

Economics of Privacy and User-Generated Content

CATHERINE TUCKER

Abstract

The Internet has allowed an unprecedented expansion of the data firms can collect and the amount of content users can upload. Understanding the forces that shape the demand and supply of this content is critical for understanding the likely evolution of the Internet. We review foundational and cutting-edge research on the economics of privacy and user-generated content (UGC), and sketch out issues for future research.

INTRODUCTION

Let us step back in time to the early 1990s, when the Internet had not yet begun to revolutionize the provision of content (news, books, TV, movies, and music), and most content was not yet held in digital form. For many people now, that world is a dim memory. Newspapers financed the provision of news content through advertising, which advertisers bought in the hopes of reaching the newspaper's readership, but there was no way to track whether the advertising had worked or not, or to target particular demographics. Publishers financed the production of physical books via publishing houses that were highly selective about which books to take on, and which distributed books through a tight network of bookstores and in a pre-specified order (hardback and then paperback). Music production was similarly controlled by music labels that produced physical vinyl records, and later CDs, and distributed them through a tight network of physical music stores, and individual writers and musicians got a small cut. Movies and TV, again financed through advertising, were released and distributed through a small number of channels and in a predetermined order, in order to maximize revenue. Although remnants of this system still exist 20 years on, the economics of privacy and "User Generated Content" indeed, of knowledge itself have changed radically and permanently. Understanding this change is the key to understanding future developments in privacy and content production.

Emerging Trends in the Social and Behavioral Sciences. Edited by Robert Scott and Stephen Kosslyn.
© 2015 John Wiley & Sons, Inc. ISBN 978-1-118-90077-2.

This change, for many working in the content industries, has been terrifying. Newspaper and TV advertising have become much less attractive to advertisers than newer, cheaper, more targeted and more trackable forms of online advertising. The result has been massive layoffs at newspapers, and a simultaneous explosion in news content created by amateurs for free on the Internet. As newspapers and TV news, both supported by untargeted advertising, have declined in quality and in their ability to finance expensive investigative bureaus, organizations such as Wikileaks have taken advantage of data digitization to provide scoops larger and more significant than any by newspapers in decades, and have done it based on a nonprofit donation model.

As older channels have weakened, user-generated content has become exponentially more important. Users now generate news, music and literary materials and find fewer barriers to reaching a market for them (Keen, 2008). The distinction between self-presentation (the materials you might “publish” to your Facebook page) and publication of creative productions is progressively breaking down. Hundreds of millions of people will happily upload information about their preferences, tastes and ideas as a mechanism of self-expression, in a way that makes that information accessible to advertisers. The level of privacy necessary to facilitate the commercial exploitation of the Internet therefore turns out to be somewhat lower than theorists had supposed. Generational shifts are also producing an Internet-using population that, relative to the Internet-using population of 1995, is much less sensitive to privacy concerns, especially where a sense (or illusion) of control over one’s personally identifiable information is maintained. Increasingly, it is not the content itself—what might have originally been termed an “artistic work” that creates profit, but the ways in which that content can be monetized, either by the platform that advertises next to it, or via live performances and syndication.

FOUNDATIONAL WORK

There are two streams of literature that are crucial for understanding the intersection between privacy and user-generated content from an economics perspective.

The first is the classic literature on the economics of privacy. Posner (1981); Stigler (1980) set out the classic view of privacy-seeking as distortion of information. Here, privacy stands for secrecy. In such models, privacy generally operates as a cost that prevents the flow of information necessary to facilitate functioning markets. (Hirshleifer, 1971) suggests that keeping information private does not “[lead] to any improvement in productive arrangements” and provides “an incentive for individuals to expend resources in a socially

wasteful way.” This view of privacy has been questioned in recent theoretical research that integrates notions of behavioral economics into theoretical models (Acquisti, John, & Loewenstein, 2008) to help explain consumers’ preferences for privacy, and also to set out why there may be a role for regulation in some cases to protect consumer privacy. Other work that has documented an economic need for privacy has used straightforward questions of price discrimination to understand consumer motivations for privacy which amount to more than simple secrecy. Hui and Png (2006) provide a nice review of this literature.

Other work has attempted to understand the tradeoffs involved in providing strong privacy protections, in that doing so can reduce otherwise desirable market outcomes such as the adoption of electronic medical records (Miller & Tucker, 2009) and lifesaving medical technologies (Miller & Tucker, 2011), the ability of small firms to enter digital markets (Campbell, Goldfarb, & Tucker, 2011; Goldfarb & Tucker, 2012a), and even the ability of municipalities and states to limit certain kinds of advertising (Goldfarb & Tucker, 2011).

The second pertinent stream of literature for understanding the economic role of user-generated content is the economics literature on platforms. This literature describes the role of network operators in facilitating the interaction between two (or more) groups of users, when each group of users values the presence of the other group (Armstrong, 2006; Rochet & Tirole, 2006). This is important for understanding user-generated content, because it seems natural to suppose that users who generate content will be more likely to value a platform to disseminate their creative work if there are other people to consume the content (Ahn, Duan, & Mela, 2011). One paper that ably makes the link between this platform literature and user-generated content is Hoffstetter, Shriver, and Miller (2009), who examine how an exogenous shift in wind speeds promotes blog postings about windsurfing, which in turn promotes participation in the network. Similarly, Albuquerque *et al.* (2010) describes this audience-feedback progress for an online magazine platform.

MORE RECENT WORK

User-generated content online can take many forms. The largest stream of research in marketing and economics has evaluated the effects of user-generated content in the forms of reviews and quality information, starting with early work such as Mayzlin (2004) and Chevalier and Mayzlin (2006). This is a very large literature, and it is hard to do it justice within the confines of a short article. Recent papers include Zhao, Yang, Narayan, and Zhao (2012), who model how consumers learn about products through

this user-generated content. There has also been work that suggests how profitable these new sources of data can be for firms. Archak, Ghose, and Ipeirotis (2011); Netzer, Feldman, Goldenberg, and Fresko (2012); Decker and Trusov (2010); Lee and Bradlow (2011) discuss how the mining of product reviews helps market research and the development of product features. Ghose, Ipeirotis, and Li (2012) discusses how such data can be used for rankings, and Tirunillai and Tellis (2012) show the linkage between this form of user-generated content and stock market returns. This type of research deals with consumer opinions, which are typically less personally sensitive than other forms of user-generated content, so there has not been much consideration given to how privacy concerns might temper the production of such user-generated content. One potentially acceptable mechanism would be the “behavior-based price discrimination” described by Fudenburg and Villas-Boas (2006), where consumers’ own reviewing behavior is used to generate personalized prices.

There has been less academic work relating to the active production of creative content such as videos and blog postings. Gal-Or, Geylani, and Yildirim (2010) studies the production of user-generated news content. Ghose and Han (2011) study the production of user-generated content on mobile networks. There is little economics work on visible user-generated content platforms such as Youtube, but the topic has received attention from computer science academics. (Cha, Kwak, Rodriguez, Ahn, 2007) provide an excellent overview of Youtube user-generated content, examining the lifecycle of uploaded videos, how much illegal content is in the system, and how the system might be made more efficient. There have also been a few papers that look directly at the intersection of user-generated content and privacy, focusing mainly on social media and business networking sites. Early examples of this are Gross and Acquisti (2005); Acquisti and Gross (2006), who investigate privacy concerns on Facebook. There is also an information systems literature that uses survey research to try and understand consumer behaviors (Xu, 2007; Xu, Teo, Tan, & Agarwal, 2012) regarding the release of information in social media settings. However, for much of the economically focused empirical work, the focus has not been predominantly on privacy concerns surrounding the content, but instead on privacy concerns surrounding the advertising that supports the content Tucker (2011, 2012). This research suggests that striking the right balance on privacy protection is essential for establishing successful advertising-supported social network platforms.

This highlights a general tension between the notions of privacy and user-generated content. In particular, the very nature of user-generated content is that it is there to be broadcast and consumed. On the other hand, the very nature of privacy concerns is that users are, for various reasons,

reluctant to have information pertaining to them consumed by others. One might initially observe that the users who are least comfortable with the content they generate being broadcast, consumed or used to target advertising, will also be the least likely users to contribute publicly to creating content. However, it is also possible that users generate content as a means of self-expression, without being fully aware of the ways that their contribution can be reproduced, disseminated or even used to harm their offline reputation. This tension is often highlighted in discussions about how age affects the consumption and broadcast of user-generated content. A survey-based literature has evolved in legal studies that asks directly about perceptions of privacy (Sheehan and Hoy (2000), Marwick, Diaz, and Palfrey (2010), etc.). For example, Hoofnagle, King, Li, and Turow (2010) show that there is a perception of increased attention to privacy over time. In general, this work, alongside more empirically based work such as Goldfarb and Tucker (2012b), suggests that people say they are more focused on privacy than they were 5 years ago. This could have benefited from the wider publicity given to shifts in commercial privacy policies for major websites in the press, and with increased discussion of mass government surveillance. Consistently with recent discussions in law and philosophy, this shift may reflect a change in consumers' perceptions over what contexts warrant privacy.

(Zittrain, 2009) (p. 212), among others, links how the kind of technology that enables user-generated content changes perceptions of what is public and what is private, arguing that changes in technology "threaten to push everyone toward treating each public encounter as if it were a press conference." It is still very early to determine how such new psychological shifts may affect consumers' economic decision-making, and we are not aware of systematic or empirical research that would give insight into it. Reflecting the legal perspectives in this stream of research, Taylor (2004) develops an economic model of information-sharing across firms that provides one explanation of why the context in which data is used may matter: if data is used to price-discriminate, and changes in technology mean that firms can sell the data to other firms even outside their sector, then many rational consumers will choose to keep information private in order to prevent future price discrimination. Prospectively, such models can potentially be used in the future to understand how the motivation to broadcast user-generated content intersects with privacy concerns.

ISSUES GOING FORWARD

This discussion illustrates that this is a very new area. I conclude with three major avenues that would be particularly useful areas for future research.

Motivation for production of user-generated content: One open question is exactly what are the intrinsic and extrinsic motivations are for the production of user-generated content. Typically research has modeled this in terms of size of audience. This makes sense give the earlier literature on network effects, but there appear to be instances where adding more content at the margin brings no clear benefit to the user. For example, there are 4815 reviews of the movie “Mean Girls” starring Lindsay Lohan on Amazon.com, of which 4650 were glowing five-star reviews. What motivated that Amazon user to post the 4650th positive review for the movie? Without understanding the motivations for users of posting user-generated content even in the absence of a likely audience, it is difficult to understand the likely distortions in the quality of the content produced. Works of Daugherty *et al.* (2008); Moe and Schweidel (2012); Godes and Silva (2011) have examined how the need to correct previous user-generated content can affect behavior, but there are too many instances where people post user-generated content that reinforces earlier user-generated content for this to be a full explanation.

Optimal privacy protections for user-generated content. There is a tension between the desire by many producers of user-generated content to expose their thoughts and creative works to the largest possible audience, and a desire to protect these original works from the eyes of undesirable readers such as bosses or an insurance agency. One answer to this dilemma is to encode into social media sites very granular privacy controls. However, understanding how best to resolve the tension of allowing both privacy controls but in a way that is not cumbersome and easy to understand represents a huge technical challenge, especially in the context of a user base that is generally not well informed about the effects of different privacy architectures on their user experience. Another, more cumbersome answer is legislation that would limit and shape the uses to which user-generated content can be put. This represents an exceptional challenge for policymakers and economists, who need to think about what incentives would lead to a welfare-improving outcome, while dealing with a rapidly changing and novel technological environment. Early efforts appear to be underway in some US states and in the European Union to forbid employers or potential employers from requiring as a condition of employment access to employees’ or candidates’ online social media profiles.

Optimal privacy protections *over time* for user-generated content. One crucial issue when it comes to privacy and user-generated content is the question of the appropriateness of privacy protection over time. The classic concern over privacy online is that I may have time-inconsistent preferences. For example, I may, as a college student think that user-generated content involving naked guitar playing is amusing and rewarding and that I do not need to impose strict privacy controls when posting it. However,

after I graduate and am pursuing a career in accountancy, such videos may come back to haunt me, and I may retrospectively wish that I could have employed stricter privacy protections to prevent the release of such materials. This illustrates a tension between the digital and persistent nature of user-generated content that makes it hard to retrospectively apply new privacy protections to it. More broadly, the predigital era enabled individuals to shift their self-presentation over time without it being possible to hold their previous incarnations or opinions against them; the digital era may increase the pressure on individuals to present a more consistent image over the course of early adulthood through to old age. Further research is needed to understand whether users do indeed have time-inconsistent preferences when it comes to privacy, or whether users initially are uninformed about the consequences of their privacy choices. Given these inconsistencies, further research is needed to understand whether there is a role for privacy protection and laws to help address such scenarios.

REFERENCES

- Acquisti, A., & Gross, R. (2006). Imagined communities: Awareness, information sharing, and privacy on the Facebook. In *6th Workshop on Privacy Enhancing Technologies*, pp. 36–58.
- Acquisti, A., John, L. K., & Loewenstein, G. (2008). What is privacy worth? Mimeo, CMU.
- Ahn, D.-Y., Duan, J. A., & Mela, C. F. (2011). An equilibrium model of user generated content. Working Papers 11–13, NET Institute.
- Albuquerque, P., Pavlidis, P., Chatow, U., Chen, K.-Y., Jamal, Z., & Koh, K.-W. (2010). Evaluating promotional activities in an online two-sided market of user-generated content. Simon School Working Paper No. FR 10-08.
- Archak, N., Ghose, A., & Ipeiritos, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, 57(8), 1485–1509.
- Armstrong, M. (2006, Autumn). Competition in two-sided markets. *RAND Journal of Economics*, 37(3), 668–691.
- Campbell, J. D., Goldfarb, A., & Tucker, C. (2011). Privacy regulation and market structure. Mimeo, University of Toronto.
- Cha, M., Kwak, H., Rodriguez, P., Ahn, Y., & Moon, S. (2007). I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, pp. 1–14. ACM.
- Chevalier, J., & Mayzlin, D. (2006). The effect of word of mouth online: Online book reviews. *Journal of Marketing Research*, 43(345–354).
- Daugherty, T., Eastin, M., & Bright, L. (2008). Exploring consumer motivations for creating user-generated content. *Journal of Interactive Advertising*, 8(2), 1–24.
- Decker, R., & Trusov, M. (2010). Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing*, 27(4), 293–307.

- Fudenburg, D., & Villas-Boas, J. M. (2006). Behavior based price discrimination and customer recognition. *Handbooks in Information Systems* (Vol. 1, pp. 377–435, Chapter 7). Amsterdam, the Netherlands: Elsevier.
- Gal-Or, E., Geylani, T., & Yildirim, T. (2010). User-generated content in news media. Technical report, working paper.
- Ghose, A., & Han, S. P. (2011). An empirical analysis of user content generation and usage behavior on the mobile internet. *Management Science*, 57(9), 1671–1691.
- Ghose, A., Ipeirotis, P. G., & Li, B. (May/June 2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science*, 31(3), 493–520.
- Godes, D., & Silva, J. C. (2011). Sequential and temporal dynamics of online opinion. *Marketing Science*, 31, 448–473.
- Goldfarb, A., & Tucker, C. (2012a). Privacy and innovation. In *Innovation Policy and the Economy* (Vol. 12, NBER Chapters.). Cambridge, MA: National Bureau of Economic Research, Inc.
- Goldfarb, A., & Tucker, C. (2012b). Privacy and innovation. *Innovation Policy and the Economy*, 12(1), 65–90.
- Goldfarb, A., & Tucker, C. E. (2011). Privacy regulation and online advertising. *Management Science*, 57(1), 57–71.
- Gross, R., & Acquisti, A. (2005). Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society, WPES '05*, New York, NY, USA, pp. 71–80. ACM.
- Hirshleifer, J. (1971). The private and social value of information and the reward to inventive activity. *American Economic Review*, 61(4), 561–74.
- Hoffstetter, R., Nair, H. S., Shriver, S., & Miller, K. (2009). Social ties and user generated content: Evidence from an online social network. Working Papers 09-28, NET Institute.
- Hoofnagle, C. J., King, J., Li, S., & Turow, J. (2010). How different are young adults from older adults when it comes to information privacy attitudes and policies? Mimeo, Berkeley.
- Hui, K., & Png, I. (2006). Economics and information systems. In *Handbooks in Information Systems* (Vol. 1, Chapter 9: The Economics of Privacy). Amsterdam, the Netherlands: Elsevier.
- Keen, A. (2008). *The cult of the Amateur: How blogs, MySpace, YouTube, and the rest of today's user-generated media are destroying our economy, our culture, and our values*. Crown Publishing Group.
- Lee, T. Y., & Bradlow, E. T. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research*, 48(5), 881–894.
- Marwick, A. E., Diaz, D. M., & Palfrey, J. (2010). Youth, privacy, and reputation. Mimeo, Harvard Law School.
- Mayzlin, D. (2004). Promotional chat on the internet. *Marketing Science*, 25, 155–163.
- Miller, A., & Tucker, C. (2011). Can healthcare information technology save babies? *Journal of Political Economy*, 2, 289–324.
- Miller, A. R., & Tucker, C. (2009, July). Privacy protection and technology adoption: The case of electronic medical records. *Management Science*, 55(7), 1077–1093.

- Moe, W. W., & Schweidel, D. A. (2012). Online product opinions: Incidence, evaluation, and evolution. *Marketing Science*, 31(3), 372–386.
- Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (May/June 2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31(3), 521–543.
- Posner, R. A. (1981). The economics of privacy. *American Economic Review*, 71(2).
- Rochet, J.-C., & Tirole, J. (2006, Autumn). Two-sided markets: A progress report. *RAND Journal of Economics*, 37(3), 645–667.
- Sheehan, K. B., & Hoy, M. G. (2000). Dimensions of privacy concern among online consumers. *Journal of Public Policy and Marketing*, 19, 62–73.
- Stigler, G. J. (1980). An introduction to privacy in economics and politics. *The Journal of Legal Studies*, 9(4), 623–644.
- Taylor, C. R. (2004, Winter). Consumer privacy and the market for customer information. *RAND Journal of Economics*, 35(4), 631–650.
- Tirunillai, S., & Tellis, G. J. (2012). Does chatter really matter? dynamics of user-generated content and stock performance. *Marketing Science*, 31, 198–215.
- Tucker, C. (2011). Social networks, personalized advertising, and privacy controls. Mimeo, MIT.
- Tucker, C. (2012). Social advertising. *Journal of Marketing Research*. Mimeo, MIT.
- Xu, H. (2007). The effects of self-construal and perceived control on privacy concerns. In *28th Annual International Conference on Information Systems (ICIS 2007)*, Montreal, Canada.
- Xu, H., Teo, H.-H., Tan, B. C. Y., & Agarwal, R. (2012). Effects of individual self-protection, industry self-regulation, and government regulation on privacy concerns: A study of location-based services. *Information Systems Research*, 23, 1342.
- Zhao, Y., Yang, S., Narayan, V., & Zhao, Y. (2012). Modeling consumer learning from online product reviews. *Marketing Science*, 32, 153–169.
- Zittrain, J. (2009). *The future of the internet—and how to stop it*. New Haven, CT: Yale University Press.

CATHERINE TUCKER SHORT BIOGRAPHY

Catherine Tucker is the Mark Hyman Jr. Career Development Professor and Associate Professor of Marketing at MIT Sloan. Her research interests lie in how technology allows firms to use digital data to improve their operations and marketing, and in the challenges, this poses for regulations designed to promote innovation. She has particular expertise in online advertising, digital health, social media, and electronic privacy. Generally, most of her research lies in the interface among Marketing, Economics, and Law. She has received an NSF CAREER award for her work on digital privacy and a Garfield Award for her work on electronic medical records.

Dr. Tucker is Associate Editor at Management Science and a Research Associate at the National Bureau of Economic Research. She teaches MIT Sloan's course on Pricing and the EMBA course Marketing Management

for the Senior Executive. She has received the Jamieson Prize for Excellence in Teaching as well as being voted “Teacher of the Year” at MIT Sloan. She holds a PhD in economics from Stanford University and a BA from Oxford University.

RELATED ESSAYS

- Intellectual Property (*Economics*), Michele Boldrin and David K. Levine
 Misinformation and How to Correct It (*Psychology*), John Cook *et al.*
 Political Advertising (*Political Science*), Erika Franklin Fowler
 Innovation (*Economics*), Adam B. Jaffe
 Technology Diffusion (*Economics*), Adam B. Jaffe
 Search and Learning in Markets (*Economics*), Philipp Kircher
 The Psychological Impacts of Cyberlife Engagement (*Psychology*), Virginia S. Y. Kwan and Jessica E. Bodford
 Network Research Experiments (*Methods*), Allen L. Linton and Betsy Sinclair
 Data Mining (*Methods*), Gregg R. Murray and Anthony Scime
 Retrieval-Based Learning: Research at the Interface between Cognitive Science and Education (*Psychology*), Ludmila D. Nunes and Jeffrey D. Karpicke
 The Impact of Learning Technologies on Higher Education (*Psychology*), Christopher S. Pentoney *et al.*
 Text Analysis (*Methods*), Carl W. Roberts
 Digital Methods for Web Research (*Methods*), Richard Rogers
 Education in an Open Informational World (*Educ*), Marlene Scardamalia and Carl Bereiter
 Emerging Trends: Shaping Age By Technology and Social Bonds (*Sociology*), Annette Spellerberg and Lynn Schelisch
 Content Analysis (*Methods*), Steven E. Stemler
 Information Politics in Dictatorships (*Political Science*), Jeremy L. Wallace