FINAL EXAM APPLICATION

Date, ______16/04/2019_____________

PhD ADMINISTRATIVE CENTRE (CADES)
UNIVERSITA' BOCCONI
Piazza Sraffa, 11
20136 MILANO

I, the undersigned

Family name	YILMAZ
Name	ERDEM DOĞUKAN
Student ID no.	1824390
Place of birth (Town/City)	SARIYER / ISTANBUL
Province (only for Italians)	
Country	TURKEY
Date of birth (day/month/year)	0 4 0 8 1 9 9 0
Permanent home address	ORTAKLAR CADDESI GORUNTU SOK DORTEL AP6 6/3
Post code	34234
Town / City	MECIDIYEKOY/ISTANBUL
Province (only for Italians)	
Country	TURKEY
Tel. no.	
Mobile no.	+90 5467214515
Personal e-mail address:	erdemdogukanyilmaz@gmail.com
Enrolled in Cycle	30

of the PhD Program in	Business Administration and Management

REQUEST

to take part in the PhD final exam.

I DECLARE

a) that the **final** title of my thesis is (**the title indicated in this form, including its punctuation and capitalization, has to be written on the thesis in an identical manner; please, do not write in capital letters**)

Three Essays on Digital Innovation

b) that my Advisor is Professor:

Family name	GAMBARDELLA	Name	ALFONSO

c) that my Co-Advisor - if any- is Professor:

Family name		Name	

d) that my PhD program curriculum is as follows (this section is only to be used by PhD students for **Business Administration and Management, Economics and Finance, Legal Studies**):

Management

e) that I have read the “Instructions and deadlines concerning thesis submission and defence applying to all PhD programs starting from and including the 29th Cycle” published on the Bocconi site;

f) that I will send my final thesis, following the online procedure, to Società NORMADEC (http://www.tesixpress.it/bocconi/index.asp) within the stated deadline.

Handwritten Signature

Stamp duty prepaid, as authorised by Inland Revenue Service – Lombardy District Office - no. 3/00042/2001, 26.3.2001

PhD THESIS DECLARATION

(use a pc to fill it out)

I, the undersigned

FAMILY NAME	YILMAZ
NAME	ERDEM DOĞUKAN
Student ID no.	1824390

Thesis title:

Three Essays on Digital Innovation

PhD in	Business Administration and Management
Cycle	30
Student's Advisor	Alfonso Gambardella
Calendar year of thesis defence	2019

DECLARE

under my responsibility:

1) that, according to Italian Republic Presidential Decree no. 445, 28th December 2000, mendacious declarations, falsifying records and the use of false records are punishable under the Italian penal code and related special laws. Should any of the above prove true, all benefits included in this declaration and those of the temporary “embargo” are automatically forfeited from the beginning;

2) that the University has the obligation, according to art. 6, par. 11, Ministerial Decree no. 224, 30th April 1999, to keep a copy of the thesis on deposit at the “Biblioteche Nazionali Centrali” (Italian National Libraries) in Rome and Florence, where consultation will be permitted, unless there is a temporary “embargo” protecting the rights of external bodies and the industrial/commercial exploitation of the thesis;

3) that the Bocconi Library will file the thesis in its “Archivio Istituzionale ad Accesso Aperto” (Institutional Registry) which permits online consultation of the complete text (except in cases of temporary “embargo”);

4) that, in order to file the thesis at the Bocconi Library, the University requires that the thesis be submitted online by the student in unalterable format to Società NORMADEC (acting on behalf of the University), and that NORMADEC will indicate in each footnote the following information:

- PhD thesis (*write thesis title*) … Three Essays on Digital Innovation……………………………………………………………………………
  ……… ;
- by (S*tudent’s family name and name)* …………..Yılmaz Erdem Doğukan ……………………………………………… ;
- defended at Università Commerciale “Luigi Bocconi” – Milano in the year (year of defence) 2019 ;
- the thesis is protected by the regulations governing copyright (Italian law no. 633, 22nd April 1941 and subsequent modifications). The exception is the right of Università Commerciale “Luigi Bocconi” to reproduce the same, quoting the source, for research and teaching purposes;

- **<u>only</u> when a separate “Embargo” Request has been undersigned**: the thesis is subject to “embargo” for (indicate duration of the “embargo”) ……… months;

5) that the copy of the thesis submitted online to Normadec is identical to the copies handed in/sent to the members of the Thesis Board and to any other paper or digital copy deposited at the University offices, and, as a consequence, the University is absolved from any responsibility regarding errors, inaccuracy or omissions in the contents of the thesis;

6) that the contents and organization of the thesis is an original work carried out by the undersigned and does not in any way compromise the rights of third parties (Italian law, no. 633, 22nd April 1941 and subsequent integrations and modifications), including those regarding security of personal details; therefore the University is in any case absolved from any responsibility whatsoever, civil, administrative or penal, and shall be exempt from any requests or claims from third parties;

7) **choose 7a <u>or</u> 7b**:

7a) that the thesis is not subject to "embargo", i.e. that it is not the result of work included in the regulations governing industrial property; it was not written as part of a project financed by public or private bodies with restrictions on the diffusion of the results; is not subject to patent or protection registrations.

Date _16/04/2019_______________

Abstract

The first essay describes how upstream works are affected from the derivative works in the digital world. To evaluate this question, I focus on the release of derivative designs in a 3D printable design community, called as Thingiverse. I find that while the release of derivative designs substantially increases the demand for the designs in general, this effect is relatively more pronounced for the artistic designs. These results suggest that derivative works can create benefits for the original works rather than hurting their success when intellectual property rights are less restrictive.

The second essay highlights when derivative designs that are results of remixing practices are more valuable for open communities. The previous literature has confirmed the role of remixing as the driver of the generativity in open communities. At the same time, however, the previous literature has argued that remixing might inhibit creativity and not necessarily beneficial for fostering innovation. Therefore, it is important to determine when derivative products that are result of remixing are valuable for the community. By bringing insights from innovation management literature, first this essay argues and shows that derivative designs, on average, create higher value when they bring together disparate products. Second, it also illustrates that designs create more value when contributions to designs are gathered over time from different users.

The third essay examines the impact of contributor reputation and reciprocity on contributions being designated for reuse. This essay argues that users with a higher reputation level have lower expected returns from allowing other users to reuse their work and are therefore less likely to give up some of their intellectual property rights. On the other hand, a high level of direct reciprocity increases users' feeling of indebtedness and pushes them to allow others to reuse their work. This essay sheds light on the dynamics of user contribution in open communities with fine grained data.

Acknowledgements

First and foremost, I would like to thank my advisor, Alfonso Gambardella. I could not have asked for more from a PhD advisor. You have guided my research and demonstrated to me the joy of solving problems. You have never failed to motivate or inspire me, and have never disappointed me in your support. I am fortunate and grateful for all your support over the years. Your strong dedication to your own work and great personality will inspire me for years to come.

Next, I would like to thank the other members of my thesis committee: Andrea Fosfuri, Cedric Gutierrez Moreno, and Chiara Farronato. I look up to them as exemplary academics that have inspired me throughout my Ph.D. journey. I would also like to thank the other professors at Bocconi who have guided me over the years, including Paola Cillo and Thorsten Grohsjean.

I would then like to thank my friends, for all the great conversations, for everything you have taught me and introduced me to over the years. Your presence in my life here made it wholesome and meaningful, so thank you Tim, Doğan, Burak, Isabella, Sepideh, Paola, Shanming, Chiara, Tatiana and Varun.

To my grandparents, aunts, uncles and cousins: Thank you for your all the care and affection to me throughout my life. To Vera: thank you for inspiring me and delighting me every day. And to the three anchors of my life; my parents and my little sister. You are the foundation on which I build everything, knowing that even if it all collapses you will always be there for me. I would like to dedicate this thesis to you.

THE BENEFITS OF REUSE:

WHEN DERIVATIVE WORK IMPROVES THE PERFORMANCE OF PARENT WORK

Erdem Dogukan Yilmaz

erdem.yilmaz@unibocconi.it

Department of Management & Technology, Bocconi University

Abstract

Advances in digital technology have considerably lowered the cost of replication and increased incentives for reuse of knowledge. Subsequently, one of the most important research questions has become how innovation should be governed in the digital era, in particular whether intellectual property rights should be strengthened or loosened. In this paper, I examine cumulative creativity and reuse through the release of derivative works on Thingiverse, an online 3D printing design community that is governed by creative commons licenses. Creative commons is an alternative innovation governance mechanism that allows individuals to freely use and build upon creations of others, as long as they give attribution to the original work. Using a matched sample of 122,774 observations for 1,338 designs and their weekly downloads during 2015-2016, I use a difference-in-differences estimator and find that the release of a derivative design increases demand for the original design, on average, by 13.2 %. However, the release of derivative designs does not impact all designs equally. Artistic designs are more likely to benefit from reuse compared to functional designs. These results suggest that, under certain circumstances, derivative works can create benefits for the original works rather than hurting their success when intellectual property rights are less restrictive.

1. Introduction

In the last couple of decades, we have witnessed how the internet has been reducing the cost of transmission and reuse of digital content. With these reductions in cost, both the creation and the distribution of content have become disaggregated. Today, not just professionals but also amateurs participate in the creation of new content and use the content of others in their own creations. Wikipedia users, for instance, tweak existing articles, while computer programmers on GitHub fork existing codes and artists remix others' musical pieces. Then, this content is distributed through a variety of intermediaries such as search engines and creative content platforms. Thus, this new way of disaggregated creation increases the variety and in many cases improves the usefulness of existing content and therefore fosters innovation.

However, this disaggregated creation might come at a cost: The ease of reusing might increase the creation of content without authorization. Furthermore, the increasing number of derivative works – works that build upon preceding works, i.e. they might hurt the property of first-generation creators as they might serve as a substitute. Consequently, this might reduce their incentives to invest in innovation. At the core of examining the reuse process for cumulative innovation, therefore, lies the understanding of whether reuse hurts the property of original creators or not and, if not, how much one can obtain from allowing others to use their own creations. Despite the relevance of reuse in the digital age, we know relatively little about the interaction between upstream and derivative works. Answering the question of whether derivative work hurts upstream work is important as it is likely to contribute to the ongoing debate on whether property rights should be strengthened or loosened in the digital era. On the one hand, content industries have persistently pressured for stronger property rights (LaPolt et al., 2014) as they claim piracy and derivative works are hurting their products' market success. On the other hand, many scholars have raised concerns about extensive property rights as they harm cumulative innovation by shrinking the public domain (Samuelson 2008, Lessig 2004). The examination of this research question is becoming even more important with the emergence of three-dimensional (3D) printing technology as the holders of utility and design patents might soon face a similar situation as companies in content industries.

In this paper, I study the effect of the release of derivative works on the demand for upstream work in a community of 3D printable designs called “Thingiverse”. Every day, hundreds of new products are introduced to the community and the community is governed with Creative Commons licenses. One of the most important features of these licenses is that they require downstream creators to give attribute to upstream work(s) that they built upon if it is required by the original creator. Thus, Creative Commons licenses enable the creators of upstream work to get credit from their derivative products. In this paper, I argue that, given this citation-like mechanism, the release of derivative works might lead to a discovery of upstream works and hence might increase the demand for them. This effect is, however, likely to depend on whether downstream work horizontally or vertically differentiates itself from the upstream work. If downstream work horizontally differentiates itself, a fraction of people discovering the upstream work through the downstream work might end up downloading and eventually 3D printing the upstream work. On the other hand, if the derivative work serves for a certain functionality and improves upstream work (i.e. vertically differentiates itself and offers a higher value proposition compared to upstream work), it is more like to serve as a substitute. By using the panel data of 669 treatment and control design pairs, I shed light on this theoretical question. I begin my analyses by conducting a difference-in-differences analysis of reuse on the demand for upstream work. I measure the demand with the number of downloads. The download measure captures the demand in a similar manner as other digital creative content industries, such as mobile phone applications, music and film streaming where the consumption of the goods entails very low or no cost. The identifying assumption in the difference-in-differences analysis rests on parallel trends in the outcome in the pre-treatment period, which suggests that in the absence of treatment the trend in the outcome among the treated and control groups would have been the same. However, in the full sample, I observe a divergent trend among the treated and control observations before the treatment, and therefore I use coarsened exact matching (CEM) procedure to ensure a balance between the treatment and control groups. By using a matched sample, I find that reuse triggers a 13.2 % percent increase in downloads.

To account for the potential substitution effect, I check whether the effect of reuse diverges among functional and artistic designs. The idea is that derivatives of functional designs are more likely to constitute a substitute compared to artistic designs. This is because functional designs are meant to serve for a pre-defined purpose and therefore their value is more objective, whereas the appeal of artistic designs is likely to differ from consumers to consumers, given that tastes for aesthetic products are likely to be heterogeneous (Krippendorff and Butter 1993). To classify designs as functional or artistic, I use category information provided by Thingiverse. As expected, I find the positive effect of reuse on the demand to be, on average, smaller for functional designs. Given that reuse appears to increase the demand for the upstream work, I turn my attention to check whether this additional demand is due to demand spillovers from the downstream work. I observe a positive correlation between the page impressions of the downstream work and the number of downloads of the upstream work. In particular, I find that the number of downloads of an upstream work in a given week increases with an increasing number of views of the derivative design in the same week. While I do not have any information about which users download which designs, this moderation effect provides some evidence that a fraction of people visiting derivative work also downloads the upstream work.

This paper makes four contributions to the literature. First, departing from prior research that estimates the impact of changes in copyright law (Danaher et al. 2014, Nagaraj, 2017), this paper provides the first empirical evidence on the relationship between reuse and performance of parent work in a setting where innovation is governed by Creative Commons licenses. I show that upstream innovators might benefit from downstream reuse when intellectual property rights are more permissive. This implication is likely to reach beyond the setting of this study as this study also indirectly test the effectiveness of creative commons as an alternative innovation governance system. In a variety of settings including books, movies and video games, the effectiveness and moral issues of digital rights management (DRM) are highly debated. At the end, if the rational for IP is to encourage people to create new things and increase the consumption rather than increasing the prices, weakening the IP might prove to be beneficial in creative industries. Therefore, this study might have important implications for policymakers in those settings by showing that, under certain circumstances, loosening IP is beneficial for the first generation of innovators.

Second, this study contributes to the open community literature. Most of the studies in open communities focus on why users reuse others' innovations (e.g. Haefliger et al. 2008, Sojer & Henkel, 2010) and why some innovations are reused while others are not (Hill & Monroy-Hernández, 2013; Kyriakou et al., 2017; Stanko 2016). Diverging from these studies, I focus on the question of how reuse affects the performance of upstream work in a setting in which innovation is governed by Creative Commons licenses. One implication of these results might be that one way in which user contributions can be spurred is by strengthening the attribution mechanism which is at the core of Creative Commons licenses.

This study also contributes to the literature on open innovation. Previous studies have emphasized the importance of an open community to make users' contributions visible and identifiable (Ma and Agarwal, 2007), so that they will maintain their contributions. One way to reinforce visibility of user contributions and encourage participation might be to strengthen the attribution mechanism. Ensuring users of getting proper attributes for their contribution might help the community to satisfy users' needs for recognition and reward.

By showing that downstream reuse actually promotes upstream work, this research also contributes to the literature on cumulative creativity. In this literature, the papers by Murray and Stern (2007) and Williams (2013) provide one of the first causal evidences that intellectual property rights hinder follow-on research. Similarly, the study by Nagaraj (2017) indicates that digital content is more likely to be re-used on Wikipedia when it has lapsed into the public domain. The findings of the study further indicate that copyrighted pages receive 20 percent less readership. In this study, however, rather than focusing on whether property rights block further innovations or whether reused content creates value, I focus on the relationship between derivative and parent work, similar to the paper by Watson (2017a). While my findings have some similarities with the findings by Watson (2017a) there are four main differences between this study and his study that might also explain why our results differ in magnitude. First, we study this question in 3D printable design settings whereas he studies a similar question in the music industry. Second, while his study provides evidence at the artist level, I provide evidence at the product level. Third, in my setting, individuals have more freedom to choose which products they will reuse and downstream re-users are obliged to give explicit reference to the upstream product. Finally, I extend his work by showing that these results might not be generalizable to settings where products are functional rather than artistic.

2. Theoretical Background

In this section, I use insights from the literature to motivate the relationship between reuse, innovation incentives and the demand.

Cumulativeness of Innovation, Knowledge Reuse and Innovation Governance

In the literature, innovation is frequently conceptualized as an evolutionary search through which different knowledge elements are recombined (e.g. Fleming 2001; Nelson & Winter 1982; Kaplan & Vakili, 2015). As posited by Nelson and Winter (1982), "the creation of any sort of novelty in art, science, or practical life— consists to a substantial extent of a recombination of conceptual and physical materials that were previously in existence". Schumpeter (1942) was one of the first scholars that emphasized the importance of the cumulative aspect of innovation. He argues that, with increasing innovation, economy-wide returns to innovation also increase as inventors rely and build on the previous ideas of others when they create their own innovations. From academic science to industrial innovation and open innovation systems, previous knowledge is always incorporated in newer generations of innovations (Boudreau & Lakhani, 2015). For instance, by analyzing US patent data and patent technology codes, Strumsky and Lobo (2015) find that inventions are rarely created from scratch but that they are rather combinations of existing innovations.

With the advent of digitization, it has become easier to share and get access to innovations created by others (Kane & Ransbotham, 2012). The term ''remixing'' has been increasingly used to describe the use of existing materials in the creation of something new. The concept of remixing is very well established in the music domain (Flath et al., 2017) and it has been subject to a controversy from the very beginning of its emergence. Managers of record companies believe that any unauthorized extraction of pre-existing content constitutes a copyright infringement. More recently, the debate about remixing has evolved from whether music producers would allow remixing to how to handle the increasing popularity of remixes (Monroy-Hernindez, 2012), and whether remixes create value (Sinnreich et al., 2009). Sinnreich et al. (2009), for instance, have argued that remixing restricts creativity by enabling people to create perfect copies and that derivative products have little value for their creators and society at large. However, there is still a lack of research concerning the effect of remixing on the upstream work and its relevance for cumulative knowledge creation.

Innovation is governed by a variety of frameworks including the patent system, intellectual property rights, academic science, open source software and Creative Commons licenses. Each system has its own rules for knowledge disclosure, access, and reuse (Boudreau & Lakhani, 2015). In the private sphere, innovation is governed by patents and intellectual property rights. Copyright is more frequently used in creative industries whereas patents are more likely to be used in research and development intensive industries such as chemicals, pharmaceuticals, and manufacturing (Boudreau et al. 2018). Copyrights provide slightly narrower IP rights that might avoid inefficiencies generated by broad patent rights. Yet, copyright has been around for quite long, around 95 years in the United States and Europe (Li, MacGarvie & Moser, 2018). In the public sphere, norms blended with different types of licenses and organizational structures govern innovation (Murray & O'Mahony, 2007). Regardless of their differences, innovation systems mainly aim at to boost cumulative knowledge creation by protecting the incentives of the first generation of innovators and at the same time make sure that they disclose detailed information to the second generation of innovators (Nelson 1959; Arrow 1962; Murray & O'Mahony, 2007). Therefore, disclosure is the first step that initiates the creation of cumulative innovation (Dasgupta & David 1994). Both private and public institutional frameworks make knowledge available to all by mandating innovators to disclose their information.

The second condition for cumulative innovation is access, which constitutes the first point of departure between private and public frameworks. The patent system, for instance, provides the right to exclude others from using and reusing innovations for a predefined term in exchange for sharing the details of innovation. However, these patented innovations can be licensed to third parties under certain conditions. Similarly, copyright restricts follow-on innovators from reuse, but the duration of rights is quite long for the content that is protected by copyrights. In recent years, copyright has been strengthened even more by the Digital Millennium Copyright Act, which reinforces the barrier for bypassing technical barriers through content re-engineering (Samuelson, 2001).

Most observers however agree that the effectiveness of copyright has been significantly reduced in the digital world. Many companies have adopted DRM systems to maintain control over digital content and devices after their sale. These companies include hardware manufacturers, publishers, copyright holders, and individuals (Zhang, 2016). However, it has been shown that DRM may actually increase piracy as it restricts the usability of digital content and in turn reduces consumers' willingness to pay for content (Sinha et al., 2010). Furthermore, DRM technologies have received strong criticism from both the public and scholars. For example, community members of creative commons have raised a global campaign against DRM and they consider DRM a misuse of copyright law by controlling the use of legally acquired content. Similarly, Belleflamme and Peitz (2014), have argued that DRM technologies go beyond preventing copyright infringement and restrict users from engaging in perfectly legal activities. Consumers should be able to make backup copies, share digital content with friends and family, access work in the public domain and use the copyrighted material for research purposes. DRM has also been subject to discussion in the European Parliament. The European member states are planning to revise some of the restrictions that DRM poses on legal users. For instance, legislative proposals include providing users with wider opportunities to use copyrighted materials in education and research.

In the public domain, access to knowledge for use is almost always granted and reuse is subject to fewer restrictions. Examples of these communities include academic science, open source software and do it yourself (DIY) communities. In these communities, properties of innovators are protected by Creative Commons licenses or GNUs. Creative Commons offer a number of different licenses that apply different rights and protections to creative work. In contrast to GNU, which grants users the right to copy, modify, and redistribute the software and where conditions of licensing are standard, licensors of Creative Commons can exercise more rights for their work depending on their preferences. Licensing under Creative Commons comprises four conditions, and licensors can choose their combinations when they license their products. These conditions include attribution, non-commercial, no derivative works and share-alike. Attribution allows licensees to copy, distribute, display and perform the work and make derivatives, but they have to give credit to the innovator. Non-commercial grants the same rights but only for non-commercial purposes. The same rules also apply for no derivative works, but licensees are not allowed to create derivatives and release it into the market. In that sense, the no derivative works license grants usage rights but does not permit to reuse and recombine original innovation in the creation of a new innovation. Finally, share-alike allows licensees to create derivative works but only under a license identical to the license that governs the original work.

Creative Commons can be understood as a more permissive version of copyright. If someone violates the terms of use that a Creative Commons license provides, the license for that person terminates. If the person keeps using the work, then he will commit a copyright infringement as he no longer has the license to use the work. The creator of the work can then take action against the violator in the same way as he would against someone who has infringed a copyrighted work. From video to images, audio and product design, a variety of organizations have adopted Creative Commons licenses. A popular photograph sharing platform called Flickr, for instance, was one of the first commercial websites to adopt Creative Commons licenses. The website provides artists with an option to choose between the Creative Commons licenses or a classical “all rights reserved” license. Similarly, the audio-sharing website ccMixter allows users to choose from a list of Creative Commons licenses when uploading their work (Monroy-Hernández, 2012).

The third condition for cumulative innovation is reward, which has a substantial effect on the first two conditions. This is often referred to as contractibility of innovation, as any innovation system should be able to assure that the originating innovator is honored and rewards are conferred upon him (Arrow, 1971; Boudreau and Lakhani, 2015). These rewards might include monetary rewards as well as reciprocal rewards such as reputation or credit, in both private and public settings (Murray and O'Mahony, 2007). Particularly for open communities, these motivations include self-interest (e.g., Lakhani & von Hippel 2003, Lerner & Tirole 2002), peer recognition, (Lakhani & Wolf 2005), reputation (Lerner & Tirole, 2002), social capital (Nambisan & Baron 2010; Wasko & Faraj 2000), external recognition (Jeppesen & Frederiksen 2006) and social exchange (Faraj & Johnson 2011). In open communities, therefore, survival of the community is highly dependent on attracting high-quality contributions from voluntary members by satisfying their motives for participation. (O’Mahony & Ferraro, 2007; Ma & Agarwal, 2007).

A recent stream of literature focuses on the question of how copyright policies affect reuse of materials that are protected by copyright. For instance, Nagaraj (2017) and Watson (2017b) find that copyright restrictions significantly reduce reuse. Watson (2017a) examines how derivative works affect the performance of upstream work given that copyright is granted for some of the original work in the music industry setting. He finds that, on average, reuse positively affects demand for upstream work. Gans (2015) examines the impact of reusing under different copyright regimes where different parties can negotiate the extent of reusing. Under the assumption that reuse harms the copyright owner, he argues that traditional copyright protection provides more incentives for both creators of the original work and downstream re-users. He further argues that fair use regimes might improve social welfare where parties discuss the extent of reuse.

Past research has generally recognized the potential impact of reuse on upstream work by observing the changes in copyright and the amount of reuse. This paper, however, takes a different perspective by examining the direct relationship between downstream and upstream work in a setting where intellectual property rights are permissive from the very beginning. In particular, it examines whether and when a derivative work harms the demand for the parent work, or whether it actually creates additional demand for it.

Decay of Attention, Information Spillover and Popularity

As digital technologies have reduced the costs of producing and distributing digital goods such as books, music, and product designs, there is a dramatic increase in the introduction of new products into the relevant markets. Therefore, consumers or users cannot be fully informed about the relevant characteristics of the products in these markets. Hence, consumers do not just need to decide about which product to purchase but also about which products they would like to learn about. Typically, consumers make their decisions in a variety of ways, including observing the behavior of others or looking at recommendations and best seller lists, as well as through information spillovers (Sorensen, 2017).

Hendricks & Sorensen (2009), for instance, focus on quantifying the lack of consumer information in the music industry by measuring the effects of the new album releases on previous albums of the same artist. By analyzing music album sales between 1993 and 2002, they find that earlier albums of artists benefit from increased demand due to the promotional activity and radio airplay around the release of a new album. This positive information spillover effect starts to pick up a couple of weeks before the new album's release, and persists for several months. Furthermore, they show that the effect is particularly large if the previous album of the artists was not a hit and if the new release is a hit. Similarly, by using data on the TV broadcasting of movies, Kumar et al. (2014) find that broadcasting of movies on television creates a positive information spillover effect on DVD sales. Their essential idea is that when a consumer sees the movie on TV, he learns about the movie and might like it enough to buy a DVD of the movie or recommend it to his friends. Similar to Hendricks & Sorensen (2009), they find this effect to be stronger for less-known movies. Taken together, these patterns suggest that many products do not realize their full sales potential due to a lack of information and that the release of associated products leads to a discovery effect.

When a new product is released, typically there is a buzz surrounding the product. However, this effect dissipates and interest in a particular product tends to decay. For instance, Zhang (2016) finds that when labels remove DRM from their music catalogues, digital music sales increase by 10%. She finds positive effect of DRM removal to be larger for albums of older vintages compared with newer ones, because the information about older vintages tends to decay. The most common characteristic of markets for creative products is that there is a large volume of new and diverse products that are constantly introduced to the market and that a large fraction of sales takes place at the initial release of the product. Plausibly, consumers might not be fully informed about all products and there are different mechanisms that emerge to facilitate product discovery.

The studies emphasize how cost of creation, production, and distribution has changed in the era of internet and show that traditional IP rights based on economic scarcity do not work effectively anymore. Also, it seems that traditional IP rights are not necessarily beneficial in some cases and not effective in stopping a flood of piracy. In the face of new technologies such as 3D printing, owners of the design and utility patents may face the same problems as companies in creative content industries: They may find themselves unable to prevent pirate production of their designs. In contrast to creative content industries, there might be different contingencies that come into play in the case of 3D printing. Besides creating attention, derivative work might create competition for upstream work. The net effect is likely to depend on whether or not derivative work offers an alternative or an improvement to the original work. If it offers an improvement (or in other words is vertically differentiated), derivative work might harm the market for the upstream work regardless of whether it creates an attention spillover or not. If it offers an alternative (or in other words is horizontally differentiated), the fraction of users seeing the derivative product might end up buying the original product. In the next section, I empirically examine whether the positive or the negative effect of reuse prevails and explore some of the contingencies.

3. Setting

I use data from Thingiverse, one of the largest online communities dedicated to 3D printable objects. Thingiverse was launched in October 2008 by MakerBot to encourage sharing of 3D printable design files as a complementary good to their 3D printers. In the community, there is a wide variety of designers that produce a number of different designs for different product categories. These categories group designs by their functional purpose and include 3D printing, art, fashion, gadgets, hobby, household, learning, models, tools and toys & games. These ten categories are further divided into seventy-nine sub-categories. Users in Thingiverse can browse designs by using different search parameters including keyword, category, tag, group, and functionality. The design files in Thingiverse are licensed under the Creative Commons license which allows users to freely share their innovations for other members of the community to print and most of the time also reuse. While most of the open hardware communities expect a contribution to open projects, Thingiverse provides a common ground from which derivatives of products may form. Furthermore, the community puts strong emphasis on providing recognition for designers. Free sharing and attribution mechanisms together facilitate the growth of the community by enabling users to build on others' innovations and at the same time keep them motivated to contribute (West & Kuk, 2016).

In contrast to other open communities where users contribute to ongoing projects, designs in the Thingiverse community are mostly the result of an individual effort, which makes the sense of ownership much stronger. An example of this is provided by the case of an eBay seller called just3dprint who has downloaded more than two thousand designs from Thingiverse and sold them for profit. While many designers in the community did not find this case shocking, they were complaining that they should have received, at least, a proper attribution from the seller. The first notice was given to the community by Louise Driggers, a designer who has posted a design of a sad emoji face with the description of the situation. Later on, many more designers have checked the eBay page of just3dprint and realized that their work was also affected. An important number of members from the community have send their complaints to both EBay and just3dprint. In response to just3dprint, Makerbot has also issued a statement over the violation of their Terms of Service, and said that they will be in consultation with their legal team to discuss next steps. With increasing complaints to eBay from both designers and Makerbot, items have been deleted from eBay. At the end, Loubie Driggers has uploaded a new emoji design, but this time with a happy face, to thank the community for their support.[1]

The website provides a variety of data such as the design id, the designer's name, the number of downloads, number of views of the webpage of the design, likes, number of times that a product is reused, the number of times the product is 3D printed, the launch date of the design, date and text of the comments, and from which product(s) the focal design is reused. When a design is reused, the algorithm of the website automatically increases a statistic called "remix it" by one unit.[2] At the same time, designers of the derivative products are required to provide a reference to parent work(s) from which they formed their designs. In case of not providing proper attribute, the derivative product will be removed from the community and designers who violate Creative Commons licenses might be banned.

There are four ways through which users might discover designs. First, they might click on the product list page where by default all of the products are listed based on their novelty. Second, users might choose to take a look at the product list page of particular categories where products are again by default listed based on their novelty. Third, they might use the search tool and conduct a search based on keywords. In this case, the platform, by default, sorts products based on the number of downloads given the keyword. Finally, if they are following a particular designer, their new designs might be fed into their news feed. The design of the platform is similar to other large digital platforms where the newest content is fed into the users' news feed when they enter to the main page and the most relevant and high quality content is shown when they specify a keyword in their search for products. As argued in the paper, reusing might be

[1] To see the discussion: https://all3dp.com/makers-uproar-poached-thingiverse-models-ebay/
[2] Refer to the appendix for snapshots from the website.

an additional mechanism through which designs are discovered as derivative designs are required to provide an explicit link to the designs that they built upon.

4. Data and Descriptive Statistics

Descriptive Statistics with Full Sample

I collected weekly data for the designs introduced to the community in the first six months of the year 2015 and observed downloads of these designs till the end of year 2016. I also track the weekly downloads of derivative designs of these upstream designs when they are released to the community. The data consists of design submissions made by over 10,000 individuals consisting of a total of 40,710 submissions. To ensure enough variation around the treatment, I use four weeks of pre-and post-treatment periods and drop all of the designs that are remixed before the first four weeks or has less than four observations after the treatment. I also dropped designs that are self-remixed i.e. reused by the original creator as the timing of reuse might be more strategic, and dropped the designs that are featured on the main page of Thingiverse as it might create a buzz for the product which might stimulate reuse. Furthermore, I dropped meta-models from the sample which are customizable and therefore make resulting products less likely to be innovative (Kyriakou et al., 2017). I also dropped designs that were reused more than once to closely examine the one-to-one relationship between upstream and derivative designs. Among these designs, 1,684 of them are treated and 39,026 of them are remained in the control group.

I measure demand with the number of downloads. Measuring demand by downloads is consistent with previous work on open innovation and platforms, which quantifies the subjective qualities and market success of creative work by visualizations, downloads or likes (Crowston, Howison, & Annabi, 2006; Liebowitz 2005). Figure 1 depicts a histogram of reuses.

As can be seen from the figure, a large portion of the reuses are observed in the first week of the launch of the design. Figure 2 plots the logarithm number of downloads over time for both treated and untreated designs. The horizontal axis in the graph is the time from the release of the design. The figure shows that there is a similar and overall decreasing temporal pattern for treated and control designs. Moreover, a comparison of the control and treatment groups reveals that demand for the designs in the treated group, on average, is higher than those in the control group. Figure 3 shows the distribution of weekly downloads. The distribution has a long right tail. Around 6 percent of the observations exceed 10, and the maximum is 15,517. The proportion of zeros is around 25.4 percent.

Figure 4 plots the number of downloads around the reuse. The horizontal axis in the graph is time from the week of reuse, where zero is the week of reuse. According to Figure 4, there is an immediate increase in the number of downloads on the week of reuse and this effect decays over time. Thus, Figure 4 provides preliminary evidence that there is an increase in the number of downloads after the reuse.

Empirical Design

In this paper, I am interested in estimating demand spillovers after reuse. In an ideal experiment, I would randomly select a sample of designs from the entire population of designs and randomly assign other designers to create derivative products based upon them and distribute these designs to relevant markets. Then, I would observe the difference between the demand for treated and control groups by tracking the changes in downloads. The present analysis differs from this ideal experiment as it relies on observational data. In this study, I measure demand spillovers after reuse by employing a difference-in-differences (DID) design.

In the empirical analysis, I use an approach similar to Hendricks and Sorensen (2009) and Kumar et al (2014) where not all products are present in each calendar time as they are released at different times to the market and where a decay of interest in products can be observed. In my analysis, I focus on the first treatment effect: the release of a derivative and its impact on demand for the upstream design. Let y^0_{it} denote the number of downloads of upstream design *i* in period *t* without treatment, let y^s_{it} denote demand for the upstream design in period t when design *i* is in the sth period of the treatment. For each design, t indexes time since the design's release, rather than the calendar time. The purpose of the regression is to estimate the average treatment effect on the treated (ATE) by simply taking the difference of y^s_{it} - y^0_{it} for each period of the treatment window. However, I have a counterfactual problem as I do not observe what would happened to the designs in the absence of the treatment. My approach is to measure this difference by using designs that are not treated at all or not yet as the control group. This approach is often referred to as "generalized differences in differences". Similar to the classical differences in differences framework, the validity of the generalized DID designs relies on the assumption that the unmeasured variables are either time-invariant -- therefore soaked up by the fixed effects or they are time-varying but group invariant. These restrictions together suggest that units in the treatment and control groups should unveil a similar set of period-specific changes and the timing of treatment should be random across treated observations. This is often referred to as "parallel trend assumption" (Angrist and Pischke 2008).

In the estimation, I use generalized a difference-in-differences framework with Poisson pseudo maximum likelihood (PPML) specification (e.g. Azoulay et al. 2010, Burtch et al. 2018). In this context, using a Poisson model is more appropriate given the low average values and skewed distribution of the dependent variables (Wooldridge, 1999). Furthermore, given that the data contains many zeros, the PPML estimator provides more reliable estimates than a log-OLS specification, especially in the presence of heteroskedasticity (Silva and Tenreyro, 2006; Silva and Tenreyro 2011). Moreover, the maximum likelihood specification of Poisson is fully robust to violation of the Poisson assumptions, as well as to unmodeled serial correlation (Wooldridge, 2010). Therefore, I use PPML as the primary specification. To estimate the effect of reuse on the upstream work, I use the following equations:

$$WeeklyDownloads_{it} = exp\ [\beta_1\ Postreuse_{it} + \delta i + \eta it + \gamma t + u_{it}] \quad (1)$$

where $WeeklyDownloads_{it}$ is the downloads of the design i at week t, $Postreuse_{it}$ is the value of the treatment for design i at week t, and δ_i is a design fixed effect, η_{it} is the age fixed effect and γ_t is the calendar time fixed effect parameter that is estimated. In the specification, design fixed effects control for time-invariant factors that characterize a design, such as complexity and 3D printability. Age fixed effects control for decreasing interest in designs after their introduction, which is important in light of the demand patterns observed in Figure 2. The age fixed effect variable takes the value of one on the week of the introduction of the design and is incremented by one for the following weeks. The age dummies allow for a flexible decay path of demand. However, it implicitly assumes that the shape of this decay path is the same across designs. Therefore, I also interact age fixed effects with design fixed effects to control for design specific time trends as an alternative specification. Finally, I introduce time-fixed effects for months of the year to control for community wide unobserved temporal trends that equally affect all designs at a particular calendar time. The coefficient of interest is β_1, which can be interpreted as the relative change of the demand for the treatment group compared to the control group, caused by the reuse. The advantage of generalized difference-in-differences is that it can improve precision and provide better fit of the model and that it does not assume all units in the treatment and control groups have the same average outcome variable. Thus, it allows the common change in the outcome variable to vary by different time periods.

To test the validity of the parallel trends before treatment assumption, I augment the above specifications with leads and lags of the treatment variable, similar to the approach by Autor (2003). Adding lead values enables me to check the validity of the parallel trends assumption. In particular, it enables me to see whether there are differences between the ex-ante trends of 'treated' and 'untreated' observations. Under the null hypothesis of no differences in pre-treatment trends, the coefficients on the lead variables should be zero. I also add lags of the treatment variable to see whether the impact of reuse accelerates, stabilizes, or mean reverts. In particular, I add indicator variables for 1, 2 and 3 weeks before reuse, weeks 0–3 after reuse, and week 4 forward. The first six indicator variables take the value of one in the relevant weeks whereas the final variable is kept to be one starting from the fourth week of reuse. As the results from Table 3 on column III indicate, however, the lead values are positive and significantly different than zero. Therefore, I use a coarsened exact matching (CEM) procedure to obtain a better set of controls for the reused designs and re-estimate the difference-in-differences model on this subset of designs.

In the matching analysis, for each treated design I created a sample of potential control designs by exactly matching on sub-category, the week of the design's release and a derivative dummy variable which indicates whether an upstream design itself is a derivative of another design or not. From this restrictive sample, I selected one control design for each treated design based on the pre-treatment popularity of the designers measured by the logarithm of number of followers one week before the treatment, the logarithm of the number of downloads and the rolling average of the logarithm number of downloads one week, two and three weeks prior to the treatment. After the matching, I dropped treated with no exact matched control group designs.

I show the summary statistics and balance checks of the treatment and the control groups in Table 4. The treated and control groups have no statistically significant differences on the matching variables. For the remainder of the analysis, I used the matched sample.

Next, I check for the potential substitution effect. To do so, I use the category information of the designs. In the innovation management literature, categories are used to define the functional boundaries of technologies and products (Utterback 1996) and a similar categorization exist on Thingiverse. In particular, I considered designs in 3D printing, tools and gadgets categorized to be functional designs. Other categories such as art, models and fashion are more likely to consist of designs with artistic value. The remaining categories, namely hobby, household, learning and toys & games are likely to contain a mixture of both functional and artistic designs. To account for the potential substation effect, I use the following specification:

$$WeeklyDownloads_{it} = \exp[\beta_1\, Postreuse_{it} * Functional_i \; + \delta i + \eta i + \gamma t + u_{it}] \quad (3)$$

Where the $Functional_i$ variable takes the value of one if it is categorized under 3D printing, tools or gadgets categories and zero otherwise. I also provide a regression analysis separately for each category in the appendix.

Next, to examine whether increasing downloads and page impressions of derivative work transform into increasing upstream work downloads, I use the following specifications:

$$WeeklyDownloads_{it} = exp\,[\beta_1\, Postreuse_{it} * LnWeeklyDerivativeViews_{jt} \; + \delta i + \eta i + \gamma t + u_{it}]$$

(4)

where $LnWeeklyDerivativeViews_{jt}$ corresponds the logarithm of the derivative works' page impressions in a given week. A positive coefficient on the interaction term would indicate a positive correlation between the demand for the upstream work and the page impressions of the downstream work.

5. Results

Baseline Results with Full Sample

The baseline estimation results are reported in Table 3. In each column, the dependent variable is the number of weekly downloads for designs and the main variable of interest is *Postreuse*, which takes the value of one for treated designs. In all columns I use robust standard errors clustered at the design level. The regression result in column (I) indicates that reused designs have, on average, a 33.3 ($e^{0.288}$ -1) percent increase in demand. In column (II), I add design specific time trends to the regression. The coefficient on the *Postreuse* variable shrinks from 33.3 percent to 12.1 ($e^{0.115}$ -1) percent. In the third column of Table 3, I add lags of the treatment variable along with the lead values to see whether the impact of reuse accelerates, stabilizes, or mean reverts. In particular, I add indicator variables for 1, 2 and 3 weeks before reuse, weeks 0–3 after reuse, and week 4 forward. The first five indicator variables take the value of one in the relevant weeks whereas the final variable is kept to be one starting from the fourth week of reuse. This approach is similar to the approach by Autor (2003). As the results in column (III) of Table 3 indicate, the coefficients on the leads are significantly different from zero, which provides evidence that there are differences in pre-treatment trends across treated and control groups before reuse. In the final column of Table 3, I also add design-specific time trends. Results are qualitatively similar to the results from the previous column. However, when I control for design-specific time trends, the level of the significance of the lead values diminishes. Given that the parallel trend assumption does not hold, I turn to the estimation with the matched sample in the next section.

Baseline Results with the Matched Sample

The matching procedure ensures that there is no ex-ante difference among the treatment and control groups. However, as matching involves trading in bias for variance, the matched sample is reduced to 669 treatment-control pairs, compared to the full sample with 40,710 designs. The regression results from the matched sample in Table 5 indicate that there is a 13.2 ($e^{0.124}$ -1) percent increase in demand after reuse. When I add design-specific time trends, the magnitude of the coefficient grows to 17.8 ($e^{0.164}$ -1) percent. Similar to the analysis with the full sample, I test the validity of the parallel trends assumption by adding leads of the treatment to the regression. The results in column (III) of Table SS indicate that I cannot reject the null hypothesis of no difference in the pre re-use trends. The results are robust to inclusion of design specific time trends as indicated in column (IV). In the week of reuse, the number of downloads increases substantially by 31 percent, and it fluctuates between 20 and 18 percent throughout the following three weeks. After the fourth week of reuse, it averages 11 percent. The coefficients on lag values indicate that the treatment effect increases during the first few weeks, and then plateaus. The results remain qualitatively similar when I add design-specific time trends.

Substitution

Given that reuse increases the demand for upstream designs, I now turn my attention to examine whether the derivative designs compete with the original designs that they are based upon. To do so, I interact the Postreuse variable with the functional design dummy variable. The sum of the Postreuse and the interaction term ($e^{0.165-0.08}$-1) is positive, which indicates that derivative work increases the demand for upstream work regardless of the type of the product. However, this effect is smaller for functional designs compared to other designs. Namely, the net effect is ($e^{0.165-0.08}$-1) 9.9 percent. In the appendix, I also present regressions with full set of category dummies.

Test of the Mechanism

Here, I examine whether there is a correlation between an increasing number of downloads of the upstream design and page impressions of the downstream design. To understand whether the increase in the number of downloads of upstream work come from the derivative work, I interact the log-transformed number of views of the derivative work with the *Postreuse* dummy. After the interaction, the coefficient on *Postreuse* variable in Table 7 gets insignificant which indicates that positive effect of reuse is fully mediated by page impressions of the derivative designs. This result is not surprising given that only a certain fraction of the users also visit the upstream work, downstream designs with a larger market are more likely to benefit upstream works.

6. Discussion

By using a CEM procedure and difference-in-differences specification, I find that reuse affects the demand for upstream work positively. Moreover, I observe the positive effect to be smaller if the design is functional in its nature. I also find that the discovery effect is moderated by the number of page impressions of the derivative designs. These findings suggest that conditional on the type of design, permissive intellectual rights might be beneficial for the original creators by creating attention for the upstream product.

These findings have important implications for both firms and policymakers. From a managerial standpoint, managers in content industries are making decisions about whether or how to protect the market for their products. These findings highlight that, under certain circumstances, it might be more beneficial for companies to allow for derivative creations.

Especially firms in creative industries where tastes of the consumers are more heterogeneous, firms might switch from traditional means of protecting intellectual property rights to a more permissive one. Creative Commons licenses might be one alternative as it requires that proper attribute is given to the product at hand, which might lead to an additional product discovery.

For policymakers, these results also point to the importance of permissive intellectual property rights for cumulative innovation. Under the current law called the Digital Millennium Copyright Act, also known as the DMCA, the judges in the United States assess the purpose and the character of reuse to determine whether a derivative work infringes intellectual property or not. They do so by examining the nature of the copyrighted work, the amount and the substantiality of the portion taken, and finally, the effect of usage on the potential market for the original piece[3]. The fact that the release of derivative products might indeed improve the market for upstream products, especially for those creative products, could lead policy makers to reconsider whether reuse should be classified as an infringement of the copyrights by focusing on the characteristics of the product. The rationale is that if the release of derivative works does not necessarily harm the market of the upstream works, the permission to reuse might be granted to downstream creators given that the release of their products increase the variety and hence potentially the consumption and welfare.

One concern with the results could be the lack of external validity. In particular, one might be concerned that Thingiverse represents an idiosyncratic setting for my analysis of the impact of reuse on demand. However, the characteristics of this setting are similar to other digital creative media settings where the demand for products decays over time and the consumption of the goods entail zero or very low cost. Moreover, the results that I find here are also similar to the

[3] https://fairuse.stanford.edu/overview/fair-use/four-factors/

emerging literature on the effect of reuse in the demand for upstream work (e.g. Watson 2017a), which further alleviates concerns regarding the generalizability of my findings. Furthermore, the potential drawback of Thingiverse in terms of generalizability also provides an opportunity to simulate a world with weak IP. In particular, the default ease of reuse in the setting is ideal for testing the effectiveness of more permissive intellectual property rights.

A few limitations remain in the paper. First, the relationship between the derivative and upstream designs are not necessarily bidirectional. A fraction of the people visiting upstream designs might be visiting downstream designs as well. However, given that there is an immediate effect of reuse on the demand for upstream work and given that there is a buzz surrounding the introduction of the products including derivative designs, the idea that a larger fraction of people are coming from downstream work to upstream work is a credible assumption. Unfortunately, my data does not contain indication of who views or downloads which design. Having this information would have provided me with a more direct test of the discovery mechanism. Second, if the purchase of the products was costly, the magnitude of positive spillover effect for artistic designs might be smaller, whereas the negative effect for functional designs might be even larger. Yet, given that 3D printing technology separates the creation of innovations from their production and distribution, purchasing a digital design file might remain relatively inexpensive similar to purchases of songs for online streaming and mobile phone applications. Indeed, even in the pre-internet era only a small fraction of the cost was conferred to artists or designers to produce content. Instead, the cost was associated with production and distribution of the physical material. In fact, there is a large stream of literature showing that creators are more likely to be intrinsically motivated. With the internet, these costs have been driven towards zero. Therefore, even in the case of substitution, consumers might purchase both of the upstream and downstream design, given that the real cost associated with the design is its production. Furthermore, the fact that the number and variety of content has increased in the era of the internet provides further evidence that weakened IP did not necessarily discourage people to create. Third, I use category information provided by Thingiverse to categorize designs as functional and artistic. However, functional designs are also likely to carry some artistic value and some of the products might be misclassified. The validity of the results on substitution therefore relies on the assumption that the percentage of falsely classified designs is negligible. Fourth, some of the designers of derivative products might not report their parents and Thingiverse might not able to identify all of those cases. However, this does not pose a threat to the validity of my analysis as I would be underestimating the effect of reuse on upstream work in this case.

References

Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Arrow, K. (1962). Economic welfare and the allocation of resources for invention. In *The rate and direction of inventive activity: Economic and social factors* (pp. 609-626). Princeton University Press.

Arrow, K. J. (1971). *General competitive analysis* (No. 04; HB171, A7.).

Autor, D. H. (2003). Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of labor economics*, *21*(1), 1-42.

Azoulay, P., Graff Zivin, J. S., & Wang, J. (2010). Superstar extinction. *The Quarterly Journal of Economics*, *125*(2), 549-589.

Belleflamme, P., & Peitz, M. (2014). *Digital piracy* (pp. 1-8). Springer New York.

Boudreau, K. J., Jeppesen, L. B., & Miric, M. (2018). Digital'Mash-Ups,'Patents, and Copyright.

Boudreau, K. J., & Lakhani, K. R. (2015). “Open” disclosure of innovations, incentives and follow-on reuse: Theory on processes of cumulative innovation and a field experiment in computational biology. *Research Policy*, *44*(1), 4-19.

Burtch, G., Carnahan, S., & Greenwood, B. N. (2018). Can you gig it? An empirical examination of the gig economy and entrepreneurial activity. *Management Science*, *64*(12), 5497-5520.

Danaher, B., Smith, M. D., Telang, R., & Chen, S. (2014). The effect of graduated response anti-piracy laws on music sales: evidence from an event study in France. *The Journal of Industrial Economics*, *62*(3), 541-553.

Dasgupta, P., & David, P. A. (1994). Toward a new economics of science. *Research policy*, *23*(5), 487-521.

Faraj, S., & Johnson, S. L. (2011). Network exchange patterns in online communities. *Organization Science*, *22*(6), 1464-1480.

Flath, C. M., Friesike, S., Wirth, M., & Thiesse, F. (2017). Copy, transform, combine: Exploring the remix as a form of innovation. *Journal of Information Technology*, *32*(4), 306-325.

Fleming, L. (2001). Recombinant uncertainty in technological search. *Management science*, *47*(1), 117-132.

Gans, J. S. (2015). Remix rights and negotiations over the use of copy-protected works. *International Journal of Industrial Organization*, *41*, 76-83.

Goldfarb, A., & Tucker, C. (2017). *Digital economics* (No. w23684). National Bureau of Economic Research.

Haefliger, S., Von Krogh, G., & Spaeth, S. (2008). Code reuse in open source software. *Management science*, *54*(1), 180-193.

Hendricks K and Sorensen A. (2009). Information and the skewness of music sales. *Journal of Political Economy*. *117*(2):324–69

Hill, B. M., & Monroy-Hernández, A. (2013). The remixing dilemma: The trade-off between generativity and originality. *American Behavioral Scientist*, *57*(5), 643-663.

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political analysis, 20*(1), 1-24.

Jeppesen, L. B., & Frederiksen, L. (2006). Why do users contribute to firm-hosted user communities? The case of computer-controlled music instruments. *Organization science*, *17*(1), 45-63.

Kane, G. and Ransbotham, S. (2012) Codification and Collaboration: Information Quality in Social Media, in Proceedings of the 33th International Conference on Information Systems (Orlando, USA, 2012); Atlanta: AIS.

Kaplan, S., & Vakili, K. (2015). The double‐edged sword of recombination in breakthrough innovation. *Strategic Management Journal*, *36*(10), 1435-1457.

Krippendorff K, Butter R (1993). Where meanings escape functions. *Design Management J*. *4*(2):30–37.

Kyriakou, H., Nickerson, J. V., & Sabnis, G. (2017). Knowledge reuse for customization: Metamodels in an open design community for 3D printing. *MIS Quarterly*, *41*(1), 315-332.

Kumar A, Smith MD, Telang R. (2014). Information discovery and the long tail of motion picture content. *MIS Quarterly*. *38*(4):1057–78.

Lakhani, K. R., & Von Hippel, E. (2003). How open source software works:“free” user-to-user assistance. *Research policy*, *32*(6), 923-943.

Lakhani, K. R., & Wolf, R. G. (2005). Why Hackers Do What They Do: Understanding Motivation andEffort in Free/Open Source Software Projects. *Perspectives on Free and Open Source Software*.

LaPolt, D., Rosenthal, J., & Meller, J. (2014). A response to professor menell: A remix compulsory license is not justified. *Colum. JL & Arts, 38,* 365.

Lerner, J., & Tirole, J. (2002). Some simple economics of open source. *The journal of industrial economics*, *50*(2), 197-234.

Lessig, L. (2004). The creative commons. *Mont. L. Rev.*, *65*, 1.

Lessig, L. (2008). *Remix: Making art and commerce thrive in the hybrid economy*. Penguin.

Li, X., MacGarvie, M., & Moser, P. (2018). Dead poets' property—how does copyright influence price?. *The RAND Journal of Economics*, *49*(1), 181-205.

Ma, M., & Agarwal, R. (2007). Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. *Information systems research*, *18*(1), 42-67.

Menell, P. S. (2016). Economic Analysis of Copyright Notice: Tracing and Scope in the Digital Age. BUL Rev., 96, 967.

Monroy-Hernández. A. (2012). Designing for Remixing: Supporting an Online Community of Amateur Creators, *Thesis*, Massachusetts Institute of Technology.

Murray, F., & O'Mahony, S. (2007). Exploring the foundations of cumulative innovation: Implications for organization science. *Organization Science*, *18*(6), 1006-1021.

Murray, F., & Stern, S. (2007). Do formal intellectual property rights hinder the free flow of scientific knowledge?: An empirical test of the anti-commons hypothesis. *Journal of Economic Behavior & Organization*, *63*(4), 648-687.

Nagaraj, A. (2017). Does copyright affect reuse? Evidence from google books and wikipedia. *Management Science*. Articles in Advance.

Nambisan, S., & Baron, R. A. (2010). Different roles, different strokes: Organizing virtual customer environments to promote two types of customer contributions. *Organization Science*, *21*(2), 554-572.

Nelson, R. R. (1959). The simple economics of basic scientific research. *Journal of political economy*, *67*(3), 297-306.

Nelson, R. R., & Winter, S. G. (1982). The Schumpeterian tradeoff revisited. *The American Economic Review*, *72*(1), 114-132.

O'Mahony, S., & Ferraro, F. (2007). The emergence of governance in an open source community. *Academy of Management Journal*, *50*(5), 1079-1106.

Reschke, B. P., Azoulay, P., & Stuart, T. E. (2017). Status spillovers: The effect of status-conferring prizes on the allocation of attention. *Administrative Science Quarterly*, 0001839217731997.

Rietveld, J., & Eggers, J. P. (2018). Demand Heterogeneity in Platform Markets: Implications for Complementors. *Organization Science*, *29*(2), 304-322.

Rosen, L. (2005). *Open Source Licensing Software Freedom and Intellectual Property Law*. Prentice Hall, Upper Saddle River, NJ.

Samuelson, P. (2008). Unbundling fair uses. *Fordham L. Rev.*, *77*, 2537.

Samuelson, P. (2001). Economic and constitutional influences on copyright law in the United States.

Schumpeter, J. (1942). Creative destruction. *Capitalism, socialism and democracy*, *825*.

Silva, J. S., & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and statistics*, *88*(4), 641-658.

Silva, J. S., & Tenreyro, S. (2011). Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator. *Economics Letters*, *112*(2), 220-222.

Sinha, R. K., Machado, F. S., & Sellman, C. (2010). Don't think twice, it's all right: Music piracy and pricing in a DRM-free environment. *Journal of Marketing*, *74*(2), 40-54.

Sinnreich, A., Latonero, M., and Gluck, M. (2009). Ethics reconfigured: How today's media consumers evaluate the role of creative reappropriation. Information, Communication & Society, 12(8):1242.

Sojer, M., & Henkel, J. (2010). Code reuse in open source software development: Quantitative evidence, drivers, and impediments.

Sorensen, A. T. (2017). Bestseller Lists and the Economics of Product Discovery. *Annual Review of Economics*, *9*, 87-101.

Stanko, M. A. (2016). Toward a theory of remixing in online innovation communities. *Information Systems Research*, *27*(4), 773-791

Strumsky, D., & Lobo, J. (2015). Identifying the sources of technological novelty in the process of invention. *Research Policy*, *44*(8), 1445-1461.

Utterback, J. (1996). *Mastering the Dynamics of Innovation*, Boston: Harvard Business School Press

Wasko, M. M., & Faraj, S. (2000). “It is what one does”: why people participate and help others in electronic communities of practice. *The Journal of Strategic Information Systems*, *9*(2-3), 155-173

Watson, J. (2017a). What is the Value of Re-use? Complementarities in Popular Music. Working Paper

Watson, J. (2017b). Copyright and the Production of Hip-Hop Music. Working Paper.

West, J., & Kuk, G. (2016). The complementarity of openness: How MakerBot leveraged Thingiverse in 3D printing. *Technological Forecasting and Social Change*, *102*, 169-181.

Williams, H. L. (2013). Intellectual property rights and innovation: Evidence from the human genome. *Journal of Political Economy*, *121*(1), 1-27.

Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics*, *90*(1), 77-97.

Wooldridge, J. M. *Econometric analysis of cross section and panel data*. MIT press, 2010.

Zhang, L. (2016). Intellectual property strategy and the long tail: Evidence from the recorded music industry. *Management Science*. Articles in Advance.

Figures and Graphs

Figure 1. Histogram of first reuses

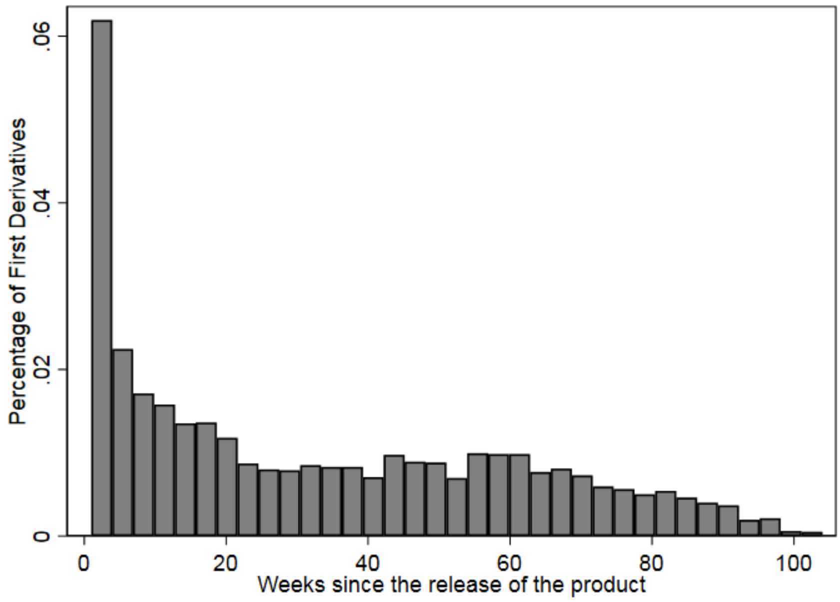
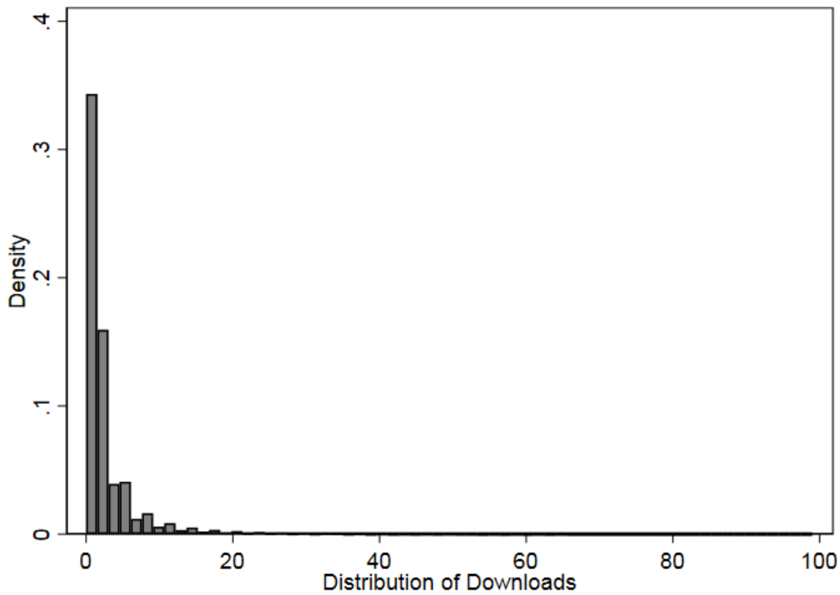


Figure 2. Histogram of downloads



Note: Histogram shows the number of designs whose maximum number of weekly downloads are smaller than hundred for the illustration purposes

Figure 3. The number of the downloads about the release of products

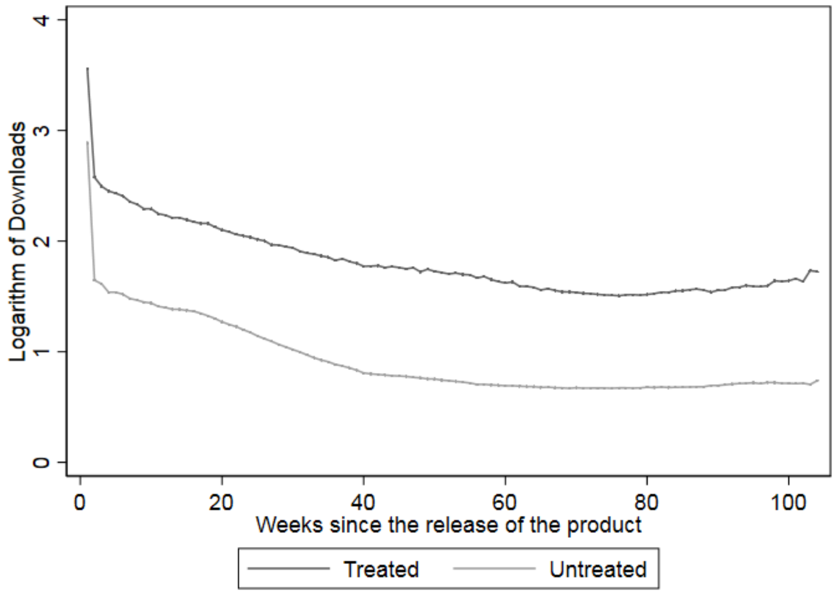
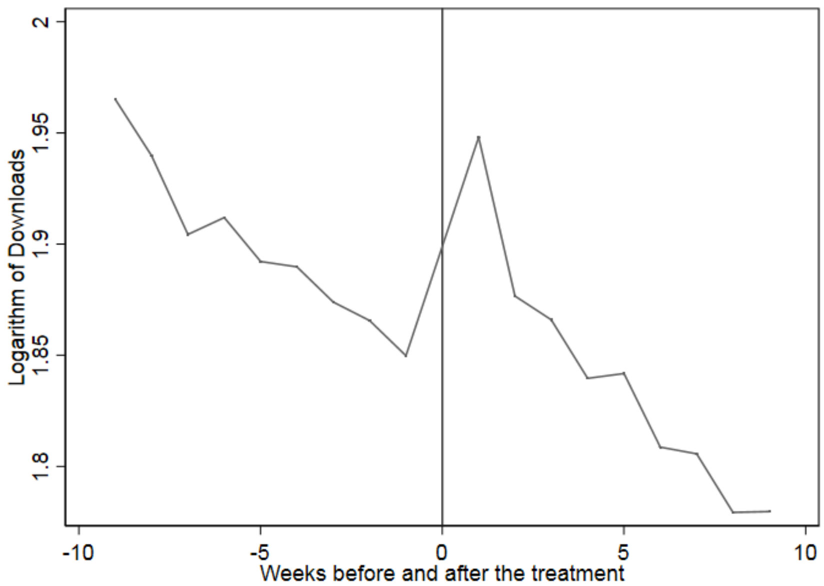


Figure 4. The number of downloads around the reuse



Note: For the illustration purposes, this graph exclude data of the designs that are reused within the first 10 weeks of after their release

Table 1. Definitions of the variables used in this study

WeeklyDownloads	Weekly downloads of upstream work
Postreuse	Dummy variable that indicates whether designs is reused at time *t*
Functional	Dummy variable that indicates whether upstream design is a functional design
LnWeeklyDerivativeViews	Logarithm of number of weekly page visits to derivative work

Table 2. Descriptive statistics with full sample

	Count	Mean	Sd	min	Max
WeeklyDownloads	3772756	3.2314	13.21	0	15517
Postreuse	3772756	0.02245	0.1401	0	1

Table 3. Baseline Results with non-matched sample -- Differences in Differences

	(I) Weekly Downloads	(II) Weekly Downloads	(III) Weekly Downloads	(IV) Weekly Downloads
Postreuse	0.288^{***}	0.115^{***}		
	(0.0283)	(0.0312)		
Postreuse_{t+1}			0.232^{***}	0.0718^{+}
			(0.0358)	(0.0369)
Postreuse_{t+3}			0.241^{***}	0.0740^{*}
			(0.0332)	(0.0342)
Postreuse_{t+2}			0.229^{***}	0.0672^{+}
			(0.0338)	(0.0350)
Postreuse_{t}			0.360^{***}	0.206^{***}
			(0.0316)	(0.0317)
Postreuse_{t-1}			0.327^{***}	0.174^{***}
			(0.0386)	(0.0386)
Postreuse_{t-2}			0.309^{***}	0.156^{***}
			(0.0383)	(0.0400)
Postreuse_{t-3}			0.297^{***}	0.145^{***}
			(0.0318)	(0.0333)
$\text{Postreuse}_{t-4\_forward}$			0.306^{***}	0.114^{**}
			(0.0316)	(0.0419)
Design FE	Yes	Yes	Yes	Yes
Age FE	Yes	No	Yes	No
Time FE	Yes	Yes	Yes	Yes
Time Trends	No	Yes	No	Yes
Log Likelihood	-6685391.0	-6588025.6	-6684384.6	-6587847.2
Pseudo R2	0.606	0.612	0.606	0.612
N	3772756	3772756	3772756	3772756

Standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4. Summary statistics for treatment and control groups after the matching

	Treatment		Control		
	Mean	SD	Mean	SD	t-stat
$LnDownloads_{t-1}$	1.341	0.701	1.331	0.705	-0.254
$LnDownloads_{t-2}$	1.329	0.745	1.311	0.731	-0.477
$LnDownloads_{t-3}$	1.327	0.748	1.308	0.748	-0.453
$LnAverageDownloads_{t-1}$	1.649	0.591	1.626	0.598	-0.708
$LnAverageDownloads_{t-2}$	1.661	0.595	1.641	0.605	-0.598
$LnAverageDownloads_{t-3}$	1.675	0.602	1.664	0.616	-0.331
$LnFollowers_{t-1}$	1.359	1.251	1.336	1.251	-0.333
Number of Designs	669		669		

Table 5. Baseline results with matched sample -- Differences in Differences

	(I) Weekly Downloads	(II) Weekly Downloads	(III) Weekly Downloads	(IV) Weekly Downloads
Postreuse	0.124^{**}	0.164^{***}		
	(0.0462)	(0.0232)		
Postreuse_{t+3}			0.0454	-0.0152
			(0.0291)	(0.0262)
Postreuse_{t+2}			0.0415	-0.0109
			(0.0299)	(0.0257)
Postreuse_{t+1}			0.0417	-0.000438
			(0.0315)	(0.0254)
Postreuse_{t}			0.277^{***}	0.240^{***}
			(0.0352)	(0.0290)
Postreuse_{t-1}			0.210^{***}	0.178^{***}
			(0.0355)	(0.0298)
Postreuse_{t-2}			0.188^{***}	0.162^{***}
			(0.0354)	(0.0304)
Postreuse_{t-3}			0.186^{***}	0.164^{***}
			(0.0394)	(0.0332)
$\text{Postreuse}_{t-4\_forward}$			0.110^{*}	0.151^{***}
			(0.0533)	(0.0302)
Design FE	Yes	Yes	Yes	Yes
Age FE	Yes	No	Yes	No
Time FE	Yes	Yes	Yes	Yes
Time Trends	No	Yes	No	Yes
Log Likelihood	-239273.2	-238187.0	-239202.0	-238174.7
Pseudo R2	0.497	0.499	0.497	0.499
N	122774	122774	122774	122774

Standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6. Substitution analysis with matched sample

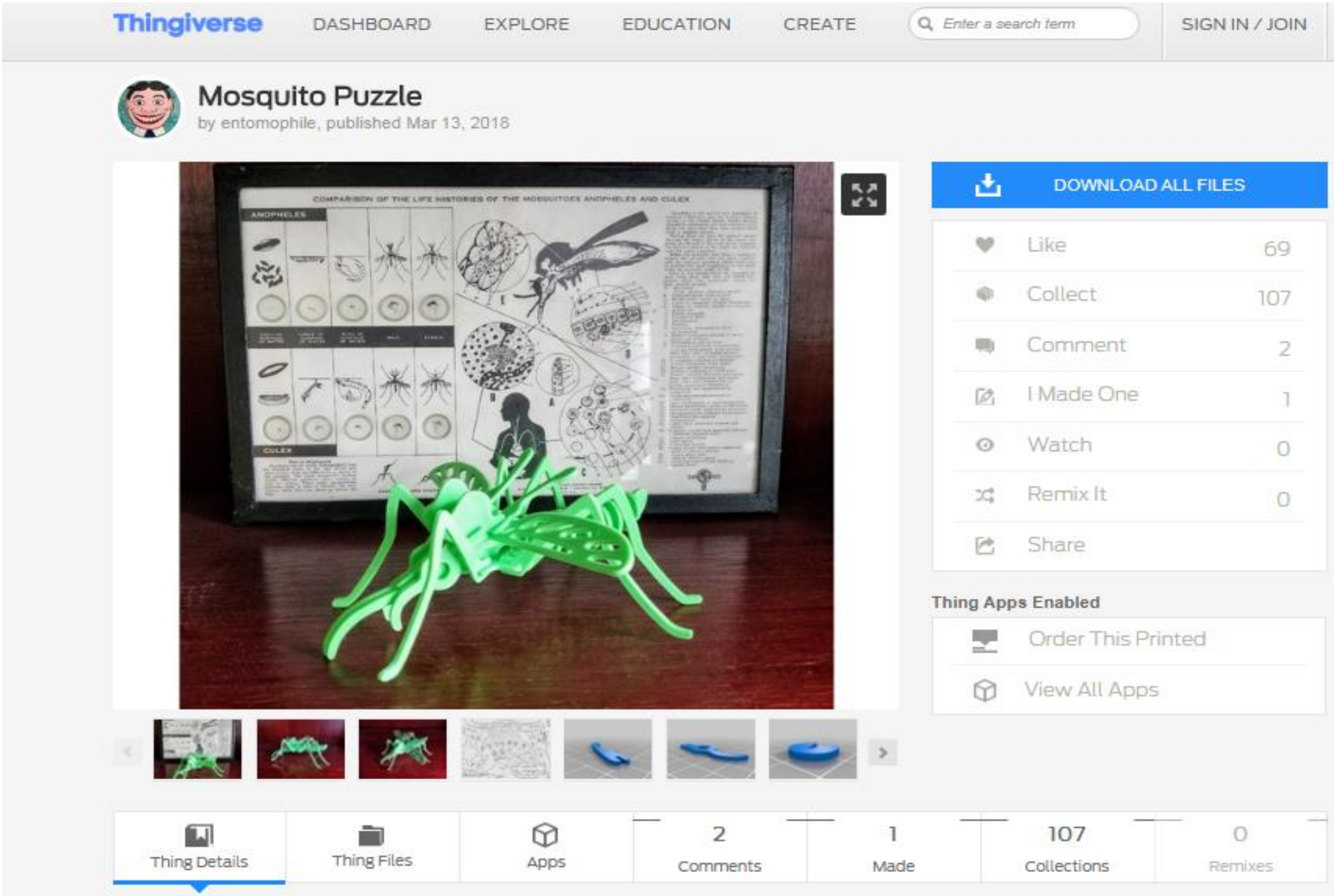
	WeeklyDownloads
Functional * Postreuse	-0.0887*
	(0.0445)
Postreuse	0.164***
	(0.0498)
Design FE	Yes
Age FE	Yes
Time FE	Yes
Log Likelihood	-239226.6
Pseudo R2	0.497
N	122774

Standard errors in parentheses
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

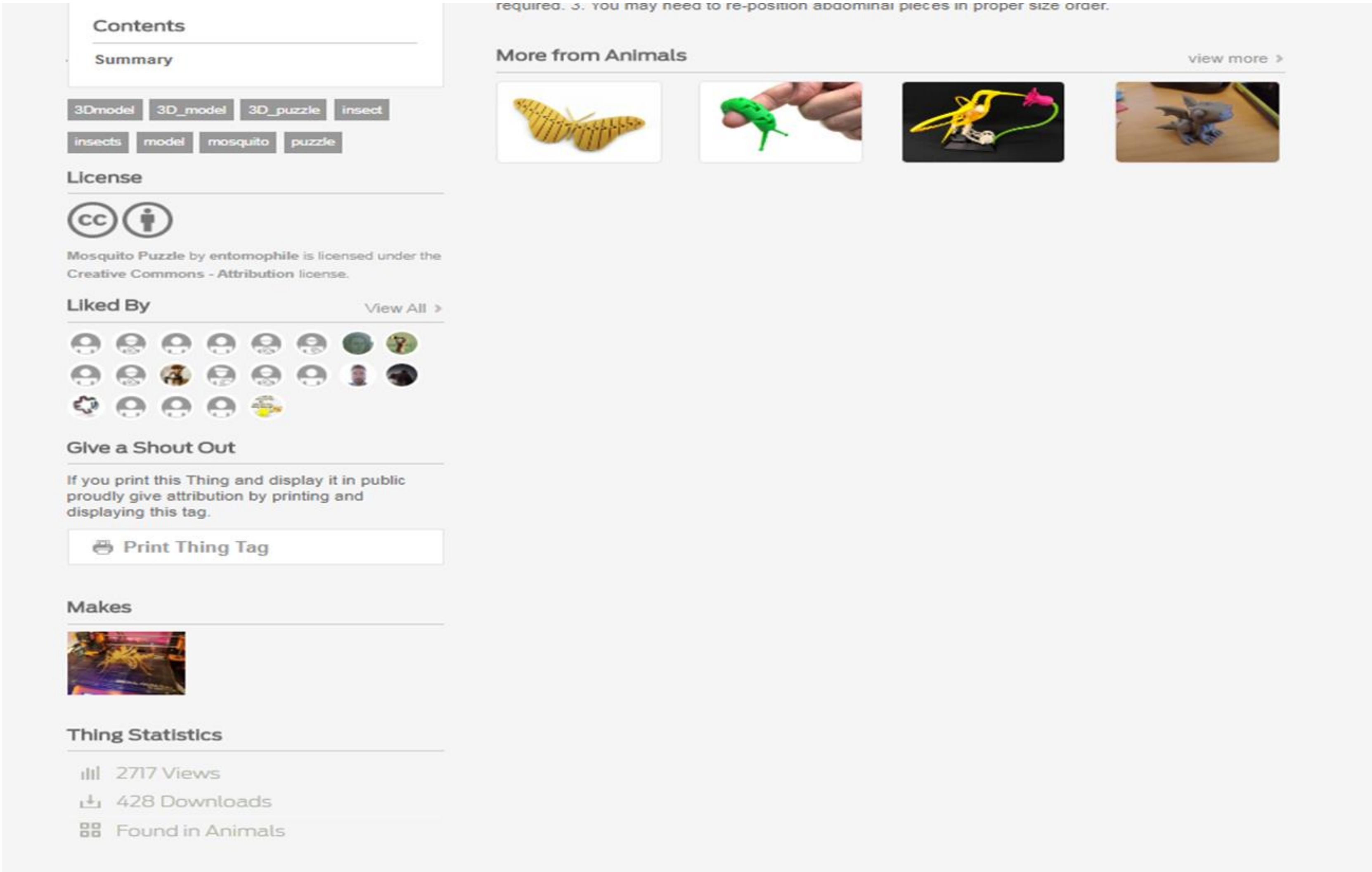
Table 7. Test of the Mechanism

	WeeklyDownloads
Postreuse	-0.0267
	(0.0690)
LnWeeklyDerivativeViews * Postreuse	0.0701***
	(0.0154)
Design FE	Yes
Age Fe	Yes
Time FE	Yes
Log Likelihood	-239051.8
Pseudo R2	0.497
N	122774

Standard errors in parentheses
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$



Note: First half of a screenshot of a design page from Thingiverse



Note: Second half of the screenshot of the design from Thingiverse

Table A1 Substitution effect with category dummies-- matched sample

	WeeklyDownloads
Postreuse (3D Printing as Baseline)	0.0782
	(0.0564)
Art * Postreuse	0.107
	(0.127)
Fashion * Postreuse	0.377^{**}
	(0.138)
Gadgets * Postreuse	-0.000737
	(0.0828)
Hobby * Postreuse	0.0103
	(0.0538)
Household * Postreuse	0.0503
	(0.0754)
Learning * Postreuse	0.202
	(0.176)
Models * Postreuse	0.302^{***}
	(0.0658)
Tools * Postreuse	-0.0666
	(0.110)
Toys & Games * Postreuse	0.0356
	(0.0829)
Design FE	Yes
Age FE	Yes
Time FE	Yes
Log Likelihood	-239046.9
Pseudo R2	0.497
N	122774

Table 2A. List of categories and subcategories

Category	**Subcategory**	**Category**	**Subcategory**	**Category**	**Subcategory**
3D Printing	3D Printer Accessories	Models	Animals	Household	Bathroom
3D Printing	3D Printer Extruders	Models	Buildings & Structures	Household	Containers
3D Printing	3D Printer Parts	Models	Creatures	Household	Décor
3D Printing	3D Printers	Models	Food & Drink	Household	Household Supplies
3D Printing	3D Printing Tests	Models	Model Robots	Household	Kitchen & Dining
3D Printing	3D Printing	Models	Model Furniture	Household	Office
Art	Art Tools	Models	People	Household	Organization
Art	Interactive Art	Models	Props	Household	Outdoor & Garden
Art	2D Art	Models	Vehicles	Household	Pets
Art	Coins & Badges	Models	Models	Household	Replacement Parts
Art	Math Art	Tools	Hand Tools	Household	Household
Art	Art	Tools	Machine Tools	Learning	Biology
Art	Scans & Replicas	Tools	Parts	Learning	Engineering
Art	Signs & Logos	Tools	Tool Holders & Boxes	Learning	Math
Art	Sculptures	Tools	Tools	Learning	Physics & Astronomy
Fashion	Accessories	Hobby	Automotive	Learning	Learning
Fashion	Bracelets	Hobby	DIY	Toys & Games	Chess
Fashion	Costume	Hobby	Electronics	Toys & Games	Construction Toys
Fashion	Earrings	Hobby	Music	Toys & Games	Dice
Fashion	Glasses	Hobby	R/C Vehicles	Toys & Games	Games
Fashion	Jewelry	Hobby	Robotics	Toys & Games	Mechanical Toys
Fashion	Keychains	Hobby	Sport & Outdoors	Toys & Games	Playsets
Fashion	Rings	Hobby	Hobby	Toys & Games	Puzzles
Fashion	Fashion			Toys & Games	Toy & Game Accessories
Gadgets	Audio			Toys & Games	Toys & Games
Gadgets	Camera				
Gadgets	Computer				
Gadgets	Mobile Phone				
Gadgets	Tablet				
Gadgets	Video Games				
Gadgets	Gadgets				

BUILDING ON THE SHOULDERS OF GIANTS: THE PERFORMANCE OF DERIVATIVE DESIGNS IN AN OPEN COMMUNITY

Alfonso Gambardella
alfonso.gambardella@unibocconi.it
Department of Management & Techology & ICRIOS, Bocconi University, and CEPR London

Erdem Dogukan Yilmaz
erdem.yilmaz@unibocconi.it
Department of Management & Technology, Bocconi University

Abstract

Research has increasingly emphasized the importance of open communities as a driver of innovation performance (Dahlander et al., 2008; Füller et al., 2008; von Krogh and von Hippel, 2006). With the digitization of knowledge- intensive tasks, some companies even go as far as building their business models and innovation activities around the open communities. What is central to these communities is free revealing, which enables users to recombine, modify and integrate the knowledge of other community members, also sometimes referred as remixing (Faraj et al., 2011). Enabled by digital technology, these remixing practices drive generativity in these communities. At the same time, however, remixing might harm innovation as using different material as a source might inhibit creativity. Therefore, it is important to determine when derivative products that are result of remixing are valuable for the community. By drawing from innovation search literature, we hypothesize that remixing is likely to create breakthrough products when derivative products (i) bring together disparate products or (ii) marginally improve products that have a higher degree of cumulativeness. We test these hypotheses in a 3D printable design community called Thingiverse and find strong correlations to support our claims.

1. Introduction

With advances in digitization, innovations have become easily accessible for reuse and recombination for different individuals across the world. To take advantage of the dynamic nature of digital innovation and harness creativity outside of the organization boundaries, organizations in many fields have invested in open communities to bring together the creativity of dispersed individuals. In these communities, the practices of reuse and recombination of innovations is increasingly called “remixing”. In such settings, the free availability of innovations, information or design and digital capabilities to remix constitute an important advantage for the companies in tapping into the knowledge of crowds, as it facilitates collective intelligence and drives the generativity in the community that is the “capacity to produce unprompted change driven by large, varied, and uncoordinated audiences” (Zittrain 2006: 1980).

Remixing can be considered a particular type of online collaboration (Faraj et al., 2011), where it can be defined as the “community’s use of an existing innovation as source material or inspiration to aid in the development of further innovations” (Lessig 2008). Remixing is critical for the sustainability of open communities, as sharing and combining knowledge contributes to the community’s worth as it fosters variation along with providing personal benefits (Jeppesen and Frederiksen, 2006; Murray and O'Mahoney 2007; von Hippel and von Krogh 2006). Remixing of other’s innovations is beneficial for the community as it enables cumulative innovation through refinement and improvement (Murray and O’Mahony 2007). Furthermore, combinations of digital artifacts may lead to new innovations that go beyond the anticipation of original creators and might generate products with extreme value (Yoo et al, 2012; Raustiala and Sprigman 2012). Yet, the generativity of digital innovation might come at the cost of originality (Hill and Monroy-

Hernandez, 2013). Using others' innovation as a source material might inhibit creativity because people tend to become trapped in the existing logic (Sinnreich et al., 2009). An increasing availability of digital goods and tools therefore, might lead to an unintended uses of innovation; it might decrease the originality of output and hurt cumulative innovation (Keen, 2007).

Although we have information on the importance of remixing and its divergent consequences, the number of empirical studies on remixing is limited and it seems that what users actually do when remixing is taken for granted and not given too much attention. In an attempt to address this limitation, we theorize that there exist different paths to creation of derivative products and certain type of paths are more likely to generate breakthrough products. By breakthrough products, we refer to the products that are downloaded by the large amount of the users and potentially create a higher value for customers compared to other products. In innovation search literature, it has been long recognized that the process of drawing from heterogeneous knowledge resources and combining them is one of the most important factors that affect breakthrough generation. It is because the combination of heterogeneous knowledge elements enables to overcome local inefficiencies of a particular knowledge (Fleming, 2001; Ahuja and Lampert, 2001). Our first argument is that, derivative goods are more likely to become breakthroughs (or hits), when they bring together disparate products. Although an invention might be novel and successful, the realization of their impact takes considerable time as it often needs to go through a refinement process (Rosenberg, 2010). Therefore, the impactful invention might be the one that builds further on the novel invention rather than being the one that introduced novelty. Building on this insight, we argue that derivative goods are more likely to become breakthroughs when they marginally improve a product with higher level of cumulativeness. By cumulativeness, we refer to order of derivative in an inheritance hierarchy chain. In this chain, each design that start from scratch has a cumulativeness score of one and their derivatives take the value of two, derivatives of derivatives take the value of three and so on. This conceptualization is similar to sequence-based software versioning schemas, where version of the software is incremented by one unit for each update.

We test our arguments in the largest 3D printable design community called “Thingiverse”. In Thingiverse, it is possible keep track of interrelated innovations as designers in the community are prompted to provide reference to upstream designs if they use them as benchmarks. The sample comprises 278,341 designs from 40968 designers that are listed under 79 different categories.[4]

Understanding the creation of breakthroughs through remixing is important as maximizing the value and number of high quality contributions, rather than the number of average outcomes is very important for the survival and the prosperity open communities (Lee and Cole, 2003; Kogut & Metiu, 2001). Furthermore, although digital technologies have heightened the role of communities and made it the central focus of many firms’ innovation activities, many of these communities have been observed to collapse as they are intensely populated by low-quality information and products (Eaton et al, 2015). In digital settings where user or customer attention is a scarce resource (Sorensen, 2017), companies should simultaneously create new content and make sure that content is valuable for the users. In essence, companies should seek to gain the advantages of generativity from distributed innovation while at the same time they might need to control generativity to make sure that it plays out in their own interests. Identifying and encouraging certain types of remixes,

[4] Our dataset also comprises uncategorized designs.

therefore, might help companies to ease this tension between generativity and originality and maintain acceptable control.

This paper makes several contributions to the literature. First, it contributes to the literature on knowledge collaboration in open communities (Von Hippel, 2005; Lerner and Tirole, 2002; Faraj et al., 2011; Dahlander and Piezunka, 2014). The previous literature on knowledge collaboration in open communities has not focused on dynamics of knowledge collaboration where contributions are iterative by their nature and recombined in complex ways (Faraj et al, 2011). This paper contributes to this area of literature by identifying different remixing practices and their associated outcomes.

Beyond the literature on open communities, this study also contributes to the large body of work on innovation search. Researchers in innovation management and economics of innovation fields have intensively examined search processes underlying the creation of products that give rise to exceptionally high impact. Some researchers find support for the argument that unfamiliar or atypical combinations of knowledge enables the possibility of exceptionally well performing outcomes with greater variance (Fleming, 2001; Katila and Ahuja, 2002; Nelson and Winter, 1982). On the other hand, a handful of researchers have argued that deep knowledge and narrow combinations in one domain yields to high impact outcomes (Gardner, 2011; Hayes, 2013; Simonton, 1999). In this paper, we find that seemingly contradicting paths in creation of breakthroughs co-exist. This results also resonates well with the findings of Kaplan and Vakili (2015) who also show that breakthroughs might be created by both distant and local search.

Finally, this study provides a more direct view into the concept of search in the literature. Most studies treat the notion of knowledge recombination in abstract or metaphorical terms, primarily through the use of citations (e.g., Carnabuci & Operti, 2013). Nevertheless, there is actually little evidence that previously cited work plays a significant role in the creation of current innovation. Patent citations may be retrospective and the result of suggestions by examiners, rather than the chronology of other innovations that the innovation at hand builds upon. By combining self-reported citations with our similarity measure, we also make a methodological contribution to the search literature.

2. Open Communities

Digital technology has lowered barriers to innovate, thus various individual actors have started to participate in innovation activity. Consequently, innovation has become more complex and distributed among different individuals (Yoo et al, 2012). Open communities, bring together these geographically dispersed individuals to support innovation activity. The first clear example of this approach has been the rise of the free/open source software movement. With the rise of the internet, programmers from all around the world find one another and collaborate with each other using different tools. This open source movement has moved beyond voluntary participation, and many firms have started to encourage and sponsor users to participate in these communities (Lerner, 2013).

In open communities, users often freely reveal their innovations to both peer users and firms without any monetary compensation. Research has shown that individuals justify their investments in the development of free innovations with benefits they derive from their intrinsic motivations (Lakhani & Wolf, 2003) and from the use of their own innovations (Gambardella, Raasch, & von Hippel, 2016). For example, survey-based research on the motives of programmers' participation to open software development projects points to the presence of these intrinsic motivations that encompass enjoyment derived from the intellectual challenge and its completion (Lakhani & Wolf, 2005). Moreover, the free revealing of a high-quality code can also increase a programmer's reputation in the eyes of his peers, which may also lead to an increase in the programmer's value on the job market (Lerner & Tirole, 2002). These studies show that the benefits of contributing to public goods outweigh the benefits that other participants derive from free-riding on the innovation outcome.

In general, communities need a critical mass to flourish because individuals function as different pieces of the puzzle and improve others' innovation performance with their feedback (Wasko, Teigland, & Faraj, 2009). Furthermore, increasing diversity in the community is also an important factor that affects the ultimate success. The varying degrees, and the diversity of knowledge and skills increase the likelihood of successful innovation outcomes (Surowiecki, 2004; Zittrain 2006). Likewise, diversity of knowledge is associated with increasing contribution (Ren et. al., 2015). However, while increasing the quantity and diversity of members, firms also need to sustain participation by existing members in order to remain successful (Dahlander & Piezunka, 2014; Iriberri & Leroy, 2009). This can be achieved through monetary awards (Boudreau & Lakhani, 2009), the creation of a shared identity (Nambisan & Baron, 2009) or improvements in the creative process (Füller, Matzler, & Hoppe, 2008).

Another important factor that affect survival and growth of open communities is knowledge collaboration (Faraj et al., 2011). In open communities, knowledge collaboration takes a variety of forms; in Wikipedia users modify and integrate the knowledge of others (Kane and Ransbotham, 2016). On ccMixter and Scratch, users remix each other's creations and on Sourceforge.net, programmers co-develop software applications (Hill and Monroy-Hernandez, 2013). Therefore, in these communities, knowledge is dynamically created by collaborative melding of ideas and contributions of others. Access to diverse knowledge in user communities enhances recombination and reuse of ideas and solutions and boosts innovation (Hargadon and Sutton, 1997). These different types of innovations represent the fundamental underlying dynamics of innovations in open communities as characterized by the generative view.

Despite the fact that remixing, is a well-established phenomenon in open community setting, research into the actual mechanism of remixing has been limited. Kyriakou et al. (2017) distinguishes between remixing for customization and remixing for innovation and particularly focus on the former one. They argue that meta-models that facilitate the creation of derivatives are particularly important for the generativity of the platform as they are easier to modify. However, while their study mostly focuses on the generativity, they do not inform us so much about the value created by the derivative products. Indeed, Hill and Monroy-Hernández (2013) have argued that parent works that are generative simultaneously lead to less novel derivatives due to their associated characteristics. In a recent paper, Flath et al. (2017) recognized different ways through which derivative products are generated. However, they do not focus on the potential divergent outcomes of different paths of remixing. In this paper, rather than focusing on the characteristics of the parent design, we focus on the different processes through which derivative products are generated as well as their resulting outcomes.

3. Hypotheses

This paper focuses on the performance of derivative products within an open community. The main promise of remixing is both generativity and increased creativity by enabling recombination. The previous literature on peer production and remixing, however, has expressed skepticism about the fundamental promise of remixing and resulting derivative products. They argue that factors associated with increased remixing are also associated with low quality work and remixing actually reduces incentives to investing in learning and innovations (Keen 2007; Hill and Monroy-Hernández, 2013). Potential explanations for these observations might be that individuals tend to fixate: they might get pre-occupied by the ideas that are embedded in the source material (Smith, Ward, and Schumacher 1993). Furthermore, the crowd usually lacks expertise in the task domain, and actors might not have enough capability or intrinsic motivations to generate something new. The users in a crowd might rather prefer to free-ride on the other's innovations which is a common observation in the open community setting (Krishnamurthy, 2002). Although, previous studies approve the promise of remixing as a generative process, they do not confirm its role as a constrictive tool for generation of creative ideas.

In this paper, however, we take into account the fact that not all of the derivative products are created equally. In some cases, there will be conformity to the features of a predecessor, and slight improvements of innovations, or sometimes a high or moderate modification of these features. In other cases, derivative products might bring together separate parent works. We hypothesize that these different types of recombinations are likely to lead to different outcomes. In particular, we argue that derivative products that bring together separate parent works or build upon designs that has a higher degree of cumulativeness are more likely to be breakthroughs.

The value of reuse and recombination has been recognized in the open innovation literature. Freely revealing and combining the code is a common practice in open-source development projects (Lakhani and von Hippel, 2004; Haefliger et al., 2008; Dabbish et al., 2012) and the literature finds that code reuse significantly improves productivity and quality in software development (e.g. Haefliger et al., 2008). However, the term we refer as remixing differs from other type of reuse in open communities. Unlike other forms of peer production, derivative designs almost always have identifiable creators (Sinnreich, 2010) and it is possible to track the evolution of work through chains of remixes (Cheliotis and Yew, 2009). Second, previous studies that examine reuse have mostly focused on more "functional" tasks that can be easily modularized and objectively evaluated. However, here we focus on a task that is at the same time both functional and creative; digital product design. The design task involves underspecified goals and operators. In contrast to well-defined problems where the initial states, evaluation functions, and transformation functions are well specified, there is a limitless amount of knowledge that may enter into a design space (Simon, 1973). The resulting outcomes must be evaluated by colleagues in the same field before it can be considered innovation in the larger sense (Csikszentmihályi, 1999; Simonton, 2010).

In this paper, we particularly focus on extreme value outcomes or breakthroughs generated through remixing. The term "breakthrough" has been used in two distinct ways in the innovation literature. The first one is related to path-breaking discontinuity in technological development that makes existing technologies obsolete (Henderson & Clark, 1990; Tushman & Anderson, 1986; Lavie, Stettner, & Tushman, 2010). This sort of breakthroughs provide basis for new technologies (Trajtenberg, Henderson, & Jaffe, 1997). Second, the term "breakthrough" has been used to refer to the impact of innovations on subsequent innovations, which is often times measured by citations (e.g. Singh and Fleming, 2010; Conti et al. 2013). In this paper, our usage of "breakthrough" is closer to the second one and aligned with the definition of "hit", "killer" or "blockbuster" products in the platform and product innovation literatures. Products of this kind are used to attract a big mass of consumers to the platform and they have become an essential factor for purchasing hardware devices.

Convergent Remixes

We first consider the diversity of parent designs used in creation of derivative designs. In their recent paper, Flath et al (2017) refer this kind of remixes as convergent remixes. The simplest form of this type of derivative designs inherits at least from two designs. These type of derivative products are, on average, likely to be less imitative as they combine knowledge embedded in different designs. A handful papers in the innovation literature has shown that breakthroughs can be produced by bringing together distant or diverse knowledge (Weisberg 1999; Taylor & Greve 2006; Singh and Fleming, 2010). This is because bringing together different bodies of knowledge and recombining them might help an individual to avoid becoming trapped in an inefficient local peak of separate knowledge elements (Ahuja & Lampert, 2001; Levinthal & March, 1993). Because creative work involves in a recombination process, and when individuals have an access to a broad set of knowledge components they can experiment at a broader scope with greater flexibility.

In the psychology literature also, it has been shown that drawing from diverse knowledge in creative thinking leads to extreme value outcomes as combining knowledge increases the chances of finding novel and useful ideas through discovery of atypical links between them (Simonton 2004). In this creative process, individuals discover or envision potential connections among different knowledge elements either deliberately or subconsciously and retain the combinations that are novel and useful (Schilling and Green, 2011). Seemingly random connections of disparate knowledge elements had led to the most famous scientific breakthroughs (Simonton, 1995). Likewise, product design literature provides corroborating insights; combining previous designs of others into the design of a new product generates more useful and novel designs (Nickerson, Sakamoto & Yu, 2011). Taken together, we expect that,

Hypothesis 1: *Derivative products that bring different parent products together are more likely to be breakthroughs*

Cumulativeness of Parent Design

Schumpeter (1942) was one of the first scholars that recognized the importance of the cumulative aspect of innovation. With increasingly cumulative innovation, economy-wide returns to innovation also increase because inventors rely and build on the previous ideas of others when they create their own innovations. Caballero and Jaffe (1993) developed a model in which new ideas are produced by using private research effort and the public stock of existing knowledge as inputs. In their model, the productivity of private inputs in research varies as a function of aggregate public knowledge. These studies highlight the importance of cumulative knowledge that is available for inventors to build upon.

Rosenberg (2010) put emphasis on the importance of the availability of knowledge in economic experimentation for technological advancement. He argues that there are many things that cannot be observed or known ex-ante and therefore, there should be experiments to determine which alternatives are worth pursuing further and which are not. In his conceptualization, learning occurs only after observing the interaction of the product with the market (Greenstein, 2012). Proceeding sequentially and incrementally is less costly and also one can keep variety of options alive that enhances spontaneous discovery (Rosenberg, 2010). Economic experiments are important for value capture as they are the way resolving market, technological, and organizational uncertainties (Stern, 2006).

Availability of experimentation is particularly important for industries such as product design, software, computer and semiconductors where innovation is sequential and complementary. Sequential means each innovation builds on the preceding one, whereas complementary means that each individual pursues different research lines in the innovation process and thus enhances the probability that innovation arises (Bessen and Maskin, 2009). Similar arguments have been made in the open community setting as well. For instance, Raymond (1999) describes "Linus' law"— "with enough eyeballs, all bugs are shallow"— which suggests that software that is a result of collaborative effort will contain less bugs and therefore will be higher quality. Furthermore, this type of collaboration might also enable users to overcome functional fixedness. Functional fixedness occurs when solvers are "focusing on some function of an object while overlooking another necessary for problem-solving" (Arnon & Kreitler, 1984, p. 11).

As a result, one might expect the value of a design to be higher with an increasing level of cumulativeness. Here, by cumulativeness, we refer to the order of parent design in a "chain" network structure. Increasing cumulativeness in parent design might help users to concentrate search to a narrow field. Thus, users might focus their interest on local anomalies in a particular design as the number of drawbacks in an original design likely to be reduced as it evolves on the remix chain. Identifying local anomalies and fine-tuning existing innovations is a way through which breakthrough innovations are generated (Audia & Goncalo 2007). Thus, while identifying and improving the overlooked aspects of the parent work, users could perpetuate the value of a parent design. Therefore, the cumulativeness of parent design is likely to be more important only if derivative design preserves the main features and improve marginally. This leads to the following hypothesis:

Hypothesis 2: *The probability of generating a breakthrough design increases with the cumulativeness of the parent design*

4. Methods

Sample and Data

3D printing technology carries a revolutionary potential due to its ability to produce rapid prototypes, reduce production costs, and create products that are self-assembled (Lipson and Kurman, 2013). The greatest promise of 3D printing technology is its ability to transform digital information into physical products. Thus, 3D printing blurs the distinction between digital and physical products. 3D printing technology enables trial-and-error experiments on physical artifacts Similar technologies have been traditionally found in areas ranging from mechanical designs to the design of integrated circuits. However, today they are moving over into individuals' daily lives via 3D printing technologies. Individuals can now easily create a replica of an existing product, create digital remixes and manufacture a physical version of this remix at their homes.

In our study, we use data from Thingiverse, one of the largest online communities dedicated to 3D printable designs. These 3D printable designs are licensed under the Creative Commons licenses or GNU General Public License. More than 97% of the designs on Thingiverse allow users to allow reusing their material and all of them allows for 3D printing. While most of the open software and hardware communities expect a contribution to open projects, Thingiverse provides a common ground from which derivatives of products may form. As design interactions are recorded in the community, it creates opportunities for research on the way designs evolve (Nambisan 2003).

In the platform, designers share necessary files to print or remix the designs along with one or more images of their designs. These images might contain snapshots of drawings in design programs or printed output. The website provides a variety of data such as the designer name, the number of downloads, views, likes, remixes, launch date of the designs, date and text of the comments, and from which product(s) the focal design is remixed. The data was web scraped by us (i.e., data was systematically harvested from the internet) on May 1, 2018. Since time is required for a design to attain a certain number of downloads, remixes, and comments once they are made public, we allow two years of lag between the posting of the designs and measurement of the variables to ensure sufficient time to measure the design's value, based on the number of downloads. The data contains all of the design submissions made to different categories by over 74,737 individuals during an eight year period, from 2008 through 2016, consisting of a total of 312,100 submissions.[5] However, we dropped individuals with single design from our dataset. The final data compromises 278,341 design- level observations of 40968 individuals.

We use three different samples in our analysis. First, we begin with the full population of designers and collect all designs introduced to the community. Second, we use the only-derivative sample. Restricting the sample allows us to more strictly compare the different processes in creation of

[5] Thingiverse claimed to have over 1 million designs by the end of 2016, however, designs that are not in our dataset neither submitted to ten major categories nor listed under the URL https://www.thingiverse.com/newest

derivative designs that lead to breakthroughs. Since there are many more non-derivative designs, this also prevents us from generating standard errors whose precision only results from oversampling of non-derivative designs. Furthermore, we are not able to calculate similarity measure for each and every design due to specific file formats in our dataset which compel us to use the third sample.

The dataset also includes the designs themselves for the derivative designs and their parents. Design files might have different formats, but users predominantly prefer STL by 89% and SCAD by 9%. The remaining two percent have a variety of formats, such as DXF, BLEND, OBJ, PLY and so on. We adopt an algorithm crated by Kazhdan et al. (2003) to measure improvement over parent designs. This algorithm provides an objective method of improvement that can assist us with understanding the differences between parent and derivative designs. In particular, the algorithm represents 3D designs with rotation and scale invariant descriptors and later summarizes data into a vector. Then, by comparing the vectors of different designs we have identified how similar the parent – derivative pairs are. Some of the designs have file formats that contain two dimensional information about the design form or are broken files, which impede us to use our measure. We dropped those designs and their observations from our dataset. Before using the algorithm however, we cleaned the data in several steps. First, we checked whether designs have single or multiple digital files. In 83% of the cases designs have a single file.[6] For the designs that have a single file, we checked whether it is watertight or not. If it is watertight, it indicates that information in the design file represents the final form of the design at hand. If it is not watertight, the information in

[6] Designs might have different file formats that represent same design. File format conversation is relatively easy with standard 3D design tools, yet users might prefer to provide designs with several files.

a single digital file might still represent a single design, contain connected components of the same design or different versions of the design at hand. To determine which case is prevalent, we first checked the face adjacency of different bodies of design in a single file. If there is face adjacency, we added up different components by face adjacency to obtain a single body. We digitally assembled components of 1289 design and randomly checked two hundred of them to decide whether they represent the final form of design. Out of two hundred designs, 196 of them were similar to the 3D printed photos or snapshots provided by designers. We visualized each of the remaining 239 designs with a single digital file, and determined whether the design files contain different versions of the same design or not. If they contained different versions, we chose the highest similarity score as the true score of similarity in our comparison of parent – derivative pairs. In the remaining cases where the design was not water tight, the design file represented a single design body with a complex form or consisted of separate design bodies that are related to each other -- for instance, the design of a chess set or landscape. If the file represented a single body, we used a similar methodology to the one that we used for other designs. On the other hand, if the design file contained information for related but separate bodies of designs, we base our evaluations on the characteristic of the derivative design. If the derivative design consisted of a single body, then we compared its shape with each design body contained in the parent design digital file. If it consisted of multiple bodies (one could imagine comparing two chess sets), we compared each piece with the closest match in the digital file of the parent design and took the average.

An analogical methodology is used for designs with multiple digital files. In some cases (96%), different design files contained different components of the same design. Accordingly, we checked whether they could be merged from adjacent faces. We again randomly checked 200 of the designs to see whether they were similar to 3D printed photos or snapshots provided by designers and in 189 of the cases they matched the photos or snapshots. Similar to the case of single digital files, some of the digital files represented different versions of the same design and we kept the parent version-derivative pair with the smallest similarity score (more than 3% of the cases). For some of the designs, different components could be merged with the main body of the design. In those cases (119 in total), we keep all of the potential combinations and again, choose the parent version-derivative pair with the highest similarity score. We used the shapely and trimesh Python packages to manipulate design files.

Measures

Dependent Variable. Our primary dependent variable, *Breakthrough,* takes the value of one if a design is in the top 5% of the distribution in terms of downloads within the same month-year, and the same sub-category, or zero otherwise. The definition of our *Breakthrough* variable is similar to breakthrough definitions in patented innovations settings where innovations have typically been measured based on the distribution of the number of "forward citations" (e.g. Conti et al, 2013; Kaplan and Vakili, 2015). Furthermore, measuring impact by downloads is also consistent with previous work on open innovation and platforms, which quantifies the subjective qualities and market success of creative work by visualizations, downloads or likes (Crowston, Howison, & Annabi, 2006; Liebowitz 2005). According to our measure, of the 278,341 designs in our sample, 17,312 are breakthroughs.

Independent Variables. Our main variable of interest is *IsDerivative* which indicates whether a design itself is a remix of another project. It is important to note that this variable only measures the materials being copied and, as a result, does not capture "conceptual" remixing as such, borrowing material from any source outside Thingiverse, or inspiration from sources inside Thingiverse without downloading and copying a project. We hypothesized that remixing is likely to create breakthrough designs when the resulting derivative design (i) brings together different parent designs or (ii) marginally remixes parent designs with a higher level of cumulativeness. To operationalize the first case, we define *NumberofParents* and *MultipleParent* variables. While the former variable is a continues variable that accounts for the number of parent designs that the derivative design derived from, the latter one is a dummy variable that takes the value of one if the derivative design has more than one parent. *Cumulativeness* is operationalized by counting the order of derivative design in the remix chain. The degree of improvement is operationalized by a *Similarity* score, which is calculated by the algorithm crated by Kazhdan et al. (2003). For each parent-remix pair, we compute the similarity index that range between zero and one.[7] We interact this variable with both *Cumulativeness* and *IsDerivative* variables to assess the effect of marginally remixing of conventional designs on the probability that design is a breakthrough.

Control Variables. We define an *Experience* variable to control for users' prior experience that might influence both the likelihood of their design being high impact and particular search choices in remixing. *Experience* is measured as the log of the cumulative number of designs by the focal user prior to submitting the current design. Furthermore, we also use a *SelfRemix* variable to control for whether the designer of the parent design and of the derivative design are same. [8] To control for the fact that users might be more or less likely to know which products designs are likely to be

[7] In reality, complete dissimilarity is impossible in design space and our index ranges from x to 1. The upper limit of 1 indicates that derivative designs are exact copies of parent designs.
[8] For the designs that have multiple parents, *SelfRemix* variable takes the value of one only if the designer of the most similar design is the same as the designer of derivative design.

appreciated by the community, we include *Tenure* as the number of days since the first design of the user was introduced. This variable also captures the general passage of time and account for possible experience accumulated outside the open community (Riedl and Seidel, 2016). We also have a variable called *Customizable*, which takes the value of 1 if the design at hand is a metamodel. These metamodels are relatively easy to modify (Kyriakou et al, 2017) as they only require the changing of parameters and thus might be associated with an increasing number of downloads. We also include the *"TimeElapsed"* variable that corresponds to the time interval between the submission of a parent design and its derivative. We also include the lagged value of the *"TimeElapsed"* variable called *"LogTimeElapsed"*.

Empirical Strategy

To estimate the relationship between remixing and the probability of producing a breakthrough design, we use a linear probability model for two reasons. First, focusing on the linear probability model enables us to estimate a model using extensive fixed effects without the exclusion of any observation. Second, the mass point of the dependent variable is far from zero or one. Also, as noted by Angrist & Pischke (2009), there is typically a little qualitative difference between the linear probability model and logit specifications.

When an individual searches for a solution to a particular problem, his search will be guided by what knowledge they possess as a result of previous searches and their natural search tendencies. To account for these differences, we use individual fixed effects along with logarithms of Experience. We also control for the sub-category of the design to ensure we account for any overall differences in breakthrough likelihood and success at the sub-category level. Finally, to control for variations in the platform at the time of design introduction, we include cohort fixed effects. The main specification that we use for the full sample has the following functional form:

$$Breakthrough_{itk} = \beta_1\, IsDerivative_{i} + \beta_2\, IsDerivative_{i} * MultipleParent_{i} +$$

$$\beta_3\, IsDerivative_{i} * Cumulativeness_{i} + Controls + \delta j + \lambda k + \gamma t + uit \; (1)$$

Where δj is individual, λk is sub-category and γt is cohort fixed effect. For the only derivative sample, we use the following specification:

$$Breakthrough_{itk} = \beta_1\, MultipleParent_{i} + \beta_2\, Cumulativeness_{i} +$$

$$Controls + \delta j + \lambda k + \gamma t + uit \; (2)$$

For the third sample that consists of derivative designs with similarity score, we interact *Cumulativeness* variable with *Similarity* score.

$$Breakthrough_{itk} = \beta_1\, MultipleParent_{i} + \beta_2\, Cumulativeness_{i} +$$

$$\beta_3\, Cumulativeness_{i} * Similarity + Controls + \delta j + \lambda k + \gamma t + uit \; (2)$$

In the alternative specification, we replace *MultipleParent* with *NumberofParents*.

5. Results

We use three different samples to identify the effect of the different remix practices on the probability that a design is a breakthrough. First, we use our full sample which allows us to observe the main effect of being a derivative design on the probability that a design is a breakthrough. Second, we use the derivative-only sample to obtain more precise results. Finally, we use our third sample which consists of derivative designs for which we were able to measure similarity. The third sample allows us to illustrate the mechanism for the second hypothesis. Although we do not argue that our analysis is free of selection bias, we believe that these analyses along with robustness tests collectively provide strong correlations between the derivative designs that are the result of different practices and their performance.

Tables 1 provides descriptive statistics whereas Tables 2, 3 and 4 provide the correlation statistics for this study's variables. All regressions use robust standard errors that were clustered at the designer level given the possibility that the choice of remixing might be correlated with designer ability. Table 5 shows the baseline results of the linear probability regressions without interactions. Model 1 includes designer, sub-category, and cohort fixed effects; model 2 adds control variables; model 3 shows the correlation between derivative designs and breakthrough variables, model 4 adds designer, sub-category, and cohort fixed effects variables in addition to the derivative variable and model 5 incorporates control variables. In all tables, the results are robust to heteroscedasticity and clustered at the designer level. While we do not offer any theory about the main effect of being a derivative, the fully saturated model, Model 5, indicates that derivative products have a 2.36 percentage points higher probability of being breakthrough than non-derivative designs.

Table 6 shows results from the linear probability regressions that test H1 and H2 by estimating the full sample. Model 1 and 2 test H1 by adding the interaction of derivative *(IsDerivative)* and multiple parent *(MultipleParent)* dummies. We focus on model 2 for the main effect of having multiple parents on the probability that a design is a breakthrough. Consistent with our first hypothesis, we find that when a derivative design combines two or more designs it has a 7.09 percentage points higher probability to be a breakthrough compared to non-derivative designs. The results are also consistent when we use our alternative measure of convergent remixes, NumberOfParents. In Model 2 of Table 7, we find that a unit increase in the number of parents increases the probability that the derivative design is a breakthrough by 0.0329 percentage points. Overall, this result is consistent with the intuition that incorporating diverse knowledge in creative thinking leads to extreme value outcomes.

Our second hypothesis concerns the effect of cumulativeness on the probability that a design is a breakthrough. Models 3 and 4 test H2 by adding the interaction between the derivative *(IsDerivative)* dummy and cumulativeness *(Cumulativeness)*. In model 3 and 4, the interaction terms are positive and significant. The results from Model 4 indicate that a unit increase in the cumulativeness of the parent design increases the probability that the derivative design is a breakthrough by 0.00841 percentage points. We also note that the main effect of the derivative dummy becomes insignificant after including the interaction term, indicating that cumulativeness is disproportionally driving the performance of derivatives. This result resonates well with the famous expression that *"A dwarf standing on the shoulders of a giant may see farther than a giant himself."*

Next, we replicate our analysis with the only-remix sample. We find consistent support for all two hypotheses across all models in tables 8 and 9. We further argued that it is more beneficial to make marginal remixes if the parent design at hand has a higher level of cumulativeness. This in turn, would allow designers to perpetuate the value of parent designs. As illustrated in Model 4, similarity has a negative first-order effect on the probability that the derivative design at hand is a breakthrough. The interaction between *Cumulativeness* and *Similarity* is positive and significant. When cumulativeness is equal to or higher than five, the interaction ameliorates the negative first-order effect of the given level of similarity.

Robustness Checks

We now turn to discussing under which assumptions the estimated coefficient on our independent variables can be interpreted as a causal effect. The identifying assumption underpinning our analysis so far has been that after controlling designer, cohort and sub-category fixed effects, we can attribute any systematic differences in probability of producing a breakthrough to different remixing practices that create superior performance because of the quality of resulting output. In an ideal setting, we would run an experiment in which designers are randomly assigned to create derivative products by combining designs with different cumulativeness while others would be assigned to create designs from scratch. Then, we would distribute those designs to the open community and observe the outcomes. Such an experiment would allow us to estimate the effect of different remixing practices by comparing the performance of derivative designs with designs in the control group. Clearly, in our setting, parent designs are not chosen randomly but the choice might be a function of the general appeal and popularity.

If we are identifying a higher likelihood for being a breakthrough for derivative designs because of parent design's inherent popularity or an increasing trend for a particular type of product, we should see a negative correlation between the length of time elapsed between the introduction of parent and derivative designs and the probability that design being a breakthrough. We report the results from a set of regression that include the variable *TimeElapsed* in Table 7. Yet, our results show no correlation. If we log transform the *TimeElapsed* variable, instead, we see a positive correlation.

Another concern might be that above models compare derivatives of very different types of parent designs. To address this concern, we introduce a parent fixed effect in Table 8. The significant and positive coefficient remains unchanged on the *MultipleParent* variable. We do not present any result for *Cumulativeness* as parent fixed effects will soak up the variation on *Cumulativeness*. Furthermore, we created a dummy variable for each parent design – designer pairs. Inclusion of these pairs into our regression do not change our main results.

6. Conclusion

In this paper, we focus on remixing of 3D printable objects as a type of knowledge collaboration in open communities. This article's primary contribution is the determination of factors that affect the performance of derivative designs. Using an innovation search lens to understand the mechanisms through which derivatives are created is an important distinction, as they have different performance implications. Our results indicate that when derivative products bring together separate designs or if they marginally improve designs with a higher degree of cumulativeness, they are more likely to be breakthroughs.

These findings make contributions to different streams of literature. Our first contribution is to the literature on knowledge production in open communities (Dahlander and Piezunka, 2014; Faraj et al., 2011). While knowledge collaboration in open communities has been examined extensively, attention on remixing has been limited. This paper contributes to this literature by showing the relationship between different remixing practices and their performance. While derivative designs carry value for their designers, their influence will be quite limited unless others appreciate them. Therefore, companies might design systems that encourage certain types of remixing and allow certain types to be posted to the community. This study also makes contributions to the literature on knowledge search and recombination. The results suggest that both bringing together separate designs (i.e. search scope) and marginally improving design with a higher cumulativeness (i.e.

search depth) have positive relationships with a design’s performance which supports earlier findings in innovation management literature.

Another contribution of the study is the adoption of a novel algorithm for measuring the degree of improvement that goes beyond patent citations. Whereas most recent attempts at content-based analysis of innovation have focused on the textual analysis of patents, this paper demonstrates that unpatentable ideas can be quantified, and exploits a data-rich setting to study how high impact outcomes might be achieved.

Our study is not without limitations. First, although we offer three robustness checks, we are limited in our ability to control for self-selection. Second, the definition (and measure) of breakthroughs may sound deterministic. However, this definition captures the idea that because of the skewed distribution of inventions, only a few are ever really valuable for the community.

References

Ahuja, G., & Morris Lampert, C. (2001). Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic management journal*, *22*(6-7), 521-543.

Angrist, J., & Pischke, J. S. (2009). Mostly harmless econometrics: an empiricists guide.

Arnon, R., & Kreitler, S. (1984). Effects of meaning training on overcoming functional fixedness. *Current psychological research & reviews*, *3*(4), 11-24.

Audia, P. G., & Goncalo, J. A. (2007). Past success and creativity over time: A study of inventors in the hard disk drive industry. *Management Science*, *53*(1), 1-15.

Bessen, J., & Maskin, E. (2009). Sequential innovation, patents, and imitation. *The RAND Journal of Economics*, *40*(4), 611-635.

Boland Jr, R. J., Lyytinen, K., & Yoo, Y. (2007). Wakes of innovation in project networks: The case of digital 3-D representations in architecture, engineering, and construction. *Organization science*, *18*(4), 631-647.

Boudreau, K., & Lakhani, K. (2009). How to manage outside innovation. *MIT Sloan management review*, *50*(4), 69.

Boudreau, K. J., & Lakhani, K. R. (2015). “Open” disclosure of innovations, incentives and follow-on reuse: Theory on processes of cumulative innovation and a field experiment in computational biology. *Research Policy, 44*(1), 4-19.

Caballero, R. J., & Jaffe, A. B. (1993). How high are the giants' shoulders: An empirical assessment of knowledge spillovers and creative destruction in a model of economic growth. *NBER macroeconomics annual*, *8*, 15-74.

Carnabuci, G., & Operti, E. (2013). Where do firms' recombinant capabilities come from? Intraorganizational networks, knowledge, and firms' ability to innovate through technological recombination. *Strategic Management Journal*, *34*(13), 1591-1613.

Cheliotis, G., & Yew, J. (2009, June). An analysis of the social structure of remix culture. In *Proceedings of the fourth international conference on Communities and technologies* (pp. 165-174). ACM.

Crowston, K., Howison, J., & Annabi, H. (2006). Information systems success in free and open source software development: Theory and measures. *Software Process: Improvement and Practice*, *11*(2), 123-148.

Csikszentmihalyi, M. (1999). 16 implications of a systems perspective for the study of creativity. In *Handbook of creativity*(pp. 313-335). Cambridge University Press.

Conti, R., Gambardella, A., & Mariani, M. (2013). Learning to be edison: Inventors, organizations, and breakthrough inventions. *Organization Science*, *25*(3), 833-849.

Dabbish, L., Stuart, C., Tsay, J., & Herbsleb, J. (2012, February). Social coding in GitHub: transparency and collaboration in an open software repository. In *Proceedings of the ACM 2012 conference on computer supported cooperative work* (pp. 1277-1286). ACM.

Dahlander, L., Frederiksen, L., & Rullani, F. (2008). Online communities and open innovation. *Industry and innovation*, *15*(2), 115-123.

Dahlander, L., & Piezunka, H. (2014). Open to suggestions: How organizations elicit suggestions through proactive and reactive attention. *Research Policy*, *43*(5), 812-827.

Eaton, B., Elaluf-Calderwood, S., Sorensen, C., & Yoo, Y. (2015). Distributed tuning of boundary resources: the case of Apple's iOS service system. *MIS Quarterly: Management Information Systems*, *39*(1), 217-243.

Faraj, S., Jarvenpaa, S. L., & Majchrzak, A. (2011). Knowledge collaboration in online communities. *Organization science*, *22*(5), 1224-1239.

Flath, C. M., Friesike, S., Wirth, M., & Thiesse, F. (2017). Copy, transform, combine: exploring the remix as a form of innovation. *Journal of Information Technology*, *32*(4), 306-325.

Fleming, L. (2001). Recombinant uncertainty in technological search. *Management science*, *47*(1), 117-132.

Füller, J., Matzler, K., & Hoppe, M. (2008). Brand community members as a source of innovation. *Journal of Product Innovation Management*, *25*(6), 608-619.

Gardner, H. (2011). *Creating minds: An anatomy of creativity seen through the lives of Freud, Einstein, Picasso, Stravinsky, Eliot, Graham, and Ghandi*. Basic Civitas Books.

Gavetti, G. (2012). PERSPECTIVE—Toward a behavioral theory of strategy. *Organization Science*, *23*(1), 267-285.

Goldfarb, A., & Tucker, C. (2017). *Digital economics* (No. w23684). National Bureau of Economic Research.

Greenstein, S. (2012). Economic experiments and the development of Wi-Fi. *Advances in Strategic Management, 29*, 3-33.

Haefliger, S., Von Krogh, G., & Spaeth, S. (2008). Code reuse in open source software. *Management science*, *54*(1), 180-193.

Hargadon, A., & Sutton, R. I. (1997). Technology brokering and innovation in a product development firm. *Administrative science quarterly*, 716-749.

Hayes, J. R. (2013). *The complete problem solver*. Routledge.

Henderson, R. M., & Clark, K. B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative science quarterly*, 9-30.

Hill, B. M., & Monroy-Hernández, A. (2013). The remixing dilemma: The trade-off between generativity and originality. *American Behavioral Scientist*, *57*(5), 643-663.

Iriberri, A., & Leroy, G. (2009). A life-cycle perspective on online community success. *ACM Computing Surveys (CSUR)*, *41*(2), 11.

Jeppesen, L. B., & Frederiksen, L. (2006). Why do users contribute to firm-hosted user communities? The case of computer-controlled music instruments. *Organization science*, *17*(1), 45-63.

Kazhdan, M., Funkhouser, T., & Rusinkiewicz, S. (2003, June). Rotation invariant spherical harmonic representation of 3 d shape descriptors. In *Symposium on geometry processing* (Vol. 6, pp. 156-164).

Kane, G. C., & Ransbotham, S. (2016). Content as community regulator: The recursive relationship between consumption and contribution in open collaboration communities. *Organization Science*, *27*(5), 1258-1274.

Kaplan, S. (2008). Cognition, capabilities, and incentives: Assessing firm response to the fiber-optic revolution. *Academy of Management Journal*, *51*(4), 672-695.

Kaplan, S., & Vakili, K. (2015). The double-edged sword of recombination in breakthrough innovation. *Strategic Management Journal*, *36*(10), 1435-1457.

Katila, R., & Ahuja, G. (2002). Something old, something new: A longitudinal study of search behavior and new product introduction. *Academy of management journal*, *45*(6), 1183-1194.

Kyriakou, H., Nickerson, J. V., & Sabnis, G. (2017). Knowledge reuse for customization: Metamodels in an open design community for 3D printing. *MIS Quarterly: Management Information Systems*, *MIS Quarterly, 41(1),* 315-332.

Keen, A. (2007). *The Cult of the Amateur: How Today's Internet is Killing Our Culture.* Crown Business, 3rd printing edition.

Krishnamurthy, S. (2002). Cave or community?: An empirical examination of 100 mature open source projects.

Kogut, B., & Metiu, A. (2001). Open-source software development and distributed innovation. *Oxford review of economic policy*, *17*(2), 248-264.

Lakhani, K. R., & Wolf, R. G. (2003). Why hackers do what they do: Understanding motivation and effort in free/open source software projects.

Lakhani, K. R., & Von Hippel, E. (2004). How open source software works:"free" user-to-user assistance. In *Produktentwicklung mit virtuellen Communities* (pp. 303-339). Gabler Verlag.

Lavie, D., Stettner, U., & Tushman, M. L. (2010). Exploration and exploitation within and across organizations. *The Academy of Management Annals*, *4*(1), 109-155.

Liebowitz, J. (2005). Linking social network analysis with the analytic hierarchy process for knowledge mapping in organizations. *Journal of knowledge management*, *9*(1), 76-86.

Lipson, H., & Kurman, M. (2013). *Fabricated: The new world of 3D printing*. John Wiley & Sons.

Lee, G. K., & Cole, R. E. (2003). From a firm-based to a community-based model of knowledge creation: The case of the Linux kernel development. *Organization science*, *14*(6), 633-649.

Lerner, J. (2013). The comingled code: Open source and economic development. *MIT Press Books*, *1*.

Lerner, J., & Tirole, J. (2002). Some simple economics of open source. *The journal of industrial economics*, *50*(2), 197-234.

Lessig, L. (2008). *Remix: Making art and commerce thrive in the hybrid economy*. Penguin.

Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic management journal*, *14*(S2), 95-112.

Marx, M., Strumsky, D., & Fleming, L. (2009). Mobility, skills, and the Michigan non-compete experiment. *Management Science*, *55*(6), 875-889.

Murray, F., & O'Mahony, S. (2007). Exploring the foundations of cumulative innovation: Implications for organization science. *Organization Science*, *18*(6), 1006-1021.

Nambisan, S. (2003). Information systems as a reference discipline for new product development. *MIS quarterly*, 1-18.

Nambisan, S., & Baron, R. A. (2009). Virtual customer environments: testing a model of voluntary participation in value co-creation activities. *Journal of product innovation management*, *26*(4), 388-406.

Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital Innovation Management: Reinventing innovation management research in a digital world. *Mis Quarterly*, *41*(1).

Nickerson, J. V., Sakamoto, Y., & Yu, L. (2011, May). Structures for creativity: The crowdsourcing of design. In *CHI workshop on crowdsourcing and human computation* (pp. 1-4).

Nelson, R. R., & Winter, S. G. (1982). The Schumpeterian tradeoff revisited. *The American Economic Review*, *72*(1), 114-132.

Partha, D., & David, P. A. (1994). Toward a new economics of science. *Research policy*, *23*(5), 487-521.

Raymond, E. (1999). The cathedral and the bazaar. *Knowledge, Technology & Policy*, *12*(3), 23-49.

Raustiala, K., Sprigman, C., & Sprigman, C. J. (2012). *The knockoff economy: How imitation sparks innovation*. Oxford University Press.

Ren, Y., Chen, J., & Riedl, J. (2015). The impact and evolution of group diversity in online open collaboration. *Management Science*, *62*(6), 1668-1686.

Ries, E. (2011). *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*. Crown Books.

Riedl, C., & Seidel, V. P. (2016). Design myopia and vicarious learning from good versus bad examples in creative design competitions. In *Academy of Management Proceedings* (Vol. 2016, No. 1, p. 12773). Briarcliff Manor, NY 10510: Academy of Management.

Rosenberg, N. (2010). Economic experiments. In *Studies On Science And The Innovation Process: Selected Works of Nathan Rosenberg* (pp. 367-389).

Rosenkopf, L., & Almeida, P. (2003). Overcoming local search through alliances and mobility. *Management science*, *49*(6), 751-766.

Rosenkopf, L., & Nerkar, A. (2001). Beyond local search: boundary‐spanning, exploration, and impact in the optical disk industry. *Strategic Management Journal*, *22*(4), 287-306.

Schilling, M. A., & Green, E. (2011). Recombinant search and breakthrough idea generation: An analysis of high impact papers in the social sciences. *Research Policy*, *40*(10), 1321-1331.

Schumpeter, J. (1942). Creative destruction. *Capitalism, socialism and democracy*, *825*, 82-85.

Simon, H. A. (1973). Applying information technology to organization design. *Public Administration Review*, *33*(3), 268-278.

Simonton, D. K. (1995). Foresight in insight? A Darwinian answer.

Simonton, D. K. (1999). *Origins of genius: Darwinian perspectives on creativity*. Oxford University Press.

Simonton, D. K. (2004). *Creativity in science: Chance, logic, genius, and zeitgeist*. Cambridge University Press.

Simonton, D. K. (2010). Creative thought as blind-variation and selective-retention: Combinatorial models of exceptional creativity. *Physics of life reviews*, *7*(2), 156-179.

Singh, J., & Fleming, L. (2010). Lone inventors as sources of breakthroughs: Myth or reality?. *Management science*, *56*(1), 41-56.

Sinnreich, A., Latonero, M., & Gluck, M. (2009). Ethics reconfigured: How today's media consumers evaluate the role of creative reappropriation. Information, Communication & Society, 12(8), 1242-1260.

Sinnreich, A. (2010). *Mashed up: Music, technology, and the rise of configurable culture*. Univ of Massachusetts Press.

Sorensen, A. T. (2017). Bestseller Lists and the Economics of Product Discovery. Annual Review of Economics, 9, 87-101.

Smith, S. M., Ward, T. B., & Schumacher, J. S. (1993). Constraining effects of examples in a creative generation task. *Memory & cognition*, *21*(6), 837-845.

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology,translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social studies of science*, *19*(3), 387-420.

Stern, S. (2006). Economic experiments: The role of entrepreneurship in economic prosperity. *Melbourne Review: A Journal of Business and Public Policy, The*, *2*(2), 53.

Surowiecki J (2004) The Wisdom of Crowds (Random House, New York).

Taylor, A., & Greve, H. R. (2006). Superman or the fantastic four? Knowledge combination and experience in innovative teams. *Academy of Management Journal*, *49*(4), 723-740.

Trajtenberg, M., Henderson, R., & Jaffe, A. (1997). University versus corporate patents: A window on the basicness of invention. *Economics of Innovation and new technology*, *5*(1), 19-50.

Tushman, M. L., & Anderson, P. (1986). Technological discontinuities and organizational environments. *Administrative science quarterly*, 439-465.

Von Hippel, E. (2005). *Democratizing innovation*. MIT press.

Von Hippel, E., & Von Krogh, G. (2015). Crossroads—identifying viable “Need–solution pairs”: Problem solving without problem formulation. *Organization Science*, *27*(1), 207-221.

Von Krogh, G., & Von Hippel, E. (2006). The promise of research on open source software. *Management science*, *52*(7), 975-983.

Wasko, M. M., Teigland, R., & Faraj, S. (2009). The provision of online public goods: Examining social structure in an electronic network of practice. *Decision Support Systems*, *47*(3), 254-265.

Weisberg, R. W. (1999). I2 Creativity and Knowledge: A Challenge t0 The0ries. *Handbook of creativity*, *226*.

Weitzman, M. L. (1998). Recombinant growth. *The Quarterly Journal of Economics*, *113*(2), 331-360.

Yoo, Y., Boland Jr, R. J., Lyytinen, K., & Majchrzak, A. (2012). Organizing for innovation in the digitized world. *Organization science*, *23*(5), 1398-1408.

Zittrain, J. L. (2006). The generative internet. *Harvard Law Review*, 1974-2040.

Tables and Figures

Table 1. Descriptive Statistics

	count	Mean	Sd	min	max
			Full Sample		
Breakthrough	278341	0.06	0.24	0.00	1.00
IsDerivative	278341	0.15	0.36	0.00	1.00
MultipleParent	278341	0.03	0.16	0.00	1.00
NumberOfParents	278341	0.19	0.55	0.00	10.00
Cumulativeness	278341	0.37	0.95	0.00	13.00
SelfRemix	278341	0.03	0.16	0.00	1.00
Experience	278341	1.75	1.28	0.00	6.00
Tenure	278341	224.22	342.16	0.00	2915.00
Customizable	278341	0.02	0.14	0.00	1.00
			Only Derivative Sample		
Breakthrough	35094	0.10	0.30	0.00	1.00
MultipleParent	35094	0.18	0.39	0.00	1.00
NumberOfParents	35094	1.30	0.85	1.00	10.00
Cumulativeness	35094	2.47	0.96	2.00	13.00
SelfRemix	35094	0.20	0.40	0.00	1.00
Experience	35094	2.05	1.28	0.00	6.00
Tenure	35094	313.90	389.76	0.00	2759.00
Customizable	35094	0.04	0.19	0.00	1.00
TimeElapsed	35094	350.83	399.43	0.00	2887.00
LogTimeElapsed	35094	4.80	1.93	0.00	7.96
			Only Derivative Sample with Similarity Measure		
Breakthrough	30188	0.10	0.30	0.00	1.00
MultipleParent	30188	0.19	0.39	0.00	1.00
NumberOfParents	30188	1.31	0.86	1.00	10.00
Cumulativeness	30188	2.51	0.99	2.00	13.00
Similarity	30188	0.74	0.12	0.42	1.00
SelfRemix	30188	0.20	0.40	0.00	1.00
Experience	30188	2.08	1.28	0.00	6.00
Tenure	30188	316.62	390.07	0.00	2753.00
Customizable	30188	0.04	0.19	0.00	1.00
TimeElapsed	30188	354.32	397.74	0.00	2887.00
LogTimeElapsed	30188	4.82	1.93	0.00	7.96

Table 2. Correlations for Full Sample

No	Variable	1	2	3	4	5	6	7	8	9
1	Breakthrough	1								
2	IsDerivative	0.0543***	1							
3	MultipleParent	0.0526***	0.383***	1						
4	NumberOfParents	0.0672***	0.824***	0.704***	1					
5	Cumulativeness	0.0524***	0.814***	0.368***	0.700***	1				
6	SelfRemix	0.0461***	0.393***	0.198***	0.360***	0.354***	1			
7	Experience	0.0472***	0.0383***	0.0485***	0.0578***	0.0375***	0.125***	1		
8	Tenure	0.0583***	0.0783***	0.0398***	0.0694***	0.0693***	0.0997***	0.535***	1	
9	Customizable	0.0463***	0.0442***	0.0302***	0.0402***	0.0398***	0.0460***	0.0309***	0.0510***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3. Correlations for Only Derivative Sample

No	Variable	1	2	3	4	5	6	7	8	9	10
1	Breakthrough	1									
2	MultipleParent	0.0734***	1								
3	NoOfParents	0.0844***	0.742***	1							
4	Cumulativeness	0.0302***	0.106***	0.0896***	1						
5	SelfRemix	0.0574***	0.0563***	0.0691***	0.0663***	1					
6	Experience	0.0592***	0.0935***	0.118***	0.0286***	0.304***	1				
7	Tenure	0.0438***	0.0247***	0.0203***	0.0222***	0.174***	0.579***	1			
8	Customizable	0.0678***	0.0283***	0.0130**	0.0128**	0.0613***	0.0461***	0.0630***	1		
9	TimeElapsed	-0.00672	0.0186***	0.0329***	-0.0251***	-0.275***	-0.0412***	0.0704***	0.0179***	1	
10	LogTimeElapsed	0.000886	0.0114*	0.00645	-0.0181***	-0.434***	-0.130***	0.0137**	-0.0136**	0.751***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4. Correlations for Only Derivative Sample with Similarity Measure

No	Variable	1	2	3	4	5	6	7	8	9	10	11
1	Breakthrough	1										
2	MultipleParent	0.0752^{***}	1									
3	NoOfParents	0.0869^{***}	0.741^{***}	1								
4	Cumulativeness	0.0349^{***}	0.107^{***}	0.0925^{***}	1							
5	SelfRemix	0.0615^{***}	0.0637^{***}	0.0766^{***}	0.0824^{***}	1						
6	Experience	0.0638^{***}	0.0841^{***}	0.110^{***}	0.0164^{***}	0.298^{***}	1					
7	Tenure	0.0461^{***}	0.0159^{**}	0.0113^{*}	0.0157^{**}	0.171^{***}	0.575^{***}	1				
8	Customizable	0.0703^{***}	0.0312^{***}	0.0123^{*}	0.0158^{**}	0.0589^{***}	0.0432^{***}	0.0685^{***}	1			
9	Similarity	-0.0453^{***}	-0.111^{***}	-0.0966^{***}	-0.0372^{***}	0.0416^{***}	-0.0409^{***}	-0.0276^{***}	0.00185	1		
10	TimeElapsed	-0.00789	0.0149^{**}	0.0321^{***}	-0.0398^{***}	-0.290^{***}	-0.0261^{***}	0.0886^{***}	0.0134^{*}	-0.0657^{***}	1	
11	LogTimeElapsed	0.00301	0.00263	-0.00191	-0.0319^{***}	-0.445^{***}	-0.123^{***}	0.0247^{***}	-0.0140^{*}	-0.0950^{***}	0.752^{***}	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5. Baseline Results -- Full Sample

	(1) Breakthrough	(2) Breakthrough	(3) Breakthrough	(4) Breakthrough	(5) Breakthrough
IsDerivative			0.0366***	0.0261***	0.0263***
			(0.00149)	(0.00205)	(0.00212)
Experience		0.00306***			0.00303***
		(0.000895)			(0.000895)
SelfRemix		0.0171***			-0.00296
		(0.00457)			(0.00481)
Tenure		-0.000160*			-0.000157*
		(0.0000635)			(0.0000634)
Customizable		0.0762***			0.0763***
		(0.00708)			(0.00707)
Constant			0.0566***		
			(0.000476)		
Designer FE	Yes	Yes	No	Yes	Yes
Sub-category FE	Yes	Yes	No	Yes	Yes
Cohort FE	Yes	Yes	No	Yes	Yes
R^2	0.221	0.222	0.003	0.222	0.223
Observations	278341	278341	278341	278341	278341

Standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6. Results -- Full Sample

	(1) Breakthrough	(2) Breakthrough	(3) Breakthrough	(4) Breakthrough	(5) Breakthrough	(6) Breakthrough
IsDerivative	0.0184***	0.0191***	0.00588	0.00629	0.00323	0.00390
	(0.00207)	(0.00213)	(0.00478)	(0.00484)	(0.00475)	(0.00481)
IsDerivative X MultipleParent	0.0525***	0.0524***			0.0508***	0.0509***
	(0.00503)	(0.00499)			(0.00502)	(0.00498)
IsDerivative X Cumulativeness			0.00839***	0.00841***	0.00640***	0.00647***
			(0.00189)	(0.00188)	(0.00188)	(0.00188)
Experience		0.00252**		0.00263**		0.00251**
		(0.000895)		(0.000896)		(0.000895)
SelfRemix		-0.00510		-0.00403		-0.00590
		(0.00482)		(0.00479)		(0.00480)
Tenure		-0.000154*		-0.000157*		-0.000155*
		(0.0000633)		(0.0000634)		(0.0000634)
Customizable		0.0763***		0.0763***		0.0763***
		(0.00705)		(0.00707)		(0.00705)
R^2	0.222	0.224	0.222	0.223	0.222	0.224
Observations	278341	278341	278341	278341	278341	278341

Standard errors in parentheses

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7. Results -- Full Sample

	(1) Breakthrough	(2) Breakthrough	(3) Breakthrough
IsDerivative X NumberOfParents	0.0328***	0.0329***	0.0323***
	(0.00402)	(0.00400)	(0.00398)
IsDerivative X Cumulativeness			0.00625***
			(0.00187)
IsDerivative	-0.0139**	-0.0131**	-0.0272***
	(0.00492)	(0.00496)	(0.00661)
Experience		0.00261**	0.00259**
		(0.000896)	(0.000896)
SelfRemix		-0.00676	-0.00753
		(0.00486)	(0.00485)
Tenure		-0.000156*	-0.000157*
		(0.0000634)	(0.0000634)
Customizable		0.0764***	0.0764***
		(0.00704)	(0.00704)
R^2	0.223	0.224	0.224
Observations	278341	278341	278341

Standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8. Results -- Only Remix Sample

	(1)	(2)	(3)	(4)	(5)	(6)
	Breakthrough	Breakthrough	Breakthrough	Breakthrough	Breakthrough	Breakthrough
MultipleParent	0.0552***	0.0546***			0.0538***	0.0533***
	(0.00573)	(0.00569)			(0.00572)	(0.00568)
Cumulativeness			0.00834***	0.00825***	0.00647**	0.00642**
			(0.00225)	(0.00224)	(0.00223)	(0.00223)
Experience		0.00523		0.00640+		0.00522
		(0.00330)		(0.00331)		(0.00330)
SelfRemix		-0.00159		-0.00278		-0.00302
		(0.00593)		(0.00597)		(0.00594)
Tenure		-0.000301		-0.000333		-0.000311
		(0.000206)		(0.000206)		(0.000206)
Customizable		0.0750***		0.0754***		0.0747***
		(0.0160)		(0.0161)		(0.0160)
R^2	0.300	0.301	0.297	0.298	0.300	0.301
Observations	35094	35094	35094	35094	35094	35094

Standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9. Results -- Only Remix Sample

	(1) Breakthrough	(2) Breakthrough	(3) Breakthrough
NumberOfParents	0.0348***	0.0347***	0.0342***
	(0.00436)	(0.00433)	(0.00432)
Cumulativeness			0.00619**
			(0.00222)
Experience		0.00580+	0.00577+
		(0.00329)	(0.00328)
SelfRemix		-0.00303	-0.00439
		(0.00595)	(0.00596)
Tenure		-0.000312	-0.000322
		(0.000206)	(0.000206)
Customizable		0.0756***	0.0752***
		(0.0158)	(0.0159)
R^2	0.302	0.304	0.304
Observations	35094	35094	35094

Standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10. Results -- Only Remix Sample with Similarity Score

	(1) Breakthrough	(2) Breakthrough	(3) Breakthrough	(4) Breakthrough	(5) Breakthrough
Similarity	-0.0955***	-0.0948***	-0.207***	-0.177***	-0.172***
	(0.0171)	(0.0171)	(0.0469)	(0.0468)	(0.0468)
Cumulativeness		0.00843***	-0.0242+	-0.0210	-0.0202
		(0.00241)	(0.0130)	(0.0130)	(0.0130)
Cumulativeness X Similarity			0.0447*	0.0377*	0.0362*
			(0.0175)	(0.0174)	(0.0175)
MultipleParent				0.0503***	
				(0.00612)	
NumberOfParents					0.0330***
					(0.00447)
Experience				0.00540	0.00582
				(0.00361)	(0.00359)
SelfRemix				-0.00200	-0.00330
				(0.00654)	(0.00653)
Tenure				-0.000272	-0.000280
				(0.000223)	(0.000224)
Customizable				0.0808***	0.0820***
				(0.0171)	(0.0170)
R^2	0.303	0.304	0.304	0.308	0.311
Observations	30188	30188	30188	30188	30188

Standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11. Robustness Check 1 -- Time Elapsed between the submissions of parent and derivative designs

	(1) Breakthrough	(2) Breakthrough	(3) Breakthrough	(4) Breakthrough	(5) Breakthrough	(6) Breakthrough
TimeElapsed	0.00000575	0.00000508	0.00000399			
	(0.00000507)	(0.00000525)	(0.00000577)			
LogTimeElapsed				0.00412***	0.00455***	0.00453***
				(0.00110)	(0.00123)	(0.00136)
MultipleParent		0.0532***	0.0502***		0.0528***	0.0499***
		(0.00568)	(0.00612)		(0.00568)	(0.00611)
Cumulativeness X Similarity			0.0380*			0.0383*
			(0.0174)			(0.0174)
Cumulativeness		0.00643**	-0.0212		0.00640**	-0.0214+
		(0.00225)	(0.0130)		(0.00223)	(0.0130)
Similarity			-0.177***			-0.174***
			(0.0468)			(0.0468)
Experience		0.00477	0.00483		0.00469	0.00478
		(0.00329)	(0.00361)		(0.00330)	(0.00362)
SelfRemix		-0.00162	-0.000900		0.00561	0.00656
		(0.00619)	(0.00682)		(0.00665)	(0.00731)
Tenure		-0.000311	-0.000272		-0.000317	-0.000281
		(0.000206)	(0.000223)		(0.000206)	(0.000224)
Customizable		0.0746***	0.0808***		0.0748***	0.0808***
		(0.0160)	(0.0171)		(0.0160)	(0.0171)
R^2	0.296	0.302	0.308	0.297	0.302	0.309
Observations	35094	35094	30188	35094	35094	30188

Standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12. Robustness Check 2 -- Parent and Parent-Designer Fixed Effects

	(1) Breakthrough	(2) Breakthrough	(3) Breakthrough
MultipleParent	0.0532***	0.0532***	0.0455**
	(0.0111)	(0.0114)	(0.0153)
Experience	0.00568	0.00568	-0.00315
	(0.00672)	(0.00675)	(0.00882)
SelfRemix	0.0315	0.0315	
	(0.0226)	(0.0216)	
Tenure	-0.000439	-0.000439	-0.000949
	(0.000449)	(0.000455)	(0.000649)
Customizable	0.0770*	0.0770*	0.0597
	(0.0318)	(0.0347)	(0.0477)
Designer FE	Yes	Yes	No
Sub-category FE	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes
Parent FE	Yes	Yes	No
Parent-Designer FE	No	No	Yes
R^2	0.617	0.617	0.539
Observations	15783	15783	6317

Standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

What Drives Users to Go Beyond Mere Contribution? – Reputation, Reciprocity and Their Effect on Users' Decision to Allow Derivative Reuse in Open Communities

Erdem Dogukan Yilmaz

erdem.yilmaz@unibocconi.it

Department of Management & Technology, Bocconi University

Abstract

As digitization proceeds, relying on open communities to gain knowledge has become increasingly important for many firms. If firms want to benefit from open communities providing knowledge and ideas, it is essential that users do not only contribute to the community but also give up some of their intellectual property rights and allow for derivative reuse of their work. In this paper, we analyze two factors that affect the decision of users to allow derivative reuse of their work. We argue that users with a higher reputation level have lower expected returns from allowing other users to reuse their work and are therefore less likely to give up some of their intellectual property rights. On the other hand, a high level of direct reciprocity increases users' feeling of indebtedness and pushes them to allow others to reuse their work. We use data from the 3D-printable design community *Thingiverse* to provide empirical support for our claims.

1. Introduction

Knowledge is a source of competitive advantage for organizations (Barney 1991; Grant 1996; Penrose, 1959). Firms can access new knowledge either by investing in innovation (Cohen and Levinthal, 1990), buying it from other companies (Arora et al., 2004), forming alliances (Lavie 2006) or engaging in open innovation (Laursen and Salter, 2006) and open communities (Jeppesen and Frederiksen 2006). Due to the ubiquity of the internet, relying on open communities to gain knowledge has become a more and more important phenomenon. Increasing digitization of design and manufacturing processes and lower costs of communication have increased the viability of user driven product innovation (Baldwin and Von Hippel 2011; Gambardella and von Hippel 2018).

While users in open communities usually do not receive any monetary compensation, previous research has shown that they are pushed to contribute to open communities by various types of intrinsic motivations, such as enjoyment derived from challenging tasks, reciprocity, recognition by peers or their interest in gaining social capital (Lerner & Tirole, 2002; Lakhani & Wolf, 2005; Wasko & Faraj, 2005). For firms, it is important to sustain continuous contribution by users with diverse backgrounds because the varying degrees and the diversity of knowledge and skills across users have been shown to increase the likelihood of successful innovation outcomes (Frey & Gallus, 2015). In this paper, we extend previous research by focusing not on the motivations of users to merely contribute to a community with their innovations, but rather on their motivations to make these innovations available for reuse by other users, who can build on these innovations and potentially improve them.

More specifically, we focus on different licensing choices of users for their designs in a three-dimensional (3D) printable design community called *Thingiverse*. This website was launched in 2008 by one of the biggest producers of 3D printers, *MakerBot Industries*, in order to ensure complementarity between digital 3D designs and *MakerBot*'s 3D printers. The community is governed with *Creative Commons* licenses that provide users with the freedom to claim different rights over their products. *Creative Commons* licenses consist of five main tools, namely "Public Domain", "Attribution", "ShareAlike", "NonCommercial" and "NoDerivative". Users on *Thingiverse* can choose among different combinations of these licenses. In this paper, we focus in particular on the choice of the "NoDerivative" license by users and argue that, as the reputation level of users increases, they are more likely to restrict others from creating derivatives based on their product. This reasoning is based on the idea that users with a higher reputation level have lower expected returns from allowing other users to reuse their work because they benefit less form the advertising effect that a reuse of their work might have. We further argue that the choice of users is also affected by the degree of reciprocity. If users use the design of others as a source material for the creation of their own designs, they are less likely to choose the "NoDerivative" license as they feel indebted to others. We use panel data on a sample of 153,484 designs from 26,880 designers and fixed-effect regressions to provide empirical support for our predictions.

Our work contributes to previous research in several ways. First, it contributes to the literature on open communities as strategic assets. Understanding the behavior of users is important as many firms have started to employ open communities as a source of external knowledge. Under the right conditions, these communities might provide a competitive advantage for firms by enabling them to tap into the knowledge of the crowd. Thus, an open community might turn into a strategic asset that is difficult to imitate (Jeppesen and Frederiksen , 2006), since users might develop new product designs of potential commercial value (Gambardella et al., 2016). Previous literature on open innovation has shown that about half of consumer innovators are willing to share their innovations without any monetary compensation and only a small portion of them acquire intellectual property (De Jong et al., 2015). By examining the licensing choices of users, this study uncovers important contingencies that affect user's propensity to pursue a more open property rights regime that allows others to reuse their work. Allowing reuse is particularly important in digital environments as it allows firms to benefit from the increased malleability of digital information and because it is a necessary condition for cumulative innovation in open communities over time (Murray and O'Mahony, 2007).

Second, this paper contributes to the literature on motivations of users to contribute to open communities. While previous research has focused on the drivers that affect the quantity and quality of users' contributions to the community, contributions are only valuable for the community as a whole if they can be reused by others and therefore allow for cumulative innovation. For this reason, we focus particularly on users' motivation to not merely contribute to the community, but to make their work available for reuse by others. We show that users' motivation to allow derivative reuse depends on their reputation and the level of reciprocity. Beyond the current setting, these contingencies also point to some of the potential obstacles that firms might face in the appropriation of value from users. If a firm pursues a strategy in which users are core innovators rather than being the creators of complementary innovations, firms will have to make sure that high-reputation users are benefitting from the contributions of others as well.

Finally, to the best of our knowledge, this paper is the first paper on *Creative Commons* licenses that shows correlations with fine-grained empirical data. Previous studies on this increasingly popular phenomenon have mostly relied on theoretical reasoning in their analyses. By using fine-grained empirical data, we are able to quantify the magnitude of the effect that factors such as reputation and reciprocity have on users' choices of different types of *Creative Commons* licenses.

2. Theory and Hypotheses

Firm and User Innovation

In order to strengthen their innovation process, an increasing number of firms employ user communities. In these communities, users freely reveal their innovations, which then serve as a basis for further innovations created by refinements and recombination (Von Hippel, 1988). For instance, a firm might pick up the most promising user-generated innovations and integrate them into future product developments, or use them as complementary goods for their original product (Jeppesen and Frederiksen, 2006). Previous research has documented the importance and magnitude of user innovation in a variety of industries, including medical devices, scientific instruments, semiconductors, software and sports equipment (Von Hippel, 2005), and has shown that users often outpace producers in the introduction of innovations (Baldwin and Von Hippel 2011). Assigning part of the innovation process to users has been argued to be especially beneficial if users are motivated to create a solution that will meet their specific situation and if it is difficult to transfer knowledge from users to the firm, i.e. when information is “sticky” (Von Hippel, 1998).

With advancements in internet technology and the digitization of design and manufacturing processes (CAD-CAM), the cost of design and communication has substantially decreased for users (Baldwin and Von Hippel, 2011). Users are now better equipped and more motivated to innovate by themselves. With the increasing diffusion of 3D printing technologies, even the production cost of manufacturing customized designs has decreased significantly. Therefore, more and more companies have launched open communities to tap into the creativity of crowds, for instance by focusing on the production of hardware tools for 3D printing and enabling users to design the actual product to be printed (West and Kuk, 2016).

An increasingly important complement to the user innovation perspective has been the rise of innovation communities, which are not directly linked to the innovation process of a specific firm. These communities typically consist of a group of individuals that are linked through a common technology and freely share their problems, solutions and designs with each other (Lakhani 2016). One of the most prominent examples for this approach has been the rise of the free/open source software movement. With the rise of the internet, programmers all around the world have started to find one another and to collaborate with each other using digital technologies. While such movements are generally created as independent projects that are based on voluntary participation, many firms have started to encourage and sponsor users to participate in these communities (Lerner and Tirole, 2005).

Why Users Contribute to Open Communities

A fundamental question in the study of innovation is what motivates innovation. While the motivation of producer innovators derives from profit, user innovators often freely reveal their innovations to both peer users and firms without any monetary compensation. Instead, research has shown that users justify their investments in the development of free innovations with benefits they derive from various types of intrinsic motivations (Lakhani and Wolf 2005).

For instance, it has been shown that users can be motivated to innovate because they benefit from the use of their own innovations (Gambardella et al., 2016). According to previous research, up to 38 percent of consumer sport enthusiasts (Lüthje et al., 2005), 26 percent of library information system users (Morrison et al., 2000), and 19 percent of Apache software users (Franke and Piller, 2003) report innovating for their own use.

Other intrinsic motivations include enjoyment derived from challenging tasks, reciprocity, recognition by peers or gaining social capital (Lakhani and Wolf, 2005; Lerner and Tirole, 2002; Wasko and Faraj 2005). For instance, survey-based research on the motives of programmers' participation in open software development projects points to intrinsic motivations resulting from enjoyment of intellectual challenges and their completion (Lakhani and Wolf 2005). In the same context, free revealing of high-quality code can also increase a programmers' reputation among their peers, which in turn may lead to an increase in the programmers' value on the job market (Lerner and Tirole 2002).

Why Users Need to Contribute

In open communities, member contribution has been identified as one of the most critical elements of survival and prosperity (Ren et al., 2007; Wiertz and de Ruyter, 2007). More specifically, both the quantity and quality of contributions need to be sustained for a community to be successful. The quantity of contributions is important because communities need to reach a critical mass to flourish. Individuals function as different pieces of the puzzle and improve the innovation performance of others with their feedback (Wasko et al., 2009). Furthermore, the varying degrees and the diversity of knowledge and skills across users have been shown to increase the likelihood of successful innovation outcomes (Frey and Gallus, 2017).

However, while increasing the quantity and diversity of members, firms also need to sustain the participation of existing members in order to survive (Dahlander et al. 2019; Iriberri and Leroy, 2009). This can be achieved for instance through quality signaling mechanisms such as user badges (Lappas et al. 2017), monetary awards (Boudreau and Lakhani, 2009), the creation of a shared identity (Nambisan and Baron, 2009) or improvements in the creative process (Füller et al., 2008). This is particularly important because participation in user communities is often not equally distributed. In general, only a few members of the community account for the vast majority of the contributions (Dahlander and Frederiksen, 2012; Lakhani and Von Hippel, 2004).

On the other hand, firms have to ensure not only the quantity but also the quality of contributions by their members. Community members have been found to value learning from others, and this learning process significantly affects their willingness to participate in the community (Nambisan and Baron, 2009). Furthermore, Porter and Donthu (2008) found that quality content is directly related to measures of shared values, respect, and trust, ultimately supporting a willingness to cooperate in innovation efforts.

The Importance of Building on Prior Knowledge

Contributing to an open innovation community generally involves revealing information freely to others. Harhoff et al. (2003) identify some of the key factors under which it is beneficial for users to reveal information. For instance, freely revealing information might allow other users to improve the innovation that is revealed, which in turn can allow the original creator of the innovation to benefit from improvements, or it might help innovators to set the standards. According to the authors, freely revealing information is particularly beneficial if rivalry is low and if users can expect reciprocity and reputation effects.

However, free revealing of information does not necessarily mean that all intellectual property rights related to an innovation are given up by the innovator. In many settings, freely revealed innovation does not directly become a public good but users can choose among different licensing alternatives, which differ in the extent to which they allow other users to reuse the focal innovation and create derivative innovations based on it.

Beyond the mere quantity and quality of contributions, the question of whether these contributions can be reused (and potentially improved) by others is one of the key factors that determine the value of contributions for an open community. Compared to merely contributing, allowing other users to reuse their work generally requires additional motivation and commitment from the original creators, but is at the same time extremely important for the development of an open community as a whole. Only if users reveal information to others and allow them to reuse it, cumulativeness of innovation in open communities is ensured (Murray and O'Mahony, 2007). For this reason, the possibility of reusing knowledge has been shown to be one of the most important features that boost innovation in open communities (Bayus 2013; Yu and Nickerson, 2011). For instance, it has been found that code reuse significantly improves productivity and quality in software development (Haefliger et al., 2008), and that cumulative remixes (i.e., reuses of reused contributions) will be more complete work (Hill and Monroy-Hernández, 2013). In other words, one of the most important premises of open communities is that the contributors in open communities are able to "stand on the shoulder of giants" (Merton, 1993) by reusing and transforming the work of predecessors.

The Importance of Building on Prior Knowledge in Digital Settings

Allowing other users to reuse a focal innovation is particular important in digital settings, where the structure of innovation costs has changed and information are highly malleable. In general, while designing a product, one has to consider different elements that define the product, as well as the interactions between them, which can result in a potentially unlimited number of options to choose from. While scarce resources have traditionally been identified as the limiting factor in evaluating these options (Weitzman, 1998), the cost of many of these resources is significantly reduced in digital settings, and the main limiting factor are intellectual property rights.

Furthermore, the increased malleability of information in digital settings creates a vast set of potential reuses of a given innovation. Because of this malleability, digital designs might acquire functions for which they were not originally intended. In the innovation literature, this phenomenon is referred to as exaptation (Andriani and Cattani, 2016). However, the high level of malleability of digital designs, as well as resulting phenomena such as exaptation, can only be fully exploited if users give up some of their intellectual property rights and allow others to reuse their innovations.

In this paper, we develop and test hypotheses on factors that affect the choice of users to not only contribute to a community, but to allow other users to reuse and build upon these contributions. More specifically, we focus on the effect of reputation and reciprocity on the choice of users to allow for such derivative reuse.

The Effect of Reputation on Allowing Derivative Reuse

A consistent finding in the open community literature is that a large share of users only contribute marginally (Dahlander and Frederiksen, 2012), while a small percentage of participants produces most of the work (Krishnamurthy, 2002). These core users play a crucial role in the sustenance of open communities. To sustain participation of these users and highlight their contributions, especially if they are high quality, open communities can adopt different strategies.

Highlighting of high-quality contributions can for example be found on *StackOverFlow*, a website where users can ask and answer questions on various topics. On this website, once a new response to a question is posted, community members can up-vote or down-vote it depending on its quality and relevance (Liu et al., 2008). Responses with the most up-votes are then displayed at the top of the page. Such ranking mechanisms do not only make it easier to obtain useful information (Ghose et al., 2014; Guan and Cutrell, 2007; Pan et al., 2007), but also offer increased visibility for respondents with higher quality which, in turn, can further motivate users to submit high quality content for different questions (Ghosh and Hummel 2014). In addition to that, *StackOverFlow* and other communities also award badges to users who have, for example, exceeded a certain threshold level of contributions (Cavusoglu et al., 2015; Grant and Betts, 2013). Other communities, such as the 3D printable design community

Thingiverse that we study in this paper, allow users to follow other users whose work they appreciate. The number of followers that each user has is prominently displayed on his or her profile and therefore clearly visible for other users in the community. Gaining up-votes, badges or additional followers ultimately increases the reputation of a user. Therefore, the best strategy for users that are seeking reputation would be to submit high-quality innovations to the community. Consistent with this reasoning, previous research has shown that increasing a users' reputation, for instance by awarding them badges, is positively correlated with the quantity of their contributions (Anderson et al., 2013; Grant and Betts, 2013).

However, as users achieve higher reputation within the community, their preferences with regards to which type of contribution path they want to pursue are likely to change. For instance, Lappas et al. (2017) find that increasing reputation significantly affects users' risk-taking propensity, performance and topical interests. More specifically, the authors draw on prospect theory to show that as users' reputation exceeds the community's mean, their motivation "to take risks that could hurt their hard-earned status" decreases significantly. Similarly, they also find that users tend to focus on a narrower set of topics as their reputation increases beyond the community's mean. The idea that the motivation and risk-taking propensity of users decrease with increasing reputation is consistent with the results of several other studies. For instance, it has been shown that users put more effort into contributing to the community if they are close to achieving a reward that increases their reputation, such as a badge (von Rechenberg et al., 2016). Once they have reached the level of contributions that is required to obtain the reward, they significantly decrease their effort (Goes et al., 2016).

In this paper, we argue that increasing reputation might not only change the motivation of users to contribute, but also their propensity to allow other users to reuse these contributions and create derivatives. While continuous user contributions are important for sustaining open communities, it is only the reuse by other users that ensures cumulativeness of innovation in open communities (Murray and O'Mahony, 2007).

A user's reputation level has a strong effect not only on the behavior of the user himself, but also on the behavior of other users in the community. For instance, previous literature has found a strong relationship between user reputation and remixing of the user's products by other users. More specifically, products of high reputation users are more likely to be used and reused by other members of the community (Jenkins 2006). Within the context of open communities, scholars have also shown that users with higher reputation often attract more people to remix their designs (Cheliotis and Yew, 2009; Hill and Monroy-Hernández, 2013). This effect has generally been attributed to higher visibility of high reputation users, which in turn increases the probability that another user will find their product and remix it.

Research has found that users can actually benefit from sharing their contributions with others, but that the magnitude of this beneficial effect depends largely on the users' reputation. In the context of the music industry, lessening an album's sharing restrictions can increase sales by 10% on average, but this effect is much larger for less popular albums and significantly reduced for top-selling albums (Zhang, 2016). This finding has been attributed to the fact that more popular albums are already well-known by a large share of the population and have therefore less to gain from increased sharing among users. Similarly, Watson (2017) shows that the release of a derivative (or remix) song to the market increases the demand for its upstream product by 3% and that this effect is particularly strong if a song by a less prominent artist is remixed. Again the author attributes this effect to the limited ability of highly prominent artists to benefit from advertising effects. Given that users with a higher reputation level have lower expected returns from allowing other users to reuse their contributions, we expect them to be more likely to restrict other users from derivative reuse.

Hypothesis 1: Users with a higher level of reputation are less likely to allow derivative reuse.

The Effect of Reciprocity on Allowing Derivative Reuse

The motivation of users to participate in open communities can be sustained through different mechanisms. Similar to academia, norms are one of the most important mechanisms that govern user communities. Previous literature has identified the norm of reciprocity to be one of the most prominent factors that affect motivation to contribute (Harhoff et al., 2003). One of the first scholars to study the concept of reciprocity was Gouldner (1960) who argued that recipients of a benefit from someone else feel "indebted to the donor" and remain so until they repay. Others have argued that this indebtedness even creates a feeling of discomfort among individuals, which in turn will push them to repay (Greenberg, 1980). More generally, reciprocity has been described as one of the most important factors that ensure maintenance of supportive exchanges among individuals (Shumaker and Brownell, 1984).

Besides reciprocity as an exchange mechanism in the relationship between two particular individuals, which is generally referred to as direct reciprocity, there is also the broader concept of generalized reciprocity (Kollock, 1999). Members of a community often know that the person they helped may never be in the position to pay back, but someone else in the community might be (Rheingold, 1993). Even though exchanges of benefits are not balanced at the individual level, they might be balanced at the level of the community (Kollock, 1999). In this case, reciprocity is not seen as a mechanism that can result in a payoff in return for providing benefits to other members of the community (Harhoff et al., 2003), but rather as a generalized norm that supports the architecture of the community and pushes individuals to pay something back to the community in the long-run (Lakhani and Von Hippel, 2004; Murray and O'Mahony, 2007).

Previous studies have found strong evidence for the existence of patterns of reciprocity in the interaction among members of online communities (Faraj and Johnson, 2011). However, the empirical evidence for a positive effect of reciprocity on contributions to an open community is somewhat mixed. For instance, it has been shown that expected reciprocity is one of the main reasons why individuals freely reveal innovation-related information in innovation communities outside of firms (Franke and Shah, 2003). Similarly, in the context of question and answer websites, the community is more likely to reciprocate if users put relatively more effort into a question (Wu and Korfiatis, 2013). While reciprocity has been found to be one of the main drivers for individuals to contribute to open source software projects, it has also been shown that individuals who are driven by a “desire to conform to the norms of the community” generally spend less effort when creating their contribution, and are less interested in its ultimate value for the community (Shah, 2006). On the other hand, several studies find no empirical support for the importance of reciprocity (Jeppesen and Frederiksen ,2006; Wasko and Faraj, 2005). For instance, Wiertz and de Ruyter (2007) find that an individual’s perception of reciprocity does not have any effect on the quality or quantity of his or her contributions. This mixed evidence might be attributed at least in part to particular characteristics of the empirical settings that have used for previous studies. Scholars have mostly studied reciprocity in contexts such as independent (Wu and Korfiatis, 2013) or firm-hosted (Wiertz and de Ruyter, 2007) Q&A websites, networks of practice (Wasko and Faraj, 2005), volunteer communities (Franke and Shah, 2003) and open source software projects (Shah, 2006). In these settings, there is no clear distinction between contributing to the community and allowing others to reuse this contribution, because every contribution also involves, at least to some extent, disclosure of knowledge and potential reuse by others.

In this paper however, we focus specifically on the effect of reciprocity on giving up property rights rather than the effect of reciprocity on contributions per se. This is particularly important given that in many settings, such as the 3D printable design community *Thingiverse* that we study in this paper, allowing other users to reuse one's contribution requires an additional step compared to merely contributing. In this step, users that have contributed to the community actively decide whether they want to allow other users to reuse their work and potentially benefit from its reuse. Allowing other users to reuse one's own work to produce derivatives based on it, is arguably a larger benefit that might in turn result in a higher feeling of indebtedness and a higher propensity of other individuals to pay back (Gouldner, 1960; Greenberg, 1980). The effect of reciprocity should therefore be more pronounced if it is not just about contributing to a community, but about actively giving up intellectual property rights and making one's own work available for reuse by others. For this reason, we argue that:

Hypothesis 2: Users are more likely to allow derivative reuse if there is a higher level of reciprocity.

2. Data and Methods

Setting

With falling hardware prices and free circulation of digital designs, 3D printing is increasingly paving its way to become a mainstream phenomenon. The term 3D printing has been increasingly used to denote machines used primarily by end users, while in industrial contexts, additive manufacturing (AM) is an already established term. Nevertheless, these terms are often used interchangeably and they both refer to a process of joining materials layer upon layer by using data from digital 3D models. Although AM has existed for a couple of decades, its technical performance has significantly increased in recent years. Due to this improvement in technical performance, the usage of AM technology has expanded from rapid prototyping into direct manufacturing. The digital and additive nature of AM provides users or companies with a higher degree of design freedom and flexibility, as it does not require any specific tool for manufacturing. As a result, AM facilitates the production of customized products and allows designs to evolve while diffusing across multiple users. The technology has already been adopted in different sectors, such as jewelry, medical applications (where personalization is regarded to be highly crucial for the human body) and aerospace (where it enhances the performance of structural components). However, many barriers still remain in the broader adoption of the technology as issues of standardization, intellectual property, certification, skills and education need to be solved.

Creating value for consumers via 3D printing requires two offerings: designs to be manufactured and hardware to manufacture those designs (West & Kuk, 2016). *Thingiverse* is one of the biggest 3D printable design communities and plays a central role in the diffusion of 3D printable design. It is owned by *Makerbot* and functions as a complementary asset for *MakerBot* as it adds value to their 3D-printer sales by offering free designs. *MakerBot* was founded in 2009 by some volunteers of an open hardware community called *RepRap* and has been built upon the 3D printers of this community. In the *RepRap* community, users have focused on creating an open hardware that is able to autonomously construct most of its own mechanical components. Although *RepRap* was an open hardware community from its early release, *MakerBot* has started to patent some of the community designs, which created a huge controversy in the eyes of many users. Nevertheless, *Thingiverse* was started in November 2008 as a complementary site to *MakerBot*. In 2013, both *Makerbot* and *Thingiverse* were acquired by *Stratasys*, which is a relatively large and established manufacturer of 3D printers. Similar to open communities, *Thingiverse* is characterized by sharing, weak intellectual property rights, and free flow of information among users. At the same time, activities in the community are to some extent interrelated with *MakerBot's* commercial interests. *Thingiverse* is advertised by *Makerbot* in its hardware sales and is a major selling point for the company.

Data

In our study, we use data that we gathered from *Thingiverse*. Sharing practices in *Thingiverse* slightly differ from other open source software projects as *Thingiverse* users can choose among different licensing options when they upload their 3D printable designs to the community. *Creative Commons* licensing is the most common one with more than 98% percent of the designs being licensed under *Creative Commons*. *Creative Commons* licenses are typically used to protect traditionally copyrighted material such as music, film, photography and literature (Moilanen et al. 2015). *Creative Commons* licenses enable users to combine different modules and create a more tailored copyright license that satisfies their personal needs. These modules are "Attribution" (imposes no restriction on what others can do except acknowledgement), "Non-Commercial Use" (the original creation cannot be used for commercial purposes), "Share Alike" (derivative creations should comply with the terms of the original creation) and "No Derivatives" (the original creation cannot be changed when it is redistributed). By combining these four modules, users can generate six different licenses: "Attribution"; "Attribution-ShareAlike"; "Attribution-NoDerivs"; "Attribution-NonCommercial"; "Attribution-NonCommercial-ShareAlike"; and "Attribution-NonCommercial-NoDerivs". Furthermore, there is also a "Public Domain Dedication" license, which means that the design is a pure public good. Therefore, *Creative Commons* licenses exhibit different degrees of openness. Apart from the "Public Domain Dedication" license, the most open license is the "Attribution" license, which only requires attribution. On the other hand, the most restrictive license is "Attribution-NonCommercial-NoDerivs", which does not allow for the creation of derivatives or for the usage of material for any commercial purposes. Regardless of the choice of license, users are free to download any of the designs on *Thingiverse* and 3D-print them as they wish.

Along with *Creative Commons* licenses, users can also choose other free licenses that are derived from the software that they use when uploading their designs. These licenses are two *GNU* licenses – "*GNU* General Public License" and "*GNU* Lesser General Public License" – and the "Berkeley Software Distribution (*BSD)* license". The "*GNU* General Public License" is similar to the "Public Domain Dedication" license in that users are free to change the original innovation and use pieces of it for new creations. Moreover, users are free to sell the innovations for a fee. The "*GNU* Lesser General Public License" and "*BSD* licenses" are similar to the "Attribution-ShareAlike" license in that users are obliged to give the recipients all the rights that are provided to them. *Table 1* shows the distribution of different licenses in our dataset between 2014 and 2016.

We collected data on a sample of 190,774 designs from *Thingiverse* to study how the licensing choices of users change depending on changes in reputation and reciprocity. The sample consists of all designs submitted to any of the 79 sub-categories, as well as non-categorized designs from 2014 to 2016. For each design in the sample, we observe the design brief that includes the name and description of the design, the creators of the design, the list and dates of previous designs of the designer, the number of downloads and views on the date of the scrape (February, 2018), the submission date and the licensing choice of users. We also observe the number of followers that each designer has on a weekly basis. We dropped 24,476 observations that belong to designers with single designs. We also dropped derivative designs whose parent designs force derivatives to comply with the terms of the parent designs. Our final dataset compromises 153,214 designs of 26,848 designers.

Measures

Our primary dependent variable, *AllowDerivative*, takes the value of one if a designer allows for the creation of derivatives and zero otherwise. In other words, if a designer chooses the "NoDerivative" attribute in his licensing choice, the *AllowDerivative* variable is equal to zero and in all other cases, it is equal to one.

Furthermore, we have three main independent variables of interest. The first one is *Followers,* which is the logarithm of the number of followers plus one and will be used to test the first hypothesis. *Thingiverse* allows users to "follow" updates from other users. A higher number of followers is likely to reflect a user's reputation for providing high-quality contributions. As we have argued in the first hypothesis, we expect the contribution behavior of users to change with their level of reputation.

The other two variables are used to measure the level of reciprocity. *IsDerivative* indicates whether a given design is itself a remix of another design. This measure is closely related to the concept of direct reciprocity, as it shows whether users decide to give something back directly in exchange for the benefit that they received from someone else. The second variable that we use to measure the level of reciprocity is *PreviousDerivatives,* which is defined as the logarithm of the number of times that the designer at hand created a derivative design before the submission of the design at hand. This variable is more related to the concept of generalized reciprocity. While users may not have benefitted from others in the case of a particular design, they may have benefitted from others in the past.

It is important to note that these last two variables only measure the material being copied and, as a result, does not capture "conceptual" remixing as such, borrowing material from any source outside of *Thingiverse,* or inspiration from sources inside *Thingiverse* without downloading and copying content from another design. We also define an *Experience* variable to control for users' prior experience that might influence both the quality of the design and the number of followers that a designer has. When we calculated the *PreviousDerivatives* and *Experience* variables, we considered the entire stock of designs and derivative designs of designers. In other words, we also considered designs and derivative designs that were introduced by the designers prior to our starting date in 2014. Finally, we also control for the logarithm of *FileSize* plus one, in kilobytes, which may reflect the complexity of the design (Yin et al., 2014) as well as the effort put into the design.

Table 2 contains descriptive statistics and *Table 3* shows the correlations between our variables. Out of 153,214 designs, 4,773 do not allow for derivatives. Choosing the "NoDerivative" license for a derivative design is therefore a rare event. Similarly, out of 12,090 derivative designs, only 147 do not allow for derivatives. The final sample does not compromise self-derivatives because all of the self-derivative observations have been automatically dropped when we dropped derivative designs whose parents have the "Attribution-ShareAlike" attribute. As can be seen from *Table 3,* there is a high correlation between *Followers* and *Experience*. However, the large number of observations in our sample reduces concerns about multicollinearity.

Empirical Strategy

To estimate the relationship between reputation, reciprocity and the probability of allowing a derivative, we use a linear probability model. Focusing on the linear probability model enables us to estimate a model using extensive fixed effects (especially designer fixed effects) without the exclusion of any observation. Also, there is typically little qualitative difference between the linear probability model and logit specifications (Angrist and Pischke, 2010). Furthermore, predicted probabilities outside the 0–1 interval are not a serious concern if the goal of the estimation is finding the independent variable's marginal effect (Wooldridge, 2015), which is precisely our goal.

When individuals make a licensing choice, their choice will be guided by their previous preferences and their preferences related to reputation and reciprocity. To account for these differences, we use individual fixed effects which allow us to difference out all time-invariant individual characteristics. We also control for the sub-category that the design belongs to in order to account for any overall differences in the likelihood of choosing a particular regime at the sub-category level. Despite this extensive set of controls, there may still be a concern that a user's preference for a particular license changes with particular time trends in the platform. To control for variations in time trends in the platform at the time of design introduction, we include cohort fixed effects (month and year of design release). The main specification that we use for the full sample has the following functional form:

$$\begin{aligned} AllowDerivative_{jitk} \\ &= \beta_1 Followers_{j,t-1} + \beta_2\, PreviosDerivatives_{j,t-1} + \beta_2\, IsDerivative_{i} \\ &+ Controls + \delta_j + \lambda_k + \gamma_t + u_{it} \; (1) \end{aligned}$$

where δ_j are individual, λ_k are sub-category and γ_t are cohort fixed effects. Robust standard errors are clustered at the designer level, and all regressions include designer, sub-category and cohort fixed effects.

3. Results

In the hypotheses section, we argued that increasing reputation should lead to a decrease in the probability of allowing derivative reuse (*H1*), whereas increasing levels of reciprocity should lead to an increase in the probability of allowing derivative reuse (*H2*). *Table 4* and *Table 5* show the results from the linear probability regressions that we used to test *H1* and *H2*. Column (1) on *Table 4* shows that a one percent increase in the number followers is associated with a 0.0061 percent decrease in the probability that a designer allows for derivatives. For the interpretation of the main effect of an increasing reputation on the probability that a designer will allow for derivative reuse, we focus on the fully saturated model in column (6) in *Table 5*. Consistent with our first hypothesis, we find that a one percent increase in the number of followers decreases the likelihood of allowing for derivatives by 0.005 percent. Given that the baseline probability of not allowing for derivative reuse in our sample is 0.03 percent, a ten percent increase in the number of followers translates into a 1.7 percent decrease relative to the baseline. Overall, the results from *Table 4* and *Table 5* are consistent with the idea that increasing reputation is an important factor that affects users' choice of allowance for derivative.

On the other hand, the results for *Hypothesis 2* are somewhat mixed. In this hypothesis, we argued that users should be more likely to allow derivative reuse if there is a higher level of reciprocity. However, as can be seen from the tables, there is no evidence for a statistically significant effect of *PreviousDerivatives* on the likelihood of allowing for derivative reuse. This result suggests that generalized reciprocity does not increase the likelihood that users allow others to reuse their work. On the other hand, the results show that there is a statistically significant effect of *IsDerivative* on the likelihood of allowing for derivative reuse. This suggests that direct reciprocity is an important factor in predicting the likelihood that users allow others to reuse their work, even though generalized reciprocity does not seem to play any role. According to *Table 5* column (5), on average, designers of derivative designs are 0.00820 percentage more likely to allow for derivatives. Relative to the baseline rate of 0.03 percent for not allowing derivative reuse, this transforms to a relative increase of 26% in the likelihood of allowing for derivative reuse for derivatives designs compared to designs that start from scratch.

4. Discussion

In the previous sections, we have argued that user reputation and the level of reciprocity are important factors in determining the extent to which users in open communities allow other users to reuse their work for the creation of derivatives. Our reasoning is based on the idea that users with a higher reputation level have lower expected returns from allowing other users to reuse their work because they benefit less form the advertising effect that a reuse of their work might have. At the same time, if there is a high level of reciprocity, users feel more indebted to the community and are therefore more likely to allow others to reuse their work for the creation of derivatives. Using data from the 3D printable design community *Thingiverse*, we have provided empirical evidence that supports most of our claims. Consistent with our expectations, the results provide strong evidence for a negative effect of reputation on users' likelihood of allowing for derivative reuse. On the other hand, we find no evidence for an effect of generalized reciprocity on users' likelihood of allowing for derivative reuse, but we do find evidence for a positive effect of direct reciprocity. While this is not entirely consistent with our expectation, it does make sense from a theoretical standpoint. Our findings can be explained by the fact that users who have directly benefitted from someone else's work perceive a higher level of indebtedness, which in turn pushes them to make their work available for reuse by others as well (Gouldner, 1960; Greenberg, 1980). On the other hand, such high level of indebtedness might not be perceived by users who have benefitted from the contribution of others in the more distant past.

Our research has allowed us to get a better understanding of what drives a user's decision on whether or not to give up intellectual property rights and allow for derivative reuse in open communities. While previous research has highlighted the importance of continuous user contribution to sustain open communities, not all contributions are equally valuable for the community as a whole. Beyond the mere quantity and quality of contributions, one of the key factors that determine a contribution's value for an open community is whether it can be reused (and potentially improved) by others. Compared to merely contributing, allowing other users to reuse their work generally requires additional motivation and commitment from the original creator, but is at the same time extremely important for the development of an open community as a whole. Only if users allow reuse of their work by others, communities can benefit from the increased malleability of information in the digital era and ensure cumulative innovation over time.

With our study, we contribute to previous research in a number of ways. First, we contribute to the literature on open communities as strategic assets. Understanding the behavior of users is important as many firms have started to employ open communities as the source of external knowledge. Under the right conditions, an open community might turn into a strategic asset. This is particularly true if the knowledge generated by this community is cumulative and thus difficult to imitate. Understanding the factors that lead users to allow reuse is therefore particularly important as it is one of the core conditions for cumulativeness of innovation in the community (Murray and O'Mahony, 2007).

Second, this paper contributes to the literature on motivations of users to contribute to open communities. Scholars have previously focused on the factors that determine the extent to which users contribute to the community. In our paper, we go a step further by analyzing not only the motivations of users to contribute to an open community, but also their motivations to make their work available for reuse by others. Our findings hold important implications for firms that want to rely on user communities to generate knowledge as they allow them to understand some of the mechanisms that ensure the willingness of high-reputation users to make their knowledge available to others.

From a broader perspective, our research highlights the important role of incentives in shaping investments in innovation. Because it is often difficult for inventors to capture the full social value of their inventions, it has long been recognized that, in the absence of an appropriate incentive mechanism, competition and imitation may lead to less innovation than is socially desirable. For firms it is therefore especially important to provide additional incentives for those individuals who already have a strong reputation for providing high-quality work. Our results also indicate that digitization does not just affect less prominent individuals (such as artists in niche genres), but that it might also amplify the success of the most popular ones and ultimately have an effect on their behavior, which is problematic for the community as a whole.

Empirically, we contribute to previous research by using for the first time fine-grained empirical data to study *Creative Commons* licenses. Previous studies in this setting have, to the best of our knowledge, mostly relied on theoretical reasoning in their analyses. By using quantitative data on *Creative Commons* licenses, we are able to provide empirical evidence for the effect that factors such as reputation and reciprocity have on the choice of different types of licenses, and at the same time assess the magnitude of these effects.

References

Anderson A, Huttenlocher D, Kleinberg J, Leskovec J. 2013. Steering user behavior with badges *Proceedings of the 22nd international conference on World Wide Web*. ACM, 95-106.

Andriani P, Cattani G. 2016. Exaptation as source of creativity, innovation, and diversity: Introduction to the special section. *Industrial and Corporate Change*. 25(1): 115-131.

Angrist JD, Pischke J-S. 2010. The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*. 24(2): 3-30.

Arora A, Fosfuri A, Gambardella A. 2004. *Markets for technology: The economics of innovation and corporate strategy* (MIT Press).

Baldwin C, Von Hippel E. 2011. Modeling a paradigm shift: From producer innovation to user and open collaborative innovation. *Organization Science*. 22(6): 1399-1417.

Barney J. 1991. Firm resources and sustained competitive advantage. *Journal of Management*. 77 99-120.

Bayus BL. 2013. Crowdsourcing new product ideas over time: An analysis of the Dell IdeaStorm community. *Management Science*. 59(1): 226-244.

Boudreau K, Lakhani K. 2009. How to manage outside innovation. *MIT Sloan Management Review*. 50(4): 69-75.

Cavusoglu H, Li Z, Huang K-W. 2015. Can gamification motivate voluntary contributions? The case of stackoverflow Q&A community *Proceedings of the 18th ACM conference companion on computer supported cooperative work & social computing*. ACM, 171-174.

Cheliotis G, Yew J. 2009. An analysis of the social structure of remix culture *Proceedings of the fourth international conference on communities and technologies*. ACM, 165-174.

Cohen WM, Levinthal DA. 1990. Absorptive capacity: A new perspective of learning and innovation. *Administrative Science Quarterly*. 35(1): 128-152.

Dahlander L, Frederiksen L. 2012. The core and cosmopolitans: A relational view of innovation in user communities. *Organization Science*. 23(4): 988-1007.

Dahlander L, Jeppesen LB, Piezunka H. 2019. How do organizations manage crowds: Define, broadcast, attract, and select *Research in the Sociology of Organizations*.

De Jong JP, von Hippel E, Gault F, Kuusisto J, Raasch C. 2015. Market failure in the diffusion of consumer-developed innovations: Patterns in Finland. *Research Policy*. 44(10): 1856-1865.

Faraj S, Johnson SL. 2011. Network exchange patterns in online communities. *Organization Science*. 22(6): 1464-1480.

Franke N, Piller FT. 2003. Key research issues in user interaction with user toolkits in a mass customisation system. *International Journal of Technology Management*. 26(5-6): 578-599.

Franke N, Shah S. 2003. How communities support innovative activities: an exploration of assistance and sharing among end-users. *Research Policy*. 32(1): 157-178.

Frey BS, Gallus J. 2017. Towards an economics of awards. *Journal of Economic Surveys*. 31(1): 190-200.

Füller J, Matzler K, Hoppe M. 2008. Brand community members as a source of innovation. *Journal of Product Innovation Management*. 25(6): 608-619.

Gambardella A, Raasch C, von Hippel E. 2016. The user innovation paradigm: impacts on markets and welfare. *Management Science*. 63(5): 1450-1468.

Gambardella A, von Hippel EA. 2018. Open source hardware as a profit-maximizing strategy of downstream firms. Centre for Economic Policy Research, London.

Ghose A, Ipeirotis PG, Li B. 2014. Examining the impact of ranking on consumer behavior and search engine revenue. *Management Science*. 60(7): 1632-1654.

Ghosh A, Hummel P. 2014. A game-theoretic analysis of rank-order mechanisms for user-generated content. *Journal of Economic Theory*. 154 349-374.

Goes PB, Guo C, Lin M. 2016. Do incentive hierarchies induce user effort? Evidence from an online knowledge exchange. *Information Systems Research*. 27(3): 497-516.

Gouldner AW. 1960. The norm of reciprocity: A preliminary statement. *American Sociological Review*. 25(2): 161-178.

Grant RM. 1996. Toward a knowledge-based theory of the firm. *Strategic Management Journal*. 17(S2): 109-122.

Grant S, Betts B. 2013. Encouraging user behaviour with achievements: an empirical study *Proceedings of the 10th Working Conference on Mining Software Repositories*. IEEE Press, 65-68.

Greenberg MS. 1980. A theory of indebtedness. K.J. Gergen, M.S. Greenberg, R.H. Willis, eds. *Social Exchange: Advances in Theory and Research* (Springer, Boston, MA), 3-26.

Guan Z, Cutrell E. 2007. An eye tracking study of the effect of target rank on web search *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 417-420.

Haefliger S, Von Krogh G, Spaeth S. 2008. Code reuse in open source software. *Management Science*. 54(1): 180-193.

Harhoff D, Henkel J, Von Hippel E. 2003. Profiting from voluntary information spillovers: how users benefit by freely revealing their innovations. *Research Policy*. 32(10): 1753-1769.

Hill BM, Monroy-Hernández A. 2013. The remixing dilemma: The trade-off between generativity and originality. *American Behavioral Scientist*. 57(5): 643-663.

Iriberri A, Leroy G. 2009. A life-cycle perspective on online community success. *ACM Computing Surveys (CSUR)*. 41(2): 11.

Jenkins H. 2006. *Convergence culture: Where old and new media collide* (NYU Press, New York, NY).

Jeppesen LB, Frederiksen L. 2006. Why do users contribute to firm-hosted user communities? The case of computer-controlled music instruments. *Organization Science*. 17(1): 45-63.

Kollock P. 1999. The economies of online cooperation. M.A. Smith, P. Kollock, eds. *Communities in cyberspace* (Routledge, London), 220-239.

Krishnamurthy S. 2002. Cave or community? An empirical examination of 100 mature open source projects. *First Monday*. 7(6).

Lakhani KR. 2016. Managing communities and contests to innovate with crowds. D. Harhoff, K.R. Lakhani, eds. *Revolutionizing Innovation: Users, Communities, and Open Innovation* (MIT Press, Cambridge, MA), 109-134.

Lakhani KR, Von Hippel E. 2004. How open source software works:"Free" user-to-user assistance. *Research Policy*. 32(6): 303-339.

Lakhani KR, Wolf R. 2005. Why hackers do what they do: Understanding motivation and effort in free/open source software projects. J. Feller, B. Fitzgerald, S. Hissam, K.R. Lakhani, eds. *Perspectives on Free and Open Source Software* (MIT Press, Cambridge, MA.

Lappas T, Dellarocas C, Derakhshani N. 2017. Reputation and contribution in online question-answering communities. SSRN.

Laursen K, Salter A. 2006. Open for innovation: The role of openness in explaining innovation performance among UK manufacturing firms. *Strategic Management Journal*. 27(2): 131-150.

Lavie D. 2006. The competitive advantage of interconnected firms: An extension of the resource-based view. *Academy of Management Review*. 31(3): 638-658.

Lerner J, Tirole J. 2002. Some simple economics of open source. *Journal of Industrial Economics*. 50(2): 197-234.

Lerner J, Tirole J. 2005. The economics of technology sharing: Open source and beyond. *Journal of Economic Perspectives*. 19(2): 99-120.

Liu Y, Bian J, Agichtein E. 2008. Predicting information seeker satisfaction in community question answering *Proceedings of the 31st annual international ACM SIGIR conference on research and development in information retrieval*. ACM, 483-490.

Lüthje C, Herstatt C, Von Hippel E. 2005. User-innovators and "local" information: The case of mountain biking. *Research Policy*. 34(6): 951-965.

Merton RK. 1993. *On the shoulders of giants* (University of Chicago Press, Chicago).

Morrison PD, Roberts JH, Von Hippel E. 2000. Determinants of user innovation and innovation sharing in a local market. *Management Science*. 46(12): 1513-1527.

Murray F, O'Mahony S. 2007. Exploring the foundations of cumulative innovation: Implications for organization science. *Organization Science*. 18(6): 1006-1021.

Nambisan S, Baron RA. 2009. Virtual customer environments: testing a model of voluntary participation in value co-creation activities. *Journal of Product Innovation Management*. 26(4): 388-406.

Pan B, Hembrooke H, Joachims T, Lorigo L, Gay G, Granka L. 2007. In google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication*. 12(3): 801-823.

Penrose E. 1959. *The Theory of the Growth of the Firm* (John Wiley and Sons, New York).

Porter CE, Donthu N. 2008. Cultivating trust and harvesting value in virtual communities. *Management Science*. 54(1): 113-128.

Ren Y, Kraut R, Kiesler S. 2007. Applying common identity and bond theory to design of online communities. *Organization Studies*. 28(3): 377-408.

Rheingold H. 1993. *The virtual community: Homesteading on the electronic frontier* (Addison-Wesley, Boston, MA).

Shah SK. 2006. Motivation, governance, and the viability of hybrid forms in open source software development. *Management Science*. 52(7): 1000-1014.

Shumaker SA, Brownell A. 1984. Toward a theory of social support: Closing conceptual gaps. *Journal of Social Issues*. 40(4): 11-36.

Von Hippel E. 1988. *The sources of innovation* (Oxford University Press, Oxford).

Von Hippel E. 1998. Economics of product development by users: The impact of "sticky" local information. *Management Science*. 44(5): 629-644.

Von Hippel E. 2005. *Democratizing innovation* (MIT press, Cambridge, MA).

von Rechenberg T, Gutt D, Kundisch D. 2016. Goals as reference points: empirical evidence from a virtual reward system. *Decision Analysis*. 13(2): 153-171.

Wasko MM, Faraj S. 2005. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly*. 29(1): 35-57.

Wasko MM, Teigland R, Faraj S. 2009. The provision of online public goods: Examining social structure in an electronic network of practice. *Decision Support Systems*. 47(3): 254-265.

Watson J. 2017. What is the Value of Re-use? Complementarities in Popular Music *NET Institute Working Paper No. 17-15*. NET Institute.

Weitzman ML. 1998. Recombinant growth. *Quarterly Journal of Economics*. 113(2): 331-360.

West J, Kuk G. 2016. The complementarity of openness: How MakerBot leveraged Thingiverse in 3D printing. *Technological Forecasting and Social Change*. 102 169-181.

Wiertz C, de Ruyter K. 2007. Beyond the call of duty: Why customers contribute to firm-hosted commercial online communities. *Organization Studies*. 28(3): 347-376.

Wooldridge JM. 2015. *Introductory econometrics: A modern approach* (Nelson Education).

Wu PF, Korfiatis N. 2013. You scratch someone's back and we'll scratch yours: Collective reciprocity in social Q&A communities. *Journal of the American Society for Information Science and Technology*. 64(10): 2069-2077.

Yu L, Nickerson JV. 2011. Cooks or cobblers? Crowd creativity through combination *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1393-1402.

Zhang L. 2016. Intellectual property strategy and the long tail: Evidence from the recorded music industry. *Management Science*. 64(1): 24-42.

Table 1. Distribution of Licenses

AllowDerivative	License		Freq.	Percent	Cumulative
1	Attribution		89,028	46.67	46.67
1	Attribution - ShareAlike		52,859	27.71	74.37
0	Attribution- NoDerivative		910	0.48	74.85
1	Attribution- NonCommercial		26,706	14.00	88.85
0	Attribution- NonCommercial- NoDerivative		4,362	2.29	91.14
1	Attribution- NonCommercial- ShareAlike		11,330	5.94	97.08
1	BSD License		391	0.2	97.28
1	GNU - Public		2,244	1.18	98.46
1	GNU Lesser General		187	0.1	98.55
1	Public Domain Dedication		2,757	1.45	100
		Total	190,774	100	

Table 2. Descriptive Statistics

	count	mean	sd	min	max
AllowDeriv	153214	0.9688	0.1737	0.0000	1.0000
Followers	153214	1.2135	1.4668	0.0000	9.4885
PreviousDeriv	153214	0.0077	0.0905	0.0000	3.2188
IsDerivative	153214	0.0789	0.2695	0.0000	1.0000
FileSize	153214	6.4556	2.1277	0.0039	14.8389
Experience	153214	1.7897	1.2675	0.0000	6.4520
N	153214				

Table 3. Correlations

No	Variable	1	2	3	4		5
1	AllowDerivative	1.0000					
2	Followers	-0.1124	1.0000				
3	PreviousDerivatives	0.0094	0.0218	1.0000			
4	IsDerivative	0.0320	0.0372	0.2287	1.0000		
5	FileSize	-0.0391	0.2055	0.0138	0.0416	1.0000	
6	Experience	-0.1054	0.7294	0.0521	0.0177	0.1487	1.0000

Table 4. Main Results

	(1) AllowDerivative	(2) AllowDerivative	(3) AllowDerivative	(4) AllowDerivative	(5) AllowDerivative	(6) AllowDerivative
Followers	-0.00613**		-0.00612**	-0.00506*		-0.00505*
	(0.00250)		(0.00250)	(0.00279)		(0.00279)
PreviousDerivatives		0.00351	0.00314		0.00398	0.00345
		(0.00352)	(0.00350)		(0.00348)	(0.00348)
FileSize				-0.00136***	-0.00137***	-0.00136***
				(0.000314)	(0.000315)	(0.000314)
Experience				-0.00196	-0.00338***	-0.00197
				(0.00124)	(0.00103)	(0.00124)
Designer FE	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Sub-Category	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.559	0.558	0.559	0.559	0.559	0.559
Observations	153214	153214	153214	153214	153214	153214

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5. Main Results

	(1) AllowDerivative	(2) AllowDerivative	(3) AllowDerivative	(4) AllowDerivative	(5) AllowDerivative
IsDerivative	0.00754***	0.00757***	0.00752***	0.00822***	0.00820***
	(0.00177)	(0.00178)	(0.00179)	(0.00179)	(0.00181)
Followers		-0.00614**			-0.00505*
		(0.00250)			(0.00279)
PreviousDerivatives			0.000463		0.000121
			(0.00355)		(0.00352)
FileSize				-0.00145***	-0.00143***
				(0.000317)	(0.000317)
Experience				-0.00339***	-0.00198
				(0.00103)	(0.00124)
Designer FE	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes
Sub-Category	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.559	0.559	0.559	0.559	0.559
Observations	153214	153214	153214	153214	153214

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$