

Reality and Perception of Copyright Terms of Service for Online Content Creation

Casey Fiesler

University of Colorado Boulder
Boulder, CO
casey.fiesler@colorado.edu

Cliff Lampe

University of Michigan
Ann Arbor, MI
cacl@umich.edu

Amy S. Bruckman

Georgia Institute of Technology
Atlanta, GA
asb@cc.gatech.edu

ABSTRACT

From amateur creativity to social media status updates, nearly every Internet user is also a content creator—but who owns that content? Policy, including intellectual property rights, is a necessary but often invisible part of online content sharing and social computing environments. We analyzed the copyright licenses contained in the Terms of Service of 30 different websites where users contribute content, then conducted a survey to match perceptions of copyright terms to the reality. We found that licensing terms vary in unpredictable ways, and that user expectations and opinions differ by license and by type of website. Moreover, the most undesirable terms, such as right to modify, appear more frequently than users expect. We argue that users care about how their content can be used yet lack critical information. Site designers should take user needs and community norms into account in creating and explaining copyright policies.

Author Keywords

copyright; intellectual property; law; mechanical turk; online communities; policy; social media; survey; terms of service; user-generated content

ACM Classification Keywords

K.4.1 [Computers and Society]: Public Policy Issues --- *Intellectual property rights*

INTRODUCTION

How much control do Internet users have over the content they post online? Researchers often consider this question in terms of privacy (e.g., [13,33]). However, also hidden within websites' terms and conditions are copyright licenses that provide important information about ownership of user-created content. For example, YouTube requires that users provide them a “*worldwide, non-exclusive, royalty-free, sublicenseable and transferable license to use, reproduce, distribute, prepare derivative*

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

CSCW '16, February 27-March 02, 2016, San Francisco, CA, USA
Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-3592-8/16/02...\$15.00

DOI: <http://dx.doi.org/10.1145/2818048.2819931>

works of, display, and perform the Content in connection with the Service and YouTube's (and its successors' and affiliates') business, including without limitation for promoting and redistributing part or all of the Service (and derivative works thereof) in any media formats and through any media channels.”¹

Considering that reading only the privacy policy of every site visited would take the average Internet user over 200 hours per year, it is not surprising that many do not take the time to read often complicated terms and conditions [16,26]. Though the readability of online privacy terms has been identified as a usability and accessibility problem [16,21], intellectual property rights like those conveyed in the YouTube licensing clause are also increasingly relevant. Whereas privacy policies cover how a site can use user-provided information and data, copyright terms cover the ownership of the original content that users provide. Copyright vests at the time of creation, which means that Internet users own their blog posts, photographs, and perhaps even tweets.² Therefore, they must give websites permission to display or use their work. These permissions are typically covered in Terms of Service (TOS), which similar to privacy policies, are likely not frequently read. As a result, many may not realize what rights in their content they grant or how their work can be used. The purpose of this paper is to determine the significance of this problem by examining the reality and user perceptions of copyright terms for content creation websites.

In recent years a number of high-profile cases have demonstrated that though people are often unaware of copyright terms for the sites they use, they do care about them when it is brought to their attention. In 2009, an “online swell of suspicion” followed changes in Facebook’s copyright license, leading a number of high-profile users to delete their accounts [34]. In 2015, an email in which a Facebook representative incorrectly stated Facebook’s copyright terms went viral, soliciting concern and outrage from photographers.³ Similarly, in 2012, Instagram users

¹ <https://www.youtube.com/static?template=terms> (as of April 16, 2015)

² There is some question about whether tweets are copyrightable, and legal scholars generally agree that some are and some aren’t, depending on factors like length, originality, and content [27]. This same standard would apply to other types of content like social media status updates.

³ <http://stopstealingphotos.com/facebook-representative-states-uploaded-content/> (posted May 12, 2015)

pushed back over a provision in the photo-sharing site's TOS that allowed use of photographs "in connection with paid or sponsored content" [22]. In the summer of 2014, media attention around published research findings caused many people to realize that the TOS for websites like Facebook and OKCupid might allow use of their data for research [30]. All of these situations resulted in comments on news articles made by users vowing to no longer post their content on these websites.

Prior research also supports that people care about how their content is used. Copyright is a frequent topic of conversation in creative communities [7], content ownership is one of the major concerns that users have about Facebook [36], and people react negatively to scenarios such as a site selling their content to another company, regardless of what they agreed to in TOS [25]. We also know from research about privacy that when information *is* presented to the user, it might misalign with expectations [19] or cause changes in behavior [35]. Despite conceptual differences between copyright and privacy, there are similar underlying issues regarding ownership and control.

This policy layer is a necessary yet often invisible part of content sharing and social computing environments like those included in this work. Policy can be as powerful a design agent as technology, so understanding user impressions of policy is essential to the growth of these environments. However, its role is often considered only as an afterthought or when it becomes a problem [14]. As both users and the research community pay increasing attention to ownership and control online, intellectual property rights are a domain deserving of further attention.

We began this research with a set of unanswered questions: Do people read intellectual property terms for sites where they post content? If (as we might guess) they do not, does it matter? Do they think that any licensing term presented to them is fine, or would they find some to be unfair? What are people's intuitions about what they are agreeing to, and how well do those intuitions match the reality? Moreover, ignorance over how content can be used could be particularly problematic if terms are not consistent across websites. If someone understands Facebook's copyright policy, does this mean that they now know how YouTube or LinkedIn can use their content as well? Are these terms boilerplate, or do they vary by site?

We first analyzed the copyright license and rights-related provisions for 30 different websites focused on user-contributed content. We determined what common provisions exist and in what frequencies, as well as readability measures for each TOS. With this information in hand, we created a survey to discern users' expectations about what these terms are for the sites they use, as well as what they think they *should* be. We surveyed 410 Mechanical Turk users and found that accuracy (how well they are able to correctly identify the copyright terms for a

website), expectations (what terms they think a website has) and opinions (whether they think the site *should* have those terms) vary by licensing term and by type of website. We found that people do care about how their content is used, and moreover, some of the most undesirable uses are far more common than they expect. In other words, there is information not being conveyed properly to users, and it is information that matters.

RELATED WORK

Within the computing community, research concerning website policies has largely focused on issues of readability and web accessibility. Though information accessibility is a major theme within HCI, established tenets of good design tend to be glaringly absent from consideration in crafting terms and conditions on websites [21].

The present study focuses on TOS, but privacy policies and end-user license agreements (EULAs) are similar in presentation. They also cover similar but substantively different content. Privacy policies disclose the ways that a website or technology can gather, use, or disclose a user's data and information. Though some EULAs also include copyright terms,⁴ these contract the relationship between a user and a piece of software rather than a website. Though there is little existing research related specifically to TOS or copyright terms, privacy policies and to a lesser extent EULAs have received a fair amount of attention in the context of usability and accessibility.

This past work has shown a surprisingly high level of complexity within different types of click-through conditions found on the web [8,16,21,23,26]. Perhaps due to this complexity, users also seem to have been conditioned not to read them. One study showed less than 1% of consumers accessing EULAs, and those who did spent a median time of 29 seconds there, far less than the time required to read them [1]. In another, less than 2% of users reported reading EULAs thoroughly with about two thirds saying that they did not read them at all [9]. Yet another study showed that most users take less than 8 seconds to click through a consent dialog [2].

High readability scores suggest that users would not understand these policies even if they read them [16]. Reidenberg et al. asked comprehension questions about privacy policies to non-expert users, knowledgeable users, and privacy experts [29]. Discrepancies between non-expert users and experts (and even among experts) led them to conclude that websites are not conveying privacy information in a way that a reasonable person could understand. Similarly, Good et al. found a "strong

⁴ Though copyright related to user content is not relevant for most pieces of software, there is one domain in particular in which EULAs and intellectual property have received attention: online games and virtual worlds. Like user-generated content websites, users are able to contribute content to games like World of Warcraft and virtual worlds like Second Life, and EULAs govern ownership in that content [12].

disconnect” between actual EULA content and user expectations. They suggested that the design of EULA and TOS makes them inaccessible to users—features like “too long,” “small font,” and “legal mumbo jumbo” [8]. In a later study they showed that if users slowed down and spent more time reading notices, they experienced less regret about their decision to agree to terms [9].

Beyond the way they are presented, the content of terms and conditions can also be problematic. Researchers have examined privacy policies in detail to analyze their relationship to actual government regulations. Content analyses of the policies of library vendors, healthcare providers, and universities have shown that provisions are not always in line with regulations or expectations [4,5,23]. One study of privacy information for mobile apps showed that some data uses surprised and discomfited users [19], and another that understanding privacy information can actually change users’ decisions about website use [35].

Some researchers have taken steps toward solving this usability issue by making terms and conditions more accessible—for example, Kelley’s privacy nutrition labels [18]. Luger suggested more transparency as beneficial, providing a visualization for the readability of terms that users come across on the web [21]. Lin et al. also suggested that some privacy concerns can be alleviated by properly informing users of the reasons for certain terms [19]. Kay and Terry introduced *textured agreements*, a visual redesign of EULAs that had some positive effect on the software agreement process for users [17].

Though most of this prior work has focused on privacy policies and EULAs, many of the same ideas apply to all online terms and conditions. Extending this research to copyright is appropriate, because copyright is traditionally a confusing policy area even outside of the problem of policy readability. Prior research has revealed patterns of misunderstandings about how copyright law applies to online content, particularly since subjective concepts such as fair use are relevant in this context [6,7]. Similarly, studies of social media use have shown striking misunderstandings of rights in one’s own content—for example, that any content posted online automatically becomes public domain [24]. Marshall and Shipman’s study of Facebook showed that there are often large discrepancies between perceived rights and the reality of ownership terms a user agrees to upon using the site [25]. For these reasons, copyright is particularly relevant to CSCW systems. Jackson et al. puts forth copyright as one of the policy areas with complex relationships to CSCW design and practice [14].

In extending previous policy usability work from privacy and HCI to copyright, we are also able to examine important issues of control and ownership in online environments. In doing so, we first confirm that TOS are just as rarely read and difficult to understand as privacy policies. Next, by looking at expectations and opinions, we

consider why this accessibility problem matters in the context of people creating and sharing content online.

TERMS OF SERVICE ANALYSIS

Methods

Previous work has found that in online creative communities, issues of copyright are complicated by appropriation and remix [6,7,32]. In creating a sample of sites to analyze, we used both remix communities (websites where content builds upon existing content, such as music mash-ups or remix videos) as well as other user-generated content (UGC) and social networking sites to ensure that we covered a variety of different types of content creation. Following our prior work examining the role of copyright in these types of online communities [7], we sampled sites along similar categories of media type: writing, music, art, and video. To maintain a systematic method of choosing websites to analyze, we used Alexa search engine rankings (as described in more detail in [7]) to find the highest ranked five websites for each category that focused on user-contributed content. We also included the most popular social networking and user-generated content websites as provided by a website that also bases its algorithm on Alexa rankings.⁵ This sample therefore covers the most common UGC and social networking sites, as well as a mix of other creative content sharing.

With 6 categories, this resulted in 30 websites in our final dataset (see Table 1). We retrieved the TOS for each website. In addition to analyzing the text, we also gathered word count and reading level information. These documents were collected and analyzed for readability in June 2013, and we also checked the copyright licensing terms for any changes in July 2014 after the survey data was collected.

For this study, we focused on just copyright licensing terms—i.e., what rights the user grants in their work. The typical format of these terms is, “You grant this website an A, B, and C license to X, Y, and Z.” The first author is a law school graduate and copyright expert, and determined which sections of the terms were relevant. The format of these licensing terms is consistent, and identifying them in the text is an objective task. However, multiple coders worked on coding the content.

Without knowing exactly what types of licenses would be present, we updated our coding scheme as we went through the set of documents and re-coded previous ones. In addition to coding licenses and rights, we coded for whether the site included plain language explanations. Some of the websites did not have any copyright terms. A comparison of independent coding with a second coder achieved an 89% inter-rater reliability using Cohen’s Kappa [20].

⁵ <http://www.ebizmba.com/articles/social-networking-websites>, <http://www.ebizmba.com/articles/web-2.0-websites> (as of June 2013)

Website	Media	Words	F-K	NE	WW	RF	P	T/S	I	M/T	B	C	A	D
Sound Cloud	Music	7961	19.8	✓	✓	✓				✓				✓
Warcraft Movies	Video	1837	19	✓	✓	✓		✓	✓	✓		✓		✓
Fanfiction.net	Writing	3661	18.3	✓	✓	✓		✓						✓
Asianfanfics.com	Writing	5823	18	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
Craigslist	UGC	5006	17.6		✓		✓	✓	✓	✓				✓
Daily Motion	Video	3223	17.6	✓		✓				✓				✓
ccMixer	Music	2693	17.3											
MySpace	Social	5486	16.9	✓	✓	✓		✓		✓				✓
YouTube	Video	3764	16.6	✓	✓	✓		✓		✓				✓
Club Create	Music	5811	16.3	✓	✓	✓	✓							
Overclocked Remix	Music	1170	15.4	✓	✓		✓	✓		✓				
Twitter	Social	3486	15.3	✓	✓	✓		✓		✓				✓
Flickr	UGC	5763	15.2	✓	✓					✓				✓
Ebaum's World	Video	4659	15.1	✓	✓	✓	✓	✓	✓	✓			✓	✓
LinkedIn	Social	7294	15	✓	✓	✓	✓	✓	✓	✓		✓		✓
DeviantArt	Art	4031	14.7	✓		✓				✓	✓			✓
IMDB	UGC	2613	14.2	✓	✓	✓	✓	✓	✓	✓				✓
Vidders.net	Video	1677	14.2											
Fanart Central	Art	2591	13.2	✓	✓					✓				✓
Elfwood	Art	2300	13	✓	✓	✓				✓				✓
Facebook	Social	4477	12.9	✓	✓	✓		✓						
Pinterest	UGC	2219	12.5	✓	✓	✓				✓				✓
Archive of Our Own	Writing	9142	12.1	✓	✓	✓				✓				✓
Harrypotterfanfiction.net	Writing	985	11.9											
Google+	Social	1691	11.5		✓				✓	✓				✓
Y-Gallery	Art	2368	10.6											
Twisting the Hellmouth	Writing	3005	9											✓
The Otaku	Art	1189	8.4											
Remix 64	Music													

Table 1: Websites with media type, word count, Flesch-Kincaid score, and indication of presence of 11 common licensing terms as of July 2014 (abbreviations noted beginning on the following page). Sorted by highest to lowest Flesch-Kincaid score.

Results

Readability

For each of the 30 TOS in our dataset, we determined word count and Flesch-Kincaid Grade Level Score using statistics automatically generated by Microsoft Word (see Table 1). This is a common readability measure that has

been employed in similar studies for privacy policies [4,11,16]. Only one site on our list (Remix64, a music remix community) did not have any TOS.

The average Flesch-Kincaid Grade Level Score (representing a U.S. educational grade level) was a college sophomore reading level of 14.8, ranging from 8.4 to 19.8.

This puts the readability of these documents roughly on par with those of privacy policies, which multiple studies have found to be in the 14-15 range [11,16]. The average scores for video sites (17.7) and for music sites (17.2) does skew higher than the other media types, which could possibly be accounted for by the additional legal complexities associated with sound. Just as reading privacy policies for all the websites one visits might take years, at an average adult reading speed of 250 words per minute, the TOS for these 30 sites would take almost 8 hours to read.

Copyright Terms

When a user submits content to one of these websites, they are typically licensing that work for use by the site—at the very least, the site must be permitted to display the work, or others would not be able to see it. The sites also license content rather than requiring a transfer of copyright. A partial exception is Club Create, a music remix platform in which users remix provided samples on the site, and the website maintains ownership in some types of that content. Sixteen of the sites explicitly state that the user retains copyright (or ownership) in the work.

Typically the license is stated in a string of legalese in the TOS, similar to the YouTube clause we quoted in the introduction. The majority of these sites include specific license and use rights provisions. Only one did not include a TOS, and another four did not have any information about licensing terms. One website, ccMixter, requires users to license their work under Creative Commons. These websites are grayed out in Table 1.

It should be noted that the absence of a license such as right to display does not mean that the site does not have the right to display a user's submitted content. An *implied* copyright license can be created by law in the absence of an official agreement. However, for this study we focused on the licenses as specifically stated—those that we *know* the user grants by agreeing to a TOS.

In our dataset, we saw the following types of licenses: revocable, irrevocable, assignable, limited, nonexclusive, paid, perpetual, royalty-free, sublicensable, transferable, unrestricted, and worldwide. The most common are also the most unsurprising and necessary. Nonexclusive (20 instances) means that the user is free to use the same content however else they like, including licensing it to others. Royalty-free (17 instances) means that the website is not required to pay the user for their work. Worldwide (20 instances) means the license is effective all over the world.

License types are typically followed by an enumeration of specific uses the website can make of the work. For example, when posting on Craigslist the user is giving the website a license “to copy, perform, display, distribute, prepare derivative works from any content that you post.”⁶ We developed the following codes to cover these different

usage rights: transmit, translate, enforce, reproduce, perform, modify, adapt, transform, index, improve, edit, distribute, display, compile, backup, analyze, advertising/promotion, commentary, commercial use, in connection with site business, use by other parties, use of name/likeness, and unspecified use.

Again, the most common codes were also the least surprising. To reproduce (17 instances), distribute (18 instances), and display (20 instances) are technically necessary in order to have the work appear on the website. There are also a large number of sites that require being able to *change* submitted work—19 instances of a right to modify, adapt, or transform. At times this may only mean something like formatting, and at others, as in the case of Craigslist, the user gives the site the right to make and distribute derivative works. Table 1 shows which websites in our data set have the most common rights and licenses.

Only five sites included plain language explanation of these copyright provisions. Pinterest's TOS includes: “More simply put, if you post your content on Pinterest, it still belongs to you but we can show it to people and others can re-pin it.”⁷ Even this level of explanation is rare.

The number of codes (copyright provisions) did not fall into a pattern with respect to media type or popularity of the site. The site with the most number of codes (26), the small fan fiction website AsianFanFics, contains provisions for the site to essentially do whatever they like with whatever is posted there without any notice or attribution to the creator.

USER PERCEPTIONS SURVEY

Methods

Following our TOS analysis, we examined the relationship between the reality (the actual content of TOS) and user expectations and opinions. Are users aware of how their work can be used, and how do they feel about those uses? We hypothesized that few have read TOS. Based on our finding of difficult readability, we also predicted low accuracy in identifying terms. We also hypothesized, based on knowledge of norms about copyright in creative communities [6], that participants would feel differently about some terms over others (e.g., less favorable about terms involving commercialization and modification).

For the survey, we chose 11 of the most common terms, leaving out those that were too overlapping or non-specific. Our final set of terms that we questioned users about were:

- **Non-exclusive license (NE):** The user can also post/use this content elsewhere.
- **Worldwide license (WW):** The license does not have any geographic restrictions as to where it is valid.
- **Royalty-free license (RF):** The website does not have to pay the user royalties for their content.

⁶ <https://www.craigslist.org/about/terms.of.use> (as of April 16, 2015)

⁷ <https://about.pinterest.com/en/terms-service> (as of April 16, 2015)

- **Perpetual license (P):** The license does not expire.
- **Transferrable/sublicensable license (T/S):** The website is permitted to transfer this license or license the content to another party.
- **Irrevocable license (I):** The user cannot terminate the license once agreed to.
- **Right to modify or transform (M/T):** The website can modify the user’s content (which could range from formatting changes to derivative works).
- **Right to create backups (B):** The website can make copies of the content for the purpose of backups.
- **Right to use commercially (C):** The website can make commercial use of the content, including selling or profiting from.
- **Right to use in advertising (A):** The website can use the content in advertisements.
- **Right to display (D):** The website can display the content (a necessary attribute to show the content on the site itself).

We asked about these license terms in plain language. For example, the question for non-exclusive license read: “If you publish content on [website], can you publish that content somewhere else?” We asked participants to give us their intuition; therefore the only possible answers were “yes” and “no”—their *expectation* (or intuition) for the term. In addition to this factual answer, we asked their *opinion* of the term. Statements like “If you publish content on [website], you *should* be able to publish that content somewhere else” were followed by possible responses on a 5-point Likert scale, from “strongly agree” to “strongly disagree.” Scales that use “agree/disagree” response categories have a tendency to create an acquiescence bias amongst respondents [31], but for our purposes these scales were optimal for efficiently asking questions across a range of sites and dimensions. The survey also included demographic questions, and asked participants about copyright training and whether they had read the TOS of each website for which they answered questions.

The survey contained a list of the 30 sites listed in Table 1 and asked participants to check which sites they had used (denoting whether they had only read them or had shared content). To ensure that the survey would take comparable time for each participant, rather than asking about the TOS for every website they indicated, the survey provided a set of questions for each of three of the websites they indicated, chosen randomly. Though our sample of websites purposefully skewed towards creative communities that value appropriation, because participants answered questions based on the websites they use, our data set is proportional to the popularity of the websites (see Table 2).

We pilot tested the survey on a group of 10 individuals recruited from our personal social networks, including 2 attorneys (who corroborated our plain language explanations). After this pilot testing we tweaked the wording of some of the questions to avoid ambiguity. We then implemented the survey online and recruited participants through Amazon’s Mechanical Turk (Mturk) crowdsourcing service. This method offers greater scalability at a lower cost and a wider demographic net than localized recruitment such as through a university. Participants on Mturk have performed comparably to laboratory subjects in traditional experiments [28], and the demographics of Mturk users can be more demographically diverse than Internet samples [3].

One known limitation of Mturk is that participants are less likely to pay attention to experimental materials [10]. However, this can be somewhat mitigated by the use of “attention checks” or screening questions to gauge attention. We included a simple, invisible attention check in our survey. The list of websites that participants could choose from included a 31st option: Mechanical Turk. Since anyone taking the survey was an Mturk user, if they did not check that box, they were told that they were not eligible to take the survey. Only 14 users were bumped from the survey in this way, suggesting that the majority of Mturk users were at least paying enough attention to read the list of sites and respond accurately about their use. Having shown from our pilot study that the survey generally took less than 5 minutes to complete, we paid Mturk participants 50 cents per survey, to ensure a rate of at least \$6/hour.

We deployed our survey as an Mturk task in May 2014, and 410 workers completed it. Because the Mturk population is primarily split between the United States and India (and therefore does not allow for a true global sample), we limited our participants to U.S. citizens to avoid this binary. This is a limitation to our study. Though these results likely cannot be generalized to a global population, previous research has also found that conversations about copyright online tend to focus on U.S. law [7].

All Mturk workers are 18 or older, and we had no other requirements for participation. Our participants were 57% male, 75% white, and represented every U.S. state. The age range was 18-82 with a mean of 31. Regarding education, 39% had a college degree, and another 36% had attended at least some college. We also asked about the kind of creative work they post online. 61% had created some type of creative work in the media of writing, music, video, or art. Only 9% of our participants reported having had formal copyright education or training.

Findings

From our 410 participants, each of the 30 websites was reported as used by anywhere between 14 (Club Create) to 391 (YouTube) participants. However, each participant completed questions for only 3 (randomized) websites that

they chose. Regardless, every one of the 30 websites studied had at least one participant surveyed for that site. There were a total of 1225 sets of questions answered. Table 2 shows these numbers per site.

For each one of these 1225 sets of questions, we asked the participant whether they had read the TOS for that particular website—11% said yes. Though this number is still low, we suspect it is also inflated due to self-reporting bias. Additionally, those who contributed to a site are slightly more likely to have read the TOS than lurkers (14% compared to 10%, $p < 0.05$).

Website	Media Type	Selected	Surveys
Facebook	Social	390	124
Wikipedia	UGC	378	120
Pinterest	UGC	336	111
Craigslist	UGC	373	106
MySpace	Social	315	97
YouTube	Video	391	96
LinkedIn	Social	284	76
IMDB	UGC	323	75
Twitter	Social	269	66
Google+	Social	282	63
Flickr	UGC	233	58
DeviantArt	Art	217	56
Daily Motion	Video	172	44
Ebaum's World	Video	164	33
Sound Cloud	Music	157	31
Fanfiction.net	Writing	88	20
Harrypotterfanfiction.net	Writing	25	8
Warcraft Movies	Video	29	6
Archive of Our Own	Writing	26	4
Asianfanfics.com	Writing	20	4
Elfwood	Art	18	4
Overclocked Remix	Music	29	4
The Otaku	Art	27	4
ccMixter	Music	15	3
Remix 64	Music	15	3
Y-Gallery	Art	16	3
Fanart Central	Art	15	2
Twisting the Hellmouth	Writing	16	2
Club Create	Music	14	1
Vidders.net	Video	15	1

Table 2: How many participants selected (reported having used) each website and how many filled out surveys on each website, ordered by how represented they are in our survey data

Accuracy

One research question we set out to answer was how aligned participants' expectations for the terms of websites were with the reality of those terms. For each set of questions, we can calculate an *accuracy* score for how many of their answers matched the terms of that website.

One complexity is that not every term included in the survey was relevant for every website. This is because the *absence* of a term does not necessarily imply its negation. For example, "royalty-free," one of the more common licenses, means that users are not paid monetary royalties. Facebook requires a royalty-free license, so when we asked participants "Does Facebook pay you for your content?" the correct answer was "no." In contrast, Google+ does *not* specify a royalty-free license. However, this does not mean that Google+ *does* pay users for their content. Though for some terms this negation may be true (for example, the absence of an irrevocable license suggests that it is revocable), we could not assume that this was the case. Therefore, *accuracy* here is how well a participant was able to correctly identify terms appearing on a site. Accordingly, the six sites without copyright terms do not have accuracy scores and were not included in these calculations.

Looking across every question as an individual data point (with an N of 6822 questions about terms that *existed* out of a total 12250 asked) our participants had an accuracy of 69%. Aggregating the data by individual participant, the mean accuracy is 67% with a median of 70%, ranging from 0 to 100. By website, the mean is 66%, ranging from 45% (AsianFanFics) to 85% (Archive of Our Own).

Licensing Term	Sites	% Accurate	N
Display	20	93.4	954
Non-Exclusive	20	88.3	875
Royalty-Free	17	87.4	796
Worldwide	20	84.5	1026
Backup	3	78.2	124
Advertising	2	73.0	37
Perpetual	7	65.6	299
Commercialize	3	57.0	86
Irrevocable	7	46.6	755
Modify/Transform	19	42.2	1056
Transferrable	12	38.2	814

Table 3: Accuracy by licensing term, as well as the number of sites with that term (i.e., how common the term is). N is the number of responses used to calculate accuracy.

We can also analyze the average accuracy per licensing term, which provides much greater variance. Table 3 shows the accuracy on all questions asked about that term. The N is much lower for uncommon licensing terms, such as right to use in advertising, because this only reflects responses for websites on which these terms appeared.

The results in Table 3 show which copyright terms misalign with user expectations. With 93% accuracy, the right to display is an intuitive one. None of these websites could show content to other users if they did not have the right to display it. The high accuracy score shows that users have an intuition for common sense terms. However, other terms such as transferrable (can the website transfer your license to someone else?), irrevocable (can you take back the license once you grant it?), and right to modify (can the site change your work?) are much less intuitive. A *low* accuracy score indicates that users frequently indicated that a term was *not* there when it was—i.e., unexpected.

The variance here also explains why accuracy scores by website are potentially not very meaningful. Because different websites include different licensing terms, some may include terms that are more confusing than others. The lowest accuracy score comes from AsianFanFics, which also has the highest number of licensing terms of any of the sites. Therefore, it is difficult to speculate that this is due to factors such as the readability, when the TOS also contains more of the “confusing” terms. Our data shows no significant correlation between accuracy and either word count of TOS (Spearman $R = 0.195$) or Flesch-Kincaid score (-0.235). However, this is not surprising considering the low number of respondents reporting having read TOS.

Interestingly, whether or not the participant reported having read the TOS also did not appear to impact accuracy. Though mean accuracy for TOS readers (69%) was somewhat higher than for non-TOS readers (66%), this difference did not rise to the level of significance ($p = 0.09$). We also measured no differences based on whether the participant reported copyright training ($p = 0.24$) or whether they were a contributor or a lurker ($p = 0.16$).

In sum, 69% accuracy suggests that perhaps Internet users do have decent intuition about many of the licensing terms that they agree to, but there is a great deal of variance by licensing term. Those with low accuracy that appear frequently, such as right to modify, present the biggest potential problem for users.

Expectations and Opinions

Our participants indicated what they thought the reality *was* for each licensing term as well as what they thought it *should* be. Regardless of whether or not they were correct, this gives us a sense for user expectations and opinions about each term. With a binary yes-or-no answer (0 for no and 1 for yes), the *expectation* score for each term tells us what proportion of participants thought that licensing term existed. An *opinion* score comes from a Likert scale (with 1

being strongly disagree, and 5 strongly agree), indicating in aggregate how much participants thought it should exist for that website. There is a positive correlation between these scores (see Figure 1). The licensing terms users think the sites have generally match what they think they should have; this suggests that to some extent users trust the sites they use to have reasonable policies.

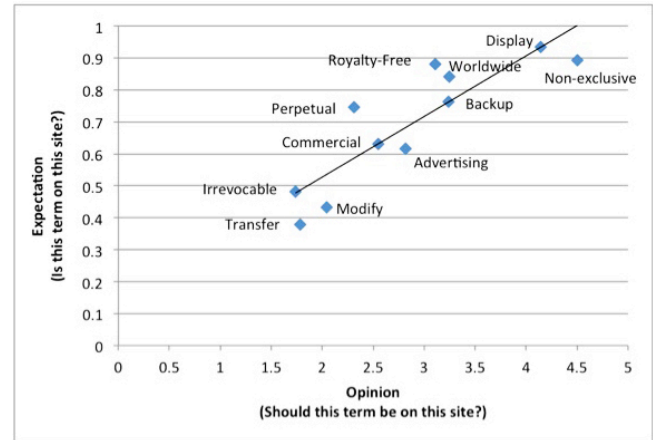


Figure 1: Positive correlation between opinion and expectation scores – opinion on a 1 to 5 Likert scale and expectation on a 0 to 1 scale representing proportion of yes and no answers.

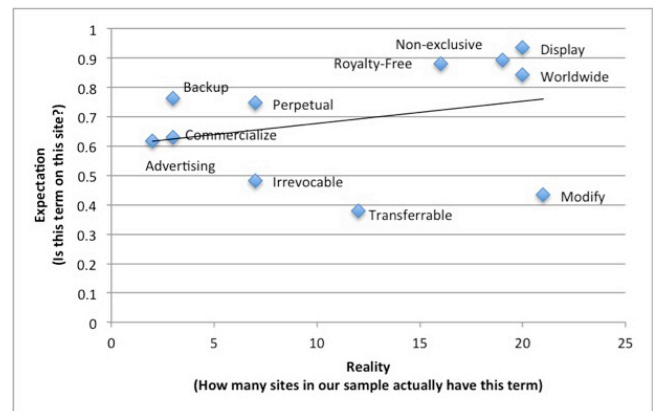


Figure 2: Correlation between expectation score (on a 0 to 1 scale representing proportion of yes and no answers) and reality (number of sites represented) for each licensing term

However, expectation does not necessarily match the *reality* of sites’ licenses; transfer and modify terms, for example, are very common. Using the number of sites in our sample that do have a certain term as a proxy for reality, the relationship between expectation (how often users expect a term to be present) and reality (how often a term *is* present) is by contrast not linear at all (see Figure 2). In other words, there are a number of terms that are often present that users would find surprising.

Just as there were no differences between contributors and lurkers on accuracy, neither were there any significant differences on expectation or opinion measures when we analyzed participants separately based on whether they had

contributed content to the site or not. This holds true for the results discussed in the next section as well.

Differences in Website Types

Prior work has suggested that content creators of different types of content (for example, art versus writing) may have different attitudes towards copyright [6]. Therefore, we expected that we might see differences in opinion based on the *type* of website participants were questioned about: music, writing, video, visual art, social media, and other user-generated content.

For each of these types of websites, we compared responses for that type (see Table 1 for a listing of type for each website) versus other responses, using t-tests to compare means. We examined both expectation and opinion. We did not see striking differences, particularly given the granularity of Likert scores (and ours are aggregate rather than single-item measures), though there were some non-random patterns. It should be noted as well that our statistical tests were assuming unequal variances, since there were unequal sample sizes—for example, 426 social media responses to 799 non-social-media responses (see Table 2).

In writing communities, which are primarily fandom-based (i.e., fan fiction sites), we see somewhat less expectation for websites to have *modify* (24% for writing versus 44% overall, $p < 0.05$) and *commercial* terms (46% for writing versus 64% overall, $p < 0.05$). This tracks to social norms in these communities around remixing permissions and noncommercialization [6]. On music and video sites, there is a small but significant difference for royalty-free licenses; participants favor these licenses somewhat *less* (2.58 for music and 2.46 for video versus 3.22 overall, $p < 0.01$), suggesting that they might be more amenable to users being paid for their work. Though we cannot make strong claims about these small differences, it may be some reflection of differing values in these communities.

In Marshall and Shipman's study of ownership values on Facebook, they suggested that the *personal* nature of Facebook content may be at the root of some of the discrepancy between perceived ownership rights and reality [25]. In our data set, social media sites do stand apart from other media types since they may contain more personal content than more creative communities. In comparing social media versus other types, we see the same pattern for nearly every licensing term: a slight favoring of the site having *less* control over their content. For example, a mean opinion score of 2.83 over 3.51 ($p < 0.001$) means that users were venturing towards "somewhat disagree" over "somewhat agree" as to whether social media sites should have a "worldwide" license. This suggests that they might favor some geographic constraints in the content use.

Interestingly, examining *expectation* data rather than opinion—what the users think the terms actually are—for some of the terms we actually see the exact opposite. On

social media sites, 42% of users versus 36% for other users think that the site has the right to transfer or sublicense content to third parties ($p < 0.05$), 54% versus 45% think that the license the user grants is irrevocable ($p < 0.01$), and 92% versus 86% think that the site does not have to pay the user royalties ($p < 0.001$). This difference suggests that though users may think that social media sites actually have *more* control over their content than other types of sites, they want them to have less. Though these differences are quantitatively small, the conceptual difference between opinion and expectation is worth noting.

Overall, while there are non-random differences between media types, the fact that the trend is still similar in terms of real numbers suggests that people have basic opinions about how any content should be used. However, these basic opinions are supplemented by at least some differences based on the website and type of content.

DISCUSSION

The first part of this study was to extend prior work on privacy policies to copyright and Terms of Service, asking whether people read TOS and whether they would understand them even if they did. The answer is no. The 11% of our participants who reported reading was somewhat higher than we expected based on prior work, but still represents a minority and may also be affected by self-reporting bias. Low readability and high word counts for TOS is on par with privacy policies, though we did find that people have good intuitions about some of the more common sense terms, such as right to display. However, low accuracy in predicting the existence of terms such as transferrable and modifiable are more problematic since these terms are common—and apparently users do not realize this.

Given these findings, our major point of reflection is: does it matter? Why is it important that users understand TOS? Is it important enough that fixing this problem should be a design goal? We argue that it does matter, for two reasons. The first is that copyright licenses are far from one size fits all. Based on our analysis of what terms exist on different websites, we see a great deal of variability. Beyond ubiquitous and common sense terms such as right to display, there are those that appear infrequently such as right to commercial use. Therefore, users cannot assume that the terms will be the same across different websites. This goes against conventional wisdom that the legalese in TOS is all boilerplate terms.

Moreover, our data shows that people do care. This follows previous work that has shown, for example, that information in privacy policies can affect purchasing behavior [35] and that Facebook users report content ownership as one of their concerns about using the site [36] while having strong opinions about that ownership [25]. Though we did not ask our participants directly about how licensing terms would affect their site use, we saw strong

differences in opinion about different terms. Clearly, copyright licenses are not one size fits all for users either.

Consider the three terms appearing as outliers in Figure 2. Transferrable, modifiable, and irrevocable are the terms that appear most often appeared when not expected. Figure 1 also illustrates that they have the lowest opinion scores, suggesting they could be the ones participants care the most about. Distaste over transferrable and modifiable licenses suggests a desire for control over how content is used: for it not to be changed nor given to someone else. Not wanting irrevocable licenses suggests a desire for choice—ability to cut ties with a website at any given point.

The fact that opinions and expectations align suggests that people generally trust these websites to protect their interests when it comes to their content. Users don't *want* to grant transferrable, irrevocable licenses to modify work, but they also don't think that they are. This misalignment between expectations and reality is a significant usability problem. Many users *are* granting these rights without realizing, and they might be unhappy if they knew.

The question as to how this harms the user would be highly dependent on the intent of the website—but the truth is that these licensing terms give them the *option* to use content in these ways. At the time of our data collection,⁸ LinkedIn's license was in part assignable, sublicensable, and irrevocable, and granted the right for them to create derivative works. Assignable and sublicensable (similar to transferrable) mean that they have your permission to give this license to some other party (for example, license your content for use on a different website). Creating derivative works could mean anything from changing the font size to producing a blockbuster film based on a blog post. Finally, irrevocable means that once you have granted these rights, you cannot un-grant them. Even if LinkedIn had no intention of selling books filled with wisdom gleaned from content posted by Mark Zuckerberg or Kim Kardashian, or featuring your profile on a new "world's worst resumes" webpage (and this is unlikely), users had given them the *option* of doing these things. The fact that they did so unknowingly is a usability problem.

Of course, with only about a tenth of users having read the TOS, is this misalignment of expectation and reality the responsibility of the website? With the TOSs we analyzed having an average of 3851 words and requiring a college sophomore reading level, users may have long been trained out of attempting to understand them. However, it is clear from our results that users *do* have opinions about how their content should be used—perhaps to the point where the licensing terms would affect their decision to use a site. Just as researchers have devised more readable modes of presentation for privacy policies [18], websites should

consider presenting their licensing terms in a readable way outside of the block of TOS legalese.

Additionally, the fact that there are differences in opinions based on media type suggests that there could be differing norms and attitudes about copyright among different types of content creators. Interestingly, the website with the highest accuracy score of participants accurately predicting their licensing terms, Archive of Our Own, is also a website that was specifically designed with the social norms of a certain creative community in mind [15].

This is not to say that site designers should alter their licensing terms based on the specifics of these findings, but that they should consider more carefully the needs of their particular community of users. Consider the amount of user research that often goes into design decisions such as where to place a button on a website or exactly how the flow of a pay transaction should be designed. How often, in contrast, do websites conduct user studies about what their copyright policies—or any part of their Terms of Service—should be?

Websites are often constrained by legal departments and required language for terms and conditions. However, in *explaining* those terms, it is important to be mindful of terms that users would find most alarming—for example, the right to modify or transform. This licensing term could mean anything from changing the font size to editing content to creating a blockbuster film. A study of data use requirements on mobile apps found that simply knowing *why* the app needed that information did much to alleviate privacy concerns [19]. An understanding of not only what the term *means* but what the site actually *means to do* would be an important step towards alleviating user concerns about uses of their content. For example, Archive of Our Own's TOS has a line after their license (including in part "modifying or adapting") that reads: "Modifying and adapting here refer strictly to how your work is displayed—not how it is written, drawn, or otherwise created."⁹ Even this amount of plain language explanation is rare. The introduction of this paper describes instances of social media users becoming alarmed when copyright terms were brought to their attention. These problems could often be avoided without a change in the policies themselves, but rather by explaining in clear terms the purpose behind them.

CONCLUSION

Online content contribution is an important component of the systems that we study within CSCW, and copyright is a necessary yet often invisible layer in these environments. Understanding user impressions of policy therefore fills a gap in this space. Towards this end, some HCI researchers have put forth the complexity of privacy policies as an important usability problem [2,5,16,18,19,21], and have also come up with technological solutions to help [18,19]. In extending the descriptive part of this work to copyright

⁸ LinkedIn's copyright terms have changed significantly since the time of this study, and no longer claim so many rights in their license. <https://www.linkedin.com/legal/user-agreement>

⁹ <https://archiveofourown.org/tos> (as of April 16, 2015)

terms, we reveal that user information deficit about online terms and conditions is generalizable beyond privacy. Moreover, the variability of licensing terms across websites, the mismatch between user expectation and reality, and differing user opinions about licensing terms reveals that this is information that matters.

We consider this an important lesson for designers of user-generated content websites who may assume that users do not read terms and conditions and therefore do not care about their content. However, there have been a number of examples in the past, such as with Facebook, Instagram, and OK Cupid, where users have demonstrated that they cared deeply [22,30,34]. Our work suggests that when users do stop to think about individual licensing terms that they might agree to, they are not ambivalent about how their content can be used. As Jackson et. al suggest, policy should be considered alongside design and practice as an important component of CSCW research [14].

Website designers could go a long way towards helping by simply including plain language explanations of their terms and intentions. Additionally, there is a potential for technological solutions. For example, researchers have had some success in automatic parsing of privacy policy text in order to present information to the user in a more readable way [18,29]. Licensing terms would be even easier to parse, since they tend to use very similar language and structure. Future research in this area could provide design and technological solutions for helping online content creators understand their rights, and also look beyond Terms of Service to further complexities of intellectual property rights. Ownership and copyright are increasingly important as content contribution continues to be a major part of life online, and we hope that future CSCW work will take up additional problems in this space as well.

ACKNOWLEDGMENTS

Thank you to Josephine Antwi, Titus Woo, Joseph Gonzales, and Jessica Feuston for their assistance on this project, as well as to colleagues (particularly Annie Antón, Eric Gilbert, and Rebecca Tushnet) and reviewers for their valuable feedback. This work was supported by NSF IIS-1216347.

REFERENCES

1. Yannis Bakos, Florencia Marotta-Wurlger, and David R. Trossen. 2009. Does Anyone Read the Fine Print? Testing a Law and Economics Approach to Standard Form Contracts. *New York University Law and Economics Working Papers* Paper 195. Retrieved from http://lsr.nellco.org/nyu_lewp/195
2. Rainer Bohme and Stefan Kopsell. 2010. Trained to Accept? A Field Experiment on Consent Dialogs. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10).
3. M. Buhrmester, T. Kwang, and S. D. Gosling. 2011. Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science* 6, 1, 3–5.
4. Mary J. Culnan and Thomas J. Carlin. 2006. Online Privacy Practices in Higher Education: Making the Grade? *Communications of the ACM* 52, 2, 126–130.
5. Julia B. Earp, Annie I. Antón, Lynda Aiman-Smith, and William H. Stufflebeam. 2005. Examining Internet Privacy Policies Within the Context of User Privacy Values. *IEEE Transactions on Engineering Management* 52, 2, 227–237.
6. Casey Fiesler and Amy S. Bruckman. 2014. Remixers' Understandings of Fair Use Online. *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '14).
7. Casey Fiesler, Jessica Feuston, and Amy S. Bruckman. 2015. Understanding Copyright Law in Online Creative Communities. *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '15)
8. Nathaniel Good, Rachna Dhamija, Jens Grossklags, et al. 2005. Stopping Spyware at the Gate: A User Study of Privacy, Notice and Spyware. *Proceedings of Symposium On Usable Privacy and Security* (SOUPS).
9. Nathaniel Good, Jens Grossklags, Deirdre K. Mulligan, and Joseph A. Konstan. 2007. Noticing Notice: A Large-Scale Experiment on the Timing of Software License Agreements. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07).
10. J.K. Goodman, C.E. Cryder, and A. Cheema. 2013. Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples. *Journal of Behavioral Decision Making* 26, 3, 213–224.
11. Mark A. Graber, Donna M. D'Alessandro, and Jill Johnson-West. 2002. Reading Level of Privacy Policies on Internet Health Web Sites. *The Journal of Family Practice* 51, 7, 642–5.
12. Sara M. Grimes. 2006. Online multiplayer games: A virtual space for intellectual property debates? *New Media & Society* 8, 6, 969–990.
13. Christopher M. Hoadley, Heng Xu, Joey J. Lee, and Mary Beth Rosson. 2010. Privacy as Information Access and Illusory Control: The Case of the Facebook News Feed Privacy Outcry. *Electronic Commerce Research and Applications* 9, 1, 50–60.
14. Steven J. Jackson, Tarleton Gillespie, and Sandy Payette. 2014. The Policy Knot: Re-integrating Policy, Practice and Design in CSCW Studies of Social Computing. *Proceedings of the 17th ACM Conference*

- on Computer Supported Cooperative Work & Social Computing (CSCW '14).
15. Anne Jamison. 2013. *Fic: Why Fanfiction is Taking Over the World*. Smart Pop, Dallas, TX.
 16. Carlos Jensen and Colin Potts. 2004. Privacy Policies as Decision-Making Tools: An Evaluation of Online Privacy Notices. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*.
 17. Matthew Kay and Michael Terry. 2010. Textured Agreements: Re-envisioning Electronic Consent. *Proceedings of the Sixth Symposium on Usable Privacy and Security (SOUPS)*.
 18. Patrick Gage Kelley, Lucian Cesca, Joanna Bresee, and Lorrie Faith Cranor. 2010. Standardizing Privacy Notices: An Online Study of the Nutrition Label Approach. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*.
 19. Jialiu Lin, Shahriyar Amini, Jason I. Hong, Norman Sadeh, Janne Lindqvist, and Joy Zhang. 2012. Expectation and Purpose: Understanding Users' Mental Models of Mobile App Privacy through Crowdsourcing. *Proceedings of the ACM International Conference on Pervasive and Ubiquitous Computing (UbiComp '12)*.
 20. Matthew Lombard, Jennifer Snyder-Duch, and Cheryl Campanella Bracken. 2002. Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability. *Human Communication Research* 28, 4, 587–604.
 21. Ewa Luger, Stuart Moran, and Tom Rodden. 2013. Consent for All: Revealing the Hidden Complexity of Terms and Conditions. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*.
 22. Matthew Lynley. 2012. Why the Web is Freaking Out Over Instagram's New Terms of Service. *The Wall Street Journal*. Retrieved from <http://blogs.wsj.com/digits/2012/12/18/why-the-web-is-freaking-out-over-instagrams-new-terms-of-service/>
 23. Trina J. Magi. 2010. A Content Analysis of Library Vendor Privacy Policies: Do They Meet Our Standards? *College & Research Libraries* 71, 3, 254–272.
 24. Catherine C. Marshall and Frank M. Shipman. 2011. The Ownership and Reuse of Visual Media. *Proceedings of the 11th ACM/IEEE-CS Joint Conference on Digital Libraries*.
 25. Catherine C. Marshall and Frank M. Shipman. 2015. Exploring the Ownership and Persistent Value of Facebook Content. *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*.
 26. Aleecia M. McDonald and Lorrie Faith Cranor. 2008. The Cost of Reading Privacy Policies. *I/S: A Journal of Law and Policy for the Information Society* 4, 543–565.
 27. Stephanie Teebagy North. 2011. Twitteright: Finding Protection in 140 Characters or Less. *Journal of High Technology Law* 11, 2, 333–364.
 28. G. Paolacci, J. Chandler, and P.G. Ipeirotis. 2010. Running Experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 5, 5.
 29. Joel R. Reidenberg, Travis Breaux, Lorrie Faith Cranor, et al. 2015. Disagreeable Privacy Policies: Mismatches Between Meaning and Users' Understanding. *Berkeley Technology Law Journal* 30.
 30. Martin Robbins. 2014. Does OKCupid need our consent? *The Guardian*. <http://www.theguardian.com/science/the-lay-scientist/2014/jul/30/does-okcupid-need-our-consent>
 31. Willem E. Saris, Melanie Revilla, Jon A. Krosnick, and Eric M. Shaeffer. 2010. Comparing Questions with Agree/Disagree Response Options to Questions with Item-Specific Response Options. *Survey Research Methods* 4, 1, 61–79.
 32. Oshani Seneviratne and Andrés Monroy-Hernández. 2010. Remix Culture on the Web: A Survey of Content Reuse on Different User-Generated Content Websites. *Proceedings of Web Science Conference (WebSci '10)*.
 33. Anna C. Squicciarini, Mohamed Shehab, and Joshua Wede. 2010. Privacy Policies for Shared content in Social Network Sites. *VLDB Journal* 19, 6, 777–796.
 34. Brian Stelter. 2009. Facebook's Users Ask Who Owns Information. *The New York Times*. <http://www.nytimes.com/2009/02/17/technology/internet/17facebook.html>
 35. Janice Y Tsai, Lorrie Cranor, Serge Egelman, and Alessandro Acquisti. 2010. The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study. *Information Systems Research* 22, 2, 254–268.
 36. Hui Zhang, Munmun De Choudhury, and Jonathan Grudin. 2014. Creepy but Inevitable? The Evolution of Social Networking. *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*.