Would you like to see these books?* response options: *Yes, show me the books,* (which was scored as 1); *No, thanks.,* (scored as 0). Participants who chose to see the three books were then shown a new prompt, which appeared directly above three images of book covers: *If you would like to visit one of the book's Amazon.com product pages, where it is available for purchase, click on the book's cover below. If you do not want to look at any of the books that have been selected for you based on some of your answers in this survey, just proceed to the next page to end the survey.* Though participants were told the books had been selected for them based on their survey responses, the same three books were shown to all participants regardless of their previous answers. Participants who clicked on one or more of the “personalized” book advertisements received a score of 1 and those who did not click on any

---

<sup>39</sup> In Experiment 1, only two of the three items were used for this scale. Yet the outcome was significant for group differences at  $p < .001$  just as in Experiment 2, which used all three items and a similar manipulation to Experiment 1. Thus, there is no reason to believe that the 2-item version measured substantively different response than the 3-item version given the construct of interest (stated preference for personalized advertising).

of the book ads shown received a score of 0. The two scores from revealing/not revealing the books and clicking/not clicking on the ads were averaged resulting in the revealed preference for personalized advertising measure.

***Stated Preference for Personalized Content.*** A 6-item scale from a prior national survey (Stevenson & Pasek, 2015) was used to assess stated preference for online content personalization. Similar to the stated preference for personalized advertising measure used before, this scaled tapped how personalized said they wanted various kinds of online media in addition to advertisements (news, social media, prices). As not all people are familiar with the practice of personalizing ads let alone news, social media, and prices, participants were first prompted with the following statement: *Some websites and apps personalize the content you see using information about you. This means the content you see differs from what others see. Information used to personalize what you see on websites and apps includes, but is not limited to, things like which websites you've visited or which apps you've used, your age, income, marital status, race/ethnicity, political affiliation, or location, purchases you've made online or in a store, which device or software you are using to access the Internet.* As this prompt contained abundant information about some of the ways online personalization is achieved, the prompt and subsequent questions were positioned in the experiment after the revealed personalization preference questions, as not to influence these related measures. Following the brief prompt about what content “personalization” refers to, participants were asked: *Indicate how personalized you would like each of the following 6 items below when you see them on websites or apps.* This was followed by the following six items presented in randomized order: *advertisements for products and services; political advertisements; discounts on products and*

*services; news stories; friends' social media posts; prices of products and services.* Response options for all six were: *not at all personalized, a little personalized, somewhat personalized, very personalized, extremely personalized.* Responses to the six types of online media were averaged to produce the measure stated preference for personalized content.

A note about the two measures of revealed preference for personalized content: These two measures were developed for Experiment 4—revealed preference for *surreptitiously-personalized* content and revealed preference for *overtly-personalized* content (described below in detail). Each measure contained four items. Each item consisted of two short 15-second video clips, only the *titles* of which were shown to participants when making their selections. As participants did not watch the videos until the very end of the survey and after all dependent variable measures were asked, there was no “contamination” of participants’ responses due to which videos they selected to watch and later watched. (Additionally, all participants, regardless of their selections, watched the same videos at the end of the survey, as described in the debrief form.) For each pair of video titles, participants were asked to select which one of the two videos they wanted to watch later on during the experiment. As participants were made to believe they had to watch each video selection later in the experiment and in order to receive payment, their selections represent a revealed preference. They were choosing what they would have to watch in the very near future, during the experiment. Also, both video titles for each question appeared directly above an image of an interface of a popular online video player to further emphasize that participants were choosing between videos that they would watch. This complimented the instructions that had already been stated. For all eight pairs of videos in these two different measures, one of the two videos was personalized for the participant in that its title included

information corresponding to a specific attribute about the participant (either participant's gender, age, location, or operating system). The other video title in the pair did not include information related to the participant, corresponding to the non-personalized option. Importantly for these two measures (for surreptitiously- and overtly-personalized content) not to contaminate one another, the four surreptitiously-personalized video pairs were always presented before the overtly-personalized ones. This was to avoid arousing participants about content personalization when measuring preference for surreptitiously-personalized content, the questions for which needed to be asked prior to measuring revealed preference for overtly-personalized content, as the latter always included a label above one of the options in each pair indicating one of the videos was "*Personalized for you based on...*" In this way, only after making their selections for content they had no reason to suspect had been personalized for them did respondents select which videos to watch from the overtly-personalized pairs of video titles.

***Revealed Preference for Surreptitiously-Personalized Content.*** Participants chose one video from each of the following four pairs of videos based only on seeing the videos' titles: *Fitness Advice for Men* vs. *Fitness Advice for Women*; *Health Tips for People Under 30* vs. *Health Tips for People Age 30 and Over*; *Advice for Weekend Road Trips West of the Mississippi River* vs. *Advice for Weekend Road Trips East of the Mississippi River*; *Computer Security Tips for Windows Users* vs. *Computer Security Tips for Mac Users*. For the gender- and age-related video pairs, whether participants selected the personalized video title containing information corresponding to them or not was determined by prior responses to questions asking participants to indicate their age range and gender identity included in a short set of demographic questions

asked at the very beginning of the experiment.<sup>40</sup> For the location- and operating system-based videos, which one of the two videos in these pairs corresponded to the participant was determined using the Qualtrics survey platform. This platform provides the option to receive metadata describing the rough location (determined by a survey respondent's IP address) and HTTP "user agent," which includes, among other technical information, the respondent's operating system. In this case, the Qualtrics-supplied US state was used to determine whether participants were taking the survey in a state east or west of the Mississippi River, which was used to score whether respondents selected the weekend road trips video for the half of the country in which they were taking the survey, and presumably lived. For the video relating to one's operating system, this was extracted from the HTTP user agent string and used to determine whether respondents selected the video corresponding to their operating system or not. Participants using neither Windows nor Mac operating systems (e.g., Linux) did not have this video selection counted in their surreptitiously-personalized revealed preference scores. Participants received a score of 1 when they selected the video title that contained information corresponding to themselves (e.g., participants in their 20s selecting the video *Health Tips for People Under 30*) and a 0 when selecting the alternative, non-personalized option (e.g., respondents completing the survey in New Mexico choosing the video *Advice for Weekend Road Trips East of the Mississippi River*). These scores were summed to produce the measure revealed preference for surreptitiously-personalized content.

---

<sup>40</sup> Asking demographic questions at the beginning of a survey, or anywhere other than the end for that matter, typically goes against best practices that suggest only asking demographic questions at the very end of a survey or interview. This is to minimize effects of identity-based priming known to occur when answering questions about oneself related to income, race, age, etc. (Jackson, 2009).

***Revealed Preference for Overtly-Personalized Content.*** Participants chose one video from each of the following four pairs, based again on seeing the videos' titles. However, in addition to seeing the video titles, this time prior to making the four selections participants were given the following additional instruction: *For the last 4 video selections, 1 of the 2 videos has been personalized for you. Meaning 1 of the 2 videos has been selected based on information about you. This personalized video is labeled "Personalized for you based on..."* Then, for each of the last four pairs of videos, in prominent red font appearing above one of the two video's titles in each pair was the label: *Personalized for you based on your (gender, age, location, computer operating system)*. Which of the two videos included this label was determined based on respondents' prior answers to the age and gender demographic questions and IP-address derived US state and HTTP user agent derived operating system, as previously described. In this way, for the overtly-personalized revealed preference measure, the video in each pair corresponding to the respondent's personal information was clearly labeled as personalized for them and based on which information. This overt personalization label appeared dynamically based on participant information above one of the following titles in each pair: *Tips for Men to Succeed at Work-Life Balance vs. Tips for Women to Succeed at Work-Life Balance; Financial Tips for People Under 30 vs. Financial Tips for People Age 30 and Over; Tips for Outdoor Adventures West of the Mississippi River vs. Tips for Outdoor Adventures East of the Mississippi River; Time-Saving Computer Shortcuts for Windows Users vs. Time-Saving Computer Shortcuts for Mac Users*. As with the surreptitiously-personalized set of video selections, participants again received a score of 1 when selecting the (labeled) personalized option and a 0 when not selecting the personalized option for each of the four pairs. These scores were summed to produce the measure revealed preference for overtly-personalized content.

**Online Privacy Concern 1.** A single item measure was included to assess individual privacy concern when using the internet. Participants were asked: *How important is online privacy to you personally?* Response options were: *not at all important, very unimportant, somewhat unimportant, neither important nor unimportant, somewhat important, very important, extremely important.*

**Online Privacy Concern 2.** A validated 16-item measure developed by Buchanan et al. (2002) assessing individual online privacy concern was also used. Participants were asked the following questions: *Are you concerned... (followed by) ...about online organizations not being who they claim they are? ...that you are asked for too much personal information when you register or make online purchases? ...about online identity theft? ...about people online not being who they say they are? ...that information about you could be found on an old computer? ...who might access your medical records electronically? ...about people you do not know obtaining personal information about you from your online activities? ...that if you use your credit card to buy something on the internet your credit card number will be obtained/intercepted by someone else? ...that if you use your credit card to buy something on the internet your card will be mischarged? ...that an email you send may be read by someone else besides the person you sent it to? ...that an email you send someone may be inappropriately forwarded to others? ...that an email you send someone may be printed out in a place where others could see it? ...that a computer virus could send out emails in your name? ...about emails you receive not being from whom they say they are? ...that an email containing a seemingly legitimate internet address may be fraudulent?* In general, *how concerned are you about your privacy while you are using the internet?* Response options for all items were: *not at all concerned, slightly concerned, moderately*



*concerned, very concerned, extremely concerned.* Responses to these 16 items were averaged to produce a measure of online privacy concern, identical that the validated scale from Buchanan.

***Opposition to Consumer Data Collection.*** A single item measure served to tap how opposed respondents were to consumer data collection on the internet. Participants were asked: *Indicate how much you agree or disagree with the following statement: Internet companies collect too much of my personal information.* Response options were: *strongly disagree, disagree, somewhat disagree, neither agree nor disagree, somewhat agree, agree, strongly agree.*

***Consumer Data Control Self-Efficacy.*** Ryan and Dzewaltowski (2002) define self-efficacy as, “the beliefs an individual has about his or her ability to engage in behaviors that lead to expected outcomes” (p. p. 491). Applying this notion of self-efficacy to the context of consumer data use in personalized advertising, a single item measure was used to gauge the degree to which respondents believed they had the ability to control how marketers used their personal information to target them with ads online. Participants were asked the following: *Indicate how much you agree or disagree with the following statement: I can control which types of personal information advertisers use to customize the online advertisements I receive.* Response options were: *strongly disagree, disagree, somewhat disagree, neither agree nor disagree, somewhat agree, agree, strongly agree.*

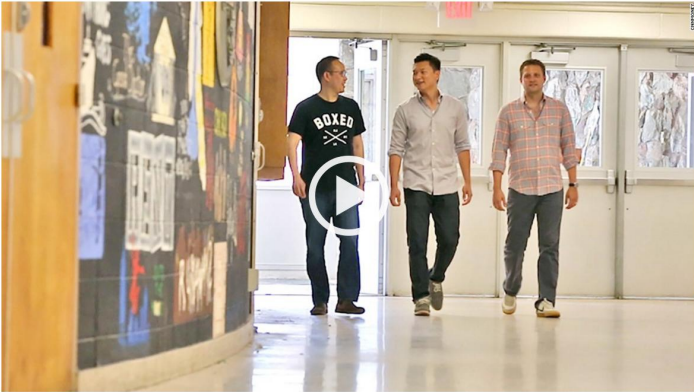
***Trustworthiness of Example Ads.*** Participants were asked to evaluate how trustworthy they found six online ads to be, which were presented within different websites or apps. Prior to rating the six ads (and six websites/apps) the following instruction was displayed: *Indicate how*

trustworthy you find the advertisements appearing on the following websites/apps and also how trustworthy you find the websites/apps themselves. In each example image, the advertisements have been labeled "AD" to distinguish the ad from the website/app on which it appears. For each of the six ads, participants were asked: *How trustworthy is the advertisement on this website?* (or "...*this app?*"). Response options were: *not at all trustworthy, a little trustworthy, somewhat trustworthy, very trustworthy, extremely trustworthy*. The six responses were averaged resulting in the trustworthiness of example ads measure. The six ads and websites/apps they were displayed within are included below.


***Trustworthiness of Example Websites/Apps.*** In addition to stating how trustworthy participants found the example ads, they were also asked to rate how trustworthy they found each of the surrounding websites or apps displaying the advertisements. Participants were asked: *How trustworthy is this website?* (or "...*this app?*"). Response options were: *not at all trustworthy, a little trustworthy, somewhat trustworthy, very trustworthy, extremely trustworthy*. The six responses were averaged resulting in the trustworthiness of example ads measure. The six ads and websites/apps they were displayed within are included below.

CNN Money Business Markets Tech Media Personal Finance Small Biz Luxury stock tickers


# Business News



This high school spawned \$2.5 billion worth of startup founders



Lux, 24-hour, and dirt cheap: Insi...



2 Credit Cards Are Offering 0% APR Until September 2017

Advertisement NextAdvisor

**AD**

Merriam-Webster

GAMES | WORD OF THE DAY | VIDEO | WORDS AT PLAY | FAVORITES Follow: f t y

Dictionary Hello! What word would you like to learn today?

An Encyclopædia Britannica Company

**AD**

Walmart  
Get 5% Cashback Bonus  
visit Discover Deals

DISCOVER  
Learn More  
Terms apply

his? her?


The Awkward Case of 'His or Her'  
Singular 'their' might be a better choice

WORD OF THE DAY  
herald  
to announce or to greet with enthusiasm  
Get Word of the Day daily email!  
Your email address


January Words of the Day Quiz  
How many of January's Words of the D...

eHow DISCOVER  Search Log In Join


# HOME DECOR & REPAIR




**DERMOLOGY ANTI AGING KIT**  
3 STEPS TO YOUNGER LOOKING SKIN  
Remove Appearance of Wrinkles  
Reduce Signs of Aging  
Now with Resveratrol  
**TRY IT RISK FREE TODAY!**



22 DIY Ways to Update Your Home on a Small Budget +6



Make a DIY Table Lamp From Marble Tiles +1




How to Install Wall Sconce Lighting +2

↑  
AD

**CNNMoney**

DOW ↑ 0.53%

- Yahoo and ABC partner for news 'alliance'
- Stocks in tug of war as Greece weighs on investors
- Tax hikes and jobs
- Mortgage help for unemployed disappears



← AD

- To fix the economy, first fix the housing market.
- Unreleased Michael Jackson song in Jeep ad
- Fewer students have jobs on graduation day

Merriam-Webster Dictionary

eloquent

**Full Definition**


**1** : marked by forceful and **fluent** expression <an *eloquent* preacher>

**2** : vividly or movingly expressive or **revealing** <an *eloquent* monument>

**el·o·quent·ly** *adverb*


**Examples**

- an *eloquent* writer and speaker, Elizabeth Cady Stanton was one of the founders of the


KITCHENSMART.  **\$50.00 off coupon**  
Kitchen Smart Appliances

← **AD**

Money 10 of 10

mango pineapple real fruit smoothie de McCafé® 

**eHow** **8 Rules of Thumb for Saving for Retirement**  
By Rose Johnson



Saving for retirement is an important endeavor that you should plan carefully. According the United States Department of Labor, more than 50 percent of Americans fail to calculate the amount needed for retirement. Several rules of thumb exist for retirement savings. Knowing these tips can help maximize your retirement savings.

← **AD**

## Appendix F

### Study-specific acknowledgements

#### *Study 1 (Chapter 3)*

I wish to thank the following individuals who supported Study 1 by providing feedback on the initial design and/or interpretation of results: Christian Sandvig, Sonya Dal Cin, Carol Crawford, Robin Means Coleman, Jeff Hancock, Tim Libert, Solon Barocas, Matt Crain, Allison Earl, Roland Gau, Michael Tschantz, and several anonymous staff members at multiple internet advertising companies. This study was generously supported by a Winthrop B. Chamberlain Grant for Graduate Student Research, Department of Communication Studies, University of Michigan.

#### *Study 2 (Chapter 4)*

I wish to thank the following individuals who supported Study 2 by providing constructive feedback and in some cases direct research assistant: Christian Sandvig, Sonya Dal Cin, Nicole Ellison, Shane Greenstein, Robin Means Coleman, Christian Fuchs, Didem Ozkul, Emily Saidel, Jack Wilson, and Diana Karge. This study was generously supported by two research grants from the Department of Communication Studies and a grant from the Horace H. Rackham School of Graduate Studies, University of Michigan.

*Study 3 (Chapter 5)*

I wish to thank the following individuals for constructive feedback on study designs and comments on conference papers versions, posters, and presentations containing either pilot and/or final results reported in Study 3: Christian Sandvig, Sonya Dal Cin, Nicole Ellison, Shane Greenstein, Leslie John, Tami Kim, Joe Turow, Jonathan Cave, Matthew Jackson, Amit Schejter, Sascha Meinrath, Josh Pasek, Yioryos Nardis, Joe Bayer, Aimee Rickman, Matt Crain, Jennifer Ihm, Colin Rhinesmith, Julia Raz, ShinJoung Yeo, Ajaay Chandrasekaran, David Hauser, Kristen Vaccaro, Karrie Karahalios, Kevin Hamilton, Jean-Christophe Plantin, Jennifer Kelley, Roland Gau, and multiple anonymous reviewers from the International Communication Association and American Marketing Association Marketing & Public Policy annual conferences. This study was generously supported by research grants from the Department of Communication Studies and the Horace H. Rackham School of Graduate Studies, University of Michigan.

## References

- Acquisti, A., John, L. K., & Loewenstein, G. (2013). What is privacy worth? *Journal of Legal Studies*, 42(2), 249-274.
- Aguirre, E., Roggeveen, A. L., Grewal, D., & Wetzels, M. (2016). The personalization-privacy paradox: Implications for new media. *Journal of Consumer Marketing*, 33(2), 98-110.
- Alwitt, L. F., & Prabhaker, P. R. (1994). Identifying who dislikes television advertising: Not by demographics alone. *Journal of Advertising Research*, 34(6), 17-30.
- American Marketing Association. (2013). Marketing, a definition. Retrieved from <https://www.ama.org/AboutAMA/Pages/Definition-of-Marketing.aspx>
- Anderson, J., & Rainie, L. (2012). The future of apps and web. Pew Research Center. Retrieved from <http://www.pewinternet.org/2012/03/23/the-future-of-apps-and-web/>
- Anderson, L. M., & Taylor, R. L. (1995). McCarthy's 4Ps: Timeworn or time-tested? *Journal of Marketing Theory and Practice*, 3(3), 1-9.
- Andrejevic, M. (2007). *iSPY: Surveillance and power in the interactive era*. Lawrence, KS: University Press of Kansas.
- Andrews, J. C. (1989). The dimensionality of beliefs toward advertising in general. *Journal of Advertising*, 18(1), 26-35.
- Arthur, L. (2013). *Big data marketing: Engage your customers more effectively and drive value*. Hoboken, NJ: John Wiley & Sons.
- Ashworth, L., & Free, C. (2006). Marketing dataveillance and digital privacy: Using theories of justice to understand consumers' online privacy concerns. *Journal of Business Ethics*, 67(2), 107-123.
- Auletta, K. (2010). *Googled. The end of the world as we know it*. London: Virgin.
- Awad, N. F., & Krishnan, M. S. (2006). The personalization privacy paradox: An empirical evaluation of information transparency and the willingness to be profiled online for personalization. *MIS Quarterly*, 30(1), 13-28.
- Bailenson, J. N., Iyengar, S., Yee, N., & Collins, N. A. (2009). Facial similarity between voters and candidates causes influence. *Public Opinion Quarterly*, 72(5), 935-961.
- Barocas, S. (2014). Data mining and the discourse on discrimination. Data Ethics Workshop, Conference on Knowledge Discovery and Data Mining, New York, NY.



- Beldad, A., De Jong, M., & Steehouder, M. (2010). How shall I trust the faceless and the intangible? A literature review on the antecedents of online trust. *Computers in Human Behavior*, 26(5), 857-869.
- Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. New Haven, CT: Yale University Press.
- Berkowitz, S. (1997). Analyzing Qualitative Data. In J. Frechtling, L. Sharp, & Westat (Eds.), *User-Friendly Handbook for Mixed Method Evaluations* (Ch. 4). Arlington, VA: National Science Foundation.
- Bhattacharjee, A. (2002). Individual trust in online firms: Scale development and initial test. *Journal of Management Information Systems*, 19(1), 211-241.
- Bogdan R. B., & Biklin, S. K. (1998). *Qualitative research for education: An introduction to theory and methods* (3rd ed.). Needham Heights, MA: Allyn & Bacon.
- Borden, N. H. (1964). The concept of the marketing mix. *Journal of Advertising Research*, 4(2), 2-7.
- Bright, A. L. (2015). Privilege in a police car: The story of my unresolved ride-along. *The Qualitative Report*, 20(5), 541.
- Buchanan, T., Paine, C., Joinson, A. N., & Reips, U. D. (2007). Development of measures of online privacy concern and protection for use on the internet. *Journal of the American Society for Information Science & Technology*, 58(2), 157-165.
- Bucy, E. P. (2003). Media credibility reconsidered: Synergy effects between on-air and online news. *Journalism & Mass Communication Quarterly*, 80(2), 247-264.
- Burchell, B., & Wilkinson, F. (1997). Trust, business relationships and the contractual environment. *Cambridge Journal of Economics*, 21(2), 217-237.
- Calo, R. (2014). Digital Market Manipulation, *George Washington Law Review*. 82(4), 995-1051.
- Carroll, J. M., & Thomas, J. C. (1982). Metaphor and cognitive representation of computing systems. *IEEE Transactions on Systems, Man, & Cybernetics*.
- Castells, M. (2013). *Communication Power*. Oxford, UK: Oxford University Press.
- Chen, S. C., & Dhillon, G. S. (2003). Interpreting dimensions of consumer trust in e-commerce. *Information Technology and Management*, 4(2-3), 303-318.
- Cohen, L. (2003), *A consumers' republic: The politics of mass consumption in postwar America*. New York: Knopf.

- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98.
- Coulter, R. A., Zaltman, G., & Coulter, K. S. (2001). Interpreting consumer perceptions of advertising: An application of the Zaltman metaphor elicitation technique. *Journal of Advertising*, 30(4), 1-21.
- Council of Economic Advisers, U.S. (2015). Big data and differential pricing, Executive Office of the President of the United States. Retrieved from <http://www.whitehouse.gov/blog/2015/02/06/economics-big-data-and-differential-pricing>
- Crain, M. (2016). The limits of transparency: Data brokers and commodification. *New Media & Society*.
- Crain, M. (2014). The revolution will be commercialized: Finance, public policy, and the construction of internet advertising. Doctoral dissertation, University of Illinois at Urbana-Champaign. Retrieved from <http://ideals.illinois.edu/handle/2142/46574>
- Culnan, M. J., & Armstrong, P. K. (1999). Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organization Science*, 10(1), 104-115.
- DeBruine, L. M. (2002). Facial resemblance enhances trust. *Proceedings of the Royal Society of London*, 269, 1307-12.
- DeBruine, L. M. (2004). Resemblance to self increases the appeal of child faces to both men and women. *Evolution and Human Behavior*, 25, 142-54.
- DeBruine, L. M. (2005). Trustworthy but not lust-worthy: Context-specific effects of facial resemblance. *Proceedings of the Royal Society of London B*, 272(1566), 919-922.
- Deighton, J., & Johnson, P. (2013). The value of Data: Consequences for insight, innovation & efficiency in the U.S. economy. Data-Driven Marketing Institute. Retrieved from <https://thedma.org/advocacy/data-driven-marketing-institute/value-of-data-study-2013/>
- Deighton, J. & Johnson, P. (2015). The value of data: Driving insight, innovation & efficiency in the U.S. economy (update to 2013 report). Data-Driven Marketing Institute. Retrieved from <https://thedma.org/wp-content/uploads/Value-of-Data-Summary.pdf>
- Derbaix, C., & Pecheux, C. (2003). A new scale to assess children's attitude toward TV advertising. *Journal of Advertising Research*, 43(04), 390-399.
- Dibb, S., & Simkin, L. (1991). Targeting, segments and positioning. *International Journal of Retail & Distribution Management*, 19(3).
- Dixon, P. (2010). The one-way-mirror society: Privacy implications of the new digital signage networks. World Privacy Forum report. Retrieved from <http://www.worldprivacyforum.org/pdf/onewaymirrorsocietyfs.pdf>.

- Díaz-Morales, R. (2015). Cross-device tracking: Matching devices and cookies. arXiv preprint. Retrieved from <https://arxiv.org/abs/1510.01175>
- DoubleClick. (2013). Real time bidding—Just how fast is it? Retrieved from <http://plus.google.com/+doubleclickdigitalmarketing/posts/UHMHAmYKZsE>
- Dutta-Bergman, M. J. (2006). The demographic and psychographic antecedents of attitude toward advertising. *Journal of Advertising Research*, 46(1), 102-112.
- Dutton, W. H. (Ed.). (2013). *The Oxford handbook of internet studies*. Oxford, UK: Oxford University Press.
- Edelman, B., Ostrovsky, M., & Schwarz, M. (2005). Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords. National Bureau of Economic Research, No. w11765.
- Ehrenreich, B. (2010). *Nickel and dimed: On (not) getting by in America*. New York: Henry Holt and Co.
- Ekelund Jr., R. B. (1970). Price discrimination and product differentiation in economic theory: An early analysis. *The Quarterly Journal of Economics*, 268-278.
- Electronic Privacy Information Center. (2016). Public opinion on privacy. Retrieved from <https://epic.org/privacy/survey/>
- Elmaghraby, W., & Keskinocak, P. (2003). Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. *Management Science*, 49(10), 1287-1309.
- eMarketer. (2014). US programmatic ad spend tops \$10 billion this year, to double by 2016. Retrieved from <http://www.emarketer.com/Article/US-Programmatic-Ad-Spend-Tops-10-Billion-This-Year-Double-by-2016/1011312>
- Eslami, M., Karahalios, K., Sandvig, C., Vaccaro, K., Rickman, A., Hamilton, K., & Kirlik, A. (2016). First I "like" it, then I hide it: Folk Theories of Social Feeds. *Proceedings of the 34rd Annual SIGCHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, 2371-2382.
- Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K., and Sandvig, C. (2015). "I always assumed that I wasn't really that close to [her]:" Reasoning about invisible algorithms in the news feed. *Proceedings of the 33rd Annual SIGCHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, 153-162.
- Evans, David S. (2009). The online advertising industry: Economics, evolution, and privacy. *Journal of Economic Perspectives*, 23(3): 37-60.

- Farahat, A., & Bailey, M. C. (2012, April). How effective is targeted advertising?. In *Proceedings of the 21st international conference on World Wide Web*, Association for Computing Machinery, 111-120.
- Federal Communications Commission. (2015). In the matter of consumer watchdog petition for rulemaking to require edge providers to honor 'Do Not Track' requests, RM-11757. Retrieved from [https://transition.fcc.gov/Daily\\_Releases/Daily\\_Business/2015/db1106/DA-15-1266A1.pdf](https://transition.fcc.gov/Daily_Releases/Daily_Business/2015/db1106/DA-15-1266A1.pdf)
- Federal Trade Commission. (2016). Big data: A tool for inclusion or exclusion? Understanding the issues. Retrieved from <https://www.ftc.gov/reports/big-data-tool-inclusion-or-exclusion-understanding-issues-ftc-report>
- Federal Trade Commission. (2015). Cross-device tracking, A Federal Trade Commission workshop. Retrieved from <https://www.ftc.gov/news-events/events-calendar/2015/11/cross-device-tracking>
- Federal Trade Commission (2016). FTC issues warning letters to app developers using 'Silverpush' code. Retrieved from <https://www.ftc.gov/news-events/press-releases/2016/03/ftc-issues-warning-letters-app-developers-using-silverpush-code>
- Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2). Stanford, CA: Stanford University Press.
- Fischer-Hübner, S., Hoofnagle, C. J., Krontiris, I., Rannenberg, K., & Waidner, M. (2011). Online Privacy: Towards Informational Self-Determination on the Internet. Manifesto from Dagstuhl Perspectives Workshop, 11061, 1(1), 1-20.
- Fox, A. (1974). *Beyond contract: Work, power and trust relations*. London: Faber & Faber.
- Gandy Jr., O. H. (1993). *The panoptic sort: A political economy of personal information*. Boulder, CO: Westview Press.
- GfK. (2014, Mar.) Survey on data privacy and trust. Retrieved from <http://www.gfk.com/Documents/GfK-Privacy-Survey.pdf>
- Ghostery. (2014). Ghostery hosts webinar to unveil findings from a new security study: Cloud-based digital marketing technology linked to widespread risks, press release (Sept. 29). Retrieved from <https://www.ghostery.com/about-us/press-releases/ghostery-hosts-webinar-unveil-findings-new-security/>
- Gigya. (2015). The 2015 state of consumer privacy & personalization. Retrieved from <https://www.gigya.com/resource/whitepaper/the-2015-state-of-consumer-privacy-personalization/>
- Goldfarb, A. (2014). What is different about online advertising? *Review of Industrial Organization*, 44(2), 115-129.

- Graepel, T., Candela, J. Q., Borchert, T., & Herbrich, R. (2010). Web-scale Bayesian click-through rate prediction for sponsored search advertising in Microsoft's Bing search engine. In *Proceedings of the 27th International Conference on Machine Learning*, 13-20.
- Greenstein, S. (2014). Baking the data layer. *IEEE Micro*, 34(4), 56-57.
- Greenstein, S. (2015). *How the internet became commercial: Innovation, privatization, and the birth of a new network*. Princeton, NJ: Princeton University Press.
- Greenwald, G. (2014). *No place to hide: Edward Snowden, the NSA, and the US surveillance state*. New York: Metropolitan.
- Guardian, the. (2013, Nov. 1). NSA Prism program slides. Retrieved from <http://www.theguardian.com/world/interactive/2013/nov/01/prism-slides-nsa-document>
- Hannak, A., Soeller, G., Lazer, D., Mislove, A., & Wilson, C. (2014, Nov.). Measuring price discrimination and steering on e-commerce web sites. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, Association for Computing Machinery, 305-318. Retrieved from <http://www.ccs.neu.edu/home/cbw/pdf/imc151-hannak.pdf>
- Hinde, R. A., & Groebel, J. (Eds.). (1991). *Cooperation and prosocial behaviour*. Cambridge, UK: Cambridge University Press.
- Hoegg, R., Martignoni, R., Meckel, M., & Stanoevska-Slabeva, K. (2006). Overview of business models for web 2.0 communities. In *Proceedings of GeNeMe*, 23-37.
- Hoffman, D. L., Novak, T. P., & Peralta, M. (1999). Building consumer trust online. *Communications of the ACM*, 42(4), 80-85.
- Hoofnagle, C. J., & King, J. (2008). What Californians understand about privacy online. Retrieved from <http://ssrn.com/abstract=1262130>
- Horrigan, J. B., & Duggan, M. (2015). Home broadband 2015. Pew Research Center. Retrieved from <http://www.pewinternet.org/2015/12/21/home-broadband-2015/>
- Howard, D. J., & Kerin, R. A. (2004). The effects of personalized product recommendations on advertisement response rates: The try this. It works! Technique. *Journal of Consumer Psychology*, 14(3), 271-279.
- Häubl, G., & Murray, K. B. (2003). Preference construction and persistence in digital marketplaces: The role of electronic recommendation agents. *Journal of Consumer Psychology*, 13(1-2), 75-91.
- Husted, B. W. (1989). Trust in business relations: Directions for empirical research. *Business & Professional Ethics Journal*, 23-40.

- Interactive Advertising Bureau. (2016). IAB's Congressional Advocacy Report. Public Policy Office of the Interactive Advertising Bureau. Retrieved from <http://www.iab.com/topics/public-policy/?topic=advocacy/>
- International Data Corporation. (2013). Always connected—How smartphones and social keep us engaged. International Data Corporation report on behalf of Facebook. Retrieved from [https://www.idc.com/prodserv/custom\\_solutions/download/case\\_studies/PLAN-BB\\_Always\\_Connected\\_for\\_Facebook.pdf](https://www.idc.com/prodserv/custom_solutions/download/case_studies/PLAN-BB_Always_Connected_for_Facebook.pdf)
- Iyer, G., Soberman, D., & Villas-Boas, J. M. (2005). The targeting of advertising. *Marketing Science*, 24(3), 461-476.
- Jackson, S. L. (2009). *Research methods and statistics: A critical thinking approach* (3rd ed.). Belmont, CA: Wadsworth.
- Kampe, K. K., Frith, C. D., & Frith, U. (2003). Hey John: Signals conveying communicative intention toward the self activate brain regions associated with mentalizing, regardless of modality. *The Journal of Neuroscience*, 23(12), 5258-5263.
- Khodyakov, D. (2007). Trust as a process: A three-dimensional approach. *Sociology*, 41(1), 115-132.
- Kim, T., Barasz, K., John, L., & Luca, M., (2015). Consumer reactance to targeted advertisements: Violating the implied social contract. Presented to American Marketing Association Marketing and Public Policy Conference, Washington, DC (June 4-6).
- Kim, Y. H., & Kim, D. J. (2005). A study of online transaction self-efficacy, consumer trust, and uncertainty reduction in electronic commerce transaction. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, IEEE, 170c-170c.
- Kirkpatrick, D. (2016). Programmatic display ad spending to reach \$22B, 67% of total spend. Marketing Dive. Retrieved from <http://www.marketingdive.com/news/programmatic-display-ad-spending-to-reach-22b-67-of-total-spend/416912/>
- Lambrecht, A., & Tucker, C. (2013). When does retargeting work? Information specificity in online advertising. *Journal of Marketing Research*, 50(5), 561-576.
- Li, X., Valacich, J. S., & Hess, T. J. (2004). Predicting user trust in information systems: A comparison of competing trust models. In *Proceedings of the 37th annual Hawaii International Conference on System Sciences*, IEEE, 1-10.
- Lin, M., Ke, X., & Whinston, A. B. (2012). Vertical differentiation and a comparison of online advertising models. *Journal of Management Information Systems*, 29(1), 195-236.
- Lincoln, Y., & Guba, E. (2000). Paradigmatic controversies, contradictions, and emerging confluences. In N. Denzin & Y. Lincoln (eds.). *Handbook of Qualitative Research* (2nd ed.) (163-188). Thousand Oaks, CA: Sage.

- Lorenzon, K., & Russell, C. A. (2012). From apathy to ambivalence: How is persuasion knowledge reflected in consumers' comments about in-game advertising? *Journal of Marketing Communications*, 18(1), 55-67.
- Luo, X. (2002). Trust production and privacy concerns on the internet: A framework based on relationship marketing and social exchange theory. *Industrial Marketing Management*, 31(2), 111-118.
- Lyon, D. (2014). Surveillance, Snowden, and big data: Capacities, consequences, critique. *Big Data & Society*, 1(2), 2053951714541861.
- Macaulay, S. (1963). Non-contractual relations in business: A preliminary study. *American Sociological Review*, 28, 55-67.
- Madden, M. (2014). Public perceptions of privacy and security in the post-Snowden era. Pew Research Internet Project. Retrieved from <http://www.pewinternet.org/2014/11/12/public-privacy-perceptions/>
- Manchanda, P., Dubé, J. P., Goh, K. Y., & Chintagunta, P. K. (2006). The effect of banner advertising on internet purchasing. *Journal of Marketing Research*, 43(1), 98-108.
- Martin, K. (2015). Privacy notices as tabula rasa: An empirical investigation into how complying with a privacy notice is related to meeting privacy expectations online. *Journal of Public Policy & Marketing*, 34(2), 210-227.
- Martin, K. (2012). Diminished or just different? A factorial vignette study of privacy as a social contract. *Journal of Business Ethics*, 111(4), 519-539.
- Marvin, G. (2015). Google AdWords turns 15: A look back at the origins of a \$60 billion business. *Search Engine Land*. Retrieved from <http://searchengineland.com/google-adwords-turns-15-a-look-back-at-the-origins-of-a-60-billion-business-234579>
- McCarthy, J. E. (1964). *Basic marketing. A managerial approach*. Homewood, IL: Irwin.
- McDermott, M. (2006). *Working-class white: The making and unmaking of race relations*. Berkeley and Los Angeles: University of California Press.
- McDonald, A. M., & Cranor, L. F. (2008). The cost of reading privacy policies. *ISJLP*, 4, 543.
- McKenna, R. (1990). Marketing is everything. *Harvard Business Review*, 69(1), 65-79.
- Means Coleman, R. (2000). *African American viewers and the black situation comedy*. New York: Garland.
- Merrick, A. (2014, Jan. 23). A death in the database. *The New Yorker*. Retrieved from <http://www.newyorker.com/business/currency/a-death-in-the-database>

- Milne, G., Mahmood Hajjat, F., Markos, E., & Pettinico, G. (2015). Assessing role of risk for specific information types on consumers' willingness to provide. Presented to American Marketing Association Marketing and Public Policy Conference, Washington, DC (June 4-6).
- Mittal, B. (1994). Public assessment of TV advertising: Faint praise and harsh criticism. *Journal of Advertising Research*, 34(1), 35-54.
- Morgan, D. L. (1997). *Focus groups as qualitative research* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Muehling, D. D. (1987). An investigation of factors underlying attitude-toward-advertising-in-general. *Journal of Advertising*, 16(1), 32-40.
- Mutz, D. C. (2005). Social trust and e-commerce: Experimental evidence for the effects of social trust on individuals' economic behavior. *Public Opinion Quarterly*, 69(3), 393-416.
- Mutz, D. C., (2011) *Population-based survey experiments*. Princeton, NJ: Princeton University Press.
- Nelson, M. R. (2008). The hidden persuaders: Then and now. *Journal of Advertising*, 37(1), 113-126.
- Nissenbaum, H. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford, CA: Stanford University Press.
- Norman, D. A. (1983). Some observations on mental models. In D. Gentner & A. Stevens (Eds.), *Mental models* (7-14). New York: Lawrence Erlbaum Associates.
- Nye, D. E. (1992). *Electrifying America: Social meanings of a new technology, 1880-1940*. Cambridge, MA: The MIT Press.
- O'Donohoe, S. (1995). Attitudes to advertising: A review of British and American research. *International Journal of Advertising*, 14(3), 245-261.
- Okazaki, S., Li, H., & Hirose, M. (2009). Consumer privacy concerns and preference for degree of regulatory control. *Journal of Advertising*, 38(4), 63-77.
- Paakki, M. K. (2008). Consumer trust in e-commerce: A feminist ethnographic study. Doctoral Dissertation, Department of Computer Sciences, University of Tampere.
- Packard, V. (1957). *The hidden persuaders*. New York: Van Rees Press.
- Page, L. (2012). 2012 Update from the CEO. Google, Inc. Retrieved from <https://abc.xyz/investor/founders-letters/2012/>



- Pasek, J., & Krosnick J. (2010). Optimizing survey questionnaire design in political science: Insights from psychology. In J. E. Leighley (Ed.), *The Oxford handbook of American elections and political behavior* (27-50). Oxford & New York: Oxford University Press.
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Cambridge, MA: Harvard University Press.
- Perrin, A. (2015). Social networking usage: 2005-2015. Pew Research Center. Retrieved from <http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/>
- Peters, C., Amato, C. H., & Hollenbeck, C. R. (2007). An exploratory investigation of consumers' perceptions of wireless advertising. *Journal of Advertising*, 36(4), 129-145.
- Petrovic, O., Ksela, M., Fallenbock, M., Kittl, C., Urban, G. L., & Zobel, R. (2003). *Trust in the network economy*. Santa Clara, CA: Springer-Verlag TELOS.
- Petrovici, D., & Marinov, M. (2007). Determinants and antecedents of general attitudes towards advertising: A study of two EU accession countries. *European Journal of Marketing*, 41(3/4), 307-326.
- Plummer, J. T. (1974). The concept and application of life style segmentation. *Journal of Marketing*, 38(1), 33-37.
- Pollay, R. W., & Mittal, B. (1993). Here's the beef: Factors, determinants, and segments in consumer criticism of advertising. *Journal of Marketing*, 57(3), 99-114.
- PwC US & IAB. (2016). Internet advertising revenue report: 2015 full year results. Retrieved from <http://www.iab.com/insights/iab-internet-advertising-revenue-report-conducted-by-pricewaterhousecoopers-pwc-2/>
- Rainie, L., & Wellman, B. (2012). *Networked: The new social operating system*. Cambridge, MA: MIT Press.
- Rainie, L., & Duggan, M. (2015). Privacy and information sharing. Pew Research Center. Retrieved from <http://www.pewinternet.org/2016/01/14/2016/Privacy-and-Information-Sharing/>
- Rao, A., & Scaruffi, P. (2013). *A history of Silicon Valley: The largest creation of wealth in the history of the planet: 1900-2013*. U.S.: Omniware Group.
- Rawson, H., & Miner, M. (2006). *The Oxford dictionary of American quotations*. Oxford & New York: Oxford University Press.
- Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge, UK: Cambridge University Press.
- ReTargeter. (n.d.). What is retargeting and how does it work? Retrieved from <https://retargeter.com/what-is-retargeting-and-how-does-it-work>

- Rotenberg, M., Barnes, K., & Gartland, C. (2015). Comment of the Electronic Privacy Information Center, U.S. Federal Trade Commission's Cross-device tracking workshop (Dec. 16). Retrieved from <https://epic.org/apa/comments/EPIC-FTC-Cross-Device-Tracking-Comments.pdf>
- Rotfeld, H. J., & Taylor, C. R. (2009). The advertising regulation and self-regulation issues ripped from the headlines with (sometimes missed) opportunities for disciplined multidisciplinary research. *Journal of Advertising*, 38(4), 5-14.
- Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist*, 26(5), 443.
- Rotter, J. B. (1980). Interpersonal trust, trustworthiness, and gullibility. *American Psychologist*, 35(1), 1.
- Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23(3), 393-404.
- Rupietta, W. (1990). Mental models and the design of user manuals. In D. Ackermann & M. J. Tauber (Eds.), *Mental models and human-computer interaction 1* (321-334). Elsevier Science Publishers B. V.
- Ryan, G. J., & Dzewaltowski, D. A. (2002). Comparing the relationships between different types of self-efficacy and physical activity in youth. *Health Education & Behavior*, 29(4), 491-504.
- Safe Shepherd. (2016). Safe Shepherd: See how it works. Retrieved from <https://www.safeshepherd.com/how>
- Samat, S., Acquisti, A., Gross, R., Pe'er, E. (2013). Visceral targeting: Using personalized face composites for implicitly targeted marketing. Presented to the 32<sup>nd</sup> Annual Advertising and Consumer Psychology Conference, San Diego, CA.
- Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, 61-71.
- Sayre, J. (1939). A comparison of three indices of attitude toward radio advertising. *Journal of Applied Psychology*, 23(1), 23.
- Schiff, A. (2015). A marketer's guide to cross-device identity. AdExchanger. Retrieved from <http://adexchanger.com/data-exchanges/a-marketers-guide-to-cross-device-identity/>
- Schlosser, A. E., Shavitt, S., & Kanfer, A. (1999). Survey of internet users' attitudes toward internet advertising. *Journal of Interactive Marketing*, 13(3), 34-54.
- Schneier, B. (2015). *Data and Goliath: The hidden battles to collect your data and control your world*. New York: W.W. Norton.

- Schwartz, J. (2001, Sept. 4). Giving the web a memory cost its users privacy. *The New York Times*. Retrieved from <http://www.nytimes.com/2001/09/04/technology/04COOK.html>
- Schwartz, P. M. (2004). Property, privacy, and personal data. *Harvard Law Review*, 2056-2128.
- Searls, D. (2012). *The intention economy: When customers take charge*. Boston, MA: Harvard Business Press.
- Seyedghorban, Z., Tahernejad, H., & Matanda, M. J. (2015). Re inquiry into advertising avoidance on the internet: A conceptual replication and extension. *Journal of Advertising*, 1-10.
- Shatnawi, M., & Mohamed, N. (2012). Statistical techniques for online personalized advertising: A survey. In *Proceedings of the 27th Annual Symposium on Applied Computing*, Association for Computing Machinery, 680-687.
- Shavitt, S., Lowrey, P., & Haefner, J. (1998). Public attitudes toward advertising: More favorable than you might think. *Journal of Advertising Research*, 38(4), 7-22.
- Shaw, E. H., & Jones, D. B. (2005). A history of schools of marketing thought. *Marketing Theory*, 5(3), 239-281.
- Sherbin, M. (2012, Oct. 5). The static website is dead - Long live personalized content. Content Marketing Institute. Retrieved from <http://contentmarketinginstitute.com/2012/10/the-static-website-is-dead-long-live-personalized-content/>
- Shugan, S. M. (2003). Defining interesting research problems. *Marketing Science*, 22(1), 1-15.
- Simonite, T. (2013). A popular ad blocker also helps the ad industry. *MIT Technology Review*. Retrieved from <https://www.technologyreview.com/s/516156/a-popular-ad-blocker-also-helps-the-ad-industry/>
- Simpson, J. A. (2007). Psychological foundations of trust. *Current Directions in Psychological Science*, 16(5), 264-268.
- SINTEF. (2013). Big data, for better or worse: 90% of world's data generated over last two years. *Science Daily*. Retrieved from <http://www.sciencedaily.com/releases/2013/05/130522085217.htm>
- Sirdeshmukh, D., Singh, J., & Sabol, B. (2002). Consumer trust, value, and loyalty in relational exchanges. *Journal of Marketing*, 66(1), 15-37.
- Smith, A. (2015). U.S. smartphone use in 2015. Pew Research Center. Retrieved from <http://www.pewinternet.org/2015/04/01/chapter-one-a-portrait-of-smartphone-ownership/>
- Solove, D. (2008). *Understanding privacy*. Cambridge, MA: Harvard University Press.

- Stefanone, M. A., Hurley, C. M., Egnoto, M. J., & Covert, J. M. (2015). Information asymmetry and social exchange: Exploring compliance gaining online. *Information, Communication & Society*, 18(4), 376-389.
- Stevenson, D. (2015). Does e-commerce personalization make all consumers vulnerable consumers? The Center for Internet and Society, Stanford Law School. Retrieved from <https://cyberlaw.stanford.edu/blog/2015/03/does-e-commerce-personalization-make-all-consumers-vulnerable-consumers>
- Stevenson, D. (2014). Locating discrimination in data-based systems. In S. Peña Gangadharan, V. Eubanks, & S. Barocas (Eds.), *Data and Discrimination: Collected Essays*. Washington, DC: Open Technology Institute, New America Foundation. Retrieved from <http://newamerica.org/downloads/OTI-Data-an-Discrimination-FINAL-small.pdf>
- Stevenson, D., & Pasek, J. (2015). Privacy concern, trust, and desire for content personalization. In *Proceedings of TPRC 43*, Arlington, VA.
- Stole, I. L. (2014). Persistent pursuit of personal information: A historical perspective on digital advertising strategies. *Critical Studies in Media Communication*, 31(2), 129-133.
- Suchman, L. (2007). *Human-machine reconfigurations: Plans and situated actions*. (2nd ed.) Cambridge, UK: Cambridge University Press.
- Suchman, L. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge, UK: Cambridge University Press.
- Sutanto, J., Palme, E., Tan, C. H., & Phang, C. W. (2013). Addressing the personalization-privacy paradox: An empirical assessment from a field experiment on smartphone users. *MIS Quarterly*, 37(4), 1141-1164.
- Tan, Y.H., & Thoen, W. (2000). Toward a generic model of trust for electronic commerce. *International Journal of Electronic Commerce*, 5(2), 61-74.
- The White House. (2012). Consumer data privacy in a networked world: A framework for protecting privacy and promoting innovation in the global digital economy. Executive Office of the President, Washington, DC. Retrieved from <https://www.whitehouse.gov/sites/default/files/privacy-final.pdf>
- The White House. (2016). Big data: A report on algorithmic systems, opportunity, and civil rights. Executive Office of the President, Washington, DC. Retrieved from [https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016\\_0504\\_data\\_discrimination.pdf](https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf)
- TRUSTe. (2015). UK consumer confidence privacy index. Retrieved from <https://www.truste.com/resources/privacy-research/uk-consumer-confidence-index-2015/>

- TRUSTe. (2016). TRUSTe/national cyber security alliance U.S. consumer privacy index. Retrieved from <https://www.truste.com/resources/privacy-research/ncsa-consumer-privacy-index-us/>
- Tsang, M. M., Ho, S. C., & Liang, T. P. (2004). Consumer attitudes toward mobile advertising: An empirical study. *International Journal of Electronic Commerce*, 8(3), 65-78.
- Tucker, C. E. (2014). The reach and persuasiveness of viral video ads. *Marketing Science*, 34(2), 281-296.
- Turow, J. (2011). *The daily you: How the new advertising industry is defining your identity and your worth*. New Haven: Yale University Press.
- Turow, J., Bleakley, A., Bracken, J., Delli Carpini, M. X., Draper, N. A., Feldman, L., Good, N., Grossklags, J., Hennessy, M., Hoofnagle, C., Howard-Williams, R., King, J., Li, S., Meltzer, K., Mulligan, D. K., & Nir, L. (2014). Americans, marketers, and the internet: 1999-2012. Annenberg Public Policy Center Report Series. Retrieved from [http://repository.upenn.edu/asc\\_papers/348](http://repository.upenn.edu/asc_papers/348)
- Turow, J., Hennessy, M., & Draper, N., (2015). The tradeoff fallacy: How marketers are misrepresenting American consumers and opening them up to exploitation. Report from the Annenberg School for Communication, University of Pennsylvania. Retrieved from [https://www.asc.upenn.edu/sites/default/files/TradeoffFallacy\\_1.pdf](https://www.asc.upenn.edu/sites/default/files/TradeoffFallacy_1.pdf)
- Turow, J., King, J., Hoofnagle, C.J., Bleakley, A., & Hennessy, M. (2009). Americans reject tailored advertising and three activities that enable it. Departmental Papers of the Annenberg School for Communication, University of Pennsylvania. Retrieved from <http://ssrn.com/abstract=1478214>
- U.S. Senate. (2013). A review of the data broker industry: Collection, use, and sale of consumer data for marketing purposes. Senate Committee on Commerce, Science, and Transportation. Staff report for Chairman Rockefeller, 1-36. Retrieved from [https://www.commerce.senate.gov/public/\\_cache/files/0d2b3642-6221-4888-a631-08f2f255b577/AE5D72CBE7F44F5BFC846BECE22C875B.12.18.13-senate-commerce-committee-report-on-data-broker-industry.pdf](https://www.commerce.senate.gov/public/_cache/files/0d2b3642-6221-4888-a631-08f2f255b577/AE5D72CBE7F44F5BFC846BECE22C875B.12.18.13-senate-commerce-committee-report-on-data-broker-industry.pdf)
- Ur, B., Leon, P. G., Cranor, L. F., Shay, R., & Wang, Y. (2012). Smart, useful, scary, creepy: Perceptions of online behavioral advertising. In *Proceedings of the 8th Symposium on Usable Privacy and Security*, Association for Computing Machinery, Washington, DC (July 11-13), 1-15.
- Vaidhyathan, S. (2012). *The Googlization of everything (and why we should worry)*. Berkeley & Los Angeles: University of California Press.
- Valenzuela, S., Park, N., & Kee, K. F. (2009). Is there social capital in a social network site?: Facebook use and college students' life satisfaction, trust, and participation. *Journal of Computer-Mediated Communication*, 14(4), 875-901.

- Van Doorn, J., & Hoekstra, J. C. (2013). Customization of online advertising: The role of intrusiveness. *Marketing Letters*, 24(4), 339-351.
- Varian, H. R. (1989). Price discrimination. In R. Schmalensee & R. D. Willig (Eds.), *Handbook of industrial organization*, vol. 1 (597-654). Elsevier Science Publishers B.V.
- van der Veer, G. C., Guest, S., Haselager, W. F., Innocent, P., McDaid, E., Oestreicher, L., Tauber, M., Vos, U., & Waern, Y. (1990). Designing for the mental model: An interdisciplinary approach to the definition of a user interface for electronic mail systems. In D. Ackermann & M. J. Tauber (Eds.), *Mental models and human-computer interaction I* (253-288). North-Holland Publishing Co.
- Walker, K. L. (2016). Surrendering information through the looking Glass: Transparency, trust, and Protection. *Journal of Public Policy & Marketing*, 35(1), 144-158.
- Warner, R. M. (2013). *Applied statistics: From bivariate through multivariate techniques*. Thousand Oaks, CA: Sage Publications.
- van Waterschoot, W., & Foscht, T. (2010). The marketing mix—A helicopter view. *Marketing Theory*, 2, 185-208.
- Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1999). *Unobtrusive measures* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Webster Jr., F. E. (1992). The changing role of marketing in the corporation. *Journal of Marketing*, 56(4), 1-17.
- Wedel, M., & Kamakura, W. (1998). *Market segmentation: Conceptual methodological foundations*. Boston: Kluwer Academic Publishers.
- Weisberg, J., Te'eni, D., & Arman, L. (2011). Past purchase and intention to purchase in e-commerce: The mediation of social presence and trust. *Internet Research*, 21(1), 82-96.
- White, T. B., Zahay, D. L., Thorbjørnsen, H., & Shavitt, S. (2008). Getting too personal: Reactance to highly personalized email solicitations. *Marketing Letters*, 19(1), 39-50.
- Williams, R. (1980). *Problems in materialism and culture*. London: Verso.
- Winner, L. (1977). *Autonomous technology: Technics-out-of-control as a theme in political thought*. Cambridge, MA: MIT Press.
- Wong, S. (1978). *Foundations of Paul Samuelson's revealed preference theory*. London & New York: Routledge.
- World Economic Forum. (2013). The internet trust bubble: Global values, beliefs and practices. Retrieved from [http://www3.weforum.org/docs/WEF\\_InternetTrustBubble\\_Report2\\_2014.pdf](http://www3.weforum.org/docs/WEF_InternetTrustBubble_Report2_2014.pdf)

- World Economic Forum. (2014). Rethinking personal data: A new lens for strengthening trust. Retrieved from [http://www3.weforum.org/docs/WEF\\_RethinkingPersonalData\\_ANewLens\\_Report\\_2014.pdf](http://www3.weforum.org/docs/WEF_RethinkingPersonalData_ANewLens_Report_2014.pdf)
- World Economic Forum. (2014). Rethinking personal data: Trust and context in user-centred data ecosystems. Retrieved from [http://www3.weforum.org/docs/WEF\\_RethinkingPersonalData\\_TrustandContext\\_Report\\_2014.pdf](http://www3.weforum.org/docs/WEF_RethinkingPersonalData_TrustandContext_Report_2014.pdf)
- Wu, T. (2015). Facebook should pay all of us. *The New Yorker* (Aug. 14). Retrieved from <http://www.newyorker.com/business/currency/facebook-should-pay-all-of-us>
- Ybarra, O., Keller, M. C., Chan, E., Garcia, S. M., Sanchez-Burks, J., Morrison, K. R., & Baron, A. S. (2010). Being unpredictable: Friend or foe matters. *Social Psychological and Personality Science*, 1(3), 259-267.
- Yu, H. (2011). A cross-cultural comparison of perceptions toward personalized e-mail advertising: Korea and the United States. *The Korean Journal of Advertising*, 22(1), 233-255.
- Yuan, S., Abidin, A. Z., Sloan, M., & Wang, J. (2012). Internet advertising: An interplay among advertisers, online publishers, ad exchanges and web users. Retrieved from <https://arxiv.org/abs/1206.1754>
- Zaheer, A., McEvily, B., & Perrone, V. (1998). Does trust matter? Exploring the effects of interorganizational and interpersonal trust on performance. *Organization Science*, 9(2), 141-159.
- Zhang, W., Yuan, S., & Wang, J. (2014). Optimal real-time bidding for display advertising. In *Proceedings of the 20th International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, 1077-1086.
- Zuckerman, E. (2013). *Rewire: Digital cosmopolitans in the age of connection*. New York: W. W. Norton.