

## Who Creates Content?

### Stratification and Content Creation on the Internet

Grant Blank  
Oxford Internet Institute  
University of Oxford

Email: [grant.blank@oii.ox.ac.uk](mailto:grant.blank@oii.ox.ac.uk)

*Abstract:* Until the Internet arrived, content creation and distribution was always an expensive, difficult process. With the Internet it is dramatically easier, faster, and cheaper. Some argue (e.g. Benkler 2006) that this will move creation out of the hands of elites and lead to wider participation in the public sphere and to enhanced democracy. This paper makes three contributions to this debate. First, it uses a national random sample of the British population. This is much broader than most prior work. Second, it creates the first evidence-based typology of Internet content creation, identifying three types named “skilled content”, “social and entertainment content” and “political content”. The implicit assumption of many researchers that only one type of content exists is not accurate. Third, using multivariate logistic regression it shows the characteristics of different populations that produce each type of content. Elites have no impact on creation of skilled content. Social and entertainment content is more likely to be created by non-elites. Only creation of political content is significantly and positively associated with elite status. These results clarify inconsistencies in prior studies. Each type of content is produced by a different kind of creator. Thus, type is more than just content; it also describes differences in who creates the content. The varying relationships between elite status and content creation suggests that it is important for future research to pay close attention to the type of content under study when considering possible democratization of creation.

Keywords: Internet; content creation; content types; mobility; elite; Oxford Internet Survey

## Who Creates Content? Stratification and Content Creation on the Internet

Grant Blank

One striking development of the Internet is the rise of an entirely new way for people to create and distribute news, opinions, entertainment and information. As a consequence of the development of blogs, personal web sites, social networking sites and the like, traditional news media have lost their institutional monopoly over the mass production and distribution of news, entertainment and opinions. Even though mass media continue to exist in the sense of a small number of large organizations producing large-circulation and large-audience publications, they have irretrievably lost the dominance that they enjoyed since the beginning of mass-circulation newspapers over 150 years ago.

I call this development “personal publishing” to distinguish it from the capital-intensive, mass-oriented corporate- and government-controlled media, which I will call “institutional publishing”. Personal publishing includes not just text (like a blog) but also music, photos, videos, books, pamphlets and other products that are available from individuals over the Internet. The important point is that with minimal cost and not much equipment these products can be created by individuals and distributed via the Internet to audiences that may be potentially very large. The Internet and small computers have combined to create an environment with a low cost of production, relatively low technical skill requirements, minimal capital needs, and low cost distribution so that it is possible for ordinary people to reach large audiences. This is new.

The implications of this development include the collapse of the old business model of traditional media organizations, and the emergence of large numbers of individuals who publish and disseminate the information and opinions that they believe are important and valuable. Many others have noticed these changes and what they all mean has been the subject of much research.

With the attractive possibilities of personal publishing in mind, one stream of early research focussed on who did and did not have access: the “digital divide” (Hoffman & Novak 1998;

Katz & Aspden 1997; van Dijk 2006). Much of this work focussed on demographic predictors of Internet use (Hoffman & Novak 1998; Howard, Rainie & Jones 2001; Katz & Aspden 1997; Loges & Jung 2001). Although personal publishing makes it possible to create and distribute content, it is not that simple. Creation takes time and energy. To convince anyone to pay attention, effective content creators needed skills like the ability to write well, to be persuasive, and to create high-quality videos, photos, or music. These skills have not become more common or easier to learn because of the Internet. The Internet does not make good writing any easier or faster. As Internet use spread, it became apparent that many people are happy to do no more than email and light web searching. They are satisfied to create content that is mostly gossipy chatter intended for friends and they do not want to go beyond posting about their personal interests and relationships (Nardi et al. 2004; Shirky 2008). Only a minority actually contribute broader content. Since access alone is not sufficient for people to become broad participants on the Internet, the conceptual focus of research has shifted to a more subtle question: who is able to make full use of the potential of the web? This focus has no standard name but is variously called the *second-level digital divide* by Hargittai (2002), the *participation gap* by Jenkins et al. (2006), *emerging digital differentiation* by Peter & Valkenburg (2006) and the *usage gap* by van Dijk (2004). These concepts highlight inequalities. This is an important topic, but it misses some issues in how stratification in the production of content is related to the type of content produced that are highlighted by the idea of personal publishing. I develop this point in the discussion, below.

Since skills like being able to write interesting and persuasive sentences tend to be associated with higher social status, is the additional participation mostly among people who are highly educated or have high incomes? If this were true then, although there would be many more people expressing opinions, there may not be much greater diversity. If content is produced mostly by high status people, then personal publishing will be dominated by the same elites as

institutional publishing. This raises several questions: Does the Internet foster a greater variety of voices and opinions? The key is stratification: Are low-status people more likely to create content on the Internet or does the Internet reproduce high-status dominance?

There are two points of view on the question of social stratification and content creation. Jenkins et al. (2006) and Benkler (2006) argue that the Internet provides opportunities for greater participation. Benkler's argument is sophisticated and worth summarizing. He points out that the Internet provides the opportunity for non-market solutions to communication and information problems, like the production and distribution of news, entertainment and opinions. These solutions increase individual autonomy in the creation and selection of information because they do not rely on very large capital investments that can only be made by governments and large media corporations. By making personal publishing possible the Internet increases individual freedom to participate in civic debates. This makes it easier for low status people to participate, he argues, and it will increase the diversity of information and opinions in the civic sphere.

Despite his theoretical persuasiveness, Benkler analyzes no data. A number of writers have analyzed relevant data. The most directly relevant studies on content creation are three multivariate studies: Correa (2010), Hargittai & Walejko (2008), and Schradie (2011a) (other studies, either bivariate or less directly focussed on content creation, will be summarized below). These studies have explored both demographic and non-demographic predictors. Although summaries (e.g. Hargittai 2008) have concluded that the Internet tends to reinforce (or at least not challenge) existing patterns of stratification, there are many details and the results for content creation are *not* consistent. If we look at income and education as measures of status the instability is marked. Education has been both not significant (Hargittai & Walejko 2008, Table 7; Correa 2010, Table 1) and significant for six of 10 dependent variables (Schradie 2011a, Table 3). Income has been included only in one paper (Schradie 2011a, Table 2) and it is significant for only two of 10 models.

This paper contributes to this conversation in three ways. First, although previous studies have used data from the United States (Correa 2010; Hargittai & Walejko 2008; Schradie 2011a), this is the first analysis of the United Kingdom. The UK has a different politics, different culture and different institutional media. Given these differences one cannot assume that research findings from other countries apply to Britain. Second, the dataset is a random sample of the entire British population, in contrast to studies of college students (Correa 2010; Hargittai & Walejko 2008). Finally, rather than assuming all content creation is the same (like, for example, Correa (2010) who sums all creation variables into a single measure) or assuming that each variable must be treated separately (e.g. Schradie 2011a), this paper attempts to identify types of creation. Different types of creation allow us to make distinctions about the effects of social status that other authors using alternative methods could not find. As we shall see, the relationship between content creation and elite social status is more complex than previous authors suppose.

### **Demographic predictors of content creation**

There are strong relationships between several demographic variables and content creation. Age is one of the most consistent predictors of all kinds of Internet activities (Dutton & Blank 2011). Older people are less likely to produce content (Jones & Fox 2009; Schradie 2011a). The political content creation studies point both ways: Age predicts political content creation in bivariate studies like Schlozman et al. (2010) but age is not significant in multivariate studies of political “Internet activity” (Kittilson & Dalton 2011:639). The studies of college students (Chen 2007; Correa 2010; Hargittai & Walejko 2008) find that older students produce less content. Notably, this effect occurs even though the age range is very limited among college students.

The effects of gender vary more than age. Females are less likely to post content online, but that effect disappears after controlling for skill and experience (Correa 2010, Hargittai &

Walejko 2008). Some find no gender differences in social network use (Chen 2007; Schradie 2011a), although they find women are less likely to post videos. Schradie (2011a) also finds women are less likely to blog, have a personal web site, participate in chat rooms, or post comments to newsgroups but these regressions do not incorporate controls for skill or experience.

Race is frequently included in American studies, but the British racial environment is very different. British racial minorities are mostly voluntary immigrants from ex-colonies, few lived in Britain before World War II, there are virtually no Hispanics, and there are large numbers of recent immigrants from Eastern Europe and the Balkans who are culturally different but not racially different or visually distinctive. Race is included in our models below, but American results provide no guidance in the UK.

This literature focuses on measures of socio-economic stratification. The three primary measures, social status, education and income, tend to be strongly, positively related to Internet *use* (e.g. Dutton & Blank 2011; Zillien & Hargittai 2009) but when we focus specifically on *content creation* the results are more ambiguous. Income is significant in only two of the 10 activities studied by Schradie (2011a, Table 2): sharing content on the Internet and posting to newsgroups. Education is positively associated with content creation in six of her 10 activities. Social status is positively associated with political content creation (Schlozman et al. 2010a), as is education (Kittilson & Dalton 2011). The studies of college students measure status in the same way, using parental education, but they reach opposing conclusions: for Correa (2010) social status is never significant; Hargittai and Walejko (2008) conclude that students with higher social status are more likely to create material but, as we shall see below, their data are more complex than their text conclusion.

Several other variables have been addressed by individual studies. Internet skills and experience are positively associated with content creation (Hargittai & Walejko 2008). Certain

forms of marital status are important. Among her 10 dependent variables, Schradie (2011a) finds that being widowed is associated with less content creation in three variables; being married is also negatively associated in four cases. Finally, living in a rural area is negatively associated in four dependent variables.

### **Beyond demographics: Psychological variables and Internet experience**

Examining this literature one is struck by the fact that there has been little attempt to explore the stability of these results after controlling for a variety of other potentially important variables, including attitudes, skills, experience and confidence. There are exceptions. Hargittai and Walejko (2008) look at skills, finding that once skills are controlled gender becomes non-significant. They look at experience measured by years online, but it turns out to be non-significant. Correa (2010) looks at psychological variables, finding that intrinsic and extrinsic motivations as well as perceived competence predict content creation. Schradie (2011a) finds that being a parent is negatively associated with content creation in three of her 10 variables.

Hassani (2006) argues that users at work may be limited in their ability to explore a range of online activities because of the presence of bosses and possible surveillance technologies designed to monitor their Internet activity. Due to privacy and the absence of regulation home users have the greatest freedom of access, and hence autonomy. Indeed, she finds that locations of use are strong predictors of Internet use.<sup>1</sup>

Many of these variables are available in our dataset, so we can address them in some detail. With this summary of the existing literature, I turn to a description of the dataset and variables.

---

<sup>1</sup> Hassani (2006) does not study content creation. Her dataset is from the October 2003 Current Population Survey and her dependent variables are participating in online transaction activities and participating in online information search activities.

## Dataset and Measurement: Oxford Internet Survey

The Oxford Internet Survey (OxIS) collects data on British Internet users and non-users. Conducted biennially since 2003, the surveys are nationally representative random samples of more than 2,000 individuals aged 14 and older in England, Scotland, and Wales. Interviews were conducted face-to-face in each respondent's home by an independent survey research company. The analyses below use the 73% of the British population who were current Internet users in 2011,  $N = 1,498$ .

The 2011 OxIS dataset has eight variables measuring Internet content creation: writing a blog, posting "writing, stories, poetry or other work" the respondent considers creative, maintaining a personal website, having a profile on a social network site (SNS), uploading pictures, uploading video or music files, sending an email or message supporting a political or social cause, and commenting on a political or social issue in a blog, tweet, or on a social network site.<sup>2</sup> The eight items were deliberately selected to encompass a diverse range of content creation and to be consistent with the existing literature. OxIS is designed to produce results comparable to other surveys and all eight content variables in OxIS duplicate items used by Correa (2010), Hargittai and Walejko (2008) and Schradie (2011a). The variables are dichotomised, where 0 means the respondent reports never having done the activity.

Among the demographic variables, race is coded as white versus non-white. Place is coded as urban versus rural. Lifestage is a three-category variable: students, labour force (employed and unemployed respondents) and retired.<sup>3</sup> Marital status has five categories: single,

---

<sup>2</sup> The SNS variable is created by asking respondents if they have a profile on five SNSs: Facebook, Twitter, LinkedIn, "dating sites", or "other". If they answered "yes" to at least one of the five, they were coded as "yes" on the SNS variable.

<sup>3</sup> A preliminary analysis showed that employed and unemployed respondents did not differ in terms of reported content creation.

married, living with partner, divorced, widowed. We also include gender, education, age and income measured as total household income before tax.

Our data contain six categories of variables related to content creation: Experience on the Internet, technical ability, bad experiences, comfort revealing personal data, confidence on the Internet, and broad technology attitudes. We discuss each in turn. Experience on the Internet: Dutton & Shepherd (2006) and Blank & Dutton (2012) find that the Internet is an “experience technology”, meaning the more people are exposed to the Internet, the more they use it. OxIS contains items asking for number of years on the Internet and also self-rated ability, named “technical ability”. Experience also includes negative experiences. Bad experiences on the Internet could influence willingness to engage in content creation because they may remind people of possible negative effects of Internet use. OxIS asks about six possible bad experiences on the Internet: SPAM, viruses, misrepresented purchases, stolen identity, requests for bank details, and accidentally reaching a porn web site. Each variable is a yes-or-no, dichotomous variable. We summed these variables to produce a “bad experiences” index, with values ranging from 0 to 6.

Some content creation, like social networking sites and writing blogs, may require revealing personal details that could allow others to identify oneself. Some may wish to keep this private. The extent to which people see this as risky may influence their creation activities. Five items ask about comfort revealing personal information: Comfort revealing an email address, a postal address, a phone number, a date of birth or a name. A principal components analysis (PCA) indicated that these also formed a single factor with a Cronbach’s alpha of 0.88 so we again used the factor scores to create a measure called “personal data comfort”.

Confidence in ability to do things on the Internet can influence willingness to produce content, as Correa (2010) finds. Five OxIS variables ask about confidence: confidence participating in an online discussion, confidence making new friends online, confidence

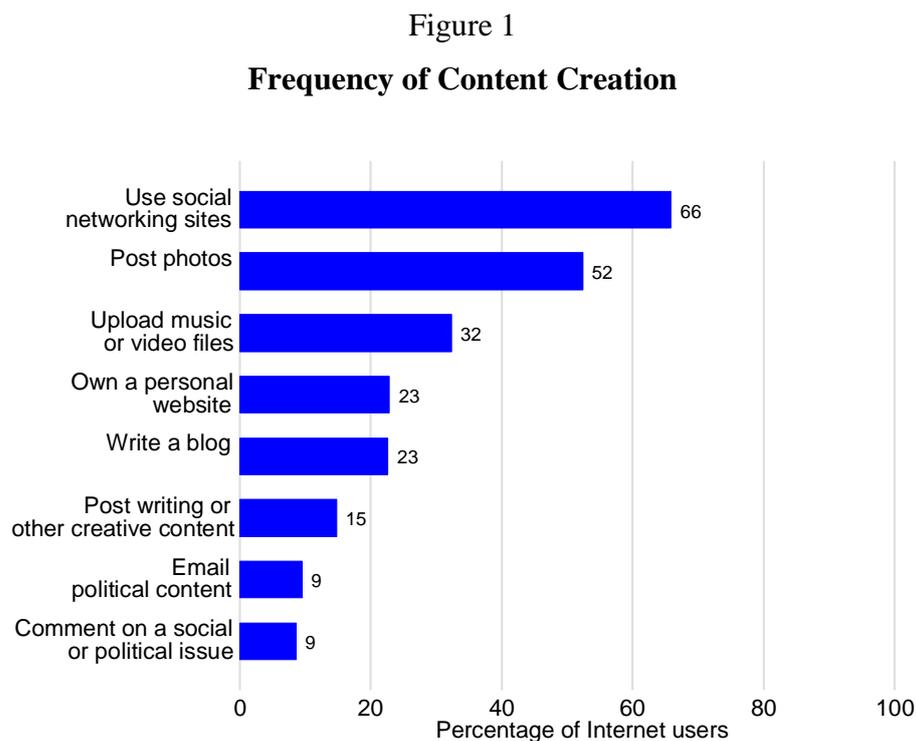
downloading music, confidence uploading photos and confidence in learning new technology. A PCA of these variables showed that they formed a single factor and Cronbach's alpha was 0.90. We summed these items to create a new continuous variable measuring "Web confidence".

Willingness to produce content can be influenced by users broad dispositions toward technology. These attitudes represent broad receptiveness to learning and using new technology. As general dispositions they represent the default point of view for people who are using the Internet. Their default point of view may be modified by their personal experiences or other factors. Nonetheless, when people are asked about technology these attitudes are the responses that they give 'off the top of their heads' (Zaller 1992). As the default perspective they influence the willingness of respondents to learn new aspects of a technology and the motivation to overcome problems. To this extent, they influence interest in content creation and motivation to learn to use them.

To measure general attitudes toward technology we again created indices with the aid of a PCA. Technology attitudes were composed of responses to likert-scaled items: openness to trying new technology, technology is making things better, plus three reverse-coded items; it is easier to do things without technology, lack of trust in technology, and nervousness around technology. The PCA had a single dominant component and the five-item index (called "technology attitudes") has a satisfactory Cronbach's alpha of 0.82. Note that none of the technology attitude items mention specific aspects of the Internet such as credit cards, product quality, trust, shopping, or accuracy of information; they are very general. The concept that this variable measures does not necessarily reflect any deep-seated conviction; instead we see it as general orientation or default point of view toward technology (Zaller 1992).

## Results

We begin by asking whether respondents participate in any content creation activities. About 79% of Internet users participate in at least one activity. Figure 1 shows the individual variables, sorted by frequency of participation. Participating on a social networking site is the most popular activity drawing 66% of users; the least popular are the two political content activities, engaging about 9% each. We speculate that social networking sites are popular because they are very easy ways to post content and communicate with friends. The political activities may be least popular due to the nature of the British politics, which is a more “professionalized” activity, involving few people and few volunteers compared to American politics.



One conclusion from Figure 1 is that there is a great deal of diversity in content creation. A key question is how diverse are the different kinds of content creation? There is some tendency to treat each type of creation as a separate category; Schradie (2011a), for example, runs separate models on each of her 10 content creation variables. This tends to make it harder to see patterns

shared across several variables. There is no reason to assume that there are actually 10 separate categories of creation, but then how many categories are there? To answer these questions about diversity of creation, I ran a PCA, see Table 1.

Table 1

**Content Creation Variables**

Factor loadings from Principal Components Analysis

Component name	Skilled Content	Social and Entertainment Content	Political Content
Write a blog	.62		
Maintain a personal website	.61		
Post writing, other creative content	.46		
Use social network sites		.67	
Post pictures		.58	
Upload video or music files		.42	
Email political content			.73
Comment on political/social issues			.68
Eigenvalues	1.93	1.84	1.44

Note: Factor loadings after varimax rotation and Kaiser normalization.

Factor loadings less than 0.15 have been omitted

OxIS 2011: N = 1,340 Internet users

The advantage of a PCA is that it categorizes the variables according to actual responses, so it provides an evidence-based set of categories. The notable result is that three clearly separate components emerge. Loading primarily on the first component are three variables: maintaining a personal website, writing a blog and posting writing, stories, poetry or other creative work. Common to these activities is that they require a certain level of technical skill, rhetorical skills and considerable personal commitment. Not everyone would have invested the effort required to master them so they are termed “Skilled content”. The second component contains visiting SNSs, posting photos, and uploading video or music files. These share a common theme of social ties and entertainment, so the component is named “Social and entertainment content”. The variables loading highest on the third component are the two political variables: sending emails or messages with political or social content and commenting on political or social issues

on web sites or discussion boards. We will call this “Political content”. Altogether these three components account for over 65% of the variance in the eight variables.<sup>4</sup> In the rest of the paper I refer to these three components as types of content creation.

One interesting result of the PCA is which variables load together. Looking first at skilled content, notice that writing, stories, poetry and other creative work loads with blogging and maintaining a personal website. This tells us something about what people do on their websites and what they write in their blogs: they tend to do various kinds of creative work. The loadings on social and entertainment content suggest that posting photos and uploading videos or music seem to be a major component of what people do on their SNSs. Finally, note that political content loads alone. It does not load with blogs, personal websites or SNSs. There is nothing in this pattern that is inherent in the technology of content production on blogs, websites or SNSs. Everyone knows of political blogs, political websites, and political content on SNSs. For that matter, photos, videos and music can have political content. The way these variables combine suggests that for most people, most of the time they spend on SNSs, blogs or personal websites is not on politically-related topics. This is consistent with the fact that relatively few people create online political content (see Figure 1). If this seems odd to academic readers, remember that political involvement is strongly influenced by education (see Table 3, below).

Using these results three new dependent variables were constructed. Each of the variables was constructed identically. Since this is the first study to identify types of content production it is necessarily exploratory and so I used the simplest possible method. Using the variables that loaded strongest on each component, a new dichotomous variable was constructed where zero

---

<sup>4</sup> To check whether Correa’s (2010) or Hargittai and Walejko’s (2008) results might be influenced by the fact that they study only students, I ran the same PCA on student Internet users only. The results are identical: three components with the same variables loading strongly on each component (table not shown).

indicates the respondent did none of the activities in that component and one indicates the respondent did at least one activity. This created three dependent variables for use in logistic regressions. Table 2 shows the percentage distributions of these variables. The distributions are quite different. Social and entertainment content is the most common, created by 75% of users. The least common is political content, produced by 13.7%.

Table 2

**Percentage distributions of Content Creation Variables**

	Yes	No	N
Skilled content	34.3	65.7	1,355
Social & entertainment content	75.0	25.0	1,359
Political content	13.7	86.3	1,351

The value of this typology is suggested by the fact that these variables show distinctions in their relationship to social status variables. Using two-way tables (not shown), income is significantly and positively related only to political content. Education is statistically significant and positively related to all three dependent variables.

**Predicting content production: Multivariate models**

We can capture the relative importance of the characteristics of content producers in multivariate models. Table 3 shows results from logistic regression models with two categories of variables, demographic variables as controls plus psychological and Internet-related variables.

Table 3  
**Predicting Content Production**  
 Odds Ratios from Logistic Regressions

	Skilled Content		Entertain. & Social Content		Political Content	
	Demog.	Psych. & Internet	Demog.	Psych. & Internet	Demog.	Psych. & Internet
Age	0.958***	0.974**	0.947***	0.970**	1.003	1.013
Female	0.868	1.191	0.943	1.347	0.818	1.044
Urban	1.294	1.401	1.218	1.131	1.013	1.046
Non-white	0.870	0.894	1.151	1.211	0.635	0.655

Income	1.028	0.957	0.949	0.856**	1.092	1.062
<i>Education</i>						
No degree	(base)	(base)	(base)	(base)	(base)	(base)
Secondary school	0.955	0.694	1.338	1.078	2.361	2.111
Further education	1.551	1.025	1.277	0.904	3.195	2.624
University degree	1.470	0.767	1.774*	1.093	7.953***	5.216**
<i>Lifestage</i>						
Student	(base)	(base)	#	#	(base)	(base)
Employed	1.650	1.901	0.947	0.762	0.395*	0.237**
Retired	2.007	2.789	0.795	0.641	0.349	0.396
Unemployed	1.174	1.378	(base)	(base)	0.446	0.439
<i>Marital Status</i>						
Single	(base)	(base)	(base)	(base)	(base)	(base)
Married	0.766	0.781	0.394**	0.359**	0.890	0.814
Living with partner	0.761	0.692	0.597	0.522	0.562	0.481*
Divorced or separated	0.622	0.497*	0.572	0.403*	0.952	0.859
<i>Internet variables</i>						
Use Internet at work		0.858		1.071		2.049*
N of devices		1.192*		1.081		1.040
Technical ability		1.812***		1.299*		1.337
Bad experiences		1.081		1.079		1.128
<i>Psychological</i>						
Technology attitudes		0.970		1.005		0.976
Personal data comfort		1.076**		1.022		1.009
Web confidence		1.099***		1.161***		1.045*
Constant	1.392	0.007***	50.053***	0.945	0.076**	0.010***
N	1053	1023	983	955	1053	1024
McFadden's R <sup>2</sup>	0.075	0.182	0.156	0.278	0.081	0.114
BIC	1471.	1331.	1109.	981.	938.	941.

Notes: \* p<.05; \*\* p<.01; \*\*\* p<.001

Omitted categories are: male, student, single, rural, white, no degree (similar to no high school diploma in the US school system).

# These variables were collinear with other variables in the model and were dropped.

We begin our interpretation with the simplest models: the three models using only demographic control variables. The only significant variable in the Skilled Content demographic model is age: content production declines with age. It is notable that nothing else is predictive. The R<sup>2</sup> is relatively small at 7.5%. For Social and Entertainment Content, production is reduced by being older and being married (married people have other things to do), while having a university degree increases production. The R<sup>2</sup> is considerably larger, 15.6%. Political Content production is dominated by two variables: being employed reduces production, while having a

university degree increases production. The impact of the university degree is dramatic with an odds ratio of almost 8!

Turning to the three models with psychological and Internet variables, in the Skilled Content model age remains significant. It is now joined by technical skill, comfort revealing personal data, confidence in ability and number of devices; all these variables increase the likelihood of production. Most of these seem related to skill: technical skill, obviously, but also confidence in abilities. A larger number of devices may be related to technical skills. Neither of the measures of status, income or education, are significant. These new variables more than double McFadden's  $R^2$  to 18.2%.

In the Social and Entertainment Content model, age and being married still significantly reduce production. Production is also reduced by being separated or divorced and by having a higher income. This is the only model where income is statistically significant, and higher incomes actually *reduce* content production. Among the psychological variables, confidence in ability and technical skills both increase production.  $R^2$  almost doubles to 27.8%.

Political Content production shows a different pattern than the other dependent variables. Being employed continues to reduce production while having a university degree increases production. Living with a partner significantly decreases creation. Among the new variables, use of the Internet at work and confidence in ability both increase creation. The McFadden's  $R^2$  increases by three percentage points to 11.4%.<sup>5</sup>

Overall, the status variables show an intriguing pattern: income is significant once: higher incomes *reduce* social and entertainment content creation. Once additional controls are

---

<sup>5</sup> BIC also increases in this model, suggesting that one cannot justify the increased model complexity due to the additional variables. The increase in BIC is not too worrisome, however, because it could be fixed by removing variables that were added but turned out to be not significant.

introduced in the form of psychological and structural variables, education is also significant only once, increasing political content creation. In the skilled content model, neither of the social status variables are significant.

## **Discussion**

This paper has three important, related empirical results: The disaggregation of content creation into types, the distinctiveness of political content and fact that the influence of social status on content creation depends on the type of content. I discuss each below in the context of the prior empirical results and theory.

The question that motivated this study is: does social stratification influence the diversity of content production? These results suggest that the answer is “it depends”. This is not an ambiguous answer—the results are quite clear—but how social stratification influences diversity of content depends on what kind of content we are talking about. For each type of content creation the answer is different.

On the one hand, skilled content creation is not influenced by either income or education. This suggests that content creation in the form of blogs, personal websites and posting writing, stories, poetry or other creative content is not the preserve of existing elites. These sorts of creation are not particularly dependent on being educated or having high income. Social status has a neutral effect on the creation of skilled content

For social and entertainment content having a high income actually reduces the likelihood of producing content. This is a remarkable result, because it says that social and entertainment content is more likely to be produced by non-elites. This result is consistent with a number of studies of content consumption: e.g. Buente & Robbin 2008; Bonfadelli 2002. This supports Benkler’s (2006) theoretical arguments about the value of increasing autonomy of creation.

To be fair to Benkler's argument about increasing autonomy, he is not so concerned with people uploading videos or writing comments on each other's walls on Facebook. His argument is more directed toward politics, and political content is much more likely to be produced by people with university degrees. Politics, at least in Britain, remains the preserve of the well-educated. It is important to emphasize again that there is actually relatively little political content: only 13.7% of Internet users produced any political content. This occurred despite the fact that the survey was in the field within a year of the 2010 general election, which should have generated more political content than in a non-election year. The item explicitly asked about creating political content *within the past year*, which encompassed the campaign, the election and the change of government that followed. This is in contrast to the pattern in the United States. Smith et al. (2009) surveyed a random sample of Americans during the 2008 presidential campaign and found that 19% of Internet users "posted material about political or social issues or a used a social networking site for some form of civic or political engagement" (p. 5). They also found that both income and education were strongly related to political engagement, saying that "the well-to-do and well-educated are more likely than those less well off to participate in online political activities" (p. 3).<sup>6</sup>

When I summarized prior work in the literature review above, I said that it shows "that the Internet tends to reinforce (or at least not challenge) existing patterns of stratification". Can these results be reconciled with prior work? I will discuss each major study in turn. Hargittai and Walejko (2008) deal with four content variables: music, poetry/fiction, film/video and artistic photography; we would classify all of them as social and entertainment content. Their

---

<sup>6</sup> Smith et al. (2009) do not create a multivariate model. The quotes reproduced in this paragraph describe relationships from two-way tables and similar graphics, which is the only analysis that they do. Schlozman et al. (2010), based on the same data, also only analyze two-way tables or equivalent graphics

conclusion is that “Students who have at least one parent with a graduate degree are significantly more likely to create content, either online or offline, than others” (p. 252). Their data are not as clear as this sentence. Although the text does not say so explicitly the sentence appears to refer to tables 3 and 4, which report “engage[ing] in creative activities” (p. 246): a *combination* of online and offline content creation. Online content creation alone is reported later in the paper in two tables in a separate section titled “Posting content online”. Looking at the social status variables in Table 6 (p. 250), which reports “the likelihood of posting content online”, only two of 25 percentages are statistically significant, and neither is a “graduate degree” variable. In Table 7 (p. 251), which reports a logistic regression on likelihood of posting content online (14 independent variables), *none* of the education coefficients are significant.<sup>7</sup> I suspect that most people would conclude that the multivariate logistic regression in Table 7 is the best result since it controls for the effect of other independent variables. This table shows that SES measured by parental education has no effect on online content creation. This is consistent with our result that education does not influence social and entertainment content creation.

Correa (2010) measures content creation with 10 items which include creating a blog, using a social networking site, uploading videos, uploading videos and posting community news, among others.<sup>8</sup> This is a mixture of skilled content and social and entertainment content. She combines the 10 items into a single scale, which, based on the results from the PCA reported in Table 1, we would not recommend. Regardless of whether they form a single scale or not, her measure of SES (parental education) is not significant. This is consistent with our result that

---

<sup>7</sup> Although Table 7 contains only respondents who create content, footnote 5 notes that “Results are robust when the analyses are performed on the entire sample.” (Hargittai & Walejko 2008:253). Logistic regression analyses on the full sample would contrast respondents who post online content vs those who do not, which is exactly what Table 3 does.

<sup>8</sup> Correa’s (2010) text on p. 77 says her scale includes 10 items, but the list in footnote 4 on p. 87 contains only 9 items.

education does not influence creation of either skilled content or social and entertainment content.

Schlozman et al. create an SES scale “based on education and family income” (2010:505). They report comparable data in two graphics using SES, one relating to political blogging and the other relating to social network use. These are graphical equivalents of two-way tables, so they are not multivariate models. Both show positive relationships, which is consistent with our results from two-way tables.

Kittilson & Dalton (2011) report a multivariate model predicting politically-oriented “Internet activity” in Table 4. This model shows a positive effect for education, which is consistent with our multivariate results for political content creation.

Schradie (2011a) uses 10 different creation variables and she reports results separately for each variable. Combining 17 PEW surveys from 2000-2008 into a single dataset, she summarizes the results from 10 logistic regressions in Table 3 (p. 157). She reports that that high school graduates are significantly less likely to produce online content than college graduates for six of her 10 content creation variables.<sup>9</sup> The six variables with significant results include variables that I classify as skilled content (e.g. personal website and share creative content) and social and entertainment content (e.g. social network site). The education results are not consistent with the British results. The summary table omits several independent variables and one of the variables omitted from the summary is her income variable. Looking at the original

---

<sup>9</sup> A puzzle is that a note at the bottom of Table 3 says it is based on Model 6 in the online supplement. Detailed Model 6 results are reported in 10 tables, Table G – Table P in the online supplement (2011b), where the coefficients for high school graduates are significantly different from college graduates for only one of the 10 content creation variables (Table P). In a personal communication, Schradie reports in a personal communication that the note is an error and Table 3 is based on Model 5 from the online supplement. Model 6 is more complex than Model 5, adding nine interaction terms. Both Model 5 and Model 6 are based on the combined 17-survey dataset. See the main text for reasons why the results based on the combined dataset may be an artifact.

tables, available in the online supplement (2011b), income is statistically significant in only one of the 10 models: social network sites, see Table P (2011b). This coefficient is negative, saying that people with higher incomes are less likely to frequent social network sites. The income results are generally consistent with the OxIS results, which are that income is significant and negative for social and entertainment content including social network sites.

The only discrepancy among prior studies lies in Schradie's results on education, and we suggest there are two possible reasons for this. The most straightforward reason is due to differences between the US and British education systems; for example, a lower proportion of the British population attend college and British higher education is much less diverse. These differences could well produce a different relationship between education and content creation.

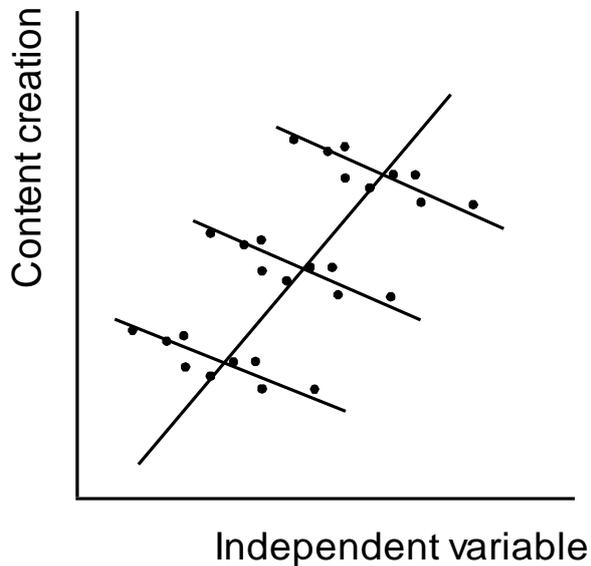
Second, Schradie combines 17 surveys across nine years into a single dataset. Far from being “replications”, as she says (p. 153), combining surveys from different years may produce different results due to a methodological artefact. During 2000-2008 “Internet time” was real. Internet use grew rapidly as new audiences with different interests in content creation joined the Internet. Content creation software improved quickly and content creation became much easier. New opportunities for content creation became available, like SNSs. In such a rapidly changing environment, combining surveys can produce coefficient estimates that are an artefact of differences between surveys.

The bivariate plot in Figure 2 illustrates the problem graphically. The vertical axis represents a type of Internet content creation. The horizontal axis can be any independent variable, perhaps income, technology attitudes or some other. The groups of points show the results of three surveys, each with a regression line. As is typical for the 2000-2008 period, creation of content is rising so each of the three surveys is higher than the previous survey. The example in the Figure shows that within each survey the coefficients—represented by the slopes of the short regression lines—are negative but, when we follow Schradie and combine the three surveys the

between-survey coefficient will be positive, as illustrated by the long regression line sloping upward. Of course, this example is drawn so that the within-survey coefficients are negative. If we keep the between-survey coefficient positive there are four additional possibilities for the within-survey coefficients. They could be (2) zero (meaning not statistically significant), (3) positive but not identical to the between-survey coefficient, (4) positive and identical to the between-survey coefficient—this is Schradie’s assumption. They could also be (5) inconsistent: some negative, some zero, and some positive. The between-survey coefficient may not be positive and the same five possibilities exist for a negative or non-significant between-survey coefficient. Schradie simply assumes without testing that the within-survey and between-survey coefficients do not differ in sign, size and significance. A stronger methodology and an appropriate test would be a multi-level model where the coefficients for each survey were allowed to vary. Without a multi-level model, her results may reflect differences across surveys rather than differences between respondents.

Figure 2

**Illustration of between-survey vs. within-survey coefficients**



In an overview of the field, Hargittai (2008:942) argues that “preliminary findings seem to suggest that ICT [sic] reinforce inequalities more than alleviating differences.” This conclusion is based on analysis of “skill-biased technological change” and not on content creation. Content creation seems to be different than technological change in general. The effect of content creation on stratification depends on what type of content you look at. Recent studies examine what I have called skilled content and social and entertainment content. This does not appear to have been a self-conscious decision, rather the authors simply overlooked any other kind of content. These two types of content are important but they are far from all.

This points to the second contribution, concerning political content: Although there is a prior multivariate study of political *participation* on the Internet (Kittilson & Dalton 2011), the current

paper appears to be the first multivariate study of political content *creation*.<sup>10</sup> Previous authors seem to have overlooked political content (Correa 2010; Hargittai & Walejko 2008) or not had political variables available (Schradié 2011a). One of the messages of this study is that political content is a different, distinctive form of content. The people who create it are a distinctive subset of the population: they are a relatively small but highly educated group; in Britain about 14% of the population. Much more research can be done on political content creation.

This brings us to the third issue: types of content. There are empirical reasons why establishing types of personal publishing is valuable. Prior work has tended to treat content creation as either “all the same” or “all different”. Proponents who assume that content is “all the same” have measured it by combining all available measures of content into a single scale (e.g. Correa 2010). This approach is clearly contradicted by the PCA results and the logistic regressions (Table 3), which show that there are important differences between types of creation. Proponents who assume that content is all different display a separate model for each and every measure content (e.g. Schradié 2011a). This extreme disaggregation makes it hard to find patterns and this is one reason that Schradié’s results are ambiguous. The payoff from constructing types is clear in the logistic regression results: how elite status influences content creation depends on the type of content.

It should go without saying that this is a cross-sectional analysis. It offers no information about whether this pattern has been stable across the 10 years of OxIS surveys. I note that the OxIS surveys have shown substantial consistency since the first survey in 2003, see Dutton and

---

<sup>10</sup> Schradié (2011a) would like to study “digital democracy” (p. 145) and the “digital public sphere” (p. 165) but, in the reproduced the question texts (2011b, Table A), not a single item contains words like “politics”, “election”, “social issues”, “democracy” or synonyms. No item explicitly measures political content. One message of the present study is that research can’t generalize from variables which are not explicitly political to political content. See Boulianne (2009) for a summary of research on political engagement, which is related to but not the same as content creation.

Blank (2011) for a detailed discussion of OxIS results across time. This suggests that the pattern described in this paper is not a sampling error artefact. It is entirely possible that this pattern emerged only as the Internet has attracted more users and become more important over the past decade. If so, attempts to replicate these results could lead to important findings.

The argument is based on British data, and it probably goes without saying that the Internet may not be the same in Germany, Brazil, China or America. This study should be replicated in other countries at other times, with a wider range of types of content. How different content creation combines into types is an important question. The three types that appear in Britain—skilled content, social and entertainment content, and political content—may not appear elsewhere. But whether or not these types described prove stable across time and countries is not the most important issue. The important point is that the Internet is not always and everywhere the same and it should not be treated as if it were. Theories must become more sophisticated.

The distinction between types of content allows a far more nuanced understanding of the effect of social status. In general, researchers have paid little attention to the whole issue of types of content. Even as sophisticated a theorist as Benkler (2006) often treats the Internet as if it had a single effect. These results suggest that this is a major theoretical oversight. It is increasingly inaccurate to assume that any part of the Internet, including personal publishing, has the same effect everywhere. There is a need for more sophisticated theories that disaggregate the effects of the Internet into categories.

Looking more closely at the categories that emerged in this paper shows the potential payoff. The types of content are more than a dependent variable. The types describe the fundamental characteristics of each instance of content, but they are more. The message of this paper is that content type is a bundle of characteristics. It is important to look at what a type contains. Each type of content is not just an outcome it is also created by its own unique production process: the characteristics of the creators of each type are different.

- Skilled content is produced by young, technically skilled people who use many devices and are comfortable revealing personal information. Social status is irrelevant.
- Social and entertainment content is produced by young, technically skilled people who are not married and have lower incomes. It is more likely to be created by non-elites.
- Political content is produced by well-educated people who are either students or use the Internet at work. They are highly educated elites.

Thus, type is not just content. The creators and the content are bound together in each type. Social status effects emerge from this complex mixture of inputs and outputs. Creation and output both matter. In this respect personal publishing is a more complex system than institutional publishing. Content creation is actually more diverse. But it is not simply diverse, it is diverse in a very structured, specific way. The structure can be traced to specific social groups who produce each type of content.

These results speak to the question we started with, is personal publishing democratising? It is well known that people from more privileged backgrounds are more likely to engage in capital enhancing activity (Bonfadelli 2002; Buente & Robbin 2008; Eynon 2009; Hale et al 2010; Zillien & Hargittai 2004). “Capital enhancing” activity generally includes interest in product or price comparison and economic or political news. Since we find political content more likely to be produced by highly educated people, our results are consistent. Personal publishing does not democratize political content creation, in this domain it reinforces existing stratification.

Skilled content is defined by how it is produced—e.g. on a personal website—rather than the actual content of the product. Thus it is not relevant to the question.

Social and entertainment content is clearly relevant. Usually social activity and entertainment are more common among low social status groups. This is especially true for popular entertainment like TV (Greenberg & Devin 1970), so these results are again consistent with prior work. But are there conditions under which the creation of social and entertainment content

could be capital enhancing? Consider the following: Social network sites include LinkedIn, which is all about networking for employment. Even use of Facebook or Twitter can expand one's social network (that is, increase bridging social capital, e.g. Steinfield et al. 2008). This can easily be capital enhancing. Through this mechanism, social and entertainment content creation can enhance social and economic mobility opportunities. To the extent that this mechanism operates existing stratification could be undermined. Whether this leads to democratisation in the public sphere is a harder question to answer. It indicates that a broader range of people is involved in creation of personalised content. Social and (particularly) entertainment content are not generally considered part of the public sphere. They certainly aren't directly related to the public sphere. However, there is an indirect relation that could be important. If involvement in social content creation expands an individual's social network and enhances their mobility chances, then this could introduce people from unusual backgrounds into elite society, and expand the range of voices available in the public sphere.

This argument goes well beyond the data available from this survey. The items in OxIS that we are using as indicators are at best only indirect evidence, and there are omissions (like no item about contributing economic or financial content), but they provide some support for the idea that content creation does not always reinforce existing stratification. Support for this argument requires survey data more directly focussed on the question of content creation and social mobility, and also qualitative data about the careers of individual content creators. The general question of the effect of the Internet on social status would benefit from career-oriented data of all kinds. I am unaware of any study of content creation that has attempted to trace careers. This is a big gap. Hargittai & Hsieh's (in press, p. 142) review article summarizes the issue: "We know especially little about the consequences of differentiated Internet uses for people's social status." This is a basic social mobility question and it requires longitudinal data. The whole issue deserves more research attention.

## Sources cited

- Benkler, Y. (2006) *The wealth of networks: How social production transforms markets and freedom*. New Haven: Yale University Press.
- Blank, G. & Dutton, W. H. (2011) 'Age and trust in the Internet: The centrality of experience and attitudes toward technology in Britain', *Social Science Computer Review*, vol. 28.
- Bonfadelli, H. (2002) 'The Internet and knowledge gaps: A theoretical and empirical investigation', *European Journal of Communication*, vol. 17, no. 1, pp. 65-84.
- Boulianne, S. (2009) 'Does Internet use affect engagement? A meta-analysis of research', *Political Communication*, vol. 26, no. 2, pp. 193-211.
- Buente, W. & Robbin, A. (2008) 'Trends in Internet information behavior, 2000-2004', *Journal of the American Society for Information Science and Technology*, vol. 59, no. 11, pp. 1743-1760.
- Correa, T. (2010) 'The Participation Divide Among "Online Experts": Experience, Skills and Psychological Factors as Predictors of College Students' Web Content Creation', *Journal of Computer-Mediated Communication*, vol. 16, pp. 1-92.
- Dutton, W. H. & Blank, G. (2011) *Next generation users: The Internet in Britain. Oxford Internet Survey 2011 Report*. Oxford: Oxford Internet Institute.
- Dutton, W. H., Shepherd, A. (2006) 'Trust in the internet as an experience technology', *Information, Communication and Society*, vol. 9, no. 4, pp. 433-51 .
- Greenberg, B. & Dervin, B. (1970) 'Mass Communication among the Urban Poor', *Public Opinion Quarterly*, vol. 34, no. 2, pp. 224-235.
- Hargittai, E. (2002) 'Second level digital divide: Differences in people's online skills', *First Monday*. 7(4). Retrieved October 31, 2011 from [http://www.firstmonday.org/Issues/issue7\\_4/hargittai/](http://www.firstmonday.org/Issues/issue7_4/hargittai/)
- Hargittai, E. (2008) 'Digital reproduction of inequality', pp. 936-944 in David Grusky (ed.) *Social Stratification*. Boulder CO: Westview Press.
- Hargittai, E. & Hsieh, Y. P. (In press) 'Digital inequality', pp. 129-150 in William H. Dutton (ed.) *Oxford handbook of Internet studies*. Oxford: Oxford University Press.
- Hargittai, E. & Walejko, G. (2008) 'The participation divide: Content creation and sharing in the digital age', *Information, Communication & Society*, vol. 11, no. 2, pp. 239-256.
- Hassani, S. N. (2006) 'Locating digital divides at home, work, and everywhere else', *Poetics*, vol. 34, pp. 250-272.
- Hoffman, D. L. & Novak, T. P. (1998) 'Bridging the racial divide on the Internet', *Science*, vol. 280, no. 5362, pp. 290-291.
- Howard, P. E. N., Rainie, L. & Jones, S. (2001) 'Days and Nights on the Internet: The Impact of a Diffusing Technology', *American Behavioral Scientist*, vol. 45, no. 3, pp. 383-404.
- Jenkins, H., Purushotma, R., Clinton, K., Weigel, M. & Robison, A. J. (2006) *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century*. Chicago: John D. and Catherine T. MacArthur Foundation.
- Katz, J. & Aspden, P. (1997) 'Motivations for and barriers to Internet usage: Results of a national public opinion survey' *Internet Research: Electronic Networking Applications and Policy*, vol. 7, no. 3, pp. 170-188.
- Kittilson, M. C. & Dalton, R. J. (2011) 'Virtual civil society: The new frontier of social capital?', *Political Behavior*, vol. 33, pp. 625-644.
- Loges, W. E. & Jung, J-Y. (2001) 'Exploring the Digital Divide : Internet Connectedness and Age', *Communication Research*, vol. 28, no. 4, pp. 536-562.
- Nardi, B. A., Schiano, D. J., Gumbrecht, M. & Swartz, L. (2004) 'Why we blog', *Communications of the ACM*, vol. 47, no. 12, pp. 41-46.

- Steinfeld, C., Ellison, N. B. & Lampe, C. (2008) 'Social Capital, Self-Esteem, and Use of Online Social Network Sites: A Longitudinal Analysis', *Journal of Applied Developmental Psychology*, vol. 29, no. 6, pp. 434-445.
- Schlozman, K. L., Verba, S. & Brady, H. E. (2010) 'Weapon of the Strong? Participatory Inequality and the Internet', *Perspectives on Politics*, vol. 8, no. 2, pp. 487-509.
- Schradie, J. (2011a) 'The digital production gap: The digital divide and Web 2.0 collide', *Poetics*, vol. 39, pp. 145-168.
- Schradie, J. (2011b) 'Online supplement' *Poetics*, vol. 39. Retrieved 15 September 2011 from <http://dx.doi.org/10.1016/j.poetic.2011.02.003>.
- Shirky, C. (2008) *Here comes everybody: How change happens when people come together*. New York: Penguin.
- Smith, A., Schlozman, K. L., Verba, S. & Brady, H. (2009) *The Internet and civic engagement*. Washington DC: Pew Internet & American Life Project. Retrieved November 1, 2011 from [http://pewinternet.com/~media/Files/Reports/2009/The Internet and Civic Engagement.pdf](http://pewinternet.com/~media/Files/Reports/2009/The%20Internet%20and%20Civic%20Engagement.pdf)
- van Dijk, J. A.G.M. (2006) 'Digital divide research, achievements and shortcomings', *Poetics*, vol 34, pp. 221-235.
- Zaller, J. (1992) *The nature and origins of mass opinion*, Cambridge: Cambridge University Press.
- Zillien, N. & Hargittai, E. 2009. 'Digital distinction: Status-specific types of Internet use', *Social Science Quarterly*, vol. 90, no. 2, pp. 274-291.