TUHH
Technische Universität Hamburg-Harburg

Driving Diffusion of Scientific Innovation - The Role of Institutional Entrepreneurship and Open Science in Synthetic Biology

Vom Promotionsausschuss der Technischen Universität Hamburg-Harburg zur Erlangung des akademischen Grades Doktor der Wirtschafts- und Sozialwissenschaften (Dr. rer. pol.)

genehmigte Dissertation

von Giulio Barth aus Münster

2018

Advisors: Prof. Dr. C. Ihl, Prof. Dr. M. G. Möhrle

Institute of Entrepreneurship, TUHH

Gutachter:
Prof. Dr. C. Ihl
Prof. Dr. M. G. Möhrle

Vorsitz:
Prof. Dr. C. Lüthje

Tag der mündlichen Prüfung: 14. September 2018

"Our victory: inevitable; our timing: uncertain."
(Drew Endy, Assistant Professor Stanford University)

Abstract

Scientific innovations need to widely diffuse to fully exploit their potential. Prior research investigated levers on the diffusion of scientific innovation with particular interest on institutions, e.g., settings of property rights. As institutional theory lacks in explaining emergence and shaping of institutions, the institutional entrepreneur approach faces these limitations. Key actors combine logics from multiple fields and convince their social context of their ideas to legitimate the creation of new institutions and shape an emerging field.

This thesis validates theories on institutional entrepreneurs and investigates the end-to process from diffusing a logic to the impact of an established institution on scientific innovations in context of the emerging synthetic biology. The field is expected to introduce the 5th revolution and characterized by the central logic of making biology an engineering discipline.

In chapter 4 theories on institutional entrepreneurs driving diffusion of institutional logics to shape an emerging field are validated. To measure the social influence mechanisms, the heterogeneous diffusion model is adapted to the institutional logic. Predictions are tested using 8.3 million article-author pairs from Scopus.com. Authors who adopt the institutional logic are identified with topic modeling methods. To control for competition effects through similar prior research, a new measure called knowledge equivalence is introduced. This study makes two contributions to the investigation of how institutional logics diffuse to shape emerging fields. Firstly, the study highlights the role of institutional entrepreneurs in diffusing institutional logics in context of scientific innovation. Secondly, to track the process of social influence the heterogeneous diffusion model is applied to the diffusion of an institutional logic.

Based on the diffused logic, property rights are shaped in a field. In context of synthetic biology, open science initiatives are motivated by the engineering approach. To assess status quo, an impact assessment of open science and its success factors is performed in chapter 5. Here a query list of 478 open science parts is matched to 104 million sequences in patent applications. This new methodology yields eleven times more hits in comparison to the count of references. Results show a moderate diffusion of open science parts and cannot validate higher characterization quality as success factor. Potential success factors are discussed in the influence model framework and recommendations for practice elaborated.

To analyze the right side of the end-to-end process, the impact of selecting between open science and commercial research tools on knowledge diffusion is examined in chapter 6. Data from research articles using the commercial Zinc Finger research tool or an open science alternative are used for analysis. Exclusion restrictions are created to correct for the endogeneity of research tool selection. Results predict lower citation rates for an explicit selection of the commercial research tool. Also, no negative effect on the commercial research tool diffusion could be observed due to co-existence of an open science alternative.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor Prof. Christoph Ihl for the continuous support of my Ph.D study and related research, for his collegiality, patience, motivation, and immense knowledge. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisor, I would like to thank my co-advisor Prof. Martin G. Möhrle for his insightful comments and encouragement.

I am also hugely appreciative of Linda Kahl from the BioBricks Foundation for her frequent inspirational discussions, coaching and the introduction to the synthetic biology community. Without her precious dedication it would not be possible for me to conduct this research.

Special mention goes to my fellow mates at the TIE for the stimulating discussions, uncomplaining support on issues with R and for all the fun we have had in the last years.

My sincere thanks also go to Drew Endy, the Endy lab and especially Conary Meyer at Stanford University, who provided me an opportunity to join their amazing team, gave me countless lessons on genetic engineering and expert input for my analyses.

I thank my forever interested, encouraging and always enthusiastic grandmother: she was always keen to know what I was doing and how I was proceeding, although it is likely that she has never grasped what it was all about.

Last but not the least, I would like to thank my beloved ones: my parents, my sister, my godfather, my aunt and uncle, Fritz, and to my girlfriend for supporting me spiritually throughout writing this thesis and my life in general.

Thanks for all your encouragement!

Contents

List of Tables

List of Figures

List of Abbreviations

BBF	BioBricks Foundation
BiOS	Biological Innovation for Open Society
BLAST	Basic Local Alignment Search Tool
BPA	BioBricks Public Agreement
CoDA	Context-dependent assembly
DIY	Do-It-Yourself
DNA	Deoxyribonucleic acid
EST	Expressed Sequence Tags
iGEM	International Genetically Engineered Machine
IPR	Intellectual Property Rights
MMNL	Mixed Multinomial Logit
OPEN	Oligomerized Pool Engineering
PAC	Pink Army Cooperative
PDC	Public Domain Chronicle
PIPRA	Public Intellectual Property Resource for Agriculture
RFC	Request for Comments
Sc2.0	Synthetic Yeast 2.0
SJR	Scientific Journal Rankings
SynBERC	Synthetic Biology Research Engineering Center
ZF	Zinc Finger
ZFA	Zinc Finger Array

ZFC	Zinc Finger Consortium
ZFP	Zinc Finger Protein

List of Symbols

α	Alpha
β	Beta
$\hat{\beta}$	Parameter estimate beta
δ	Delta
$\mathbb{D}$	Set of subjects at time t_i
η	Eta
γ	Gamma
λ	Lambda
$\tilde{l}$	Observed sample of l
μ	Mu
π	Pi
σ	Sigma
$\boldsymbol{z}$	Vector z
$\approx$	Approximately
$<$	Smaller then
$>$	Larger then
L	Likelihood function
LL	Log-likelihood function
ln	Natural logarithm
$P(X\|Y)$	The probability of X given Y

Chapter 1

Introduction

There are only two possibilities of raising the output of economy. One can increase inputs of the productive process or think of how to get more output from the same amount of inputs (Abramovitz, 1956; Rosenberg, 2004). This process, better known as innovation, is therefore a dominant force in economic growth (Rosenberg, 2004). Moreover, innovation is essential to help address global challenges, such as climate change and sustainable development (OECD, 2007).

A prerequisite for innovation is knowledge creation, the so-called scientific innovation, which is generally but not exclusively done by scientists (Pejovich, 1996). Scientific innovations need to be both created and diffused to fully exploit its potential in innovative technologies (Rogers, 2003).

Researchers have validated the impact of institutions such as property rights on the diffusion of scientific innovation, e.g., contract-based property rights on "OncoMice" reduced their application in experiments by 20-40% (Murray et al., 2007). However, property rights and institutions in general have to be created and established before they can influence the diffusion of scientific innovations. The institutional entrepreneurship approach (Friedland et al., 1991) proposes theories on how institutional entrepreneurs diffuse institutional logics in their

social context to legitimate the creation of new institutions. In the context of scientific innovation examples are the Bayh-Dole Act, which was enacted by the US Congress to commercialize science (Mowery et al., 2001) and the platform sci-hub, which removed all barriers in the way of science (Bohannon, 2016). This thesis aims to investigate the comprehensive process from diffusion of an abstract institutional logic by institutional entrepreneurs in an emerging field to the impact of created institutions on the diffusion of scientific innovation.

These studies are of particular interest in a high potential industry like the emerging field of synthetic biology. The field emerges since the early 2000's and attracts scientists from multiple disciplines such as biology, chemistry, physics and software engineering (Raimbault et al., 2016). Despite a high variety of definitions (SCHER et al., 2014) a common aim of scientists in the field is to make biology an engineering discipline (Brent, 2004) and "be able to build just about anything from biology" (Keasling, 2013). This central engineering approach can be seen as fundamental logic and its standardization, decoupling, and abstraction principles (Endy, 2009) support the development of open science initiatives.

The study in chapter 4 investigates antecedents of institutions, i.e., how institutional entrepreneurs act to diffuse institutional logics to shape behavior in an emerging organizational field. In institutional theory organizational fields are characterized by a set of institutions shaping the behavior, but there are lacks in understanding emergence of new institutions (Leblebici et al., 1991) and limitations of isomorphism in explaining agency and change (DiMaggio, 1988; Scott, 2001). The concept of institutional logics (Friedland et al., 1991) and the institutional entrepreneur approach (DiMaggio, 1988; Garud et al., 2002; Campbell,

2004; Lawrence et al., 2004; Maguire et al., 2004) face these limitations. Institutional logics are linked to institutional orders and provide legitimacy for creating and modifying institutions. Institutional entrepreneurs are agents of change, motivated by self-interest, and combine and diffuse institutional logics from different institutional orders to create and shape emerging fields.

Previous research discusses the role of institutional entrepreneurs in the emergence of new organizational fields (e.g. Lawrence et al., 2004; Maguire et al., 2004), develops a model of multiwave diffusion of institutions linked by an institutional logic (Shipilov et al., 2010) or investigates the influence of structural holes on their activities (Ritvala et al., 2009). Empirical studies analyze the role of institutional entrepreneurs in diffusing product innovation and institutions in industries (e.g. Baron et al., 1986; Guler et al., 2002). However, there are rare attempts to validate theories on institutional entrepreneurs and to understand the strategies of institutional entrepreneurs in diffusing institutional logics to shape an emerging scientific field in particular (Pacheco et al., 2010). For example, the way in which institutional entrepreneurs convince their social context and use their resources to shape an emerging field.

Three research questions are derived to validate theories on institutional entrepreneurs in context of diffusing institutional logics to shape an emerging field.

1. Do institutional entrepreneurs use social influence mechanism to diffuse an institutional logic?

2. Does reputation of institutional entrepreneurs increase their infectiousness

on their social context?

3. Are institutional entrepreneurs the active drivers of diffusion and not susceptible towards social influence?

The emerging field of synthetic biology has been chosen for investigating the impact of institutional entrepreneurs on the diffusion of an institutional logic out of three reasons. First, the interdisciplinary field has immense significance in the future world introducing the 5th revolution and its primary players are already considered highly influential in modern society (Peccoud, 2016; Esquire, 2008). Second, the engineering approach can be seen as institutional logic that shapes the emerging field. And third, in a case study certain leadership roles, i.e., institutional entrepreneurs, could be identified in the field (Raimbault et al., 2016), however, research lacks on how these actors influenced their social context.

To measure the social influence mechanisms, the heterogeneous diffusion model is adapted to the institutional logic to incorporate three factors regarding the infectiousness of prior adopters, the effect of social proximity and individuals' susceptibility towards these communication channels. A new measure of social proximity, called knowledge equivalence, is introduced to acknowledge authors' original research areas before adopting the logic.

Predictions are tested using 8.3 million article-author pairs from Scopus.com published by 153 thousand authors which have been active in the broad field of genetic engineering. Authors who adopt the institutional logic are identified with topic modeling methods. To control for competition effects through similar

prior research, a new measure called knowledge equivalence is introduced.

All three research questions can be validated in the study; institutional entrepreneurs use social influence to convince their context of the institutional logic, they are more infectiousness with increased prior reputation and they are not susceptible towards social influence.

The study makes two contributions to the investigation of how institutional logics diffuse to shape emerging fields. Firstly, the study highlights the role of institutional entrepreneurs in diffusing institutional logics in context of scientific innovation. Secondly, to track the process of social influence the heterogeneous diffusion model is applied to the diffusion of an institutional logic. The empirical data validates that institutional entrepreneurs use their resources to convince their social context of an institutional logic. This way they shape the emerging field and grow the social collective supporting their logic.

Legitimated by an established institutional logic, institutions can be created (Shipilov et al., 2010) and the diffusion of scientific innovation controlled. This impact of established institutions is investigated in chapter 5 and 6.

To better understand the complexity of how institutions and settings of property rights, in particular, are crucial to diffuse scientific innovations, a broader review of prior research is needed. Property rights, such as patents, have been created to establish a market of ideas (Nelson, 1959; Arrow, 1962; Merges et al., 1990; Merges et al., 1994; Fosfuri et al., 2001; Gans et al., 2000), because competitive markets are supposed to under-incentivize innovation due to the public good character of ideas (Nelson, 1959; Arrow, 1962). Recent research investigates effects of property rights on cumulative innovations (Argyres et al., 1998; Krimsky,

2004; David, 2001). Here, the hypothesis "tragedy of anti-commons" (Heller et al., 1998) is postulated. It states that intellectual property rights might reduce the flow of information and inhibit cumulative research (Heller et al., 1998; David, 2000; Lessig, 2002; David, 2004). However, scholars have expressed consensus that negative aspects of property rights are not caused by the established system but by licensor's behavior (Walsh et al., 2003; Murray, 2006; Murray et al., 2007). The effect of property rights on follow-on research was quantitatively analyzed in multiple studies (Murray et al., 2007; Williams, 2013; Murray et al., 2016) finding 10-40% decrease of follow-on research, if early stage innovations are protected with private property rights or contracts.

In their conclusion, Murray et al. (2016) consider the possibility that researchers are using alternative open technologies to avoid patent protection without further analyzing this hypothesis.

In their paper Murray et al. (2016) started a new research question investigating the effect of open science on the type of follow-on research. They validate an increase of creativity and diversity of follow-on research after giving access to early-stage research to academia. Studies in both streams analyze the effects on proprietary technologies and research tools. Competition between commercial and open science research tools are not yet explained.

In a third stream, researchers analyzed predictors for knowledge diffusion measured in citation rates. While publishing in open accessible journals is not correlated with citation rates (Eysenbach, 2006; Davis et al., 2008; Gaule et al., 2011), Piwowar et al. (2007) validated a positive effect when researchers share detailed result data, as follow-on research can access all needed information and build

on it directly.

Building on these three streams of prior research, i.e., the effect of property rights on diffusion of knowledge, the impact of open science research tools on type of follow-on research and predictors for citation rates, the question arises which impact the selection between open science and commercial research tools has on knowledge diffusion of a paper.

Three questions are therefore essential in regards to property rights in synthetic biology.

1. What impact does open science have in the field of synthetic biology?
2. What are success factors for open science initiatives?
3. How does selection between competing open science and commercial research tools affect knowledge diffusion of a paper?

To assess current impact of open science in synthetic biology, the diffusion of open science parts is investigated in chapter 5. By matching a query list of 478 open science sequences published by the BIOFAB project in 2010 to 104 million sequences in patent application data collections, a sophisticated analysis of re-usage is elaborated. This methodology of matching DNA sequences is new in property rights research, where patent-paper pairs citation references are standard practice to measure diffusion (Furman et al., 2011; Murray et al., 2007; Williams, 2013; Murray et al., 2016). Inventors of patent applications are personally contacted to confirm re-usage and provide feedback on further implementations. The BIOFAB parts are well suited for this analysis as they cover both newly created parts, which are analyzed to assess diffusion of open science

parts, and already existing parts with improved characterization quality, which are monitored to evaluate for effects on diffusion. Characterization quality is seen by the community as an essential factor to promote open science parts because adoption costs are lowered (Kahl et al., 2013).

Results show a moderate diffusion of new created BIOFAB parts and cannot identify an effect of higher quality characterization of re-usage on open science parts. Three advances are elaborated. The new methodology of measuring diffusion of innovation by matching DNA sequences with patent databases can be adopted to validate prior studies and extend to further research questions. Using the case of BIOFAB parts, eleven times for re-usages could be identified than with standard practice of counting references. Second, the moderate impact of open science postulated in survey results (Kahl, 2015) could be validated using the case of BIOFAB parts. Implementation of the sequences into a software was noted by an inventor, a more extensive diffusion of the parts must be assumed. Third, new research questions regarding success factors for open science technologies can be built on this study using the influence model (Asavathiratham et al., 2001).

Question three is investigated in chapter 6 using the case of zinc finger (ZF) research tools. ZF was chosen because of three reasons. First, ZF research tools were leading for gene-editing DNA before the rise of CRISPR and many innovative advancements relied on them (Chandrasekharan et al., 2009). Second, as the company Sangamo held patents on the core tool and did not disclose all relevant innovations, access was only possible through buying the expensive commercial research tool or agreeing to license agreements (Chandrasekharan et al., 2009).

Third, a consortium of academics released an open science alternative research tool (Chandrasekharan et al., 2009). The prediction is tested using 396 peer reviewed research articles, corresponding article metadata and the authors' social networks. Regression results validate that published studies using the commercial research tool are predicted to have lower citation rates. The study follows Murray et al. (2016) in evaluating effects of open science on follow-on research and validates benefits of selecting open science research tools for knowledge diffusion. In addition, no negative effect on the diffusion of the commercial research tool could be observed due to co-existence of an open science alternative. The remainder of the thesis consists of five chapters. Firstly, the institutional entrepreneur approach and the impact of institutions on diffusion of scientific innovations is reviewed in chapter 2. In chapter 3 the field of synthetic biology is explored, an understanding for property rights in synthetic biology and an overview on 28 important organizations and initiatives establishing open science in synthetic biology is given, and foundational mechanisms in genetics is explained. Chapter 4, 5 and 6 present the three studies discussed. Conclusions for results of the studies are summarized in 7. Bibliographic references and additional analyses results are given.

Chapter 2

Theoretical foundations: Institutional Entrepreneurship and Institutions

Theoretical foundations in institutional entrepreneurship and institutional theory are used for the three studies in this thesis. The following chapter summarizes key implications from prior research in institutional theory regarding emergence of new institutions and their effect on diffusion of scientific innovations.

2.1 Organizational Field and Institutions

In the research area of institutional theory, scholars try to understand the behavior of individuals and organizations. The scope of research relies on investigating the role of institutions in shaping this behavior (Thornton et al., 2008). Studies in institutional theory have used several units of analysis, such as the institutional field (Meyer et al., 1977; DiMaggio et al., 1991), the societal sector (Meyer et al., 1992) or the institutional environment (Orru et al., 1991). The finally accepted term for the constellation of actors constituting this central unit of analysis is ”Organizational field” (Scott, 1991; Wooten, 2006).

An organizational field is the domain where an organization's behavior is structured by the grid of relationships within which it was embedded (Warren, 1967; Wooten, 2006). The behavior of actors in an organizational field is guided by institutions, a structure to provide stability and collective meaning to social behavior (Scott, 1995). Institutions are needed to determine appropriate actions (Zucker, 1977; Meyer et al., 1977).

Institutions and organizational fields are interrelated in their development process. Patterns of social interactions help to produce and reproduce common understandings and practices to form institutions that define the field, at the same time institutions shape these patterns of social interactions (DiMaggio et al., 1983).

Fligstein (1997) sub-classified organizational fields into emerging, mature and stable, or into fields in crisis. Emerging fields, in particular, are characterized by a lack of institutional practices and, thus, by "fluid relationships, conflicting values and an absence of clearly identifiable norms" (Hardy et al., 2008, p. 201). Similar fields in crisis are dominated by tensions and contradictions (Fligstein et al., 1996) and thus are an easier domain for change (Greenwood et al., 2002; Greenwood et al., 2006).

The study of institutions has proceeded in multiple waves beginning with empirical analyses of organization and institutional environments (Selznick, 1949; Selznick, 1957) and with theoretical discourses on how institutions are integrating divers organizations through rules, contracts and authority (Parsons, 1956). Institutions are products of human design (DiMaggio et al., 1991) and develop in the process of institutionalization: Institutions are first seen as social facts and

by transforming from one actor to the other they receive rule-like or taken-for-granted status and get institutionalized (Zucker, 1977). Prior research concentrates on three types of institutions; practices (Leblebici et al., 1991), standards (Greenwood et al., 2002) and policies (Garud et al., 2002). An embedded institution provides legitimacy, which is seen as having higher value than efficiency (Meyer et al., 1977), thus, seen from the new institutional theory perspective, normative and mimetic forces are dominating (Wooten, 2006).

Investigating change of insistutions, Meyer et al. (1977), Zucker (1977) and DiMaggio et al. (1983) emphasize the role of culture and cognition in institutional analysis and isomorphism, which means commonality in form and function (DiMaggio et al., 1983). This effect of isomorphism was first observed in organizational structures on a societal level (Meyer et al., 1977), then extended to organizational fields (DiMaggio et al., 1983). DiMaggio et al. (1983) highlight coercive, normative and mimetic sources of isomorphism.

Critics claim that isomorphism rejects rationalism (Hoffman et al., 2002). Consistently, it leads to organizational fields being static in configuration, unitary in make up, formed around common technologies and, finally, coercive, normative and mimetic influences are forcing towards homogeneity (DiMaggio, 1995; Greenwood et al., 1996). Scholars pushed for the so-called "cognitive turn" (Lindenberg, 1998; Meindl et al., 1994). They introduced institutional logics as schema to guide behavior of field members.

2.2 Institutional Logics

Institutional logics are cultural and cognitive processes, material practices and symbolic constructions. They thus function as link between institution and action and effects of isomorphism are undermined by effects of institutional logics on individuals and organizations. This leads to a model with rational and mindful behavior and actors who are able to shape and change institutions individually (Friedland et al., 1991; Haveman et al., 1997; Thornton et al., 1999; Scott, 2000).

2.2.1 Definition of Institutional Logics

Institutional logics were defined by two separated scholars, namely Friedland et al. (1991) and Jackall (1988) and then harmonized by Thornton et al. (1999).
Friedland et al. (1991) identified conflicting practices and beliefs in modern western society and defined institutional orders, such as political democracy, state bureaucracy and capitalism, as origin. Each institutional order has a central logic, which has guiding principles and provides actors with identity. The logic thus defines mean-end of individual behavior and is constitutive for individuals, organization and society pushing the symbolic dimension into focus. They see inter-institutional contradictions between systems, e.g., market, family and profession (Friedland et al., 1991).
Jackall (1988) defines logics as complicated and experimentally constructed set of rules that actors create to make behavior predictable. The normative dimension of institutions is emphasized and intra-institutional conflicts are in focus

(Thornton et al., 2008).

Thornton et al. (1999, p. 804) build on both perspectives and defines institutional logics as 'the socially constructed, historical patterns of material practices, assumptions, values, beliefs, and rules by which individuals produce and reproduce their material subsistence, organize time and space, and provide meaning to their social reality'. Logics thus provide a 'link between the individual agency and cognition and socially constructed institutional practices and rule structure' (Thornton et al., 2008, p. 101). In this definition of institutional logics, all three dimensions of institutions, i.e., coercive, cognitive and normative ones, are simultaneously considered and integrated (Thornton et al., 2008). An alternative approach of Scott (1995) considers the three dimensions separately.

Regarding the relation between institutional logics and organizational fields, Thornton et al. (2008) postulates three hypotheses. Institutional logics stem from institutional orders, they are locally instantiated and enacted in organizational fields. Organizational fields are the space where "institutional logics get played out" (Thornton et al., 2008, p. 119).

2.2.2 Five Principles of Institutional Logics Meta-Theory

The meta-theory of institutional logics describes how institutions shape heterogeneity and stability and change both individuals and organizations. It builds on five principles (Thornton et al., 2008).

The first principle is about embedded agency (Seo et al., 2002; Battilana, 2006; Greenwood et al., 2006). Embedded agency describes the interplay of individ-

ual agency and institutional structure, i.e. individuals or organizations seek power or status, however mean-ends of interest and agency are both enabled and constrained by institutional logics (Giddens, 1984; Sewell Jr, 1992). Interests, identities, values and assumptions of individuals and organizations are thus embedded in institutional logics (Thornton et al., 2008). Condition for embedded agency is a partial autonomy of individuals and organizations in shaping social structure and action (Friedland et al., 1991).

Translating the concept to the three levels of society, i.e., institutions, organizations and individuals, intra- and cross-level effects turn up (Berger et al., 1967). On each level there are intra-level effects. On the institutional level, contradiction and interdependency take place. Organization relations are characterized by conflicts and coordination and individuals are both competing and negotiating (Thornton et al., 2008). In addition, cross-level effects occur. While individuals are embedded in organizations which, in turn, are embedded in institutions, these ones are socially constructed by actions of individuals and organizations. Cross-level effects, in particular, are critical and can lead to change and institutional entrepreneurship (Battilana, 2006; Greenwood et al., 2006).

The second principle proposes society as an inter-institutional system, the so-called concept of institutional orders (Friedland et al., 1991; Thornton et al., 2008). Thornton et al. (2008) rate the view of society as an inter-institutional system being the most important concept developed by Friedland et al. (1991). The inter-institutional systems of societal sectors all have a different batch of expectations for social relations and behavior. Institutional logics and the contradictions between them can be seen as a source of heterogeneity and agency (Thornton

et al., 2008). The concept of institutional logics allows organizational fields to be dynamic, neither characterized by homogeneity nor driven by isomorphism. Any context can be influenced by contending logics of different societal sectors (Thornton et al., 2008).

Institutional analyses need to consider two implications of this principle. First, no institutional order is, a priori, determined to be dominant in a societal sector. And second, categories of knowledge are both actively shaping by institutional logics and, in turn, designed by them. Thus, history and development of categories of knowledge need to be considered to avoid endogeneity (Thornton et al., 2008; Friedland et al., 1991).

The third principle emphasizes the equality of material and cultural characteristics of institutional orders (Friedland et al., 1991). This builds on both Becker (1974) claiming the involvement of family and religion in the consumption of goods and on Granovetter (1985) developing models which are shaped by culture. There is no dominance of either characteristic, but the interplay between them results in the development and change of institutions (Thornton et al., 2008). The meta-theory of institutional logics, thus, does not evaluate rationality or irrationality of actions, but tries to understand how contradictions and conformity of institutional logics, having both cultural and material characteristics, influence behavior in society (Thornton, 2002).

Culture, in particular, is needed for actors to realize economic and political struggles and to identify appropriate behavior (Thornton et al., 2008). Both symbolic and normative components of culture are considered responding to critics of early neo-institutional theory, which criticized dominating normative

and mimetic forces (Wooten, 2006). DiMaggio et al. (1991) underlined symbolic and cognitive dimensions of logics, but under-estimated the normative influence (Hirsch, 1997; Mizruchi et al., 1999). Ignoring the effect of norms would under-socialize the view of individual and organizational behavior (Granovetter, 1985). The consideration of norms implies simultaneous co-existence of multiple dominant and subsidiary norms and, thus, a probabilistic view to adhering to norms of behavior (Thornton et al., 2008).

The fourth principle claims multi-level setting of institutions (Thornton et al., 2008). While Friedland et al. (1991) had a focus on societal level and the influences on individuals and organizations, modern meta-theory takes place on multiple levels such as organizations, the organizational field or markets. Theoretical mechanism, i.e., effects on individuals and organizations, can happen on levels being different from the main anomaly investigated in an analysis (Thornton et al., 2008). On organizational field level, scholars, in particular, emphasize the second principle meaning the co-existence of contradictory institutional logics (Kitchener, 2002; Greenwood et al., 2006; Lounsbury, 2007).

The fifth principle considers the key assumption of the historical contingency of influencing effects, enforced by larger environments, on the behavior of individuals and organizations. Multiple scholars validated the temporal dimension of an institutional logic meaning and importance (Thornton et al., 1999; Scott, 2000; Lounsbury, 2002; Zajac et al., 2004; Meyer et al., 2006). It is therefore important in the studies of institutional logics to analyze whether assumed theories are universal through time and space or particular to the environment (Thornton et al., 2008).

2.2.3 Institutional Logics Shape Behavior

Based on the five principles of meta-theory, a certain mechanism through which institutional logics shape behavior of individuals and organizations can be developed. Thornton et al. (2008) summarize four kinds mechanism, namely collective identity, embedded agency, social classification and attention.

Collective identities and identification are mostly analyzed. Here, individuals and organizations identify with the collective identity of an institutionalized group, e.g., profession, industry or organization (Tajfel et al., 1979; March et al., 1989). Polletta et al. (2001, p. 285) define a collective identity as "individual's cognitive, moral, and emotional connection with a broader community, category, practice, or institution".

Social groups form collective identities through interaction and communication (White, 1992). Individuals who identify with a collective identity presumably collaborate with the social group, accept their rules on behavior and protect the rules against other identities (Tyler, 1999; Brickson, 2000; March et al., 1989; Kelman, 2006; Tajfel et al., 1979; White, 1992).

In the field of empirical research, scholars validated that actors can participate in multiple social groups simultaneously, e.g., professions, gender and social movements (,e.g., Abbott, 1988; Fine, 2001; Cerulo, 1997; Benford et al., 2000). Several scholars have analyzed elemental types of collective identities. They validated that collective identities can also emerge, e.g., from organizational forms or market competitors (Haveman et al., 1997; White, 1992). Through the process of institutionalization, a collective identity develops its specific institutional logics

(Jackall, 1988). Thus, any identification with a collective implies the identification with its institutional logics (Thornton et al., 2008).

As illustrated in the concept of embedded agency, the seek of actors for status and power is facilitated and constrained by institutional logics providing rules and mean-ends (Fligstein et al., 1996; Thornton et al., 1999; Lounsbury, 2002). Regardless of institutional logics, power and status exist as well as differences between individuals. However, only due to institutional logics are source and consequences of these differences understood (Jackall, 1988; Thornton et al., 1999; Lounsbury, 2002).

In social classification DiMaggio (1997) identified a key mechanism through which logics shape individual cognition. This is based on the causal chain of institutional logics influencing social and organizational categories (Douglas, 1986; Searle, 1995) and on the importance of social categories in framing individual cognition (Rosch, 1975; Medin, 1989). Systems of classifications, e.g., of social actors and organizational forms, are socially constructed and linked to institutional logics (Mohr et al., 1997; Haveman et al., 1997; Thornton et al., 2008). Due to this link, a change of institutional logics implies a creation of new categories and might change their meaning (Rao et al., 2003; Breiger et al., 2004; Haveman et al., 1997; Rao et al., 2005; Ruef, 1999).

Institutional logics can influence behavior through two kinds of mechanism related to structuring attention (Ocasio, 1997). They can create a compilation of values that assign legitimacy and importance to issues. A second mechanism is seen in giving actors a certain understanding of their interests and identities. Thus, they direct the attention of actors on issues and solutions which are in line

with institutional logics (Thornton, 2002).

2.3 Institutional Entrepreneurship

A change of institutional logics is facilitated by the inter-institutional system principle developed by Friedland et al. (1991). In this system, logics are defined by cultural differentiation, fragmentation and by contradiction both between and within institutional orders (DiMaggio, 1997; Friedland et al., 1991). Processes of change are initiated by structural overlap and institutional entrepreneurship (Thornton et al., 2008). While structural overlap is mostly seen in settings where mergers and acquisitions bring together organizations with divers institutional logics or organizations crossing multiple organization fields and therefore get in contact with multiple institutional logics (Thornton, 2004; Greenwood et al., 2006), the effect of institutional entrepreneurship has shown a high variety in appearance and is mostly driven by individuals and groups of actors.

DiMaggio et al. (1991, p. 14) claim that ”new institutions arise when organized actors with sufficient resources see in them an opportunity to realize interests that they value highly” and Maguire et al. (2004, p. 657) extend the definition describing institutional entrepreneurship as ”activities of actors who have interest in particular institutional arrangements and who leverage resources to create new institutions or to transform existing ones”. With the development of this new concept a bridge was built between old and new institutionalism and many scholars ran empirical studies for further analysis (Thornton et al., 2008).

Institutional entrepreneurship in organizational fields is initiated and driven by

institutional entrepreneurs, agents of change, who are motivated by self-interest. They can be individuals (,e.g., Fligstein, 2001b; Kraatz et al., 2002; Lawrence et al., 2004; Maguire et al., 2004; Dew, 2006) or organizations (e.g., Déjean et al., 2004; Demil et al., 2005; Garud et al., 2002; Hensmans, 2003; Leblebici et al., 1991). Central to organizational fields is the legitimacy of an organization, which can be acquired from the environment by adhering to conforming practices (DiMaggio et al., 1983).

New organizational forms, e.g., products, services or technology need legitimacy to be accepted by social context (Ashforth et al., 1990; DiMaggio, 1988; Durand et al., 2005; Rao, 1994). To acquire this, any institutional entrepreneur pursues to convince her context of necessity, validity and usefulness of her idea (Snow et al., 1992). Therefore to build an environment that enacts with her ideas and claims is a challenge for the institutional entrepreneur. These processes can take place in existing environments (Suddaby et al., 2005) or in emerging fields (Maguire et al., 2004). An institutional entrepreneur has to consider both dimensions of environments, the material as well as the symbolic one, and will im- and export practices and symbols from one to the other institutional order (Thornton et al., 2008).

Holm (1995, p. 398) postulated the paradox of embedded agency asking how actors can ”change institutions if their actions, intentions, and rationality are all conditioned by the very institution they wish to change”. Both, research on the initiation of change and the implementation by institutional entrepreneurs are reviewed in the following two sections.

2.3.1 Determinants for Inducing Institutional Change

Hardy et al. (2008) distinguish four clusters of existing literature on institutional entrepreneurship investigating the initiation of change.

The first cluster of literature analyzes properties of institutional entrepreneurs as pre-requisite to initiate change. Both cognitive psychology and critical realism developed hypotheses on how institutional entrepreneurs identify opportunities for change (Hardy et al., 2008). In cognitive psychology, institutional entrepreneurs are of an "analytically distinguished social type" (Beckert, 1999, p. 786) and they are able to reflect about the status quo of their institutional environment, to observe drawbacks and to translate this cognition into actions designing new institutions (Beckert, 1999; George et al., 2006).

Critical realism, in general, emphasizes the interaction with socially constructed institutions which declare appropriateness and define which actors have the right to reflect (Scott, 1995). Mutch et al. (2006) favor critical realism and hypothesize institutional entrepreneurs to be reflecting in isolation from others' concerns to focus on the influence of context (Mutch et al., 2006; Mutch, 2007; Leca et al., 2006).

The second cluster argues that the limited number of social positions of actors in the field, the so-called subject positions, are enablers to become institutional entrepreneurs (Maguire et al., 2004; Battilana, 2006; Oakes et al., 1998). Within these subject positions, struggles and maneuvers over power and success take place (Oakes et al., 1998). Organizational fields are "structured systems of social positions" (Oakes et al., 1998). Actors do not have power on their own but they

occupy positions with power (Hardy et al., 2008). Institutional entrepreneurs in subject positions therefore possess "specific qualities" (Meyer et al., 2006). Occupying high subject positions with high power and connectivity is assumed to lower the ability to reflect and, thus, initiate change (Holm, 1995).

However, multiple studies found several cases where even highly central actors, i.e., high subject positions, could initiate change (Townley, 2002; Lee et al., 2002; Greenwood et al., 2002; Greenwood et al., 2006; Rao et al., 2003). An explanation is that even central actors have access to divers institutional logics and are therefore able to initiate change (Hardy et al., 2008). At the same time it is assumed that actors in low subject positions are not having enough resources to drive a change. Similar to the first assumption, scholars identified cases where actors in low subject positions successfully drove changes (Hensmans, 2003; Leblebici et al., 1991; Lounsbury et al., 2003).

A third cluster values initial field conditions as an important factor for institutional entrepreneurship. They argue that institutional entrepreneurship is frequently accompanied by specific field conditions, i.e., stimuli of uncertainty or tensions and contradictions (Hardy et al., 2008). The stimulus of uncertainty is the focus of the economic approach, whereby actors try to reduce uncertainty and solve problems (Coase, 1937; Pfeffer et al., 1978). In situations of uncertainty, actors cannot calculate probabilities or define rational strategies (Knight, 1921; Beckert, 1999). Therefore, institutional entrepreneurs design institutions to make actors' behavior predictable (Coase, 1937; Williamson, 1985; North, 1990).

Tensions and contradictions as source of institutional entrepreneurship are investigated by a second group of scholars (Sewell Jr, 1992; Zilber, 2002; Seo et

al., 2002; Greenwood et al., 2002; Rao et al., 2003; Dorado, 2005; Greenwood et al., 2006). They argue that in a field, stability is a superficial phenomena and differences among actors are only 'temporarily resolved by socially negotiated consensus' (Greenwood et al., 2002). Several institutions, which are all in conflict, can arise within a single field. These tensions and contradictions are then cause for institutional change.

A fourth cluster translates the field conditions into the state of a field as important influencer for institutional change. Fligstein (1997) sub-classified organizational fields into emerging, mature and stable, or into fields in crisis. Emerging fields, in particular, are characterized by a lack of institutional practices and, thus, by "fluid relationships, conflicting values and an absence of clearly identifiable norms" (Hardy et al., 2008, p. 205).

The field conditions are dominated by tensions and contradictions (Fligstein et al., 1996). Members show a mutual interest in interaction, but only uncoordinated actions take place (Hardy, 1994; Trist, 1983). In these "potential networks" (Gray, 1985), "proto-institutions" (Lawrence et al., 2002) are narrowly diffused and institutional entrepreneurs are strategic and opportunistic (DiMaggio, 1988; Fligstein, 1997) as they foresee considerable rewards for success (Hardy, 1994). Similar observations are made in fields in crisis which are therefore an easier domain for change (Fligstein et al., 1996; Hoffman, 1999; Fligstein, 2001a; Greenwood et al., 2002; Greenwood et al., 2006).

2.3.2 Strategies to Drive Change

After initiating change, institutional entrepreneurs try to acquire legitimacy for their ideas (DiMaggio, 1988). They enact in a variety of strategies to persuade their social context of their innovation and institutionalize it (DiMaggio, 1988; Garud et al., 2003; Hardy et al., 2008). Three strategies are described in the following, namely mobilizing resources, constructing rationales by using framing and theorization, and establishing relations.

According to DiMaggio's definition of institutional entrepreneurs, 'sufficient resources' play a significant role (DiMaggio, 1988). The more legitimacy and power in terms of resources, knowledge and limited social network positions an institutional entrepreneur has, the better she can drive change and shape her context (Suchman, 1995; Beckert, 1999; Lawrence, 1999; Maguire et al., 2004; Foucault, 1972; Bourdieu et al., 1992; Hoffman, 1999). A wide range of resources is investigated by means of prior research (Hardy et al., 2008), e.g., political, financial and organizational resources (Beckert, 1999; Greenwood et al., 2006), material resources (Lawrence et al., 2006; Monteiro et al., 2015), cultural resources (Creed et al., 2002), affiliations and networks (Montiel et al., 2009; Ritvala et al., 2009; Raffaelli et al., 2014), as well as discursive resources (Hardy et al., 1999; Hensmans, 2003; Lawrence et al., 2004; Maguire et al., 2006).

Regarding material resources, scholars argue that they are mobilized to both, dominating others and negotiating support for the entrepreneur's institutionalization project. Domination of others is characterized by the ability of an institutional entrepreneur to control rewards and punishments. In negotiations,

potential supporters need to perceive tangible and intangible benefits (Colomy, 1998).

A second strategy to drive change uses constructed rationales, which is the focus for social movement theory. Institutional entrepreneurs can frame new arrangements being more agreeable to a wide context (Rao, 1998) or theorize their ideas, i.e., build chains of cause and effect from their abstract categories (Greenwood et al., 2002) to induce their innovations to the social context. They frame changes in a way to generate collective action (Benford et al., 2000; Lounsbury et al., 2003; Garud et al., 2002), e.g., by describing problems with existing practices to legitimate their new ones (Strang et al., 1993; Greenwood et al., 2002). Building a collective action frame follows a defined structure of punctuation, elaboration and motivation (Snow et al., 1986; Creed et al., 2002; Hardy et al., 2008). In these phases, the problem and its importance are identified, a diagnosis and counter activities elaborated and a call for action issued (Creed et al., 2002; Misangyi et al., 2008).

Using appropriate frames, the chance for the institutionalization of their ideas can be increased (Gray et al., 2015). For building a context to their ideas, alternative logics are utilized (Seo et al., 2002) and adjusted to, e.g., rules of society (Haveman et al., 1997) and cultural accounts (Creed et al., 2002). They position their ideas using existing categories and systems (Hargadon et al., 2001) and build on available discourses (Hardy et al., 1999; Lawrence et al., 2004).

With framing and theorization institutional entrepreneurs discredit the status quo and claim the necessity and validity of their ideas to make support reasonable for others (Rao, 1998). In doing so, institutional entrepreneurs partially

choose the way of discursive interventions with other actors, e.g., share ideas and participate in collective sense making (Boxenbaum et al., 2005; Edelman et al., 1997) and, partially, process with unilateral actions only (Dew, 2006; Garud et al., 2002; Fligstein, 2001a; Demil et al., 2005; Maguire et al., 2004; Fligstein, 1997).

A third strategy is building new relations to organize collective action (Dew, 2006; Aldrich et al., 1994; Garud et al., 2002; Lawrence et al., 2002; Wijen et al., 2007). Scholars observed profound political and social skills of institutional entrepreneurs (Perkmann et al., 2007) and, thus, the "ability to motivate cooperation of other actors by providing them with common meanings and identities" (p.397 Fligstein, 1997). By occupying subject positions, they have the ability needed to exercise power (Fligstein, 2001b).

2.3.3 Institutional Entrepreneurship in Institutional Economics

Studies in the research area of institutional theory use empirical event history data (Anand et al., 2004; Durand et al., 2005; Greenwood et al., 2006; Leblebici et al., 1991; Garud et al., 2002; Holm, 1995) except Sine et al. (2009) who validated the influence of social movements beyond their targets using a quantitative approach.

Institutional entrepreneurship in institutional theory focuses on the process of change and sees institutional entrepreneurs as agents of change motivated by self-interest (DiMaggio, 1988). In contrast to that, institutional entrepreneurship in institutional economics focuses on the outcome of a change (Pacheco et al.,

2010). Here, institutional entrepreneurs are motivated by economic benefit, only (La Croix et al., 1990). Thus, new institutions are created when benefits are dominating costs for an entrepreneur (Alston et al., 1999; Anderson et al., 1975; Demsetz, 1967; North et al., 1970).

Institutions are constantly in a status of transformation and a central factor to this dynamic are exogenous shocks, e.g., in change of demand, technology or culture (Alston et al., 1996; Anderson et al., 1975; Bromley, 1989; Finbow, 1993; Ogilvie, 2007; Ruttan, 2006; Tan, 2005; Zerbe et al., 2001). Institutional economics concentrate on lower levels of environment; formal rules, such as property rights, and governance institutions, e.g., contracts, are subject to institutional entrepreneurship (North et al., 1970). In institutional economics, mechanisms of change are less subject to individual institutional entrepreneurs but driven by collective interest groups and employ resources and strategies. Interest groups influence the direction of institutional change to their benefit and create formal institutions (Binswanger et al., 1997; Goldberg, 1974; Harris, 1997; Higgs, 1996; Krueger, 1988; Ruttan, 2006; Tan, 2005).

2.4 Impact of Institutions on Innovation Diffusion

Impact of institutions and, in particular, of property rights on the diffusion of innovation have a long history in research. Competitive markets are supposed to under-incentivize innovation due to the public good character of ideas (Nelson, 1959; Arrow, 1962). Property rights, such as patents, have thus been created to establish a market of ideas (Nelson, 1959; Arrow, 1962; Merges et al., 1990;

Merges et al., 1994; Fosfuri et al., 2001; Gans et al., 2000). In addition, they incentivize inventors to publish a detailed description of their technology and, thus, facilitate both commercialization of the idea and follow-on innovation (Scotchmer, 2004; Thursby et al., 2007; Merton, 1973; Kitch, 1977; Thursby et al., 2001; Hellman, 2007). Hence, patents are seen as by-products of scientific work (Murray, 2006) and increase the ROI of research investments (Williams, 2013).

Empirical research on the impact of property rights on isolated innovations validates positive effects on publication outcome (Fabrizio et al., 2008; Looy et al., 2003; Breschi et al., 2007; Czarnitzki et al., 2007; Stephan et al., 2007) and analyzes which breadth and length of patents minimize the cost of monopoly distortion (Gilbert et al., 1990; Klemperer, 1990). Negative effects are identified in reduced quality of research output due to property right incentives (Henderson et al., 1995; Trajtenberg et al., 1997; Czarnitzki et al., 2009).

Innovation occurs through the interaction of multiple actors (Freeman, 1989; Freeman, 1994; Lundvall, 1992; Nelson, 1993; Nelson et al., 1993; Mansfield et al., 1996; Mansfield, 1995; Mowery et al., 1999; Dosi, 2000) and field perspective is needed to capture levers for boosting social value. Therefore, recent research analyzes focuses on cumulative innovations and potential hindering by early-stage property rights. Cumulative innovations can be both sequential early stage and follow-on research as well as multipurpose innovation, i.e., patent-paper pairs. The framework of follow-on research is based on Dasgupta et al. (1994) and was elaborated in Stokes (2011). Exception of this binary categorization developed for sequential innovations are multipurpose innovations pushed further by Bayh-Dole Act in 1980 (Murray et al., 2007).

Potentially limiting factors in cumulative innovations were highlighted by several scholars (Argyres et al., 1998; Krimsky, 2004; David, 2001). The hypothesis "tragedy of anti-commons" (Heller et al., 1998), an effect imitating the so-called tragedy of commons, is postulated. It states that intellectual property rights might reduce the flow of information and, thus, inhibit cumulative research (Heller et al., 1998; David, 2000; Lessig, 2002; David, 2004). Owners of property rights can use their blocking power and scarce resources, i.e., innovations can become under-utilized. Scholars have expressed consensus that negative aspects of property rights are not caused by the established system but by licensors' behavior (Walsh et al., 2003; Murray, 2006; Murray et al., 2007).

Scotchmer (1991) theorizes on the effect of patent breadth on incentives for follow-on research and possible counter adjustments with prior agreements or licenses. When costs for an innovation exceed its standalone benefits, follow-on innovations have to cover the difference. This calculation highlights that follow-on innovations have to pay for early stage innovator loss and incentivize follow-on innovator. Broader patent protection shifts these benefits to early stage innovators, narrow protection favors follow-on innovators.

If early-stage innovators are favored three issues are seen. First, the whole social value is collected by one research firm leading to monopoly pricing. Second, over-investing due to patent races can occur. Third, full potential of early-stage innovation might not be exploited in follow-on innovations, because only a limited pool of researchers and, thus, a limited pool of knowledge is involved in the innovation process (Scotchmer, 1991; Green et al., 1995; Bessen, 2004). Van Overwalle (2010) elaborates on these questions and discusses how exceptions,

policies and patent pools can extend the range of instruments.

The effect of patent protection on follow-on research was quantitatively analyzed in multiple studies. Williams (2013) investigates the effect of initial contract-based property rights on the re-usage of sequenced human genome issued by Celera. She measures a 20-30% decrease of follow-on research and validates a lasting delay, if sequences are initially protected. Similarly, Murray et al. (2016) measure 20-40% reduction of follow-on research analyzing initially patent protected genetic engineered mice. In their conclusion, Murray et al. (2016) consider the possibility that researchers use alternative open-science research tools to avoid patent protection without further analyzing this hypothesis.

In the same publication, Murray et al. (2016) investigate the effect of open-science on the type of follow-on research. They validate an increase in creativity and diversity of follow-on research after giving open access to research tools. A higher diversity of researchers and more exploratory topics within follow-on innovations are observed. At the same time public ownership did not reduce incentives to run early stage research.

Next to follow-on research, cumulative innovation can occur in the form of multipurpose innovations. The Bayh-Dole Act in 1980 made this form popular by incentivizing researchers to patent their innovation as by-products to their scientific publications (Henderson et al., 1995). Various scholars share concerns about this trend (Heller et al., 1998; David, 2001; Campbell et al., 2002; Straus et al., 2002; Walsh et al., 2003; Walsh et al., 2005). Murray et al. (2007) validated this concern in a quantitative study. They analyzed the effect of patents complementing research output and found a 10-20% decrease of forward citations on papers

with a linked granted patent.

Chapter 3

The Emerging Field of Synthetic Biology

This thesis analyzes the impact of institutional entrepreneurs and institutions on the diffusion of scientific innovation. Three studies are performed in context of the emerging field of synthetic biology. The following chapter provides foundations on the field, current landscape of organizations establishing open science, a review of the debate on property rights in the field and a summary on important genetic mechanisms.

3.1 Structure, Challenges and Potential of the Field

The interdisciplinary field of synthetic biology has emerged since the early 2000s. Despite a high variety of definitions (SCHER et al., 2014), a common aim of scientists in the field of synthetic biology is to turn biology into an engineering discipline (Brent, 2004) and to "be able to build just about anything from biology" (Keasling, 2013). SCHER et al. (2014, p. 33) proposes the definition of synthetic biology as follows: "application of science, technology and engineering to facilitate and accelerate the design, manufacture and/or modification of

genetic materials in living organisms." An overview of major research areas, challenges in large scale collaboration projects and estimated economic potential will be given in the following.

Four research areas have been developed emerging from multiple disciplines such as biology, chemistry, physics as well as from software engineering and by using mostly renowned tools though with a distinct aim (Raimbault et al., 2016). Figure 3.1 gives an overview of the four research areas.

First, BioBrick engineering, mostly driven by a software engineering vision,

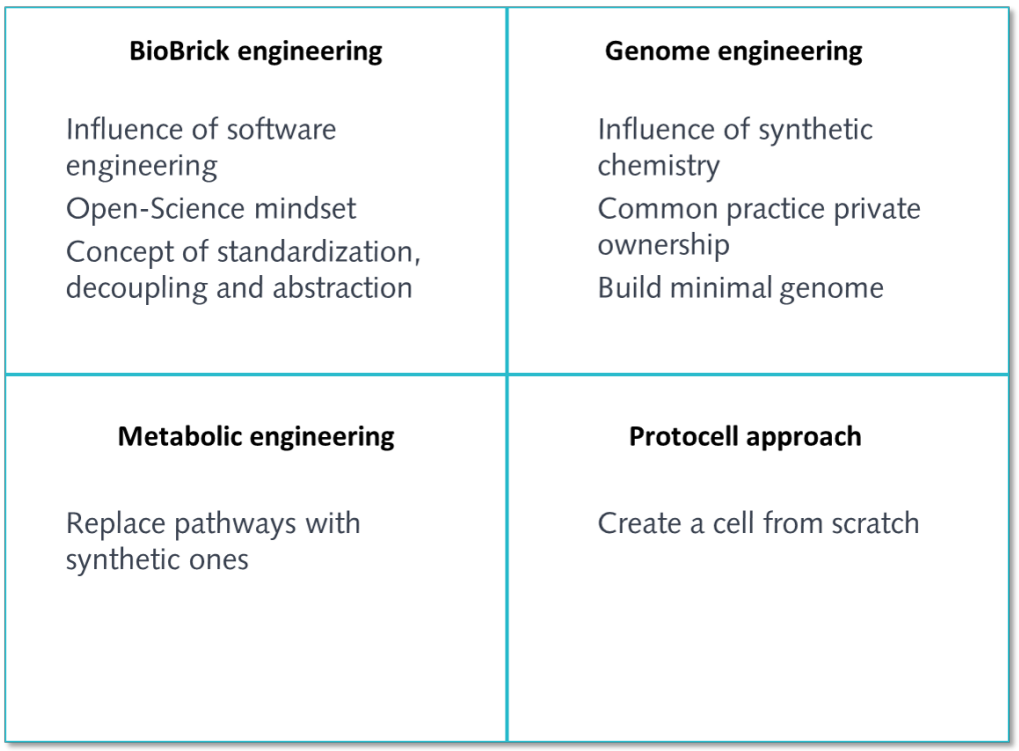


Figure 3.1: Four research areas in Synthetic Biology

tries to introduce standardization, decoupling and abstraction to increase efficiency and to enable a broad application development. BioBricks, the core technology of this area, are standardized biological parts used to build new genetic systems (Endy, 2005).

Standardization is mainly pursued in terms of assembly and measurement standards. To ensure continuous improvements, they are regularly reviewed, e.g., assembly standards are recently changed to be compatible with modern genetic engineering tools (iGEM Foundation, 2003). Second, genome engineering, taking on a synthetic chemistry vision, tries both, to assemble long DNA chains and to build a minimal genome (e.g., Gibson et al., 2010; Hutchison et al., 2016). Third, metabolic engineering attempts to replace pathways with synthetic ones and, thus, to create highly efficient alternatives to established production processes (e.g., Galanie et al., 2015; D'Espaux et al., 2017). Fourth, the protocell approach seeks to create a cell from scratch in order to have an operating system for further application development (e.g., Luisi et al., 2006; Dogterom, 2017). The first two research areas, BioBrick and genome engineering, are the largest to date (Raimbault et al., 2016) and their different visions and origins lead to a two-culture-situation in synthetic biology. On the one hand, there is the community of BioBrick engineering who follows a vision of collaborative design and engineering and, thus, fosters sharing of information and materials and public ownership. On the other hand, there are molecular biologists who brought in the practice of patenting genes and of targeting private ownership (Calvert, 2012).

Both communities are yet joined by collective gatherings such as the SBX.0 conference series, organized by the BioBricks Foundation (BioBricks Foundation, 2017b), in an effort to establish a dense network and to encourage the exchange of information and to initiate collaboration projects (Smolke, 2009). The most recent SB7.0 conference took place in Singapore in 2017 and participants from

various research fields presented their ongoing research and debated about important topics, e.g., bio-safety, bio-security, ethical concerns and regulations (BioBricks Foundation, 2017b). Being confronted with these crucial topics from government and organizations outside the field, the community is encouraged to discuss, to form a dense network and to support their common ideas.

Apart from inter-community initiatives to collaborate, researchers push intra-community collaboration projects with great interest. A huge step towards innovation has been accomplished through large-scale collaboration projects, e.g., the Human Genome Project and Yeast 2.0 (Kelavkar, 2001; Richardson et al., 2017). There are, however, still many isolated and parallel research initiatives and collaboration is complicated due to complexity in cross-lab processes (Rai, 2005). Recently announced projects, e.g., BaSyC and Genome-Project Write projects, will have to incentivize distant researcher teams to collaborate and cooperate by adding value and a common target to share resources (Dogterom, 2017; Boeke et al., 2016). Two initiatives of the BioBricks Foundation, namely OpenMTA and bionet, try to find solutions to legal, organizational and technical issues by sharing information and material (Kahl, 2017; Liddicoat et al., 2016). A detailed review of initiatives fostering open science will be given in 3.2.

Despite crucial unanswered questions to address and the significant need for basic research to perform, several applications are ready for commercialization. Applications in synthetic biology are developed for numerous markets such as pharmaceutical and diagnostic, agricultural, chemical and biofuel (Singh, 2015). The capitalization of innovative applications came to its first peak in 2013, when Jay Keasling's synthetic arteminsimin was produced on a large-scale basis by

Sanofi (Peplow, 2013). The economic potential of the emerging field of synthetic biology was recently analyzed by Singh (2015). They measured a total market value of 3 billion USD in 2013 and projected 44.2% CAGR reaching a total market volume of 38.7 billion USD in 2020. This speed of growth is driven by declining costs of DNA sequence and synthesis, a growing community and large investments by governments and private organizations, e.g., the U.S. government funded more than 820 million USD in synthetic biology research programs in the period from 2008-2014 (Wilson Center, 2015).

3.2 Organizations and Initiatives to Establish Open Science

Scholars fear potential patent thickets in synthetic biology and, in particular, in BioBrick engineering, if basic parts are patented (Oye et al., 2009) and, thus, want the ”stuff of life” (Endy, 2010) to remain open to the public. Several organizations endeavor to establish a community sharing research results, open science parts and research tools to accelerate follow-on research and application development.

For structuring the list of initiatives, a distinction between legal entities promoting open science, frameworks enabling open source and collaborative projects is shortly described. This distinction was built on the work of Johnson (2009). Framework is a basic conceptual, social and technical structure that guides and directs resources into creating value (Talukdar, 2014). Being common for an open source framework are incentives for people to contribute. In addition, the

community has open access to data without discrimination and receives attribution for its work. It can re-use products, but needs to give advancements back to the public (Ardal et al., 2011). The overview in 3.1 lists 9 legal entities, 16 frameworks and 4 collaborative projects which are meant to establish open science. This list is based on discussions with the experts Linda Kahl (BioBricks Foundation) and Jenny Molloy (OpenPlant) and does not claim completeness.

Table 3.1: List of Frameworks, Legal Entities and Collaborative Projects

Name	Founding Year	Type
Addgene	2004	Legal Entity
BioBricks Foundation	2006	Legal Entity
Cambia	1992	Legal Entity
Genspace	2009	Legal Entity
iGEM Foundation	2003	Legal Entity
La Paillasse	2011	Legal Entity
OpenTrons	2011	Legal Entity
PIPRA	2003	Legal Entity
Sage Bionetworks	2009	Legal Entity
Addgene MTA and IMTA	2004	Framework
Addgene repository	2004	Framework
BioBrick Standard	1999	Framework
BiOS Licenses	2005	Framework
BPA1.0	2009	Framework
iGEM Competition	2003	Framework
iGEM Repository	2003	Framework
OpenMTA	2017	Framework
OpenTrons - Protocol Library	2011	Framework
OpenWetWare	2008	Framework
Public Domain Chronicle	2017	Framework
RFC-Process	2011	Framework
SBOL	2011	Framework
SBOL Visual	2013	Framework
Synapse	2009	Framework
Synthetic Yeast 2.0	2011	Framework
OpenPlant	2014	Collaborative Project
Pink Army Cooperative	2009	Collaborative Project
SynBERC	2009	Collaborative Project
Zinc Finger Consortium	2005	Collaborative Project

Highly influencing legal entities and collaborative projects are shortly reviewed and their initiatives are explained in connection with open science, i.e., Addgene, BioBricks Foundation, Cambia, Genspace, iGEM Foundation, La Paillasse, Open-

Trons, PIPRA, Sage Bionetworks, OpenPlant, Pink Army Cooperative, Zinc Finger Consortium and SynBERC.

Addgene

The nonprofit plasmid repository Addgene was established in 2004 to accelerate research and development in biotechnology. Facing cost reduction pressure in academic research, repositories like Addgene encourage sharing and the collaboration of scientific samples to avoid redundant efforts at multiple sites (Chandras et al., 2009; Parsons et al., 2013; Baker, 2014; Herscovitch et al., 2012; Seiler et al., 2013; Crouzier, 2014).

Four major features are developed by Addgene, in particular, to reach this target; a plasmid archive, standardized quality control, associated plasmid information and educational materials. Plasmids are double-stranded DNA which can be introduced into, e.g., bacterial cells, to express genes independently of the genome DNA (Lederberg, 1952). With modern genetic engineering tools, plasmids are easily generated, introduced, stably stored, infinitely reproduced and, thus, are useful to a broad range of studies. These benefits make plasmids best for archives and for sharing of scientific results.

Addgene holds more than 50000 plasmids (Addgene, 2017) which have been identified in three ways. First, the Addgene team screens publications for relevant plasmids. Second, the team contacts laboratories to identify potential reagents. Third, a growing community pro-actively uses the Addgene repository and sends in reagents. Key regions of plasmids are sequenced before storage to

resolve discrepancies. Validated data is stored and associated information is linked, which is both compiled during quality control and provided by depositing scientists.

A plasmids data page, thus, contains essential details for re-usage such as the type of backbone, conditions of growth, gene inserts, cloning information, useful protocols and sequence maps. Collecting technical questions from the community, Addgene scientists continuously publish educational materials, e.g., on genetic engineering tools and collections. In order to ease browsing the large database, collections such as BioBrick Public Agreement Plasmids or Zinc Finger Consortium reagents are built with special pages. Sending out more than 115 000 plasmids in 2014, Addgene has a high impact on forming open science in synthetic biology (Herscovitch et al., 2012; Kamens, 2015).

BioBricks Foundation

The BioBricks Foundation (BBF) was founded as a public-benefit organization in 2006. It was an attempt to ensure that all information needed to build BioBricks is freely available, that the new technology is used in an ethical manner and that collaborative research is fostered. Initiatives run by the BBF include organizing conferences and online platforms to build the community, developing enabling standards, elaborating legal forms and contracts for sharing of parts and developing software to ease up material exchange. The BioBrick Public Agreement (BPA), OpenWetWare, BBF Request for Comments (RFC), bionet, Public Domain Chronicle (PDC) and OpenMTA are shortly described as examples of initiatives

and products driven by the BBF.

The BBF developed the BioBrick Public Agreement (BPA), a framework inspired by the Open Source Software, to support the open sharing of biological parts. With this agreement the contributor promises not to assert any existing or future property rights against users who promise to give attributions if requested and to use the parts responsibly. However, a downstream use of the parts is free for patenting (Smolke, 2009).

Right from the beginning in 2007, the program OpenWetWare has been a web-based wiki-style resource for sharing know-how and data such as scientific protocols. OpenWetWare has more than 28,000 webpages and nearly 16,000 users (Liddicoat et al., 2016). A highly efficient procedure to propose a standard or to describe the best practices is the BBF Request-For-Comments process (RFC). By using an RFC, people can interact with the community and develop or co-develop their ideas in teams. The soft- and hardware bundle called bionet was launched in 2016 and starts with a self-cost inventory management for labs. The developers aim to create an online platform to gain cross-lab transparency on stored material and to automate the physical exchange of reagents (Liddicoat et al., 2016).

In October 2017, the BBF announced the Public Domain Chronicle (PDC). The PDC builds on the BPA and helps researchers to secure scientific findings and methods for the public domain. By uploading information on the finding and by completing two legal forms, researchers claim discoveries for the commons and open the contribution up for redistribution, processing, and search. This way, contributions become public prior art, immediately (BioBricks Foundation,

2017a). The OpenMTA counters the highlighted complex process of material transfer agreements. It reduces this complexity similarly to a framework contract between organizations allowing material exchange and it needs only short repetitive fixations for actual exchanges (Kahl, 2017).

Cambia

Cambia, founded in 1992, is a non-profit research institute located in Canberra, Australia. Founder Richard Jefferson defined its mission as follows: establishing open science in early stage technologies in life sciences and, thus, both develop new technologies and pioneer new business models. By doing so, he counters potential limitations through a broad scope of patents and restrictive licensing which hinders access to enabling technologies (Overwalle, 2009).

Due to the support given by the Rockefeller foundation and IBM, Cambia established the Biological Innovation for Open Society (BiOS) Initiative in 2005. For providing enabling technologies for applications in agriculture in particular, the BiOS covers three levels of engagement (Jefferson, 2006). First, TheLens (former known as PatentLens) (Cambia, 2006) establishes transparency in the landscape of intellectual property. Second, BioForge develops enabling technologies and third, BiOS licenses gives legal framework for sharing technological advancements (Jefferson, 2006). TheLens and BiOS licenses will be further described in the following.

TheLens is a free online resource covering patent information from major patent offices in all fields of technology. It is an alternative to the commercial platforms

such as Derwent, Delphion and MicroPatent. In addition to full text and information on patent families, all associated electronically available sequences are stored in the database. Tutorials on intellectual property rights and information on patent policies and practices reduce researchers' efforts to run research on potential infringements and on the future re-usage of their innovations. By allowing analyses carried out by distributed users, TheLens helps to reduce the costs of initial legal research and, thus, adds up to professional legal advice (Dennis, 2004).

BiOS licenses are an example of patent based commons. In order to ensure open science for upstream technologies but also property rights on downstream applications, Cambia applied for patents on their enabling technologies and created a special license for re-use. Licensees have to agree to give back all advances they have developed on these technologies as well as all associated information (Rai et al., 2007). This copyleft style "grant back mechanism" (Pénin et al., 2008) has been adopted from open source licenses established in software industry. Due to two weak points, the BiOS licenses were not adopted by major collaborations such as PIPRA. First, licensees using a combination of BiOS technologies and other technologies had to report back to BiOS concerning all advances made to all technologies no matter where the origin was. Second, BiOS licenses did not allow any changes due to contracts with third parties. This is difficult to align with the current practice of publicly funded research (Pénin et al., 2008).

Genspace

Genspace is a non-profit organization which started in Brooklyn, New York, in 2009. The Co-Founders' (Ellen Jorgensen, Nurit Bar-Shai, Oliver Medvedik and Dan Grushkin) aim is to involve society into advances of biotechnology. In addition, they give academics the ability to "do things [they are] interested in immediately" and not to do "what your advisor [in university] tells you to do" (Mosher, 2010). In 2010 they opened the first 'government-compliant' community biotech laboratory and, thus, a new chapter in the do-it-yourself (DIY) biology movement which operated mostly in garages and self-built laboratories. Members get access to a fully equipped synthetic biology facility and to professional scientific mentorship at subsidized prices. Inside the lab, they provide adult education, science courses at high school and college level and a dialog form to allow debates about the social and ethical dimensions of biotech (Genspace, 2009). Genspace has successfully raised up several biotech startups including Opentrons. Multiple laboratories all over the world followed the example of Genspace, e.g., La Paillasse in Paris, BioCurious in the Bay Area, London Biohackerspace or Labitat in Copenhagen.

iGEM

The student competition, organized by the iGEM Foundation, is seen as the most important community-building activity in the BioBrick engineering field proclaiming its 'standardize, decouple and abstract' approach (Calvert, 2012). iGEM is a way to introduce this mindset to a large group of followers. As exper-

imenting is considered an effective way to learn, student teams are directed to use, re-use and develop BioBricks and to create innovative genetic solutions for real-world problems (Smolke, 2009).

The iGEM competition was started in 2003 as an internal competition at MIT by Drew Endy, Tom Knight, Randy Rettberg, Pamela Silver and Gerry Sussmann with the aim to shape "the ideology, values and culture" of synthetic biology (Smolke, 2009, p. 1101). Since then iGEM has grown exponentially, establishing over 300 international multidisciplinary teams with 4432 participants from all over the world in 2016. The iGEM jamboree event is held in fall when all teams come together in Boston to present their projects and on that occasion, the winners are awarded. It has become the largest event in the field of synthetic biology (iGEM Foundation, 2017b).

The competition is successful in three dimensions: community-building, implementing a value system and developing innovative prototypes.

During their summer-long project, students are working in teams with great enthusiasm and get encouraged to continue working in the field of synthetic biology (Mitchell et al., 2011). The fast feedback-loop and opportunities to propose and realize improvements turn it into an outstanding environment in academia which inspires students and shows the attractiveness to work in this emerging field (Smolke, 2009). Apart from developing prototypes, teams have to raise funds for the needed resources by themselves, e.g., request university support, grants or sponsorships and develop project management skills. For making iGEM a lasting experience and to activate an ongoing exchange in the "iGEMer"-community, iGEM alumni has recently been established to intensify

regular exchange after iGEM participation (iGEM Foundation, 2017a).

At the same time, students learn a lot about the value system and how the BioBrick engineering approach can work, i.e., how sharing and cooperation can improve development cycles. The "participant-based 'get' and 'give' approach" tries to foster an open science mindset. On this occasion, participants learn numerous success factors of collaboration projects, e.g., the quality of parts and documentation (Smolke, 2009, p. 1101). The iGEM committee is successful in introducing preferred characteristics and aims through an innovative award system. Apart from awards for an excellent idea and prototype, teams can gain awards for, e.g., entrepreneurship, cooperation, the improvement of existing parts and the development in human practices (iGEM Foundation, 2017c). In fact, setting collaboration with other teams a mandatory requirement to receive a silver medal in 2015 heavily increased the collaboration between teams (Santolini, 2017). The decisive criteria are published up front, thus, teams are aware of them and can act accordingly.

Most prominent in the outside communication are the innovative prototypes developed by the teams. Projects developing, e.g., synthetic horse crab blood, field diagnostic devices and software for genetic design underline the excellence of recent iGEM teams (iGEM Foundation, 2017d). A reason for designing this competition concerning prototyping lies in the vision of BioBrick engineering. Scientists in this research area aspire to run large scale projects with highly efficient exchange of information and data to find solutions to a vast amount of real-world problems and to develop tools that support engineering biology. iGEM is an effective way to crowd source capacities and competences and to

enable multidisciplinary students to develop innovative prototypes which, otherwise, would need a major R&D department (Smolke, 2009).
In 2016, the BioBricks Foundation added another level in their cooperation with iGEM and sent their own team called BBF Allstars over there. The team followed the mission of the BioBricks Foundation to work on the meta level and to develop advances on how iGEM as a platform should be run to solve problems. In 2016, the team developed genetic constructs for multiple genes with quantitatively defined and predictable gene expression (BioBricks Foundation, 2016).

La Paillasse

A second organization within the DIY community is "La Paillasse", the largest biohackerspace in Europe located in Paris, France. The laboratory was founded in 2011 by Marie-Sarah Adeniss, Adrien Clavairoly, Marc Fournier and Thomas Landrain to establish an environment similar to Genspace in New York. In order to give society an introduction into biotechnology and to form a space for creativity they provide workshops, events, a space for debate and interaction along with technical, scientific equipment (La Paillasse, 2011).

OpenTrons

An example of open hardware movement, proclaiming open science in BioBrick engineering is OpenTrons, a manufacturer of automated pipetting founded in 2011 and established in Brooklyn, New York (Gewin, 2013). The startup

stems from Genspace and got support by Haxclr8tr in China, Kickstarter and Y-Combinator (Y-Combinator, 2017; Buhr, 2016). Using open source hardware and software, the team around co-founder Will Canine builds affordable robots to reduce waste in standardized wet lab protocols. In 2016, more than 50 robots were used by labs worldwide (Buhr, 2016).

Building on their hardware environment, OpenTrons established an open accessible protocol library to further engage in sharing best practices and scientific advances. The protocol library covers, e.g., basic pipetting procedures such as plate consolidation and mapping, as well as PCR, purification and normalization. The community is asked to interact and to enlarge and optimize the library (OpenTrons, 2011; Buhr, 2016).

PIPRA

The Public Intellectual Property Resource for Agriculture (PIPRA) is a collaboration among multiple agricultural universities to ensure access to enabling technologies in agricultural biotechnology (Benkler, 2004). In this field, 24% of intellectual property on upstream technologies was held by the public sector and a high level of fragmentation occurred, thus, general access to enabling technologies was highly restrictive. Multiple universities, e.g., the University of California and North Carolina State University, therefore created PIPRA to manage collectively public sector intellectual property to support both US and developing country agriculture. Three short-term objectives were set. First, the review of current public sector patenting and of licensing practices to identify

potential for efficiency gains. Second, the development of a database of public sector intellectual property assets, corresponding to TheLens. Third, the development of consolidated technology packages (Delmer et al., 2003).

Sage Bionetworks

Sage Bionetworks was co-launched by Stephen Friend in 2009 as a non-profit organization to increase openness and, thus, the efficiency of biomedical research. Based at Fred Hutchinson Cancer Research Center in Seattle, Washington, Sage persuades both scientists and patients to share biomedical data and associated models (Altshuler et al., 2010). Friend recognizes established initiatives to introduce open science and the sharing of data but values the open sharing of applied models more important to increase efficiency of drug development. In order to move network analysis from animals to humans, an enormous database and computational models are needed.

For this shift, drug development has to be taken to large scale collaboration projects, both involving scientists and patients. Friend tries to run this biomedical revolution in five steps. First, a platform called "Synapse" (Sage, 2009) is established to enable data and disease model sharing. Collaborators of Sage have to agree to share all data and models one year after a project was accomplished. Second, pilot studies on collaborative development of disease models are performed. Third, to ensure that data can be shared, legal consent is gathered so that patients can control who uses their data. Fourth, patient data recorded in clinical trials are requested from companies to share. Finally, academics and

companies are supposed to collaborate in the initial testing of potential drugs until it shows safety and efficacy in the second phase of the clinical trial (Kaiser, 2012).

Synthetic Yeast 2.0

Synthetic Yeast 2.0 (Sc2.0) is an international genome engineering initiative whose aim is to build the world's first synthetic eukaryotic genome (Juhas et al., 2013; Perkel, 2012; Nawy, 2011; Dymond et al., 2011; Enyeart et al., 2011). The Sc2.0 project will lead to the construction of the first eukaryotic and, simultaneously, to the largest synthetic genome with 14 Mb organized in 16 chromosomes (Juhas, 2015). This large scale project is collaboratively performed by 9 labs worldwide and work is divided by chromosome (Synthetic Yeast 2.0, 2017).

OpenPlant

OpenPlant is a joint initiative of the University of Cambridge, John Innes Centre and the Earlham Institute. It started in September 2014 and is funded by the Biotechnology and Biological Sciences Research Council and by the Engineering and Physical Sciences Research Council as part of the UK Synthetic Biology for Growth programme. A close cooperation with the BBF is reported (Liddicoat et al., 2016). The initiative develops standards for the DNA assembly in plants and for open science upstream technologies and aims to establish a system of frameworks for the open exchange of upstream technologies and DNA parts (OpenPlant, 2016).

Zinc Finger Consortium

In order to counter the monopolist in zinc finger technologies called Sangamo, a group of academics formed the Zinc Finger Consortium (ZFC) to enable all academic researchers to use zinc finger technology and to accelerate research (Fu et al., 2013; Joung et al., 2015). They published protocols, customized proteins, a database consolidating multiple collections of Zinc Finger Proteins (ZFP) and Zinc Finger Arrays (ZFA) and software tools for the design and to optimize binding efficiency (Fu et al., 2013). In 2008, the Zinc Finger Consortium published a collection compromising ZFAs, reagents for modular assembly and for the OPEN methodology, openly accessible on the non-profit repository Addgene (Zinc Finger Consortium, 2016; Herscovitch et al., 2012). In addition, the consortium created the ZifDB database with several collections including proprietary ZFAs of Sangamo (Fu et al., 2009). Openly available software tools such as Zifit, first published in 2007 (Sander et al., 2007), support the identification of potential ZFPs and the design and the engineering process. Joung, one co-founder of the consortium, rates their efforts as "game changer. It gives academics the ability to make these proteins without going to Sangamo" (Kaiser, 2008).

Pink Army Cooperative

The Pink Army Cooperative (PAC) is an example of a community-based approach to develop biological solutions. The crowd-funded venture was founded by Andrew Hessel and two more co-founders. Doubting the success of the current system to develop drugs, the PAC tries to use a community-based approach

to develop individualized therapies for cancer, opening the process to all contributors (Nelson, 2014; Altshuler et al., 2010). Using freely available software and biological parts, the community is asked to develop specific cancer treatments on an individual basis and, thus, to counter the standard approach of drug development fostering a scalable solution (Nelson, 2014).

SynBERC

The Synthetic Biology Research Engineering Center (SynBERC) was a multi-international research project with a ten-year grant from the National Science Foundation in the years 2006-2016. In its mission statement, three targets are pointed out, namely a) developing BioBricks and assembling into integrated systems for special applications, b) educating and training new experts in synthetic biology and c) increasing the popularity of synthetic biology by educating the public about opportunities and challenges (Torrance et al., 2013).

Being the most popular professional developer and provider of BioBricks and sharing these parts in an open registry, SynBERC helps to increase the quality of available BioBricks. In 2010, the BIOFAB project, run by SynBERC, the BBF and by the Lawrence Berkeley National Laboratory, published 478 openly accessible BioBricks on Addgene. In the past 9 years SynBERC succeeded in building extensive industry partnerships and formed a cluster of nearly 50 corporates, start-ups, non-profits and industry associations which meet semi-annually for knowledge sharing (SynBERC, 2006).

3.3 Property Rights in Synthetic Biology

Property rights in synthetic biology are subject to a long debate (,e.g., Kumar et al., 2007; Rai et al., 2007; Maurer, 2009; Oye et al., 2009; Calvert, 2012; Torrance, 2011; Torrance et al., 2013; Nelson, 2014). The background will be explained by sketching the history and the current situation of property rights in synthetic biology, major complications and current initiatives to find a solution.

3.3.1 History and Current Situation

To give a background to the current situation of property rights in synthetic biology, three historic phases of gene patenting will be highlighted.

In the first phase until the early 80s, patents were only filed for research tools, not for genes or organisms. A prominent patent was filed by Cohen and Boyer in 1974, who transferred DNA fragments from one organism to another one (Hughes, 2001; Torrance, 2011).

In 1982, a first patent on genetically modified organisms started the second phase of gene patenting, where isolated and purified DNA sequences became subject to patents (Baxter et al., 1982; Calvert et al., 2011). Even though, with the discovery of the DNA in 1953, a change of cell understanding from the mendelian cell as a chemical compound to molecular genes being also a carrier of information, took place (Keller, 2000; Dupré, 2004; Moss, 2004), patent law still viewed DNA sequences simply as chemical compounds. This decision lead to a patent race on a significant scale. On the one hand, the pure quantity of issued patents rose to thousands per year in the early 2000s (Torrance, 2011),

on the other hand, the quality of patent applications began to decrease. An example of this particular quality were expressed sequence tags (EST) in the patent applications of J. Craig Venter in 1991 and in 1992. Accordingly, the applicants knew about sequences but still lacked the understanding of their full function (Cornish et al., 2003). The second phase ended in 2010 with the Myriad case, when patents of the Myriad company on isolated human genes were invalidated (Calvert et al., 2011).

Meanwhile, synthetic genes and constructs are subject to patents. It is common practice to use so-called "percent identity language", i.e., creating multiple claims in patent applications with an increasing coverage of claimed sequence, e.g., 80%, 85%, 90% (O'Brien et al., 2013). It is then up to the patent offices to review patent applications and to select an appropriate coverage limit to grant the patent (O'Brien et al., 2013). In all research areas in synthetic biology, patent applications are common. Large companies issue patents in line with their business models and small companies have to issue patents to attract venture capital companies to invest (Calvert, 2012). However, since the Myriad case was decided by the US Supreme Court, patent claims to biological sequences are subject to heightened scrutiny (Jefferson et al., 2015). Equivalently, the European Patent Office decides according to the decisions taken by the European Court of Justice in 2010. Here, the patent claiming of DNA sequences had been declared limited by requirement of functionality as disclosed in the patent text (Jefferson et al., 2015).

3.3.2 Debate on Property Rights in Synthetic Biology

The question on which type of property rights is most fruitful for which type of innovation has fascinated multiple scholars in research (Kumar et al., 2007; Rai et al., 2007; Maurer, 2009; Oye et al., 2009; Calvert, 2012; Torrance, 2011; Torrance et al., 2013; Nelson, 2014). For cumulative innovation, prior research identified limitations through patent protection in the close-related field of biotechnology (Furman et al., 2011; Murray et al., 2007; Williams, 2013; Murray et al., 2016). Collectively, experts in synthetic biology see benefits from open science in upstream technologies (Kumar et al., 2007; Endy, 2009).

In order to analyze common habits of researchers in academics and industry regarding patent protection Kahl (2015) prepared a survey at the conference SB6.0. In the survey, industry claims nontransparent property right status and highlights experienced delays and blocking in research. In contrast to that, most academics do not review probable property rights before using research tools and material for their research. In the academic world, bigger hazard is seen in legal and material issues regarding material exchange, e.g., material transfer agreements which are concluded among various organizations in order to share biological material are viewed to be too complex and, thus, delaying research.

Multiple scholars acknowledge the current structural mismatch both of patent law and of the technical, ethical and social development of synthetic biology (Calvert, 2012; Rai et al., 2007; Torrance, 2010). Torrance et al. (2013) review potential alternative intellectual property regimes to be applied to synthetic biology, such as copyright, trade secrecy and trademarks, but there is no current shift

away from patent protection observable. In both major research fields, genome and BioBrick engineering, patents are applied to methods and DNA sequences. In genome engineering, the most prominent actor J Craig Venter has a long history of patent applications at his disposal, e.g., patents on human genome segments, applications on ESTs (which were denied patent protection), a synthetic chassis called "Minimal bacterial genome" or the Gibson method (Calvert, 2012).

Even though scholars fear potential patent thickets in BioBrick engineering, if patents are issued on basic parts (Oye et al., 2009; Wellhausen et al., 2009), companies apply for patents to secure ROI (Calvert, 2012). Some scholars are debating whether there can be an ecology of both cultures or one will dominate. While Endy (2009) argues for "a rich, fully diverse ecology of commercial and public benefit use from the outset", Haselhoff (2010) sees the field of synthetic biology just before a tipping point, whether it will become an open science or private property industry.

3.4 Foundational Mechanisms in Genetics

In 1970, Francis Crick postulated that the "central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred back from protein to either protein or nucleic acid" (Crick, 1970, p. 562). The central dogma explains how the DNA, i.e., nucleotide sequences, is transcribed and translated into proteins. Basically, transcription is the process of replicating the information contained

in a DNA sequence in the form of RNA. The transcription rate depends on multiple factors related to the cell and some of them are defined by sequences surrounding the DNA sequence. In essence, transcription starts in the promoter region, a DNA sequence being designed to bind polymerase which replicates the information in the RNA. The polymerase is forced to stop transcription and to fall off the DNA sequence by terminator regions. The intensity of promoter and terminator sequences and, thus, the reliability and frequency of transcription processes can be influenced by the promoter and terminator being used. E.g., specific promoters, such as the J23119 promoter, can be used to ensure a high transcription activity and double terminators have a high reliability in stopping the transcription (Alberts et al., 2013; Setubal et al., 1997).

RNA is translated by ribosomes into proteins. This process is structured in the codon system, i.e., three bases of the RNA are translated into one specific amino acid (Nirenberg et al., 1965). Start and stop codons control initiation and end of the translation into amino acids (Alberts et al., 2013; Setubal et al., 1997). Long sequences of RNA are therefore not translated into proteins but used to direct and guide this process. Due to redundancy of genetic code, the so-called codon degeneracy, multiple codons code for same amino acids (Lagerkvist, 1978). This can lead to diverse DNA sequences coding for the same protein and making patenting of DNA sequences more complex. The folding process of proteins is not further described in this introduction.

In BioBrick engineering, genes from different organisms are identified and assembled with standard parts such as promoters and terminators to design new combinations of functional proteins (Shetty et al., 2008). The assembly of parts

can be performed by using multiple methods. Casini et al. (2015) give a comprehensive overview on DNA assembly standards. Scientists have to select appropriate assembling methods considering their genetic design and the prerequisites of the BioBricks, they made use of.

Recent research tries to increase the efficiency of biological engineering, e.g., a higher consistency of translation can be reached by engineering tactics such as overlapping translation regions to prevent ribosomes to fall off. This strategy, found in nature, was adapted by Mutalik et al. (2013) who developed a bi-cistronic design for reliable gene expression levels.

Chapter 4

Study I: How Institutional Entrepreneurs Drive the Diffusion of Institutional Logics

4.1 Introduction

Understanding the dynamics of emerging organizational fields "is an important next step in the development of institutional theory" (Lawrence et al., 2004, p. 690). The theory proposes that organizational fields are characterized by a set of institutions that shape the behavior of individuals and organizations (DiMaggio et al., 1983). However, the theory is limited in explaining the emergence of new institutions (Leblebici et al., 1991), the effects of agency and the change of organizational fields (DiMaggio, 1988; Scott, 2001).

These limitations are faced by the concept of institutional logics (Friedland et al., 1991; Thornton, 2002) and by the institutional entrepreneur approach (DiMaggio, 1988; Garud et al., 2002; Campbell, 2004; Lawrence et al., 2004; Maguire et al., 2004). Institutional logics are linked to higher institutional orders and provide legitimacy for creating and modifying institutions (Thornton, 2002). Institutional entrepreneurs are agents of change, motivated by self-interest. They combine

and diffuse institutional logics from different institutional orders to create and shape emerging fields.

Previous research discussed the role of institutional entrepreneurs in the emergence of new organizational fields (e.g. Lawrence et al., 2004; Maguire et al., 2004), developed a model of multiwave diffusion of institutions linked by an institutional logic (Shipilov et al., 2010) or investigated the influence of structural holes on their activities (Ritvala et al., 2009). Empirical studies analyzed the role of institutional entrepreneurs in diffusing product innovation and institutions in industries (e.g. Baron et al., 1986; Guler et al., 2002). However, there are rare attempts to validate the theories on institutional entrepreneurs. In particular, the strategies of institutional entrepreneurs in diffusing institutional logics to shape an emerging scientific field are not investigated yet (Pacheco et al., 2010). For example, the way in which institutional entrepreneurs convince their social context and use their resources to shape an emerging field.

This paper addresses the outlined gap by using the case of the research field of synthetic biology. The field emerges from multiple disciplines such as biology, chemistry, physics as well as from software engineering, and scientists in the field have the common aim to turn biology into an engineering discipline (Brent, 2004). The field has been chosen to investigate the impact of institutional entrepreneurs on the diffusion of an institutional logic for three reasons. First, the interdisciplinary field will have vast significance in the future world and will introduce the 5th revolution while its major players are already considered highly influential in modern society (Peccoud, 2016; Esquire, 2008). Second, the underlying engineering approach is the institutional logic that shapes the emerg-

ing field. As this cultural shift in life sciences introduces engineering principles such as standardization, decoupling, abstraction and product-development circles (Endy, 2005; Peccoud, 2016), a high level of uncertainty for researchers regarding costs and benefits of adopting the logic is present. And third, in a case study certain leadership roles, i.e., institutional entrepreneurs, were identified (Raimbault et al., 2016). However, research has not yet validated how these actors influence their social context.

In order to measure social influence mechanisms, the heterogeneous diffusion model is adapted to the concept of institutional logics to incorporate three factors regarding the infectiousness of prior adopters, the effect of social proximity and the susceptibility towards these communication channels. A new measure of social proximity, called knowledge equivalence, is introduced to acknowledge authors' original research areas before adopting the logic.

The empirical data validates that institutional entrepreneurs use their resources to convince their social context of an institutional logic. This way, they shape the emerging field and grow the social collective supporting their logic. This paper makes two contributions to the investigation of how institutional logics diffuse to shape emerging fields. Firstly, the study highlights the role of institutional entrepreneurs in diffusing institutional logics in connection with scientific innovation. Secondly, to track the process of social influence, the heterogeneous diffusion model is applied to the diffusion of an institutional logic.

The remainder of the paper consists of five sections. Theoretical foundations on institutional fields, constituting logics and institutional entrepreneurs are discussed and the heterogeneous diffusion model to measure social influence

presented. After this, data and methods of the study are provided. Results of the analysis are presented and implications on research are put forward. Finally, limitations and potential paths for further research are discussed.

4.2 Theoretical Background

4.2.1 The Institutional Entrepreneurship Approach

In the research area of institutional theory, scholars try to understand the behavior of individuals and organizations in an organizational field (Scott, 1991; Wooten, 2006). Organizational fields are defined as a "set of organizations that constitute a recognized area of life, are characterized by structured network relations, and share a set of institutions" (Lawrence et al., 2004, p. 321). However, organizational fields often do not consist of one dense community (DiMaggio et al., 1983). Rather there are various communities with different degrees of identification with the field, the so-called core and peripheral communities (Glynn, 2008).

The behavior of actors in an organizational field is guided by the set of institutions, a framework to provide stability and collective meaning for social behavior (Scott, 1995). Prior research concentrates on three types of institutions; practices (Leblebici et al., 1991), standards (Greenwood et al., 2002) and policies (Garud et al., 2002). Institutions and organizational fields are interrelated in their development process. Patterns of social interactions help to produce and reproduce common understandings and practices to form institutions that define the field, at the same time institutions shape these patterns of social interactions (DiMag-

gio et al., 1983). Fligstein (1997) sub-classified organizational fields into emerging, mature and stable, and into fields in crisis. Emerging fields, in particular, are characterized by a lack of institutions and thus by ”fluid relationships, conflicting values and an absence of clearly identifiable norms” (Hardy et al., 2008, p. 205).

To explain the change of institutions, scholars in the neo-institutional theory emphasized the role of isomorphism (Meyer et al., 1977; Zucker, 1977; DiMaggio et al., 1983). There are three main types of isomorphism: normative, coercive and mimetic. As to normative isomorphism, behavior is driven by pressure brought about by professions. Coercive isomporhism considers the influence of related organizations one is dependent on and of cultural expectations from society. Mimetic isomorphism occurs when an organization imitates the behavior or structure of another organization with high legitimacy (DiMaggio et al., 1983). Therefore, the concept diminishes consciously strategic choices (Friedland et al., 1991). Also, the emergence of new institutions is not explained (Leblebici et al., 1991) and there are limitations in considering the role of agency (DiMaggio, 1988; Scott, 2001).

The theory of institutional logics faces these limitations (Friedland et al., 1991). It proposes society as an inter-institutional system, the so-called concept of institutional orders (Friedland et al., 1991; Thornton et al., 2008). In this concept, all institutional orders, i.e., the market, the corporation, the professions, the state, the family, and religions, have a different batch of institutional logics (Friedland et al., 1991). These logics are defined by cultural differentiation, fragmentation, and by contradiction both between and within institutional orders (DiMaggio,

1997; Friedland et al., 1991). Logics, thus, provide a "link between the individual agency and cognition and socially constructed institutional practices and rule structure" (Thornton et al., 2008, p. 101).

In an interplay of individual agency and institutions, the so-called embedded agency (Seo et al., 2002; Battilana, 2006; Greenwood et al., 2006), individuals or organizations seek power or status, however, mean-ends of interest and agency are both facilitated and constrained by institutional logics (Giddens, 1984; Sewell Jr, 1992). Therefore, institutional logics provide guiding principles for creating appropriate institutions to specific problems and rhetorics accounting for why organizations or individuals should adopt new institutions (Kono et al., 1998; Lawrence et al., 2006; Thornton, 2004).

Fundamental for the theory is the equality of material and culture characteristics of institutional orders (Friedland et al., 1991). This builds on both Becker (1974) who claims the involvement of family and religion in the consumption of goods and on Granovetter (1985) who develop models which are shaped by culture. There is no dominance of either characteristic, but the interplay between them results in the development and change of institutions (Thornton et al., 2008). In this definition, all three dimensions of institutions, i.e., coercive, cognitive and normative, are simultaneously considered and integrated (Thornton et al., 2008). The theory of institutional logics, thus, does not evaluate rationality or irrationality of actions, but tries to understand how contradictions and conformity of institutional logics, having both cultural and material characteristics, influence behavior in society (Thornton, 2002).

While Friedland et al. (1991) had a focus on societal level and the influences on

individuals and organizations, modern meta-theory takes place on multiple levels such as organizations, the organizational field or markets. Also, theoretical mechanism, i.e., effects on individuals and organizations, can happen on levels being different from the main anomaly investigated in an analysis (Thornton et al., 2008).

Besides the need for specification on which level an analysis takes place, scholars emphasize the temporal dimension of meaning and importance of an institutional logic (Thornton et al., 1999; Scott, 2000; Lounsbury, 2002; Zajac et al., 2004; Meyer et al., 2006). It is therefore essential in the studies of institutional logics to analyze whether assumed theories are universal through time and space or particular to the environment (Thornton et al., 2008).

Institutional logics provide guiding principles for creating institutions which in turn define the behavior inside an organizational field. Therefore, a change to institutional logics can modify or create organizational fields. This process of change is called institutional entrepreneurship (Hardy et al., 2008). It is initiated and driven by institutional entrepreneurs, agents of change, who are motivated by self-interest (DiMaggio, 1988). They can be individuals (,e.g., Fligstein, 2001b; Kraatz et al., 2002; Lawrence et al., 2004; Maguire et al., 2004; Dew, 2006) or organizations (Déjean et al., 2004; Demil et al., 2005; Garud et al., 2002; Hensmans, 2003; Leblebici et al., 1991). Institutional entrepreneurs are of an ”analytically distinguished social type” (Beckert, 1999) and they are able to reflect about the status quo of their institutional environment, to observe drawbacks and to translate this cognition into actions designing new institutions (Beckert, 1999; George et al., 2006).

After initiating change, institutional entrepreneurs try to acquire legitimacy for their ideas (DiMaggio, 1988). They enact in a variety of strategies to persuade their social context of their innovation and institutionalize it (DiMaggio, 1988; Garud et al., 2003; Hardy et al., 2008), namely mobilizing resources, constructing rationales by using framing and theorization, and establishing relations.

Institutional entrepreneurs have strong positions with wide legitimacy and status to persuade their social context (DiMaggio, 1988; Garud et al., 2003; Maguire et al., 2004; Hardy et al., 2008). The more legitimacy and power in terms of resources, knowledge and limited social network positions institutional entrepreneurs have, the better they can drive change and shape their context (Suchman, 1995; Beckert, 1999; Lawrence, 1999; Maguire et al., 2004; Foucault, 1972; Bourdieu et al., 1992; Hoffman, 1999). Regarding material resources, scholars argue that they are mobilized to both, dominating others and negotiating support for the entrepreneur's institutionalization project. Domination of others is characterized by the ability of an institutional entrepreneur to control rewards and punishments. In negotiations, potential supporters need to perceive tangible and intangible benefits (Colomy, 1998).

Besides mobilizing resources, institutional entrepreneurs construct rationales to persuade their social context. Here, they theorize their ideas, i.e., build chains of cause and effect from abstract categories (Greenwood et al., 2002), and frame changes in a way to generate collective action (Benford et al., 2000; Lounsbury et al., 2003; Garud et al., 2002), e.g., by describing problems with existing practices to legitimate their new ones (Strang et al., 1993; Greenwood et al., 2002). Building a collective action frame follows a defined structure of punctuation,

elaboration and motivation (Snow et al., 1986; Creed et al., 2002; Hardy et al., 2008). In these phases, the problem and its importance is identified, a diagnosis and counter activities elaborated and a call for action issued (Creed et al., 2002; Misangyi et al., 2008). Institutional entrepreneurs use appropriate frames to increase the chance for the institutionalization of their ideas (Gray et al., 2015). They use alternative logics to build a context to their ideas (Seo et al., 2002) and adjust them to, e.g., rules of society (Haveman et al., 1997) and cultural accounts (Creed et al., 2002). Also, they position their ideas using existing categories and systems (Hargadon et al., 2001) and build on available discourses (Hardy et al., 1999; Lawrence et al., 2004). With framing and theorization they discredit the status quo and claim necessity and validity of their ideas to make support reasonable for others (Rao, 1998).

As a third strategy, institutional entrepreneurs establish new relations to organize collective action (Dew, 2006; Aldrich et al., 1994; Garud et al., 2002; Lawrence et al., 2002; Wijen et al., 2007). Scholars observed profound political and social skills of institutional entrepreneurs (Perkmann et al., 2007) and, thus, the "ability to motivate cooperation of other actors by providing them with common meanings and identities" (p.397 Fligstein, 1997).

Prior empirical research performed case studies on the emergence of organizational fields (e.g. Van de Ven et al., 1993; Powell et al., 1996; Lenway et al., 2001; Garud et al., 2002) and analyzed how institutional entrepreneurs diffuse institutions in governments and industries (e.g. Baron et al., 1986; Guler et al., 2002). Ritvala et al. (2009) elaborated a comparing case study on bridging attempts of institutional entrepreneurs between unconnected network structures.

Shipilov et al. (2010) developed a model of multiwave diffusion of institutions linked by an institutional logic. Here, the authors validated that the institutionalization of practices is eased by the former diffusion of an institutional logic. Grodal (2017) investigated how core and peripheral communities shape social and symbolic boundaries. They built a comprehensive model to structure the influences of communities on shaping boundaries and distinguish between the status of the organizational field. However, strategies and impact of institutional entrepreneurs on the diffusion of institutional logics in context of scientific innovations have not yet been investigated in prior empirical research.

This study extends the literature and aims to validate theoretical foundations of institutional entrepreneurship. It analyzes how institutional entrepreneurs use their resources to diffuse an institutional logic of an emerging organizational field in context of scientific innovation.

Three hypotheses can be derived:

Hypothesis 1: Institutional Entrepreneurs use social influence mechanisms to convince their social context of an institutional logic.

Hypothesis 2: Institutional Entrepreneurs have a high reputation and use this resource to convince their social context of an institutional logic.

Hypothesis 3: Institutional Entrepreneurs have a high reputation and are not susceptible towards social influence from their peers.

The hypotheses are tested in context of the emerging field of synthetic biology. Even though multiple definitions exist for the high potential multidisciplinary field, the common aim of making biology an engineering discipline can be seen

as institutional logic for the organizational field. In an exploratory case study, Raimbault et al. (2016) identified multiple key actors who try to establish this logic. The field is therefore suitable for this type of analysis.

4.2.2 The Heterogeneous Diffusion Model

Diffusion models analyze influences of pressure from social context and, thus, the diffusion of behavior and attitudes in a population (Greve, 1998; Rogers, 2003; Fichman, 2004; Hall, 2004). Early diffusion models, e.g., developed by Bailey (1975), are based on population level:

$$P(t) = [a + bs(t)]n(t) \tag{4.1}$$

with a equal to the effect of externals factors, b equal to the effect of spreaders, $s(t)$ as number of spreaders and $n(t)$ as individuals at risk. These models have three limitations (Strang et al., 1993):

1. All actors are equally susceptible to external factors
2. All links between actors and adopters are equally likely and contagious
3. Contagiousness of contacts are constant over time

The heterogeneous diffusion model (Strang et al., 1993) analyzes adoption on individual level and considers spatial and temporal heterogeneity. Spatial heterogeneity enables the analysis of individual adoption behavior, temporal heterogeneity allows influences to vary over time. The formula for the individual hazard rate at time t to adopt innovation is given by:

$$r_n(t) = exp(\alpha x_n) + exp(\beta v_n) \sum_s exp(\gamma w_s + \delta z_{ns}) \tag{4.2}$$

with x equal to the intrinsic rate of adoption of individuals at risk, v equal to the susceptibility of individuals at risk to intrapopulation linkages, w equal to the infectiousness of prior adopters, z equal to the social proximity between individuals at risk and prior adopters as well as coefficients α, β, γ and δ, respectively. The three factors measure how easily information is transmitted (social proximity), how information on adopters' actions effect individuals at risk (infectiousness) and how much individuals at risk are influenced by intrapopulational linkages (susceptibility). Intrinsic motivation is independent of the social context and, thus, is used as control (Strang et al., 1993).

Social Proximity

Social proximity is the social distance between actors. It is used to measure how easily information is transmitted and how relevant it is (Coleman et al., 1966). Socially proximate adopters reduce uncertainty about new behaviors or attitudes more efficiently by transmitting rich and relevant information in high volume (Davis et al., 1997). Thus, they provide a frame for evaluation and interpretation of information in uncertain situation by the individual at risk (Leenders, 2002). There are two approaches to analyze social proximity (Marsden et al., 1993):

1. Social cohesion: Paths between actors
2. Structural equivalence: Similarity of actors' roles

Social cohesion is the number, length and strength of paths between actors (Marsden et al., 1993). Direct ties have opportunities to communicate and share information in high frequency, thus, their social cohesion is the strongest (Burt, 1982).

Shared activities are opportunities to develop common attitudes (Homans, 1961). The longer the path, the less volume and fidelity the information transmits (Shannon, 1949). As a consequence, indirectly connected actors have less influence (Burt, 1982).

In modern times with easy access to information, personalized information by direct ties is highly valued (Rogers et al., 1981). Especially in uncertain situations, e.g., due to paradigm shifts, individuals discuss with their proximates and rely on their innovation experience (Rogers, 2003). Out of these discussions, social pressure on individuals to adopt a new behavior or attitude rises (Rogers et al., 1981). It is therefore more likely to adopt an institutional logic that has been adopted by social peers. This mechanism of social cohesion leading to common attitudes and behaviors among directly connected people has been validated in prior research for, e.g., adoption of corporate governance or merger-acquisition practices by direct partners (Davis et al., 1997; Haunschild, 1994).

In academic and corporate research, shared social activities happen in co-authoring an article and being affiliated to same organization. Here, opportunities for communication and social cohesion arise. In discussions, costs and benefits of adopting the engineering approach are evaluated.

In general, structural equivalence is present, if two actors occupy a similar role (Marsden et al., 1993) and the intensity of competition between two researchers increases with the possibility of substitution (Burt, 1987). In contrast to social cohesion, structural equivalence is based on symbolic rather than direct communication (Burt, 1987). Its significant effects have been validated in several studies (e.g. Bothner, 2003; Davis et al., 1997; Galaskiewicz et al., 1991; Burt, 1987; Flig-

stein, 1985). In the most prominent study on structural equivalence, Burt (1987) analyzed the diffusion of a new drug among physicians. The adoption rate increased in case social equivalent physicians adopted the drug prior (Burt, 1987). The emerging interdisciplinary field of synthetic biology provides opportunities for relevant research results with higher citation rates and additional grants (Yegros-Yegros et al., 2015). This expectation of reward increases competition between researchers, who are therefore more likely to adopt when structural equivalent researchers already adopted the new institutional logic.

In this study structural equivalence based on similarity of network relations (Marsden et al., 1993) is considered and a new measure "knowledge equivalence" is developed. In academia, researchers are similar with regard to the extent of their individual knowledge stock (Weinberger et al., 2007). Because of complementaries within same research fields, researchers have a higher incentive to adopt innovation, if researchers, who published articles with similar topics, already adopted it. Therefore, this study introduces knowledge equivalence to consider competition between authors. In order to distinguish between social cohesion and structural equivalence, regression models will control for structural equivalence measured both with network ties and knowledge equivalence.

Infectiousness

The infectiousness of prior adopters is driven by how heavily information about a prior actor's adoption affects individuals at risk. Prior adopters can, thus, be

more or less influential, based on individual success attributes such as size, performance or status (Greve, 2005). In innovation adoption research, three reasons are put forward why infectiousness varies with success. First, researchers with high status attract more attention and, thus, more information is available to individuals at risk (Greve, 2005). Second, the adopted innovation may be interpreted as cause for success (Greve, 2005) and third, an innovation by researchers with high status receives legitimation and is more likely to be adopted by others (DiMaggio et al., 1983; Meyer et al., 1977). In the case of synthetic biology, Raimbault et al. (2016) identified key actors, who all have a high reputation.

Susceptibility

Susceptibility describes how much an individual at risk is affected by information about others' adoption of an innovation (Greve, 2005). Institutional entrepreneurs reflect on the status quo of their institutional environment, observe drawbacks and translate this cognition into actions designing new institutions (Beckert, 1999; George et al., 2006). They combine institutional logics from divers institutional orders and persuade their social context and institutionalize their idea (DiMaggio, 1988; Garud et al., 2003; Hardy et al., 2008). Institutional entrepreneurs have, therefore, low susceptibility towards their social context.

Intrinsic rate of adoption of individuals at risk

Researchers are more or less likely to adopt an institutional logic based on their individual characteristics, these are considered as intrinsic rate in the heteroge-

neous diffusion model (Strang et al., 1993). To identify the impact of institutional entrepreneurs in diffusing the institutional logic, the intrinsic rate factors are used as control variables. The intrinsic rate of adoption of individuals at risk is accommodated by prior reputation, tenure, prior knowledge diversity and prior affinity towards the field.

Prior research shows that individual efforts are aiming towards achieving goals and each failure increases individual encouragement to search for new behavior or attitudes (Cyert et al., 1963). Successful research reduces incentives to change behavior (Greve, 2005), while slack can diminish this effect (Cyert et al., 1963). A weak negative effect of prior reputation is assumed following work of Singh (1986).

With increased experience and knowledge, individual's gain self-efficacy and incentives to change behavior shrink (Bandura et al., 1985). Researchers with higher tenure are therefore less likely to adopt the new logic.

Adoption of the logic, especially in case of the interdisciplinary field of synthetic biology, may be driven by the author's prior knowledge diversity.

Prior knowledge of research and development teams is a critical input to interpret and finally implement decisions (Hargadon, 1998; Dixon, 1999). Similar effects of an individual's knowledge diversity can be derived. An author's capability to adopt an institutional logic may increase with her prior knowledge diversity. Thus these researchers may be more likely to adopt the institutional logic (Xi et al., 1988; McGuire, 1999; Kane, 1975; Andrews, 1990)

Individuals with a high affinity for the logic are considered to be more likely to adopt. A high affinity is for example expressed by tutoring or mentoring a

university team at an annual student competition.

4.3 Data and Methods

4.3.1 Sample Construction

When it comes to test the hypotheses, a core dataset consisting of publications on synthetic biology and additional information about authoring researchers is constructed in six steps.

First, articles matching keywords used in Bailey et al. (2012) (”synthetic biology” OR ”synthetic genome” OR ”genetic engineering” OR ”novel genetically organisms” OR ”DNA synthesis” OR ”synthetic genomics”) are downloaded from Scopus publication database. The keywords cope a broader set of articles then a strict definition of synthetic biology which was used in Raimbault et al. (2016). These additional datapoints are used in the analysis as control observations and to improve quality of network variables. Scopus publication database is preferred over WebofScience due to a larger dataset of matching articles in June 2016 (Scopus: 55842 articles, WebofScience: 19713, Timeframe 2001-2016).

Second, for the 156k authors in this core corpus, who are identified with Scopus Author ID, all assigned articles are downloaded. This results in a database of 8.6 million article-author pairs.

Third, affiliation strings are matched in following matching procedure in hierarchical order and consecutive process:

Information on corresponding affiliation is extracted from strings in database with keywords, e.g., University, School, College, Hospital, LTD, GmbH, and

more than 200 names of recurring organizations, e.g., csiro, rwth, eth. Map-Package in R (Loecher, 2016) is used to extract available information on the affiliation's city and country from the text. This information is enriched with Google Maps API coordinates and Bing search results. Affiliation entries are then matched by string distance algorithm, coordinates and Bing search results. To reduce matching errors, the string distance algorithm "Full Damerau-Levenshtein distance" with penalty of 4 is used on affiliation strings grouped by author and city. This stepwise matching procedure results in 159k unique affiliation entries.

Fourth, language recognition techniques are used to identify articles publishing in synthetic biology, as the keywords from Bailey et al. (2012) cope a broader set than synthetic biology articles only. Title and abstract of articles having more than 50 words are selected for the analysis. The R package text2vec (Selivanov, 2016) is applied to identify 50 topics in the corpus. This results in a table with all articles and the assigned ratio of the 50 topics. The topics and linked articles are reviewed with experts from the field. In these discussions, eight topics are assigned to synthetic biology and a sum over these topics is calculated for each article. To select articles in synthetic biology, a threshold for the sum is determined based on sample checks and two true-positive controls. First, all articles including the keyword "BioBrick", representing an important technology in the engineering approach, need to be true-positive. Second, the complete list of articles discussed in Drew Endy's Lecture at Stanford on synthetic biology have to be selected. Articles with a higher ratio than the threshold are marked as synthetic biology publications. In the fifth step, data on an annual student

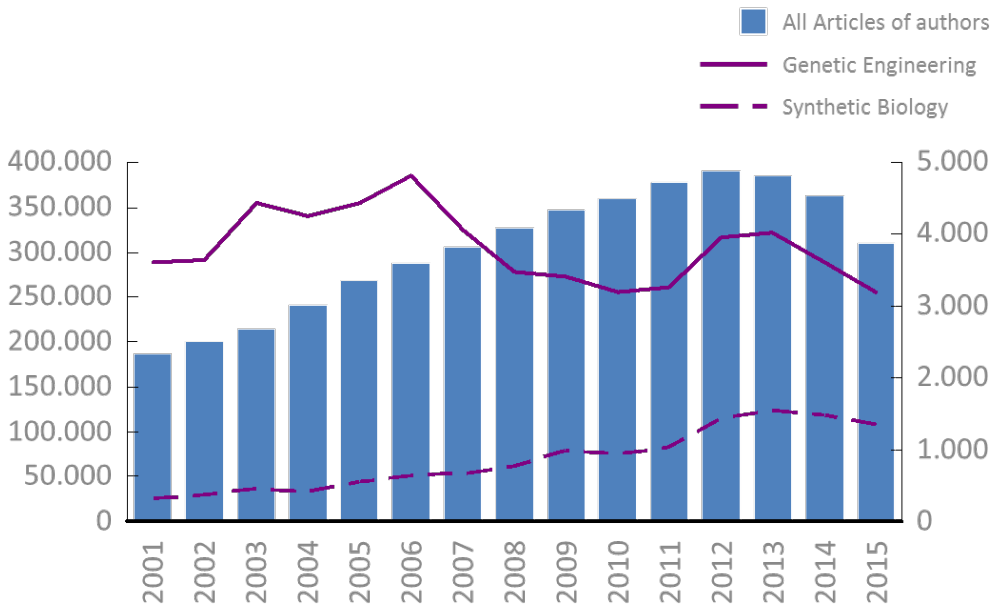


Figure 4.1: Amount of total research articles in corpus, flagged with keyword search and identified with topic modeling

competition is added to the database. The names and the active years of 911 tutors have been crawled from iGEM.org. 530 authors were matched with authors in Scopus Database using name matching and controlling for affiliation. Also in this step, the string distance algorithm "Full Damerau-Levenshtein distance" with penalty of 4 is chosen. 435 authors could be identified within the full corpus of which 235 adopted the logic after tutoring a team.

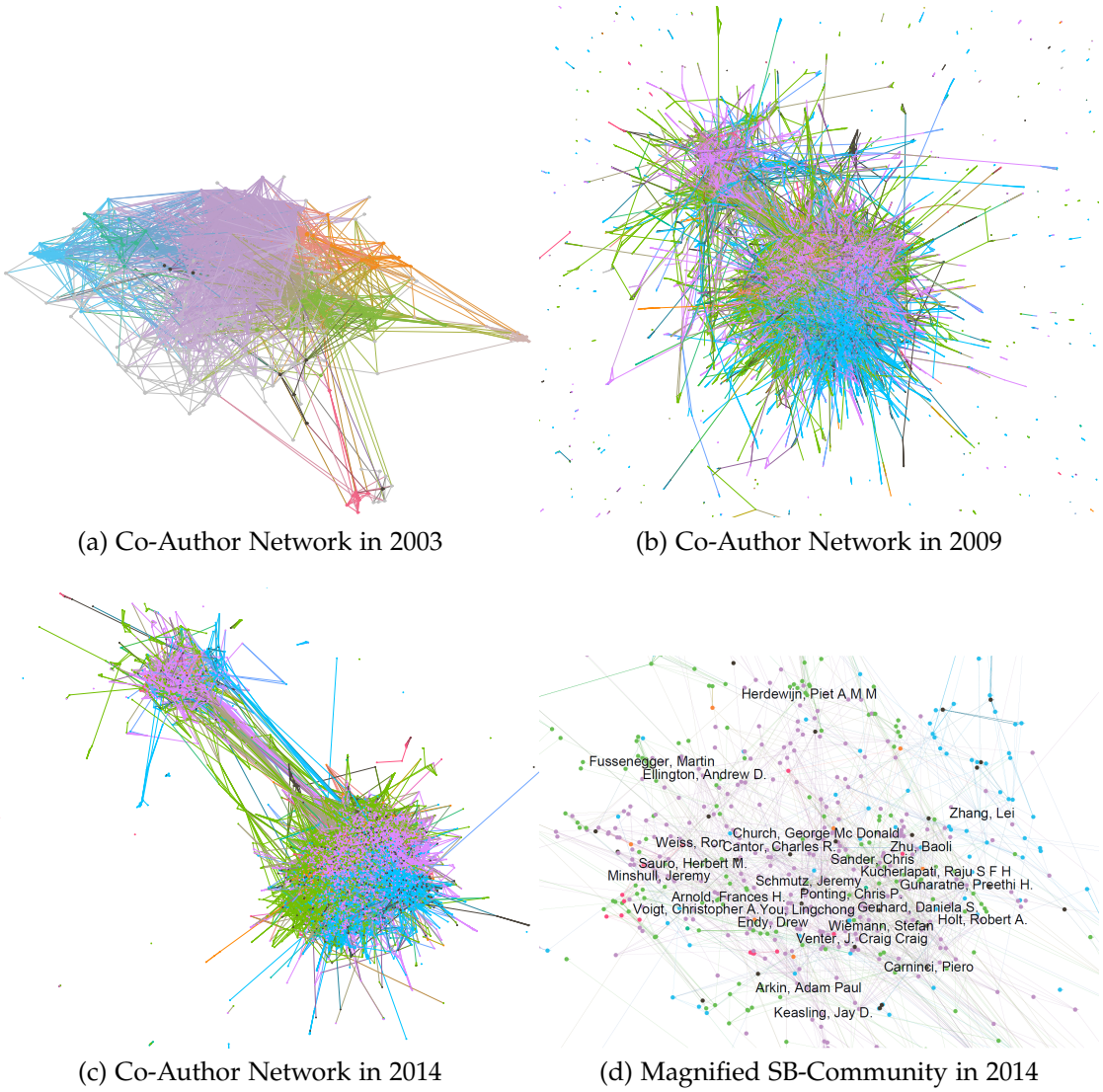
Due to incomplete data in 2016, the selected database to test the hypotheses is limited to the years 2001-2015. The final corpus includes 153727 authors and is visualized in figure 4.1. In a final step observations per author are selected for the year adopting the logic or the last year of appearance in the corpus for authors not adopting the logic.

Co-author networks visualizing authors in synthetic biology are displayed in figure 4.2. A) shows the co-author network of synthetic biology in 2003 with

assigned names of researchers having high eigencentrality values, e.g., David H. Haussler, Francis S. Collins and Richard A L Gibbs. In total, 3,263 authors are assigned to the synthetic biology network.

The graph B) visualizes data until 2009. Two observations become clear; first, the extension of the network results in 12,735 authors and second, the starting split of a core community on the top left. While all authors in the network published an article which is related to the engineering approach, a smaller group of researchers acts in a dense separated community.

The separation is more present in (C), which shows the co-author network in 2014. The network counts 27,143 authors in 2014 who published at least one article assigned to the engineering approach using natural language processing. The core community is magnified in figure (D) and labels for authors with a high network degree are displayed. The community around Drew Endy, J. Craig Venter, George Mc Donald Church, Jay D. Keasling and Christopher A. Voigt build the core community and separate from the larger peripheral community working on metabolic engineering and developing methods for synthetic biology, e.g., Jennifer Doudna studying CRISPR Cas9. Both communities are working on engineering biology, thus, the network visualization validates the natural language processing approach.



(a) Co-Author Network in 2003

(b) Co-Author Network in 2009

(c) Co-Author Network in 2014

(d) Magnified SB-Community in 2014

Figure 4.2: Development of co-author networks in the emerging field of synthetic biology

4.3.2 Measurement

DV: Adaption of Logic

The analysis aims to assess, how social influence affects adoption, and to evaluate the ability of institutional entrepreneurs to drive change by convincing their social context to adopt their ideas. Despite the fact that the prosperous field has already evolved since the early 2000s, there is still a high level of uncertainty in scientific, technical, economical and ethical dimensions. A researcher confronted

with the logic of this emerging field will have a need to rely on peers' experience to evaluate benefits and costs of adoption.

The engineering approach is identified in research articles with natural language processing from a corpus of articles assigned to the broader field of genetic engineering. For each author in the large corpus, all articles per publication year are reviewed for content related to the engineering approach. If an author publishes an article related to the approach, she is assumed to have adopted the logic.

To measure the hazard of adoption a cox regression is selected as econometric model (Cox, 1972). Herefore, both a dependent variable and an indicator variable are needed (Klein et al., 2003). The dependent variable specifies ordered event times, an indicator variable determines whether a subject is experiencing an event. Details on cox regression and the mathematical derivation are given in 4.3.3. The dependent variable "Years since 2000" accounts for time passed since the emergence of synthetic biology. The indicator variable "Synbio Article" is coded "1", if an author publishes an article assigned to the engineering approach in the current year and "0", if published articles miss the engineering approach, e.g., general biology research of transcription or protein binding, biochemistry of nucleotides and research on RNAi.

Explanatory Variables

The selected explanatory variables are predictors measuring social cohesion, infectiousness and susceptibility. All variables are based on a prior 5 year moving time window to account for declining influencing effects (Eisenhardt et al., 2007;

Singh et al., 2016; Sampson, 1997).

Social cohesion is the number, length and strength of paths between actors (Marsden et al., 1993). Two social distances can be measured using publication data by analyzing co-authorships and affiliations to identical organizations. In order to compute these measures, prior co-authors who previously published a "Synbio Article" are grouped by author. Similarly, previous authors, who were priorly assigned to the actor's current affiliation are identified.

Infectiousness of an institutional entrepreneur is assumed to vary with prior reputation. Reputation can be measured with citations of research articles (Piwowar et al., 2007). Interaction effects of social cohesion measures, i.e., prior co-authors and colleagues at the affiliation, and sum of prior citations of peer authors, who priorly published a "Synbio Article", are calculated to investigate infectiousness.

Susceptibility describes how much an individual at risk is affected by information regarding adoption of an innovation by peers (Greve, 2005). Institutional entrepreneurs are assumed to be low susceptible as they combine institutional logics from divers institutional orders and persuade their social context of their idea and institutionalize it (DiMaggio, 1988; Garud et al., 2003; Hardy et al., 2008). Interaction effects of social cohesion measures and prior citations of the actor at risk are calculated to investigate the variation of susceptibility towards social cohesion due to prior reputation.

Accounting for Reflection Problem

Manski (1993) describes the reflection problem as source of endogeneity when estimating social influence. Reflection occurs because of three observations concerning causal interpretation of social influence. First, actors build social ties based on homophily, e.g., when they share similar attitudes, interests, beliefs (McPherson et al., 2001). Second, many characteristics can cause homophilous tie formation of which most are unobserved ones. Third, these unobserved variables are correlated with the similarity of actors' behavior (Singh et al., 2013). In the field of synthetic biology an author might adopt the logic out of unobserved characteristics, e.g., behavioral or cultural, and then build social ties based on theses similarities. To account for the reflection problem two principles are introduced (Singh et al., 2013). First, all variables potentially influencing the actor on her behavior are calculated for prior 5 year time window before the event. Second, following Bramoullé et al. (2009) unobserved effects common to all actors belonging to same group are introduced. Practically, binary variables accounting for the location of the author's affiliation are introduced on continent-level.

Accounting for bias of used thresholds

Declaration of an article as "Synbio Article" is based on natural language processing and a manually set ratio threshold of specified topics. To account for potential bias due to threshold selection (35%), a robustness check with a second threshold (55%) is executed.

Control variables

Control variables include predictors considering geographic location, constraint, intrinsic rate of adoption, structural equivalence, knowledge equivalence and affiliation characteristics. Influencing variables are based on prior 5 year moving time window to account for declining influencing effects (Eisenhardt et al., 2007; Singh et al., 2016; Sampson, 1997). Synthetic biology is dominated by US research, followed by European and Asian (Raimbault et al., 2016). To control for this geographical effect and the reflection problem (Manski, 1993), the continent of an assigned affiliation is included as control variable.

In order to control for network structure, the constraint of a researcher is introduced. This measure increases, if peers communicate a lot to one another directly (dense network) or, if they distribute information indirectly via a central node (hierarchical network) (Burt, 2004). Constraint is used to control for network structure and measured with the following formula (Burt, 2004):

$$C_i = \sum_j c_{ij}, i \neq j \tag{4.3}$$

with C_i equal network constraint of actor i, c_{ij} equal i's dependence on it's tie j:

$$c_{ij} = (p_{ij} + \sum_q p_{iq} p_{qj})^2, i \neq q \neq j \tag{4.4}$$

where p_{ij} equal the ratio of actor i's network time and energy spent on tie j. p_{ij} equal the extent of direct relations and $\sum_q p_{iq} p_{qj}$ equal the magnitude of indirect relations via contacts q.

Researchers are more or less likely to adopt an innovation based on their individual characteristics, these are considered as intrinsic rate in the heterogeneous

diffusion model (Strang et al., 1993). Intrinsic rate of adoption of individuals at risk is investigated with prior reputation, tenure, prior knowledge diversity and prior affinity towards the logic. Reputation is measured with citations of research articles (Piwowar et al., 2007), tenure is calculated by subtracting earliest from focal year of analysis, knowledge diversity of focal actor is measured by an averaged row-wise cosine similarity of a document-term-matrix containing her prior article abstracts. A high affinity can be seen in tutoring or mentoring a university team at an annual student competition. Prior participations at an annual student competition as a tutor are summed for each author.

Reputation of an affiliation is calculated by the sum of prior citations of affiliated publications and affinity is measured by the sum of prior authors at an affiliation who tutored a team at an annual student competition.

To ensure social cohesion measures are not confounded with unobserved structural equivalence, two variables are included: structural equivalence based on similarity of network ties (Burt, 1982) and knowledge equivalence as to similarity of research topics before adopting the logic. The similarity of network ties is calculated with the cosine similarity of network relations between focal author and authors priorly active in Synthetic Biology and then averaged for each author. Co-authors are deselected from this calculation to separate the effect from social cohesion (Leenders, 2002). Basis for this calculation is network data from prior 5 years. Knowledge equivalence to authors who adopted the logic is calculated by averaging the cosine similarity of the document-term-matrices from article abstracts. The articles of authors 5 years prior adopting the logic are compared with articles of focal authors in last 5 years. Co-authors are deselected

from this calculation to separate the effect from social cohesion (Leenders, 2002) Observations with missing data concerning network measures or knowledge diversity are possible in two cases. Newcomers, who do not have data on a prior network, or authors, who published articles with small abstract not considered in topic evaluation, are selected for data cleaning. Missing data is handled in two steps. First the value is set to the variable's minimum and second the binary control variable "Missing Data Control" is set to '1'.

4.3.3 Econometric Model

Econometric models of durations analyze length of time spent in a given state before a transition to another state, e.g., duration of being unemployed or alive or without health insurance (Klein et al., 2003). In the setting of logic diffusion, the adoption of the logic is interpreted as transition to another state.

Given a dataset with durations that may be complete, truncated, or censored, non-parametric, parametric and semi-parametric models are used (Klein et al., 2003). If all subjects have same characteristics and just receive divergent treatment, non-parametric models can be applied. In case of individual characteristics, covariates are introduced and parametric or semi-parametric models selected (Klein et al., 2003). Parametric models need pre-specified shapes of survival function, e.g., Weibull or exponential. The main issues of parametric modeling are the dependence on the correct shape for consistent parameter estimates and the variety of parametric models that are available (Klein et al., 2003). To avoid the prediction of a distribution, Cox (1972) proposed a semi-parametric

approach, the so-called proportional hazard model. It estimates the hazard function with independent censoring and needs only weak distributional assumptions (Cameron et al., 2005).

Given a dataset of triples $(T_j, \delta_j, \mathbf{Z}_j(t))$, j = 1,...,n with T_j time on study for jth subject, δ_j is the event indicator for jth subject (δ_j = 1 event occurred, δ_j = 0 lifetime is right-censored) and $\mathbf{Z}_j(t) = Z_{j1}(t)$, ..., $Z_{jp}(t)^t$ is the vector of covariates for the jth individual at time t. Covariates in $Z_{jk}(t)^t$, k = 1,..., p can be both time variant and fixed. Following Klein et al. (2003) the model derivation is described in the following. The proportional hazard model divides the conditional hazard rate $h(t|\mathbf{Z})$ into two separate functions:

$$h(t|\mathbf{Z}) = h_0(t)c(\boldsymbol{\beta}^t, \mathbf{Z}) \tag{4.5}$$

where h_0 is called the baseline hazard and is a function of t. The baseline function describes how the risk of an event per time unit changes over time at baseline levels of covariates. $c(\boldsymbol{\beta}^t, \mathbf{Z})$, the effect parameters, are a function of $\mathbf{Z}$ and thus describe variations of hazard with explanatory covariates. A commonly used choice for the effect parameters is the exponential form:

$$c(\boldsymbol{\beta}^t, \mathbf{Z}) = exp(\boldsymbol{\beta}^t \mathbf{Z}) = exp\Big(\sum_{k=1}^{p} \beta_k Z_k\Big) \tag{4.6}$$

With the exponential form, coefficients are easily interpretable and $exp(\boldsymbol{\beta}^t \mathbf{Z}) > 0$ can be ensured. All hazard functions $h(t|\mathbf{Z})$ of form 4.5 are proportional to the baseline hazard and the scale factor $c(\boldsymbol{\beta}^t, \mathbf{Z})$ is not an explicit function of t. β can therefore be estimated without simultaneously specifying the functional form for baseline function $h_0(t)$ (Cox, 1972; Cox, 1975; Cameron et al., 2005). Because only the second term is assumed to have parametric form, this function is called

semi-parametric. The separation is shown by comparing two individuals with covariate values $\mathbf{Z}$ and $\mathbf{Z}^*$:

$$\frac{h(t|\boldsymbol{Z})}{h(t|\boldsymbol{Z}^*)} = \frac{h_0(t)exp[\sum_{k=1}^{p}\beta_k Z_k]}{h_0(t)exp\big[\sum_{k=1}^{p}\beta_k Z_k^*\big]} = exp\big[\sum_{k=1}^{p}\beta_k(Z_k - Z_k^*)\big] \tag{4.7}$$

The ratio in (4.7) is a constant, thus hazard rates are proportional. Let $t_1 < ... < t_D$ be ordered event times, D the group of all failed subjects at time t_i and $Z_{(i)k}$ the kth covariate of subject which fails at time t_i. The risk set at time t_i, $R(t_i)$, is the group of subjects not failed or censored just before t_i. The probability that a subject fails at time t_i, assuming no ties in dataset, is given by (Klein et al., 2003):

$$\begin{aligned} &P[\text{individual dies at}t_i|\text{one death}t_i] \\ &\qquad = \frac{P[\text{individual dies at}t_i|\text{survival to}t_i]}{P[\text{one death at}t_i|\text{survival to}t_i]} \\ &\qquad = \frac{h[t_i|\boldsymbol{Z}_{(i)}]}{\sum_{j\in R(t_i)} h[t_i|\boldsymbol{Z}_j]} \\ &\qquad = \frac{h_0(t_i)exp(\boldsymbol{\beta}^t, \boldsymbol{Z}_{(i)})}{\sum_{j\in R(t_i)} h_0(t_i)exp[\boldsymbol{\beta}^t, \boldsymbol{Z}_j]} \\ &\qquad = \frac{exp[\boldsymbol{\beta}^t, \boldsymbol{Z}_{(i)}]}{\sum_{j\in R(t_i)} exp[\boldsymbol{\beta}^t, \boldsymbol{Z}_j]} \end{aligned} \tag{4.8}$$

Multiplying (4.8) over all failures we receive the partial likelihood function:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{D} \frac{exp[\boldsymbol{\beta}^t, \boldsymbol{Z}_{(i)}]}{\sum_{j\in R(t_i)} exp[\boldsymbol{\beta}^t, \boldsymbol{Z}_j]} \tag{4.9}$$

Cox (1972) obtains parameter estimates, $\hat{\beta}$, by maximizing a partial log-likelihood function:

$$LL(\boldsymbol{\beta}) = ln[L(\boldsymbol{\beta})] = \sum_{i=1}^{D}\sum_{k=1}^{p}\beta_k Z_{(i)k} - \sum_{i=1}^{D} ln\Big[\sum_{j\in R(t_i)} exp[\boldsymbol{\beta}^t, \boldsymbol{Z}_j]\Big] \tag{4.10}$$

The partial maximum likelihood estimates are then found by maximizing (4.10). The efficient score equations are determined by taking partial derivatives with respect to the $\boldsymbol{\beta}$'s with $U_b(\boldsymbol{\beta}) = \delta LL(\boldsymbol{\beta}) \;/\; \delta\beta_h$, h = 1,...,p. The p nonlinear equations

can be solved numerically, e.g., using the Newton-Raphson technique (Klein et al., 2003).

In the present dataset, events are grouped by years with a high ratio of ties which may lead to bias in the cox partial log-likelihood function. Alternative partial likelihoods by Breslow (1974) and Efron (1977) and by a discrete-time, hazard-rate model developed by Cox (1972) are used for datasets with ties. The partial likelihood composed Efron (1977) is superior in performance and speed (Singer et al., 2003; Allison, 1995) and therefore adopted for this study. The partial likelihood provides approximations of the exact marginal log likelihood.

Let $t_1 < ... < t_D$ denote D distinct, ordered, event times. Let then d_i be number of failures in t_i and $\mathbb{D}_i$ the set of subjects fail at time t_i. Sum of vectors $\mathbf{Z}_j$ over all subjects who fail at time t_i is described by $\mathbf{s}_i = \sum_{j \in \mathbb{D}_i} \mathbf{Z}_j$. The exact marginal log likelihood by Efron (1977), using the annotation of Klein et al. (2003), is:

$$L_{Efron}(\boldsymbol{\beta}) = \prod_{i=1}^{D} \frac{exp(\boldsymbol{\beta}^t \boldsymbol{s}_i)}{\prod_{j=1}^{d_i} \left[\sum_{k \in R_i} exp(\boldsymbol{\beta}^t \boldsymbol{Z_k}) - \frac{j-1}{d_i} \sum_{k \in \mathbb{D}_i} exp(\boldsymbol{\beta}^t \boldsymbol{Z_k}) \right]} \tag{4.11}$$

The cox model uses proportional hazard model assumptions that the hazard function can be separated into a baseline hazard function and a function describing the effect parameters. The baseline function describes how the risk of an event per time unit changes over time at baseline levels of covariates. The effect parameters describe variations of hazard with explanatory covariates. The cox model uses this separation to drop out the baseline function before estimating β (Cox, 1972). Parametric models use pre-selected distributions for the baseline function to estimate β (Klein et al., 2003). To account for bias in estimating with cox model, parametric models using Weibull and Exponential distributions are

calculated and compared to coefficients from cox model. In addition a logit model is calculated disregarding the hazard calculation.

4.4 Results

Descriptive statistics for key variables are reported in table 4.1, regression models to test the hypotheses are in table 4.2 and marginal effects on interaction effects regarding infectiousness and susceptibility are visualized in graphs 4.3 and 4.4. The variables colleagues at affiliation, co-authors, reputation and affinity towards the engineering approach of author and affiliation are log transformed because they were heavily skewed. Structural equivalence and knowledge equivalence are standardized and scaled with 100. In table 4.2 Model 1 and 2 report

Table 4.1: Descriptive statistics for Synthetic Biology community database

Variable	Obs	Mean	Std. Dev.	Min	Max
Synbio Article	153727	.174	.379	0	1
Years since 2000	153727	12.201	3.824	1	15
Reputation	153727	2.176	1.9	0	8.72
Tenure	153727	11.972	8.263	0	24
Constraint	153727	.259	.239	0	1.696
Knowledge Diversity	153727	.901	.211	0	1
Affinity towards SynBio	153727	.002	.038	0	1.099
Reputation of Affiliation	153727	6.048	3.379	0	11.471
Affinity of Affiliation towards SynBio	153727	.457	.738	0	3.219
Structural Equivalence	153727	.167	.538	0	100
Knowledge Equivalence	153727	42.804	8.292	0	100
Missing Data Control	153727	.248	.432	0	1
Co-Authors in SynBio	153727	.304	.586	0	4.984
Reputation of Co-Authors in SynBio	153727	1.26	2.253	0	9.456
Authors in SynBio at Affiliation	153727	2.005	1.745	0	7.924
Reputation of Authors in SynBio at Affiliation	153727	2.617	2.113	0	9.036

cox regression coefficients and standard-errors for controls-only and the main effects model for hypothesis 1. Model 3-6 include interaction effects to investigate hypotheses 2 and 3.

Regarding control variables the dummy coded locations Northamerica and Europe indicate a geographical influence. A significant negative effect of author's reputation and higher tenure on adopting the logic can be validated. Prior knowledge diversity of actors is correlated positive with adopting the logic and showing high affinity has positive implications as well. Affiliation characteristics both show negative effects. Prior reputation of affiliated organization and its' affinity towards the engineering approach have negative estimates. Competition among structural or knowledge equivalent actors has a positive effect.

These results are consistent with prior research and corroborate our assumptions on intrinsic rate of motivation and influences of peers. The negative estimator of affiliation's affinity is counter-intuitive. However, correlation effects between affinity of the author and this variable explain the negative effect.

4.4.1 Hypotheses Tests

Hypothesis 1 predicts that institutional entrepreneurs use social influence mechanism to diffuse an institutional logic. Models 1 and 2 validate this hypothesis for both social cohesion measures. The variables prior co-authors and colleagues at affiliation are both statistical significant and have positive estimators. Thus, peers influence a focal author's choice to adopt the institutional logic.

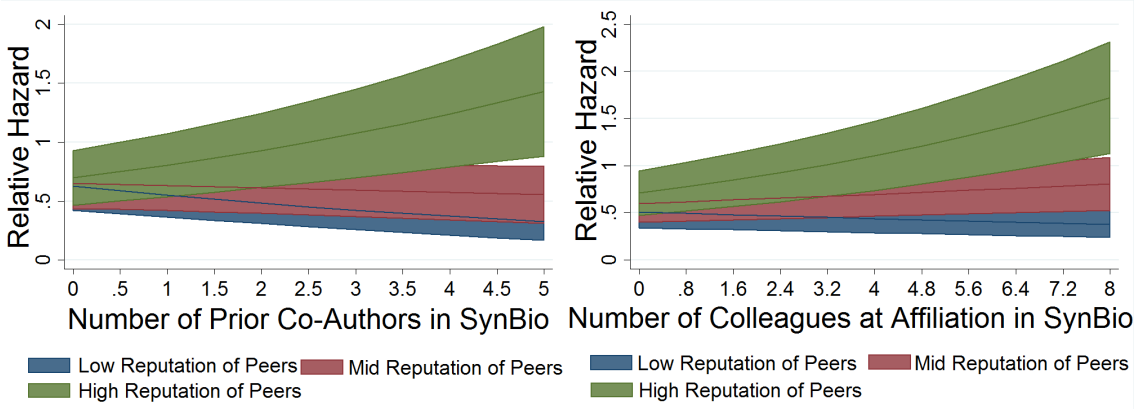
Hypothesis 2 assumes that institutional entrepreneurs use their high reputation to increase their infectiousness. Models 3 and 4 validate this hypothesis showing a positive interaction effect with z-values above 13. Marginal effects are visualized in 4.3 for prior co-authors (a) and for colleagues at affiliation (b). In-

Table 4.2: 6 models using cox regression for infectiousness and susceptibility

	(1) _t	(2) _t	(3) _t	(4) _t	(5) _t	(6) _t
Africa	0	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)	(.)
Asia	−0.0894	−0.208	−0.209	−0.174	−0.205	−0.209
	(−0.54)	(−1.26)	(−1.27)	(−1.06)	(−1.25)	(−1.27)
Australia	0.173	0.0229	0.0101	0.0519	0.0367	0.0406
	(1.01)	(0.13)	(0.06)	(0.30)	(0.21)	(0.24)
Europe	0.339*	0.137	0.134	0.153	0.142	0.147
	(2.06)	(0.83)	(0.81)	(0.93)	(0.86)	(0.89)
Latinamerica	−0.634***	−0.662***	−0.675***	−0.644***	−0.664***	−0.655***
	(−3.68)	(−3.84)	(−3.92)	(−3.74)	(−3.86)	(−3.80)
Middle-East	−0.555**	−0.654***	−0.660***	−0.638***	−0.670***	−0.667***
	(−3.25)	(−3.83)	(−3.87)	(−3.74)	(−3.93)	(−3.91)
Northamerica	0.598***	0.340*	0.323	0.299	0.336*	0.336*
	(3.62)	(2.06)	(1.95)	(1.81)	(2.04)	(2.04)
Reputation	−0.693***	−0.700***	−0.702***	−0.699***	−0.872***	−0.883***
	(−119.39)	(−121.63)	(−122.17)	(−121.74)	(−116.98)	(−87.52)
Tenure	−0.0434***	−0.0477***	−0.0473***	−0.0477***	−0.0457***	−0.0473***
	(−40.51)	(−43.98)	(−43.64)	(−44.01)	(−41.88)	(−43.60)
Constraint	−0.217***	−0.0517*	−0.0421	−0.0536*	−0.0318	−0.0402
	(−8.55)	(−2.05)	(−1.68)	(−2.13)	(−1.27)	(−1.60)
Knowledge Diversity	0.378***	0.294***	0.305***	0.299***	0.335***	0.304***
	(11.61)	(9.01)	(9.36)	(9.18)	(10.26)	(9.31)
Affinity towards SynBio	1.184***	1.046***	1.072***	1.035***	1.050***	1.072***
	(12.61)	(11.10)	(11.37)	(10.98)	(11.15)	(11.38)
Reputation of Affiliation	−0.0233***	−0.0762***	−0.0754***	−0.0728***	−0.0750***	−0.0720***
	(−11.38)	(−32.71)	(−32.40)	(−31.18)	(−32.20)	(−30.78)
Affinity of Affiliation towards SynBio	−0.219***	−0.390***	−0.386***	−0.402***	−0.389***	−0.387***
	(−22.47)	(−31.67)	(−31.34)	(−32.21)	(−31.57)	(−31.34)
Structural Equivalence	0.0651***	0.0557***	0.0553***	0.0562***	0.0611***	0.0555***
	(25.35)	(16.33)	(15.06)	(16.89)	(20.09)	(16.61)
Knowledge Equivalence	0.0211***	0.0213***	0.0212***	0.0214***	0.0216***	0.0212***
	(28.80)	(28.86)	(28.75)	(29.07)	(29.31)	(28.80)
Missing Data Control	−0.490***	−0.332***	−0.350***	−0.341***	−0.431***	−0.353***
	(−27.75)	(−18.47)	(−19.48)	(−18.97)	(−23.77)	(−19.60)
Co-Authors in SynBio		0.236***	−0.132***	0.244***	−0.310***	0.204***
		(11.46)	(−3.73)	(11.84)	(−11.82)	(9.94)
Reputation of Co-Authors in SynBio		0.0493***	0.0296***	0.0459***	0.0998***	0.0536***
		(9.40)	(5.51)	(8.75)	(18.46)	(10.30)
Authors in SynBio at Affiliation		0.0671***	0.0682***	−0.0548***	0.0671***	0.0104
		(9.07)	(9.23)	(−4.69)	(9.06)	(1.33)
Reputation of Authors in SynBio at Affiliation		0.115***	0.113***	0.0784***	0.116***	0.119***
		(23.34)	(22.89)	(13.79)	(23.39)	(24.12)
Co-Authors in SynBio × Reputation of Co-Authors in SynBio			0.0784***			
			(13.51)			
Authors in SynBio at Affiliation × Reputation of Authors in SynBio at Affiliation				0.0350***		
				(13.73)		
Co-Authors in SynBio × Reputation					0.249***	
					(45.76)	
Authors in SynBio at Affiliation × Reputation						0.0725***
						(24.12)
Observations	153727	153727	153727	153727	153727	153727

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

fectiousness is thus validated for both social cohesion measures displaying no social influence by authors with low reputation and increasing positive social influence with number of socially proximate authors with high reputation, i.e., institutional entrepreneurs.

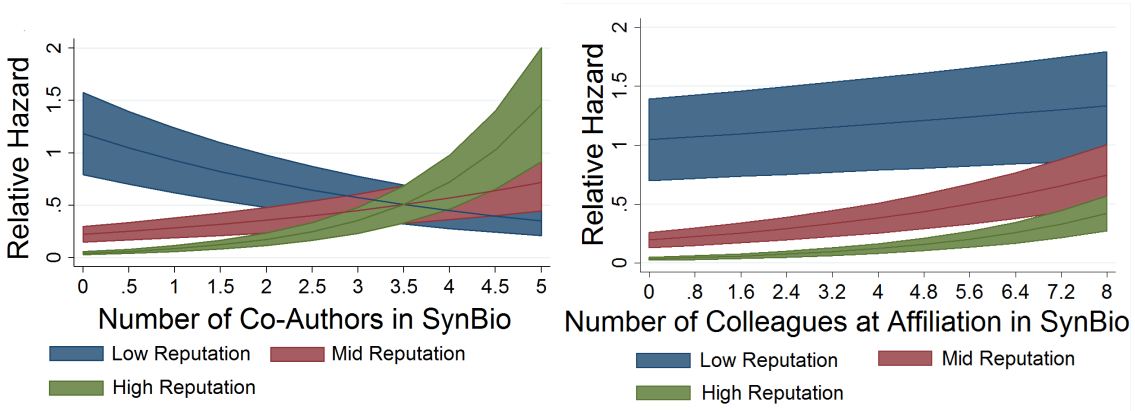


(a) Infectiousness of Prior Co-Authors (b) Infectiousness of Colleagues at Affiliation

Figure 4.3: Marginal effects for infectiousness predictors with 95% confidence interval

Hypothesis 3 is derived from activities of institutional entrepreneurs reflecting status quo of their environment and actively influencing their social context, not being highly susceptible towards their social context. Models 5 and 6 identify an overall positive interaction effect between reputation and both co-authors and colleagues at affiliation. However, the marginal effects visualized in 4.4 identify significant effects for co-authors (a) and for authors at same affiliation (b) supporting the hypothesis 3. The figure displays a strong correlation of social influence and reputation. Institutional entrepreneurs with high reputation are only susceptible to social cohesion at a very high level of social influence. In other words, strong social pressure is needed to make an institutional entrepreneur rely on socially proximate. Authors with low reputation have a high

hazard at very low level of social influence. With high levels of social influence the risk decreases for susceptibility towards co-authors. These actors might be susceptible to social influence instantly at low social pressure and if not, there are strong unobserved reasons why not to join and continue with this decision. The marginal effects of the interaction effect with colleagues at affiliation are differing only for low reputation. Institutional entrepreneurs are less susceptible towards social influence from colleagues at their affiliation, in contrast to that authors with low reputation are easily influenced and the effect increases with number of authors at their affiliation, who already adopted the logic.



(a) Susceptibility towards Prior Co-Authors (b) Susceptibility towards Colleagues at Affiliation

Figure 4.4: Marginal effects for reputation on susceptibility with 95% confidence interval

4.4.2 Model Validation and Robustness Check

To validate the selection of a semi-parametric cox regression, parametric hazard models are applied to the dataset and a logit regression is used. Table A.2 visualizes estimation results. Results with parametric models and logit regression

are confirming conclusions from the cox regression. A threshold of 35% to assign a paper to synthetic biology was iteratively specified with defined tests. To check for robustness the regression is validated with a higher threshold of 55%. Results are displayed in displayed in table A.1. All explanatory variables have equal output.

4.5 Discussion and Implications

Prior research exists on activities of institutional entrepreneurs in emerging fields, diffusion of institutions and product innovation in industry (e.g. Van de Ven et al., 1993; Powell et al., 1996; Lenway et al., 2001; Garud et al., 2002; Baron et al., 1986; Guler et al., 2002; Ritvala et al., 2009). The context of scientific innovation was not yet analyzed and the diffusion of institutional logics by institutional entrepreneurs required validation.

This study extends the literature and provides a quantitative validation of theories on how institutional entrepreneurs drive diffusion of institutional logics in context of scientific innovation to shape an emerging field. The heterogeneous diffusion model is applied to scientific innovation and institutional logics to measure the impact of social influence mechanisms and validate theories on characteristics and strategies of institutional entrepreneurs.

Results validate that institutional entrepreneurs use social cohesion to convince their social context considering both co-authors and colleagues at their affiliation. The high infectiousness of institutional entrepreneurs due to prior reputation increases social influence on socially proximate actors and thus enables them to

compel others with the institutional logic. Characteristics of institutional entrepreneurs being not susceptible towards social influence are validated as well. This study is first to investigate these effects in context of scientific innovation. Transforming these results in a call for action, policy makers should be aware of institutional entrepreneurs in their seeding strategies. Identifying, empowering and directing these key players enables the diffusion of an institutional logic and support the emergence of a new field.

4.6 Limitations

Limitations can be seen in three areas. First, the consideration of publication data only, second, combination of causal mechanisms and third, limitations in data processing.

The analysis is based on research articles accessible on Scopus, thus external actors such as mass media, consultants, funding agencies or professional communities and events like conferences are not considered. However, research suggests that these external actors shape how actors interpret innovation (Wejnert, 2002; Strang et al., 1998). Second, even though causal mechanisms, e.g., information transmission, observation and learning are considered, the distinction and separation between them are not possible based on the existing dataset. Empiric analysis observing and interrogating actors might help to accomplish this step. Third, data processing is limited by resources, e.g., selection of articles with natural language processing can be improved with a larger training data set and Scopus author identifier was used to disambiguate and collect all assigned

articles. The error rate of Scopus author identifier algorithms can influence results.

Future research opportunities are seen both in adopting the approach to alternative fields and in deep diving the present results. Scholars can use the large scale approach and validate results at alternative fields to confirm universality of findings.

Deep dives on the study results can be qualitative analyses expanding the work of Raimbault et al. (2016) and analyzing the dynamics of the core and peripheral communities building on Grodal (2017). In particular case studies and interview series capturing perspectives of institutional entrepreneurs on their strategies to drive change and influence their social context are important to confirm and expand the results of this study.

Chapter 5

Study II: Assessing the Impact of Open Science in Synthetic Biology

5.1 Introduction

With the legitimacy of established logics institutional entrepreneurs create institutions such as property rights in an emerging field, which are potential levers for the diffusion of innovation (,e.g., Merton, 1973; Kitch, 1977; Merges et al., 1990; Scotchmer, 1991; Henderson et al., 1995; Fabrizio et al., 2008; Czarnitzki et al., 2009). Quantitative studies analyzed the effects of property rights in cumulative innovations (Murray et al., 2007; Williams, 2013; Murray et al., 2016). They identified negative effects on the diffusion of scientific innovation and an increased creativity and diversity of research using open science research tools. In synthetic biology and BioBrick engineering, in particular, early stage innovations are, inter alia, basic biological parts by means of which more complex systems, i.e., follow-on research, can be developed. Therefore, experts fear potential patent thickets, if basic parts are patented (Oye et al., 2009) and they want the ”stuff of life” (Endy, 2009) to remain open to the public to improve both, speed and quality with regard to the development of applications (Endy,

2009; Maurer, 2009).

When running a survey in 2013, Kahl (2015) reviewed the current re-usage of basic parts from open science in the field of synthetic biology. While the majority of researchers claimed not to re-use parts from open databases, they maintained the low quality of these parts to be a reason for their behavior. The benefit of adopting these parts are dominated by the costs for integration and trouble shooting. A higher level of quality would therefore increase re-usage.

This study shortly reviews the status quo of open science in synthetic biology and fills two gaps in prior research. First, the current impact of open science in the research area of BioBrick engineering is lacking quantitative validation. Second, though success factors of open science are hypothesized in prior research, their analytical validation is still outstanding. To perform these analyses, a new methodology to measure the diffusion of early stage innovation in follow-on research has been developed. A query list of open science parts, published by the BIOFAB project in 2010, is matched to patent data. The BIOFAB parts are suitable for this analysis as they cover both, newly created parts and already existing parts with an improved characterization quality.

This analysis is run in three sequential steps. First, a query search using BLAST and automated filters is used to pre-select matches. Second, a manual review is performed to validate trends in time-series of formerly existing parts and of the potential re-usage of newly created ones. Third, inventors of patent applications are personally contacted to provide feedback on their experiences. With the new methodology eleven times more re-usages are identified in comparison to count of references in patent texts. Results show a moderate diffusion of new created

BIOFAB parts, but an effect of a higher quality on re-using open science parts, cannot be identified. Probable reasons for these results are discussed, follow-on questions for research enumerated and applicable solutions for practice elaborated.

5.2 Organizations and Initiatives Establishing Open Science

To provide an overview of the landscape of open science organizations, initiatives of the five selected organizations iGEM, BBF, SynBERC, Addgene and Cambia have been reviewed. A full list of identified organizations is drawn up in 3.2. To structure the overview, initiatives are clustered by four enablers to drive change.

Foster a mindset of collaborative research is the first enabler. Via the iGEM competition, with its medals honoring collaboration (iGEM Foundation, 2017c), conferences such as the SBX.0, co-organized by the BBF (BioBricks Foundation, 2017b) and by semi-annual knowledge sharing coordinated by SynBERC (SynBERC, 2006), multiple initiatives aim at promoting a mindset of collaborative research.

Develop open science parts as been identified as second enabler. The iGEM registry covering thousands of BioBricks is the most prominent initiative. Unfortunately, a common feedback from outside as to these parts claims a lack of the characterization of parts and thus insufficient quality for re-usage. While in the software industry, most startups could not be initiated without using the

open science code, startups in synthetic biology often have to buy parts and technologies (Kahl, 2015). With the idea of countering this development, the iGEM foundation introduced rewards for improving part documentation into the judging criteria of their competition (iGEM Foundation, 2017c). In addition, in 2010, the BIOFAB project run by SynBERC, the BBF and Lawrence Berkeley National Laboratory, published 478 open science BioBricks. Both, newly created parts meeting high quality standards have been developed and popular parts have been measured and documentation has been improved (SynBERC, 2006).

Third, both, physical and legal access to parts needs to be ensured. Biological parts can be openly accessed via the iGEM parts registry and via Addgene. As multiple researchers claim issues for the transfer of material, the BBF recently announced that three initiatives were intended to solve these problems. The software bionet has created an online platform to gain cross-lab transparency on stored material and to automate physical exchange. The OpenMTA counters the complex process of agreements concerning material transfer: a framework contract allows material exchange between two organizations, only short repetitive fixations are then required for actual exchanges (Kahl, 2015; Kahl, 2017). The PDC enables researchers to secure scientific findings for the public domain by making a contribution to become public prior art, immediately (BioBricks Foundation, 2017a).

Fourth, there is gain in transparency on the patent protection of parts and technologies. Cambia's TheLens is a free online resource covering patent information from major patent offices. Tutorials on intellectual property rights and information on patent policies and practices reduce researchers' efforts to run

research on potential infringements and on the further re-usage of their innovations (Dennis, 2004; Cambia, 2009). The current state of the second enabler, developing open science parts, will be analyzed in this study, quantitatively.

5.3 The Basic Local Alignment Search Tool (BLAST)

The basic local alignment search tool (BLAST) is one of the most popular sequence analysis tools available in the public domain. There are multiple programs in the BLAST suite. All programs have a word-based search procedure in common. The algorithm finds short matches between two sequences and then tries to extend from these ”hot spots” to capture full alignment regions. The output table supplies details on aligned sequence fragments and multiple measures such as conformity, measuring equality of sequences, and coverage, i.e., which percentage of the query sequence was found in the subject one. In addition, BLAST provides statistical information to evaluate the significance of results summarized in the so-called ”expect” value, i.e., the false-positive rate. Speed and sensitivity of programs vary mostly due to the word-size when it is about the search for hot spots and the way of extending (McGinnis et al., 2004). MegaBLAST and BLASTN are the important programs for nucleotide-nucleotide searches. MegaBLAST is the fastest program with a large word-size and a greedy gapped extension algorithm. It is used mostly for similar sequences, e.g., from the same organism. BLASTN uses an 11-base contiguous word to initiate extensions and, thus, is more sensitive and slower. It is recommended to adapt the choice of program to specific search sequences and to use default parameters for

searches (McGinnis et al., 2004).

5.4 Prior Research on Open Science in Synthetic Biology

Several scholars have analyzed the impact of patent protection on the diffusion of biological parts in follow-on research (Furman et al., 2011; Murray et al., 2007; Williams, 2013; Murray et al., 2016). A comprehensive review is given in chapter 2. Common to all prior findings is the limiting character of patent protection as to the re-usage of technologies and parts, both in case of initial as well as delayed protection. In addition, prior research has used citations to recognize the re-usage of basic technologies.

Central to the BioBrick engineering approach are basic parts which are built to be re-used in follow-on research. Therefore, leading scholars in this field favor open science for basic parts. When running a survey in 2013, Kahl (2015) first reviewed the current re-usage of basic parts from open science in the field of synthetic biology. They worked out perspectives of the community to analyze the current procedure of re-usage and to identify limitations. While the majority of researchers claimed not to re-use parts from open databases, they maintained the low quality of these parts to be a reason for their behavior. Discussion with experts from the field support these perspectives, e.g., Claes Gustafsson (DNA2.0), Andrew Hessel (PinkArmyCooperative) and Camille Delebecque (SynBio Consultancy).

Summarizing prior research, scholars measured the limitation of property rights

on follow-on research in biotechnology using citation data and running qualitative surveys on the current impact of open science in the field of synthetic biology. There has not yet been a study accessing this impact with the help of quantitative methods and analyzing its success factors. To address these gaps, this study applies a new approach measuring the re-usage of DNA sequences by matching query sequences in patent applications.

With this methodology two hypotheses are to be verified:

Hypothesis 1: Open science parts are little diffused into patent applications

Hypothesis 2: A higher level of quality will increase the popularity of open science

5.5 Data and Methods

5.5.1 Build Databases

Three databases are built to verify the hypotheses. Two of the three databases, a query list and a subject collection, are collections of DNA sequences. The query list is equipped with 478 DNA sequences of BIOFAB parts, both, newly created (428) and only with improved characterization (50). BIOFAB parts were released in 2010 on the Addgene repository and they are mostly short. Whereas promoters have 36-49 base pairs, terminators have 11-220 base pairs. They are of high quality and were published by SynBERC.

The query search is performed against 104,724,452 electronically available nucleotide sequences in patent applications in the years 2001-2015, which were published on patentlens.org (PatSeq DB Release 201611). Patent data is used for this study for three reasons. First, the majority of innovations in synthetic

biology as well as in BioBrick engineering, is secured by patent applications. Both large companies use patents following their business models and small companies need to apply for patents in order to attract investors (Calvert, 2012). Second, patent offices request a list of used sequences for patent applications. In the patent examination, applications are reviewed and the supplementary data is taken into consideration (Cook-Deegan et al., 2010). Third, the open science database TheLens holds all electronically available DNA sequences of major patent offices (Cambia, 2009). Analyzing the re-usage of BioBricks is, therefore, best performed by matching these DNA sequences with the supplementary data of patent applications.

Patent applications are published after 18 months (Henkel et al., 2007), thus, earliest effect of BIOFAB project can be expected in 2011. The third database covers meta data of patents and the full text of 500 matched patents which have been collected from Google patents and merged with the subject collection.

5.5.2 BLAST Search and Validation Steps

As to the finding of matches of BIOFAB sequences in patent applications, three main steps are performed. A programmed search (1) of DNA sequences in a large database with automated filters is followed by a manual review (2) and a personal request to patent inventors for feedback (3). In the first step, BLASTN is used to find matching DNA fragments with a filter on the expected value of 1e-6 as default setting. BLASTN was mostly used, whereas additional checks with BLASTX to scan for matches in proteins did not extend the result list.

MegaBLAST algorithm was neither used to ensure high quality in search results considering the short length of BIOFAB parts. The full code can be found in the appendix.

143,729 matches can be found in the full application database for all BIOFAB parts. The result list is filtered with four rules.

1. Delete duplicates
2. Clean embedded overlaps
3. Apply threshold for conformity
4. Apply threshold for coverage

After deleting duplicate matches, embedded regions are cleaned. Embedded regions can occur due to two reasons. First, because one BIOFAB part is a subsequence of another one. Second, because BlAST finds a match for a small region of a sequence and a second match for a larger one. Both would bias the count of matches and are therefore cleaned by keeping the one result with a higher matching score.

The third rule and the fourth one are associated to thresholds developed iteratively together with experts from BBF and SynBERC. Matches below conformity thresholds are cleaned just as matches below coverage thresholds are. While the conformity threshold is fixed as to all sequences, coverage thresholds are adjusted to sequence lengths. In this study, conformity is limited to a minimum of 80%. Coverage thresholds decrease in 5 ranges adjusting to the length of query sequence in base pairs:

1. Sequence length maximum 50 base pairs, coverage threshold is set to 92%
2. Sequence length maximum 150 base pairs, coverage threshold is set to 90%
3. Sequence length maximum 500 base pairs, coverage threshold is set to 80%
4. Sequence length maximum 1000 base pairs, coverage threshold is set to 70%
5. Sequence length above 1000 base pairs, coverage threshold is set to 60%

After these filters, 2,051 matches for promoters and 5,303 matches for terminators in USPTO applications are found. As promoters and terminators from BIOFAB database have short sequence lengths and as some are highly similar to natural DNA fragments, it is quite likely that sequences are not used on purpose or that they are part of longer sequences. 511 patent matches have been selected for manual review to observe trends for priorly used parts or to validate the re-usage of new parts. Here, the procedures of the two analyses differ. Trends for priorly existing parts are reviewed with sample checks due to resource limitations. In order to validate the re-usage of newly created parts, all matches in patents are manually reviewed. The full text of patents is checked by means of three questions:

1. Did applicants use the BIOFAB part with the specified function?
2. Did applicants use a BIOFAB part on purpose?
3. Did applicants refer to the BIOFAB database?

First, full texts of patents are tested for the description of the matched sequence and of the function explained. As BIOFAB promoters and terminators are mostly

part of a genetic construct, the associated genes and the purpose of using the untranslated area are reviewed. E.g. BIOFAB promoter apFAB600 is a constitutive version of a naturally occurring inducible ptrc promoter. In this case, a review was needed to validate the function of the sequence. Second, hints are searched in texts of patents for proving whether authors used the sequences on purpose, i.e., naming the promoter or terminator explicitly in their description of the genetic construct. Third, there is a check on any reference towards BIOFAB or towards the registry of biological parts as source of the BioBricks.
In the third step, patent inventors who re-used newly created BIOFAB sequences are approached personally to provide feedback to the re-usage and experiences with the parts.

5.6 Results

The BIOFAB collection covers 478 parts, of which 428 have been newly created and 50 priorly used. To monitor the diffusion of parts specifically contributed by the BIOFAB project, 428 newly created parts are searched in patent applications. 50 priorly used sequences are investigated for the effect of improved quality of parts. Table A.3 in appendix lists all BIOFAB parts with the name and the information whether or not the parts were newly created based, both, on prior appearance in the patent database as well as on feedback from the BIOFAB team.
Figure 5.1 summarizes yearly re-usages of new created BIOFAB parts in patent applications by type. The first usage of the parts is identified in 2013, when all three types were in total 4 times used in patent applications. In the year

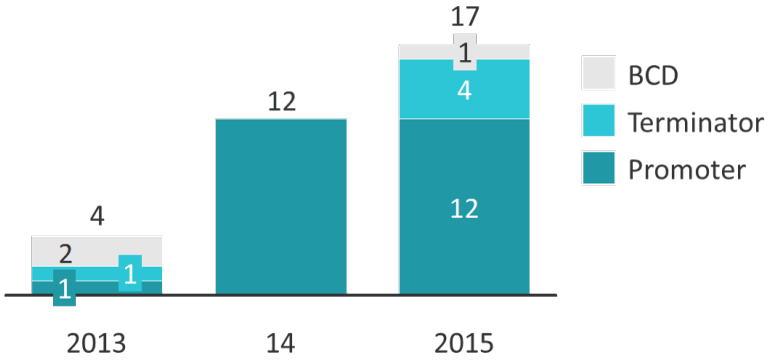


Figure 5.1: Re-usage of BIOFAB sequences in patent applications

2015, 17 re-usages of BIOFAB sequences could be identified. While terminators and complete BCD elements were seldom implemented, promoters were more frequently used in patent applications. Validation for the diffusion of BIOFAB

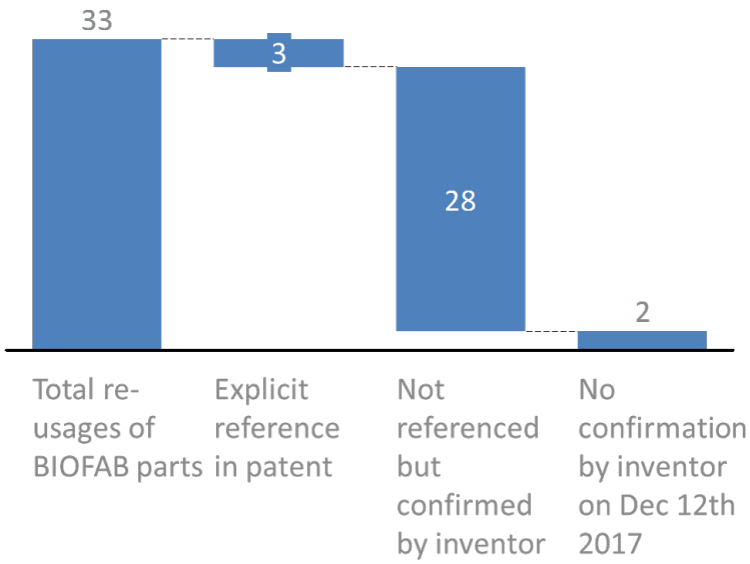


Figure 5.2: High ratio of missing references to BIOFAB sequences

parts is elaborated by explicit references in patent texts, by social distance, i.e., whether inventors, who are listed in the patent, are associated with the BIOFAB project, and by personal confirmation of inventors for re-usage of BIOFAB parts. As to the social distance to BIOFAB members, co-authorships in patents and

publications are considered. Inventors who priorly co-authored scientific work with members from BIOFAB are categorized as associated with the project. All inventors are then personally approached to confirm their re-usage of BIOFAB parts. Figure 5.2 highlights that 90% of re-usages were not referenced but per-

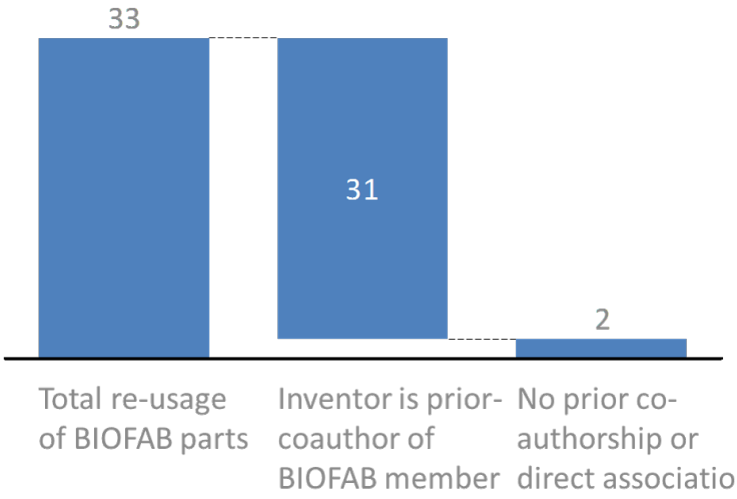


Figure 5.3: Majority of inventors re-using BIOFAB parts are connected to the project

sonally confirmed by inventors. 3 of 33 re-usages were explicitly referenced in patent text, 28 re-usages could be confirmed personally by the inventors, 2 confirmations are outstanding by December 12th 2017. Eleven times more references could be identified compared with standard methodology counting references in patent text.

To analyze the width of diffusion, social distances of inventors to BIOFAB members are explored. Figure 5.3 highlights the dominance of inventors closely connected to the BIOFAB project. 2 inventors who are not connected to the BIOFAB project have not confirmed their re-usage yet. An important feedback was provided from DNA2.0. Even though 85% of new created promoters were used

officially only by DNA2.0, the company implemented BIOFAB parts into their sequence design program which would have enabled other researchers to have access to the sequences.

While testing the second hypothesis, an assumed impact of the quality of parts on the diffusion is measured before and after the release of the BIOFAB parts. In this test, only these parts are reviewed which have been used in patent applications before being published in the BIOFAB release.

As to the promoters, the characterization by the BIOFAB team shows no impact

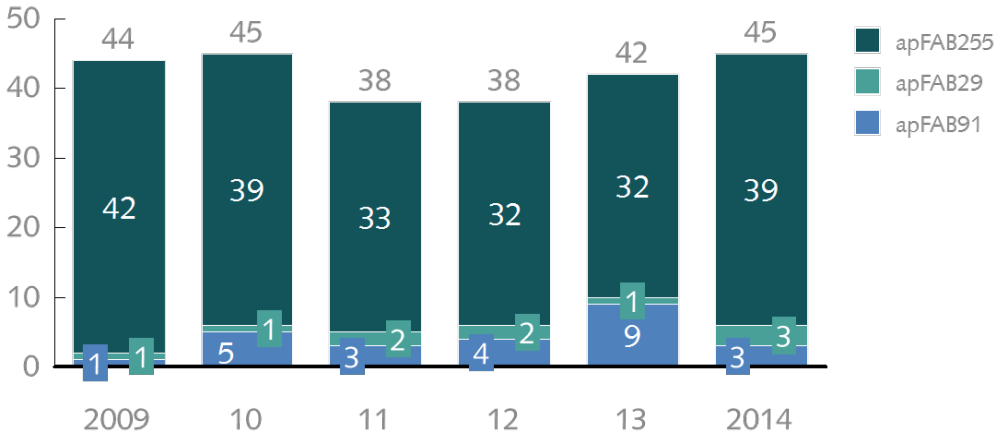


Figure 5.4: Timeline for the re-usages of characterized promoters

on diffusion, visualized in figure 5.4. None of the three characterized promoters, which had been used before the release in 2010, shows any significant change in diffusion after the publication.

Taking the terminators into consideration, figures 5.5 and 5.6 visualize the diffusion of the 10 most often used terminators characterized by BIOFAB.

The 5 mostly used parts in the first illustration show no effect as to diffusion rates. This observation is confirmed by a detailed analysis of the most popular

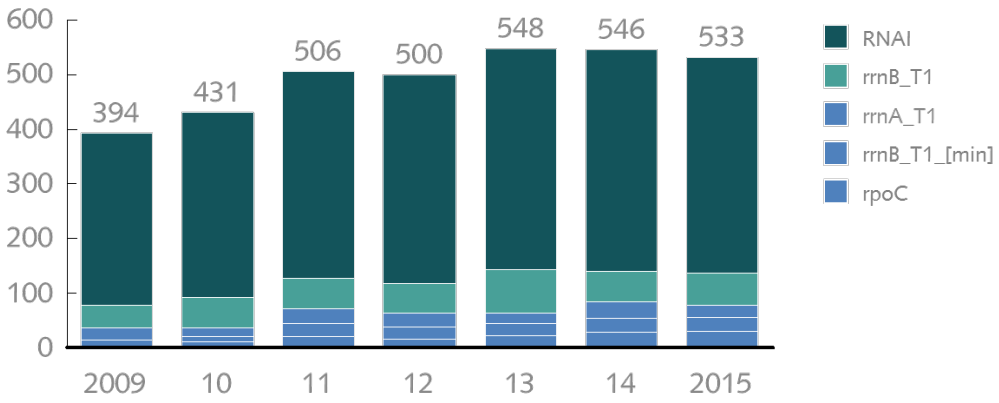


Figure 5.5: BIOFAB Terminators Characterized Top 5

terminator called ”RNAI” making up 84.30% of all re-usages. 98.26% of these re-usages adopt a version of ”RNAI” with one codon change in comparison to the BIOFAB release, i.e., one base pair is changed while codon is still coding for leucine. In addition, there has been no increase in the re-usage of the ”RNAI” terminator released by BIOFAB after 2011.

In the second graph, the diffusion of the double terminator called ”EOU_double_term”

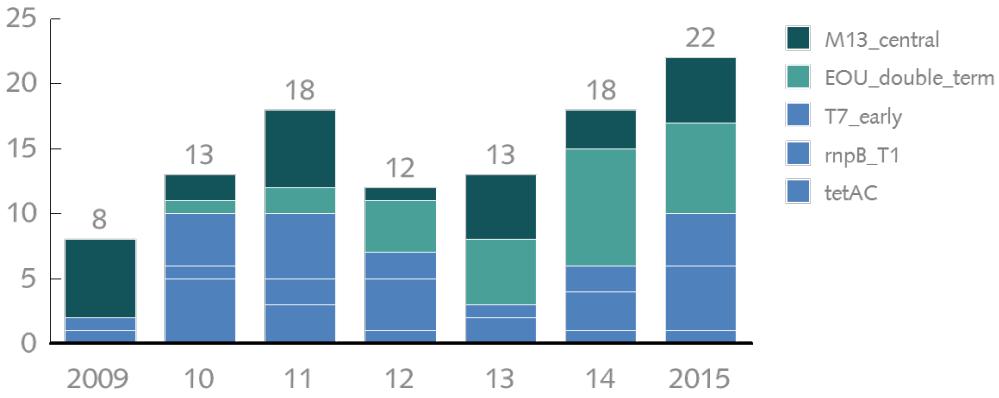


Figure 5.6: BIOFAB Terminators Characterized Top 5-10

shows an increase in 2013 and in 2014. The full text analysis of patents reveals that applicants refer to BioBrick Bba_B0015 published on the registry of parts in 2003 but not to BIOFAB. The national background of an applicant has no effect

on the results.

5.7 Discussion

This study was motivated by two gaps in research. First, qualitative research exists on the impact of open science in early stage innovations in the field of synthetic biology, but quantitative research applied to validate these hypotheses was lacking. Second, scholars asked the community for reasons of the assumption of quite a low diffusion of open science but did not verify success factors to increase the current impact of open science (Kahl, 2015). These gaps can be filled by introducing the new methodology of measuring the re-usage of basic biological parts in patent collections. Conclusions from this study are of great importance for practice, thus, potential solutions to identified issues will be discussed.

5.7.1 Implications for Research

Prior research made use of citation data and of the voluntary records of gene tests to measure the diffusion of early stage technologies in follow-on research (Murray et al., 2007; Williams, 2013; Murray et al., 2016). This study elaborates two advances. First, a new methodology to measure the re-usage of biological parts by matching DNA sequences. Second, an analysis of the current impact of open science in the field of synthetic biology. This study takes advantage of an alternative approach measuring re-usage by matching DNA sequences with

patent data. Using the case of BIOFAB parts, eleven times more re-usages can be identified. Scholars can adopt this methodology to assess the diffusion of other query lists or to validate prior research.

Second, current impact of open science in the field of synthetic biology and BioBrick engineering, in particular, has priorly been assessed via one qualitative study (Kahl, 2015). The results of this survey have now been confirmed by quantitatively analyzing the diffusion of open science parts in follow-on research. Personal feedback of inventors helped to confirm the re-usage and provided hints on additional re-usages which are not captured in this analysis.

5.7.2 Implications for Practice

This study has two major implications for practice. First, it helps to measure the impact of open science initiatives using the case of BIOFAB parts. Second, the effect of quality improvements is assessed.

It is only the consideration of references to BIOFAB in patents that leads to the assumption that their diffusion is very low, e.g., just 3 references of BIOFAB sequences are made in patents. When extending the search by matching DNA sequences with patent application collections, the number of re-usages increases by far, thus eleven times more re-usages can be validated. The measured diffusion helps initiatives to incentivize for further engagements and the feedback exchange with inventors provided the BIOFAB team with appreciation for their work and input on optimization potential.

The effect through quality improvements is assessed by observing the re-usages

of parts before and after the release of characterization by the BIOFAB team. As no positive effect through BIOFAB publication could be validated, even though survey results and experts pointed out to quality as a success factor, the community needs to reflect on potential reasons and to find solutions in case of further engagements.

A possible way to explain why researchers did not adapt the parts of higher characterization quality can be the influence model (Asavathiratham et al., 2001). This model structures reasons to adopt an innovation in four dimensions. First, community must have some knowledge about the parts and about how to re-use them. Considering this probable reason, communication of the BIOFAB release might not have reached a certain size of researchers or instructions have been insufficient. Second, the re-usage of BIOFAB parts might not have brought in enough value-add for the community. In this case, potential solutions may be that people have to reflect upon which parts are really needed and which details and which quality are best suited for adoption. Third, the prior adoption of parts by role models can persuade the community to re-use parts for their re-usage. And fourth, the established systems and routines need to allow the re-usage of parts, e.g., alignment with job incentives, partnerships with popular suppliers or an interactive framework enabling the community to take part in the process of the selection, characterization and optimization.

The recent release of the OpenMTA has considered these four dimensions. People have knowledge about the OpenMTA due to the presentation at the SB7.0. Addgene identified highly requested parts to increase the value-add for the community. Leading researchers are ambassadors for the project and, thus, become

role models, and the BBF announced a partnership with Twist, a popular synthesizer of genetic sequences, to embed the OpenMTA within existing processes (Forsyth, 2017). The success of the OpenMTA should be monitored and the importance of enumerated levers should be evaluated.

5.8 Limitations

Limitations can be seen in three areas. First, the consideration of patent data itself, second, inaccuracy in matching process and third, limitations in query data quantity.

As innovative follow-on research is expected to be secured with patent application in the majority of cases and as this study tries to assess the impact of open science parts in follow-on research, a certain part of re-usages will not be considered in this data. In addition, the assumed higher quality of BIOFAB parts is not backed by interrogations with the community. A questionnaire confirming a higher quality of BIOFAB parts would encourage a feedback.

The matching process is divided into an automated and a manual step. While thresholds set in the automated step are developed iteratively and, thus, represent the global optimum in the cost-benefit balance of the analysis, they can still lead to false-negative results. In the manual review, characteristic phrases in patent texts are used to validate re-usage. Prior training how to read patents given by experts and their sample checks increase the significance of results, but manual review can still be subject to inaccuracy.

The query data considers 478 BIOFAB parts and 84.3% of matches in patent appli-

cations are driven by the "RNAI" terminator. The two hypotheses are therefore only analyzed using time-series data. A larger quantity of query parts with a higher rate of diffusion is needed to enable sophisticated methods and, thus, to establish statistically validated hypotheses.

Chapter 6

Study III: Impact of Commercial vs Open Science Research Tools on Knowledge Diffusion

6.1 Introduction

The impact of property rights on the scientific progress has motivated many scholars to find factors which can improve the exploitation of innovations and increase their social value (,e.g., Merton, 1973; Scotchmer, 2004; Williams, 2013; Murray et al., 2016). Property rights were created to establish a market of ideas and to incentivize the exploitation of innovations (Nelson, 1959; Arrow, 1962; Merges et al., 1990; Merges et al., 1994; Fosfuri et al., 2001; Gans et al., 2000). Early research focused on isolated innovations to analyze the costs and benefits of property rights (Merton, 1973; Thursby et al., 2007; Scotchmer, 2004; Henderson et al., 1995; Trajtenberg et al., 1997; Czarnitzki et al., 2009). Empirical research found both positive and negative effects of property rights on isolated innovations, such as an increased outcome quantity at a lower quality (Fabrizio et al., 2008; Czarnitzki et al., 2009). In this setting, patents were seen as by-products of research to calculate ROI for scientific projects and as help to mone-

tize developments (Murray, 2006).

Recent research has investigated the effect of property rights on cumulative innovations. These include both multipurpose and sequential innovation and show special dynamics in terms of property right protection. In context of sequential innovation in biotechnology, research tools are subject to several studies as they mean the input into the development process of new products (Heller et al., 1998; Walsh et al., 2003; Pénin et al., 2008). Quantitative studies found a higher diffusion of open science research tools in follow-on research than of proprietary tools and an increased efficiency in their application (Furman et al., 2011; Murray et al., 2007; Williams, 2013; Murray et al., 2016). These effects are explained by the ease of access, by a higher intensity of the collaborative development and by a larger pool of interacting researchers having specialized knowledge.

Little focus was laid on the impact of protected research tools on the type of follow-on research. Murray et al. (2016) were first to investigate these effects. They analyzed the type of follow-on research which uses open science research tools and found an increase in creativity and diversity of follow-on research. A different stream of researchers analyzed predictors on knowledge diffusion evaluating the impact of sharing full data (Piwowar et al., 2007) and of publishing in open accessible journals (Gaule et al., 2011). The effect of selecting from competing open science and commercial research tools on the knowledge diffusion of published articles is not yet understood.

This study deals with this limitation using the case of zinc finger nucleases (ZFNs). ZFNs were chosen because of three reasons. First, ZFNs were a leading research tool for gene-editing DNA sequences from 2004 to 2012 and many

innovative advancements relied on using ZFNs (Chandrasekharan et al., 2009). Second, as the company Sangamo holds patents on the core research tool and did not disclose all relevant innovations, access was only possible through buying expensive kits or through agreeing to license agreements (Chandrasekharan et al., 2009). Third, a consortium of academics released an open science alternative. They published protocols, reagents and multiple software tools to enable researchers to use the research tool without license agreements or to buy commercial kits from the manufacturer Sigma (Chandrasekharan et al., 2009).

The prediction is stated that using an open science research tool has a positive effect on the knowledge diffusion of a published paper. This is tested using 396 peer reviewed research articles, corresponding article meta data and authors' prior co-authorships. The usage of open science and of commercial research tools is categorized on the basis of keyword search and manual review. After controlling for article, author and for affiliation variables, published studies using the commercial research tool are predicted to have lower citation rates, which validates the prediction. This result provides several advances contributing to research and practice.

The remainder of this study consists of six sections. Section 2 reviews prior research on property rights in innovation theory and section 3 provides an overview of zinc finger and of complications for scientists due to patent protection. In section 4 data and methods are introduced. Section 5,6 and 7 present results, implications for research and practice and limitations of the study.

6.2 Theoretical Background

Property rights and their influence on the diffusion of technology have a long history in research. Property rights, such as patents, have been created to establish a market of ideas (Nelson, 1959; Arrow, 1962; Merges et al., 1990; Merges et al., 1994; Fosfuri et al., 2001; Gans et al., 2000). They enable the commercialization of an idea (Scotchmer, 2004; Thursby et al., 2007; Merton, 1973; Kitch, 1977; Thursby et al., 2001; Hellman, 2007) and are seen as by-products of scientific work (Murray, 2006).

The empirical research validates positive effects on the quantity of publications concerning isolated innovations (Fabrizio et al., 2008; Looy et al., 2003; Breschi et al., 2007; Czarnitzki et al., 2007; Stephan et al., 2007) and analyzes which breadth and length of patents minimize the cost of monopoly distortion (Gilbert et al., 1990; Klemperer, 1990). Negative effects are identified with a reduced quality of research output due to incentives created by property right (Henderson et al., 1995; Trajtenberg et al., 1997; Czarnitzki et al., 2009). Innovation occurs through the interaction of multiple actors (Freeman, 1989; Freeman, 1994; Lundvall, 1992; Nelson, 1993; Nelson et al., 1993; Mansfield et al., 1996; Mansfield, 1995; Mowery et al., 1999; Dosi, 2000). Therefore, recent research has focused on cumulative innovations and on potential hindering by early-stage property rights.

Potentially limiting factors in cumulative innovations were highlighted by several scholars (Argyres et al., 1998; Krimsky, 2004; David, 2001). The following hypothesis "tragedy of anti-commons" (Heller et al., 1998) was postulated. It states that intellectual property rights might reduce the flow of information and

inhibit cumulative research (Heller et al., 1998; David, 2000; Lessig, 2002; David, 2004). These effects have been predicted by multiple scholars, in particular, for research tools in biotechnology as they are inputs into the process of developing new products (Heller et al., 1998; Walsh et al., 2003; Pénin et al., 2008). Nevertheless, scholars have expressed consensus that negative aspects of property rights are not caused by the established system, but by a licensor's behavior (Walsh et al., 2003; Murray, 2006; Murray et al., 2007).

Prior quantitative research on the diffusion of open science can be clustered around three streams: (1) the influence of property rights on the diffusion of research tools into follow-on research, (2) the effect of open science on the type of follow-on research and (3) essential predictors for citation rates of research articles.

The effect of property rights and accessibility on follow-on research has been quantitatively analyzed in three studies. Furman et al. (2011) validated a substantial amplification in knowledge production through direct accessibility to research tools. Williams (2013) investigated the effect of initial contract-based property rights on the re-usage of sequenced human genome issued by Celera. She measured a 20-30% decrease of follow-on research and validated a lasting delay of follow-on research, if research tools were initially protected. Similarly, Murray et al. (2016) measured a reduction of 20-40% of follow-on research analyzing initially patent protected genetically engineered mice. In their conclusion, Murray et al. (2016) considered the possibility that researchers use alternative open science research tools to avoid patent protection without further analyzing this hypothesis.

The second stream has recently been initiated by Murray et al. (2016) who investigated the effect of open science research tools on the type of follow-on research. They validated an increase of creativity and diversity of follow-on research, i.e., more exploratory topics and higher variety of researchers, after open access to the research tool had been granted. At the same time, public ownership has not reduced incentives to run early stage research.

The third stream of prior research analyzes factors which potentially influence the knowledge diffusion of a research article (Falagas et al., 2013). Researchers commonly aim to publish articles that will attract colleagues to build on their work (Cole, 2000; Piwowar et al., 2007; Bornmann et al., 2008). They use the systematic citation process to acknowledge this re-usage and, thus, the way how their scientific work builds on upward innovation and knowledge (Murray et al., 2007). In addition to mapping the information flow and to acknowledging prior work, citation counts are regarded as an indicator of scientific impact which, in turn, may be associated with career development. Frequently, research funding and promotion decisions are legitimated with citation counts and, finally, a salary-increase dollar value has been assigned (Diamond Jr, 1986).

Falagas et al. (2013) generated an overview of studies and clustered the list by the variable which was subject to the study. Several variables concerning article characteristics, team setup, affiliation reputation, journal impact, and access to publication as well as to complete result data were analyzed (Habibzadeh et al., 2010; Jacques et al., 2010; Lokker et al., 2008; Falagas et al., 2013; Petersen et al., 2014; Laurance et al., 2013; Eysenbach, 2006; Davis et al., 2008; Gaule et al., 2011). While publishing in open accessible journals is not correlated with citation rates

(Eysenbach, 2006; Davis et al., 2008; Gaule et al., 2011), Piwowar et al. (2007) validated a positive effect on citation rates, when researchers share detailed result data, because follow-on research can access all needed information and can build on it directly.

In summary, prior research on cumulative innovation has explored the impact of both contract and patent-based protection on research tools regarding their quantitative and qualitative diffusion in follow-on research. They found evidence that open science research tools lead to an increased activity of follow-on research and to a higher diversity and creativity. Prior studies focused on the proprietary character of research tools as limiting factor in scientific progress. However, commercial research tools and, in particular, the competition with open science research tools have not been subject to analysis. In this case, the selection from competing open science and commercial research tools can have an important effect on knowledge diffusion. Similarly to sharing detailed result data and, thus, to giving access to all needed information for follow-on research (Piwowar et al., 2007), using open science research tools might as well ease the re-usage of scientific results and the increase of citation rates. This study therefore states the following hypothesis:

Prediction: The selection of an open science research tool against a competing commercial one has a positive impact on the diffusion of research results.

6.3 The Case of Zinc Finger Nucleases

Zinc fingers bind proteins to specific regions of the DNA to turn on and off genes. They were discovered by Aaron Klug's group at the Laboratory of Molecular Biology in Cambridge, U.K., in 1987 (Klug et al., 1987). Kim et al. (1996) were first to report that they could attach zinc fingers to DNA-snipping enzymes and cut free floating DNA at specific locations. Bibikova et al. (2002) and Porteus et al. (2003) succeeded in using the research tool in drosophila and in human cells and, thus, showed the potential to run genetic engineering at specific locations in any organism. Customized engineered Zinc Finger Proteins (ZFP) which are fused to a FokI cleaving protein, are called Zinc Finger Nucleases (ZFNs). The research tool promised early on to have a high potential being and molecular biologist Matthew Porteus of the University of Texas Southwestern Medical Center in Dallas foresaw that it is "a phenomenal research tool", which "could change the way we do science" (Kaiser, 2005).

For preparing a cleavage at a specific location, researchers can use three procedures. First, they can modularly assemble multiple proteins to a zinc finger array (ZFA) (Fu et al., 2013). This technique of building ZFNs is the fastest and easiest one though of low reliability (Ramirez et al., 2008), because ZFPs are not modular and depend, both, on context and neighbors. Rules for combination and optimization processes are needed to increase the effectiveness of labor and the efficiency of ZFNs. Second, matching ZFPs can be identified via selecting methods. In this case, ZFPs are selected which bound to a given DNA sequence from a pool of randomly engineered ZFAs. This process is more reliable, but

needs to be performed stepwise, which takes about 1 month of labor, and a high level of expertise. Third, researchers can re-use verified ZFNs of other researchers which is the most efficient method (Fu et al., 2013).

After the initial research of Bibikova and Porteus, ZFNs were developed by the firm called Sangamo and, to some degree, separately by a group of academics. Sangamo, found in 1995, identified the potential of ZFNs and transformed the initially fragmented patent pool into a patent thicket by purchasing and licensing intellectual property rights (IPR). The company acquired Gendaq, which was founded by Aaron Klug in 2001, and assured exclusive license agreements with Matthew Porteus in 2003 and with Dana Carroll in 2004 to develop an "alphabet" of zinc fingers (Wade, 2009; Scott, 2005). In addition, they licensed patents from MIT, Scripps, Harvard and John Hopkins. The company combined intellectual property of foundational procedures for the ZFP field.

Chandrasekharan et al. (2009) analyzed the intellectual property landscape and found little possibility of workarounds. In particular, three patents, which cover rules to modularly combine ZFPs and to optimize binding quality, on the dominant modular assembly are in focus due to two reasons. First, the successful generation of efficient ZFNs using the modular approach needs specific protocols and information on rules which are developed by Sangamo. Second, even though academics reported to use proprietary materials without considering license agreements (Walsh et al., 2007), these patents did not completely disclose all needed information and researchers could not use the proprietary research tool. For academic researchers the limitation was, therefore, not due to patent protection, but due to the lack of access (Chandrasekharan et al., 2009).

With the aim to use the research tool, researchers could buy a set of ZFNs or agree on licensing the proprietary platform. The licensed manufacturer called Sigma Aldrich Inc. charged researchers 25,000 USD for a set of ZFNs (Chandrasekharan et al., 2009). Multiple researchers claimed that Sangamo was running a selective strategy for collaboration partners (Chandrasekharan et al., 2009). They would accept licensing agreements with partners who are not active on Sangamo's market of pharmaceutical therapeutics (Kaiser, 2008). The most prominent example is the startup called Phytodyne which failed due to rejected license agreements with Sangamo (Scott, 2005). "The Sangamo patent was essential, and price was a big factor" said Dan Voytas, co-founder of Phytodyne (Scott, 2005, p. 917). Sangamo is "inhibit[ing] the technology from proliferating" claimed Carlos Barbas from the Scripps Institute (Kaiser, 2005).

Sangamo had to invest heavily into consolidating the IPR and into developing the research tool without having market ready applications yet (Pearson, 2008). They potentially used the strategy of secrecy around their platform to attract private capital (Mann et al., 2007). In addition, the small company was reliant on income through license agreements (Chandrasekharan et al., 2009). The patent portfolio was, therefore, important for the success of the company (Scott, 2005; Kaiser, 2005).

The inhibit of innovation can be assessed by searching re-usages of ZFP sequences which are defined in Chandrasekharan et al. (2009) and in patent applications with PatSeqFinder (Cambia, 2009). While ZFPs which are not claimed in granted patents, are re-used 205 times in further patent applications, claimed ones are not re-used any longer. The observed blocking character of claiming

sequences supports the concern of Chandrasekharan et al. (2009) that the full exploitation of the research tool was inhibited through Sangamo's strategy. In order to counter this development, a group of academics formed the Zinc Finger Consortium (ZFC) to enable all academic researchers to use ZFNs (Fu et al., 2013; Joung et al., 2015). They published protocols, customized proteins as well as a database consolidating multiple collections of, both, ZFPs and ZFAs, and software tools for the design and for optimization (Fu et al., 2013).

The development of wet lab protocols can be divided into three phases. In 2005, the consortium published an article with instructions on the modular assembly of ZFNs (Wright et al., 2005). The fast and easy process was not adequately reliable yielding less than 50% of binding (Ramirez et al., 2008) and was, therefore, not comparable to Sangamo's commercial research tool.

In 2008, the oligomerized pool engineering (OPEN) methodology was published (Maeder et al., 2008). This method of selection yielded higher reliability comparable to proprietary products of Sangamo, but it took about 1 month to build a ZFN and only experts could perform the protocol (Fu et al., 2013).

In 2011, the consortium finally published the software driven context-dependent assembly (CoDA) combining the advantages of modular assembly and of selection methods. Using pre-verified ZFPs, the efficiency was less than of the OPEN methodology (Moore et al., 2012), however, it was practicable for most researchers (Fu et al., 2013).

Intending to make access for researchers easier and to combine resources, the Zinc Finger Consortium published a collection of ZFAs and reagents for modular assembly and the OPEN methodology on the non-profit repository called

Addgene in 2008 (Zinc Finger Consortium, 2016; Herscovitch et al., 2012). In addition, the consortium created the ZifDB database with several collections including proprietary ZFAs of Sangamo (Fu et al., 2009). Open accessible software tools such as Zifit, first published in 2007 (Sander et al., 2007), support the identification of potential ZFPs and even the design and engineering process. Joung, a co-founder of the consortium, rates their efforts as "game changer. It gives academics the ability to make these proteins without going to Sangamo" (Kaiser, 2008).

Both, the collection on Addgene and the ZifDB include hints about potential infringements of Sangamo's proprietary rights (Chandrasekharan et al., 2009). Sangamo has not yet opened any lawsuit, as they see a higher distribution of zinc finger through academia due to the co-existence of an open science alternative (Pearson, 2008). The increased use of the research tool would lead to more downstream applications. For the commercialization of downstream inventions, inventors will mostly require rights to use commercial research tool of Sangamo (Chandrasekharan et al., 2009).

6.4 Data and Methods

6.4.1 Sample Construction

With the intention to test the prediction, a dataset consisting of 396 peer reviewed research articles, article metadata and of authors' prior co-authorships are constructed in six steps.

First, all articles related to Zinc Finger are downloaded from Scopus using the

keyword search ”Zinc Finger Protein”, ”ZFP”, ”Zinc Finger”, ”Zinc Finger Binding Protein”, ”ZFN” and ”Zinc Finger Nuclease” on title, abstract and article keywords based on Chandrasekharan et al. (2009). Metadata of 23,790 articles including bibliographic and descriptive materials were downloaded in October 2016 to identify the usage of research tools and to build control variables for an analysis. Scopus is preferred over WebofScience due to a larger dataset of articles on synthetic biology as the one in June 2016 (Scopus: 55,842 articles, WebofScience: 19,713, Timeframe 2001-2016).

While ZFNs are mostly used to perform genetic engineering, the term ”ZFP” often occurs in articles whose authors analyze different combinations of proteins and their functions and not genetically engineer organisms. Steps two to four are therefore performed to build a database of articles using ZFNs to genetically engineer organisms.

Second, articles not naming ZFN explicitly, but still performing genetic engineering, are selected based on elaborated keywords. In an iterative procedure, a sample of articles, which genetically engineer organisms using ZFN, is selected, keywords are inspected and a larger corpus of articles is identified by these keywords and is reviewed later. The resulting list of keywords is as follows: ”protein-protein interactions”,”randomized zinc finger”, ”Synthetic zinc finger”, ”cleavage domain”, ”designed zinc finger”, ”engineered zinc finger”, ”Fok I”, ”FokI”, ”DNA cleavage”, ”DNA modification”, ”DNA strand breakage”, ”gene editing”, ”gene targeting”, ”genetic engineering” and ”Targeted genomic rearrangements”. This step results in 2,950 articles.

Third, 505 articles, which review technologies and do not perform follow-on

research, are deselected from the analysis. Fourth, abstracts are reviewed manually to unselect 1,651 remaining articles not using Zinc Finger research tools for genetic engineering, e.g., research articles which only describe the research tools.

For the selected 794 articles, 396 full texts could be downloaded from Scopus. The selection of articles in steps one to four are visualized in figure 6.1.

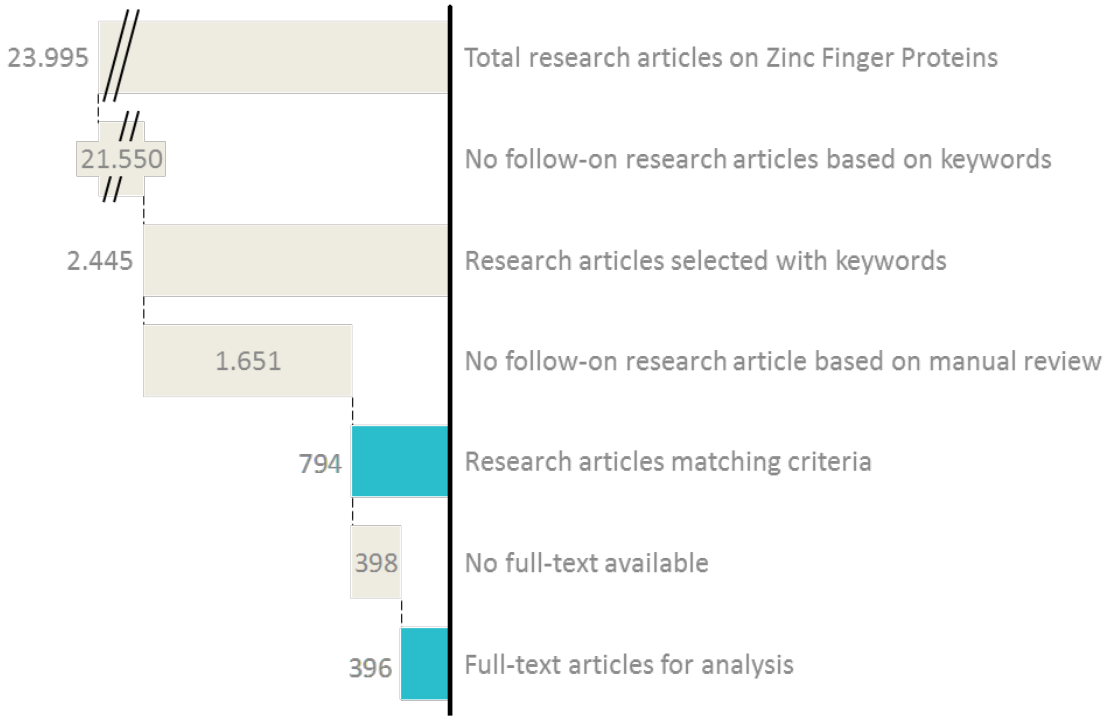


Figure 6.1: Selected articles which use Zinc Finger research tools for genetic engineering

In the fifth step, material and method sections of full texts are screened to identify the selected research tool, first with keywords, then, by manual review. Articles using the commercial research tool are selected by means of ”Sigma” or ”compoZr”. The open science research tool is linked to ”addgene”, ”zf kit”, ”finger consortium”, ”zf consortium”, ”zf-models consortium”, ”Consortium Mod-

ular Assembly", "ZiFit". The third category of research articles is named "Self-development" as teams use self-developed ZFNs or do not provide explicit information. In total, 44 research papers note the use of the open science research tool, 94 explicitly use the commercial research tool and 258 are categorized as self-development.

In summary form, all citations of the papers in scope by category result in 1,661, 3,004 and 14,484 total citations of researchers using open science, commercial research tool or self-developments. Summary statistics on article citations per category are given in table 6.1. Citations per category are not dominated by single publications. The box-plot A.1 in the appendix confirms this observation on author-level. In a final step, author affiliation and co-authorships are extracted from Scopus.

Table 6.1: Citations of research articles using ZFN categorized by research tool

Variable	Obs	Mean	Std. Dev.
Citations Open Science	44	37.750	42.656
Citations Commercial	94	31.958	47.806
Citations Self-Development	258	56.140	98.447

6.4.2 Measurement

DV: Citations

This study investigates how the selection of open science research tools or commercial ones influence the diffusion of published studies. Researchers use the systematic citation process to acknowledge this diffusion and, thus, thy influence the way how their scientific work builds on upward innovation and knowl-

edge (Murray et al., 2007). Although, citations are an imperfect and a noisy indicator (Seglen, 1997), this study measures the knowledge diffusion of a scientific article by counting its citations. Citation counts are downloaded from Scopus considering all peer reviewed research articles in the database. Due to the skewness of the variable, a log transformation is performed.

Research Tool Selection

Research tool selection is a multinomial variable with the three values named "Open Science Research Tool", "Commercial Research Tool" and "Self-Development". "Open Science Research Tool" is selected, if authors explicitly note that they had used the open science research tool released by the Zinc Finger Consortium. "Commercial Research Tool" refers to articles using the commercial research tool owned by Sangamo and manufactured by Sigma.

"Self-Development" papers use self-developed ZFNs or do not provide information on the research tool. In order to prepare for analysis, information on the research tool is searched in the method and material sections of the full texts. It is assumed that authors who use the commercial research tool and, thus, spend an extensive amount of resources on buying ZFNs in order to receive high quality will note this extra effort in their publication. Authors using the open science research tool provided by the ZFC might not always cite the source. Articles categorized as "Self-Development" and not providing information on the research tool used to prepare ZFNs might therefore, to some amount, use the open science research tool and be falsely categorized. In total, 94 research articles in the

database explicitly use the commercial research tool, 44 note the use of the open science alternative and 258 are categorized as ”Self-Development”.

Exclusion Restrictions

Due to the endogeneity of the research tool selection, exclusion restrictions are defined. They need to affect the research tool selection and must not correlate with article citations. In the present case, variables concerning social influence are used to identify contacts to authors who used one of the two research tools during the prior 5 years (Eisenhardt et al., 2007; Singh et al., 2016; Sampson, 1997). Two binary variables for each of the two research tools are created considering prior experience with a certain technology among prior co-authors or among colleagues at the same affiliation. ”Prior Co-Author Commercial” is coded ”1”, if a prior-co-author has published an article using the commercial research tool in the previous 5 years and ”0” if otherwise. ”Prior Author at Aff Commercial” is coded ”1”, if a former colleague at the same affiliation has published an article using the commercial research tool in the previous 5 years and ”0” if otherwise. Equivalent variables are created for the open science research tool.

Accounting for the Reflection Problem

Manski (1993) describes the reflection problem as source of endogeneity when estimating social influence. It is about the problem of the causal interpretation of social influence due to three observations. First, actors build social ties

based on homophily, e.g., when they share similar attitudes, interests or beliefs (McPherson et al., 2001). Second, many characteristics can cause homophilous tie formation most of which are unobserved. Third, in turn, these unobserved variables are correlated with the similarity of actors' behavior (Singh et al., 2013). In the context of ZFNs, an author might select a research tool due to unobserved characteristics, e.g., behavioral or cultural ones, and then build social ties based on these similarities. In order to account for the reflection problem, all variables potentially influencing the actor's behavior are calculated for 5 years prior to the event.

Controls

In order to control the effects on the citation rate, ten control variables are defined. The metadata of an article is used for five control variables having been validated in prior research (Lokker et al., 2008; Falagas et al., 2013; Jacques et al., 2010; Laurance et al., 2013; Petersen et al., 2014).

First, age controls for the time having passed since the publication of an article in years when citations accumulated over time. Due to the diminishing effect over time, squared values of age are added to the regression (Wooldridge, 2013). Second, count of authors sums up the number of authors who co-authored an article as a larger amount of co-authors can increase the visibility of an article (Falagas et al., 2013). Third, page count is the length of an article measured in the number of pages (Lokker et al., 2008). Fourth, title length counts the characters of a title (Jacques et al., 2010; Habibzadeh et al., 2010) and fifth, the number

of references counts the references of an article (Lokker et al., 2008).

Concerning team setup, the sum of all authors' prior citations of an article is used to measure the prior reputation of authors. In this case, all prior citations of an author listed on Scopus are considered. A higher prior reputation increases the number of citations of an article (Petersen et al., 2014). Due to heavy skewness, this variable is log transformed.

Prior experience within the author team is considered by the variables "Prior Internal Commercial" and "Prior Internal Open". These variables code "1", if an author of the article published any article using the certain research tool in previous 5 years and "0" if otherwise. The last control variable measures the prior reputation of affiliated organizations by summing up all prior publications of the affiliation on Scopus (Laurance et al., 2013). If multiple organizations are affiliated to one article, the maximum is used.

6.4.3 Econometric Model

Intending to test the prediction, a multinomial treatment regression model is used based on Deb et al. (2006b) and Deb et al. (2006a). The model uses a simulated maximum likelihood to calculate multinomial treatment effects on a dependent count variable. The indirect utility of selecting the jth treatment, j = 0, 1, 2,... ,J is described by the formula:

$$EV_{ij}^{*} = z_i'\alpha_j + \delta_j l_{ij} + \eta_{ij} \tag{6.1}$$

with i observations, j research tool selection, z_i exogenous variables, α_j associated parameters, η_{ik} i.i.d. error terms and l_{ik} latent factors with unobserved

characteristics common to a team's research tool selection and outcome. δ_j are associated parameters and l_{ik} are assumed to be independent of η_{ij}. j = 0 is the base group without a loss of generality and EV^*_{i0} = 0. With d_j binary variables representing observed treatment choice $\boldsymbol{d}_i = (d_{i1}, d_{i2}, ..., d_{iJ})$ and latent factors $\boldsymbol{l}_i = (l_{i1}, l_{i2}, ..., l_{iJ})$, the multinomial probability distribution g to explain the probability of treatment can be defined as follows:

$$Pr(\boldsymbol{d}_i|\boldsymbol{z}_i, \boldsymbol{l}_i) = \boldsymbol{g}\Big(\boldsymbol{z}_i'\alpha_1 + \delta_1 l_{i1}, \boldsymbol{z}_i'\alpha_2 + \delta_2 l_{i2}, ..., \boldsymbol{z}_i'\alpha_J + \delta_J l_{iJ}\Big) \tag{6.2}$$

Assuming g has a mixed multinomial logit (MMNL) structure, the probability can be rewritten as:

$$Pr(\boldsymbol{d}_i|\boldsymbol{z}_i, \boldsymbol{l}_i) = \frac{exp(\boldsymbol{z}_i'\alpha_j + \delta_j l_{ij})}{1 + \sum_{k=1}^{J} exp(\boldsymbol{z}_i'\alpha_k + \delta_k l_{ik})} \tag{6.3}$$

The outcome equation for the author group of an article i = 1,....,N is then formulated as:

$$E(y_i|\boldsymbol{d}_i, \boldsymbol{x}_i, \boldsymbol{l}_i) = \boldsymbol{x}_i'\beta + \sum_{j=1}^{J} \gamma_j d_{ij} + \sum_{j=1}^{J} \lambda_j l_{ij} \tag{6.4}$$

where x are exogenous variables with associated parameters β and γ_j denotes treatment effects relative to base group 0. d_{ij} are binary treatment choices and l_{ij} are unobserved factors with associated parameters λ_j. A positive λ_j means that treatment and outcome are positively correlated through unobserved characteristics, thus, there is positive selection. The assumption that f is the normal density allows the following consequence:

$$f(y_i|\boldsymbol{d}_i, \boldsymbol{x}_i, \boldsymbol{l}_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left(ln(y_i) - \mu_i\right)^2}{2\sigma^2}} \tag{6.5}$$

where $\mu_i = E(y_i|\boldsymbol{d}_i, \boldsymbol{x}_i, \boldsymbol{l}_i) = \boldsymbol{x}_i'\beta + \gamma\boldsymbol{d}_i' + \lambda\boldsymbol{l}_i'$

From this, a joint model can then be constructed:

$$\begin{aligned} Pr(y_i, \boldsymbol{d}_i|\boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{l}_i) &= f(y_i|\boldsymbol{d}_i, \boldsymbol{x}_i, \boldsymbol{l}_i) \times Pr(\boldsymbol{d}_i|\boldsymbol{z}_i, \boldsymbol{l}_i) \\ &= f(\boldsymbol{x}_i'\beta + \gamma\boldsymbol{d}_i' + \lambda\boldsymbol{l}_i') \times \boldsymbol{g}(\boldsymbol{z}_i'\alpha_1 + \delta_1 l_{i1}, ..., \boldsymbol{z}_i'\alpha_J + \delta_J l_{iJ}) \end{aligned} \tag{6.6}$$

The latent factors l_{ij} are assumed to be independently and identically distributed draws from the standard normal distribution. Their joint distribution $\boldsymbol{h}$ can be integrated from the joint density.

$$Pr(y_i, \boldsymbol{d}_i|\boldsymbol{x}_i, \boldsymbol{z}_i) = \int \left\{ f(\boldsymbol{x}_i'\beta + \gamma\boldsymbol{d}_i' + \lambda\boldsymbol{l}_i') \times \boldsymbol{g}(\boldsymbol{z}_i'\alpha_1 + \delta_1 l_{i1}, ..., \boldsymbol{z}_i'\alpha_J + \delta_J l_{iJ}) \right\} \boldsymbol{h}(\boldsymbol{l}_i) \mathrm{d}\boldsymbol{l}_i \tag{6.7}$$

The integral in (6.7) has no closed form. This can be addressed by following Gourieroux et al. (1996) and by using a simulation-based estimation:

$$\begin{aligned} Pr(y_i, \boldsymbol{d}_i|\boldsymbol{x}_i, \boldsymbol{z}_i) &= E\left\{ f(\boldsymbol{x}_i'\beta + \gamma\boldsymbol{d}_i' + \lambda\boldsymbol{l}_i') \times \boldsymbol{g}(\boldsymbol{z}_i'\alpha_1 + \delta_1 l_{i1}, ..., \boldsymbol{z}_i'\alpha_J + \delta_J l_{iJ}) \right\} \\ &\approx \frac{1}{S} \sum_{s=1}^{S} \left\{ f(\boldsymbol{x}_i'\beta + \gamma\boldsymbol{d}_i' + \lambda\tilde{\boldsymbol{l}}_{is}') \times \boldsymbol{g}(\boldsymbol{z}_i'\alpha_1 + \delta_1 \tilde{l}_{i1s}, ..., \boldsymbol{z}_i'\alpha_J + \delta_J \tilde{l}_{iJs}) \right\} \end{aligned} \tag{6.8}$$

$\tilde{\boldsymbol{l}}_{is}$ is the sth draw of a pseudorandom number from the density $\boldsymbol{h}_j$. The log-likelihood function is then given by:

$$\begin{aligned} LL(y_i, \boldsymbol{d}_i|\boldsymbol{x}_i, \boldsymbol{z}_i) \approx \sum_{i=1}^{N} \ln \Bigg[\frac{1}{S} \sum_{s=1}^{S} \Bigg\{ & f(\boldsymbol{x}_i'\beta + \gamma\boldsymbol{d}_i' + \lambda\tilde{\boldsymbol{l}}_{is}') \times \\ & \boldsymbol{g}(\boldsymbol{z}_i'\alpha_1 + \delta_1 \tilde{l}_{i1s}, ..., \boldsymbol{z}_i'\alpha_J + \delta_J \tilde{l}_{iJs}) \Bigg\} \Bigg] \end{aligned} \tag{6.9}$$

In the computational model Halton sequences, described in Train (2003), are used instead of pseudorandom draws due to speed improvements. The equation (6.9) of logged simulated likelihood is maximized using a Newton-Raphson algorithm with numerical derivatives (Deb et al., 2006a).

6.5 Results

Descriptive statistics are displayed in table 6.2. Results of the multinomial treatment regression model are provided in tables 6.3 and 6.4. The model was estimated with 100 Halton draws based on Train (2003).

Table 6.2: Descriptive statistics for ZFN analysis database

Variable	Obs	Mean	Std. Dev.	Min	Max
Citations	396	2.956	1.434	0	6.282
Age of article	396	4.578	3.101	0	12
Age squared of article	396	30.553	36.309	0	144
Count of authors	396	8.043	4.981	1	29
Page count	396	9.056	3.226	2	21
Title length	396	101.634	29.884	35	243
Number of references	396	42.616	17.107	7	118
Authors' prior citations	396	9.165	1.303	0	12.27
Prior publications at affiliation	396	51482.36	44664.07	0	188146
Research tool selection	396	2.54	.687	1	3
Prior internal commercial	396	.225	.418	0	1
Prior internal open	396	.144	.351	0	1
Prior co-author commercial	396	.424	.495	0	1
Prior co-author open	396	.396	.49	0	1
Prior author at aff commercial	396	.273	.446	0	1
Prior author at aff open	396	.144	.351	0	1

Table 6.3 shows the regression results of the open science and of the commercial research tool selection. Both regressions include exclusion restrictions on peer and affiliation level and relevant control variables for the selection: prior experience with a certain research tool, count of authors and authors' prior citations. The selection of the open science research tool is negatively influenced by a prior experience with the commercial research tool but positively by the number of authors. The selection of the commercial research tool is, both, negatively influenced by a prior team experience with the open science and the commercial research tool. Social influence is statistically relevant for co-authors. Experience of prior co-authors with the open science research tool, decreases the likelihood

to select the commercial research tool. In contrast, experience of prior co-authors with the commercial research tool has a positive influence.

Table 6.3: Regression estimates for 1st step of multinomial treatment regression

	(1) Open Science	(2) Commercial
Prior co-author commercial	0.376	1.599***
	(0.72)	
Prior co-author open	0.223	−0.886*
	(0.39)	
Prior author at aff commercial	−0.826	0.510
	(−0.99)	
Prior author at aff open	0.476	0.215
	(0.87)	
Prior internal commercial	−3.496*	1.968***
	(−2.06)	
Prior internal open	0.768	−2.392**
	(1.15)	
Count of authors	0.168**	0.108**
	(2.84)	
Authors' prior citations	−0.246	−0.189
	(−1.40)	
Constant	−1.202	−0.339
	(−0.79)	

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6.4 displays estimates of the research tool selection and of the control variables on article citations. Compared to self-development, the selection of open science (-0.382) and of commercial (-0.640) research tools has negative effects on the citation rate. Age has a positive, but decreasing effect. A shorter title length, a longer list of references and a larger group of co-authors are positively correlated with citations. Prior reputation of both, affiliated institutions

and author group, has positive coefficients, as well.

Table 6.4: Regression table for 2nd step of multinomial treatment regression

	(1) Citations
Open Science Research Tool	-0.382^{**} (-2.60)
Commercial Research Tool	-0.640^{***} (-6.15)
Age of article	0.764^{***} (13.44)
Age squared of article	-0.0447^{***} (-8.86)
Page count	0.00783 (0.48)
Title length	-0.00814^{***} (-5.51)
Number of references	0.00976^{***} (3.32)
Count of authors	0.0756^{***} (8.97)
Prior publications at affiliation	0.00000296^{**} (3.13)
Authors' prior citations	0.141^{***} (5.49)
Constant	-0.681^{*} (-2.28)

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6.5.1 Prediction Test

The prediction proposes that the selection of the commercial research tool has a negative effect on the citation rate of a paper. Regression estimates in table 6.4 validate this prediction. While both article groups using the open science

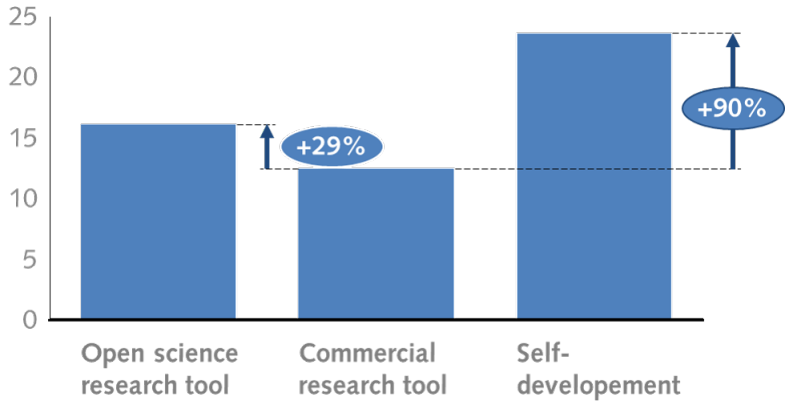


Figure 6.2: Prediction of citation rates for research articles categorized by used research tool

research tool and the commercial one correlate with lower citation rates in comparison with self-development, two observations lead to the validation of the prediction. First, the research tool selection is transformed into binary variables coding ”1” and ”0” and the coefficient of the commercial research tool selection is estimated as -0.640, while the one of the open science research tool values -0.382. Predictions considering the log transformation of the dependent variable are visualized in figure 6.2 for the selection of the open science research tool, of the commercial one and for self-development. Articles using the open science research tool are predicted to receive 16 citations, ones using the commercial research tool 12. This results in a 29% increase of citations when using the open science research tools for gene-editing.

Second, scientists using the open science research tool might not always give reference to the ZFC and are included in the self-development category. These research articles are predicted to have 24 citations. The predicted citation rate for articles using the open science research tool might, therefore, be even higher in practice. The difference to articles using the commercial research tool would

further increase.

6.5.2 Model Validation and Robustness Check

In order to validate the use of the multinomial treatment regression model, two steps are performed. First, an OLS model and lambda values from the multinomial treatment model are checked for the effects of endogeneity concerning the research tool selection. Second, exclusion restrictions are validated for exogeneity and strength.

The OLS models in table 6.5 include two binary variables ”Open Science Research Tool” and ”Commercial Research Tool’ coded ”1”, if authors of articles used open science or commercial research tool and ”0” if otherwise.

In both models, research tool selection is not statistically relevant. An inconsistent parameter estimation might be caused by endogenous regressors leading to biased estimators. Lambda values of the multinomial treatment regression model are visualized in table 6.6. Both, lambda for open science and commercial research tool selection, are statistically relevant and indicate the underestimation of treatment effects. A model considering treatment effects is therefore appropriate for the present case.

In a second step, exclusion restrictions are tested for exogeneity and strength. The estimation technique of exclusion restrictions is described in the following.

$$y = \beta_0 + \beta_1 x + u \tag{6.10}$$

Table 6.5: OLS Regression table for ZF Research Tool Selection

	(1) CitationsLog	(2) CitationsLog
Open Science Research Tool	−0.130	
	(−0.78)	
Commercial Research Tool		0.101
		(0.81)
Age of article	0.811***	0.811***
	(14.14)	(14.16)
Age squared of article	−0.0492***	−0.0488***
	(−10.13)	(−10.17)
Page count	0.0176	0.0161
	(0.90)	(0.82)
Title length	−0.00538**	−0.00550**
	(−3.14)	(−3.20)
Number of references	0.00593	0.00617
	(1.60)	(1.65)
Count of authors	0.0661***	0.0653***
	(5.49)	(5.42)
Prior publications at affiliation	0.00000230	0.00000224
	(1.93)	(1.90)
Authors' prior citations	0.161***	0.164***
	(3.56)	(3.63)
Constant	−1.230**	−1.284**
	(−2.71)	(−2.80)
Observations	396	396

t statistics in parentheses

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

Table 6.6: Model statistics for multinomial treatment regression

lnsigma	
Constant	-1.437^{***}
	(-4.56)
lambda Open Science Research Tool	
Constant	0.345^{***}
	(6.15)
lambda Commercial Research Tool	
Constant	0.960^{***}
	(17.78)
Observations	396

t statistics in parentheses

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

In standard regression (6.10) one assumption is that regressors are uncorrelated with the errors in the model, i.e., x has only a direct effect on y through β_1x, thus, x and u are independent causes for y. In case of a correlation between x and u, i.e., Cov(x,u) is not equal zero, y can be increased by x directly and indirectly by u via x. x is an endogenous regressor, the OLS estimator is, therefore, biased and inconsistent for β.

In order to generate an exogenous variation of x and to reduce bias, an exclusion restriction z is created. Changes in z are associated with variations in x but to not lead to changes in y. Two conditions for exclusion restrictions have to be validated. First, they have to be exogenous on the dependent variable showing Cov(z,u) equal 0. Second, they have to be relevant as to explaining the endogenous regressors with Cov(z, x) not equal 0 (Cameron et al., 2005). As a consequence, exclusion restrictions are only considering LATE, i.e., local average treatment effects (Bascle, 2008). In the present case, the dependent variable is the citation rate and a potential endogenous predictor is the selection of the

research tool.

Following (Walker et al., 2011), exclusion restrictions measuring social influence are built. These variables are assumed to influence the choice of the research tool but to not correlate with the citation rate of articles. Two binary variables, explained in 6.4.2, are created for each of the two research tool selection options. Models with exclusion restrictions are built stepwise based on (Bascle, 2008). The full Stata output is printed in the appendix A.4.

First, a 2SLS estimation is used and a heterogeneity test (Pagan et al., 1983) is performed to evaluate, if robust estimators are needed. Then, exogeneity and strength of the exclusion restrictions are tested. They are exogenous rejecting Anderson canonical correlation LM statistic with a p-value of 0.0835. Testing for weakness with Cragg-Donald Wald F statistic reveals less than 30% maximal IV relative bias following Stock et al. (2002). In the case of weak exclusion restrictions LIML or CUE methods are recommended (Bascle, 2008). By using the LIML method, Anderson canonical correlation LM statistic is consistent with 2SLS and Cragg-Donald Wald F statistics are below 10% maximal LIML size (Stock et al., 2002). Exclusion restrictions are strong enough. A final check of the CLR interval confirms the validity (Bascle, 2008). The coverage-corrected confidence interval for the research tool selection is between 0.167 and 2.374 at a p-value of 0.0168 conforming the selection of exclusion restrictions.

To check for robustness, the SCImago Journal Rank indicator (SJR) is included as control variable. The journal ranking indicators are downloaded from Scopus.com and split into four equally large groups. The regression table is visualized in table A.4 in the appendix. The research tool selection variables are

statistical significant and the gap between selecting the open science or the commercial research tool is larger if controlled for the journal impact factor.

6.6 Discussion

Prior research focused on the diffusion of a research tool in follow-on research and evaluated the effect on creativity of follow-on research. The influence of selecting from competing open science and commercial research tools on the knowledge diffusion of published studies was not evaluated. The results of this study involve several implications for research and practice.

6.6.1 Implications for Research

Prior research analyzed the effect of property right protection on the diffusion of a technology in follow-on research regarding quantity and type of follow-on research. Scholars listed multiple positive factors of open science, e.g., exploitation of the full potential of innovations (Bessen, 2004), and hypothesized that open science research tools might impact the diffusion of proprietary research tools (Murray et al., 2016). There was a lack of understanding the effect of selecting from competing open science and commercial research tools on the diffusion of published studies.

This study fills this gap by exploring the adoption of competing research tools and by analyzing the diffusion of published studies based on research tool selection using the case of Zinc Finger Nucleases. Results validate a negative effect

of using commercial research tools on the diffusion of published studies, i.e., articles using the commercial research tool are predicted to have lower citation rates. Therefore, this study follows Murray et al. (2016) in evaluating effects of open science on follow-on research and validates open science as a factor that influences knowledge diffusion positively.

This positive effect can have multiple explanations, e.g., the feasibility of replicating experiments and the method of building on prior research or on the needed know-how to adopt open science research tools. The feasibility of replicating experiments as an influencing factor for building on prior work was validated by Piwowar et al. (2007), who showed a positive effect when researchers share detailed result data.

In addition, applying the open science research tool requests a higher level of know-how. A team of researchers experienced in genetic engineering is a prerequisite for adopting open science research tools but, as well, for a high quality research output which, in turn, is correlated with an increase in citation rates (Nieminen et al., 2006). The variable "Authors' prior citations" aims to control for this potential effect by considering the prior reputation of the authors.

6.6.2 Implications for Practice

Accelerating research in the high-potential research area of synthetic biology is a main goal of governments and industries (Peccoud, 2016; Raimbault et al., 2016; Singh, 2015). Potential effects of competition can influence the diffusion of research tools. With the co-existence of an open science alternative, the dif-

fusion of a commercial research tool could decline. However, Sangamo did not see any challenge in the open science alternative (Kaiser, 2005) and did not start any lawsuit (Chandrasekharan et al., 2009). They predicted an increase of downstream applications with the co-existence of an open science research tool (Pearson, 2008). These applications would mostly require rights to use the commercial research tool of Sangamo (Chandrasekharan et al., 2009).

This prediction could not be sophistically analyzed with the given dataset due to the small quantity of articles using the open science or commercial research tool. However, observing the time series of articles categorized by the research tool being used in figure 6.3, no severe negative effect of the open science research tool can be seen. The ratio of articles using the commercial research tool even increased after licensing the manufacturer Sigma Aldrich and the release of the open science alternative in 2008. The release of an open science alternative had, at least, no negative effect on the adaption of the commercial research tool.

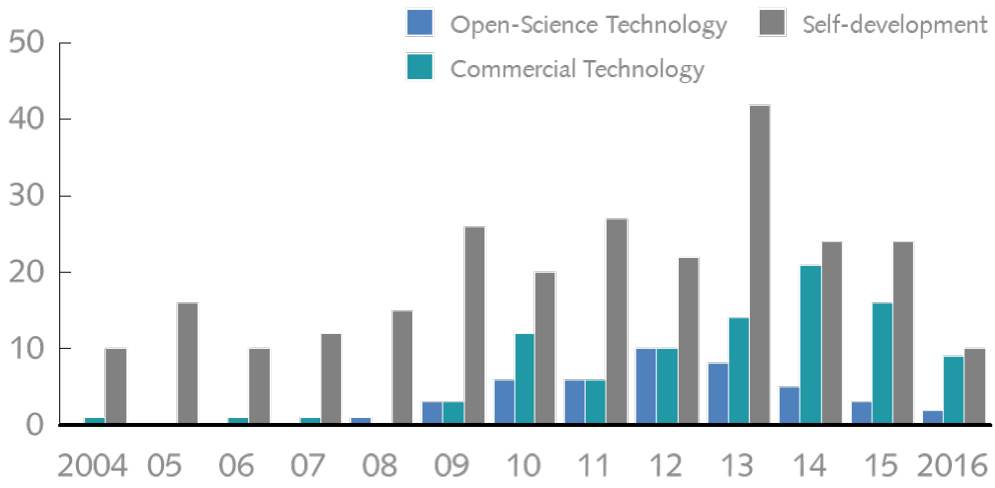


Figure 6.3: Usage of ZF research tools from 2004 until 2016

6.7 Limitations

Limitations can be seen in two areas. First, regarding the consideration of publication data only and second, regarding the limitations of the size of the sample. The analysis uses research articles accessible on Scopus, thus, external actors such as mass media, consultants, funding agencies or professional communities and events like conferences are not considered. However, research suggests that these external actors shape the way how actors interpret innovation and, thus, the selection of technology (Wejnert, 2002; Strang et al., 1998). In addition, variables considering the quality of articles are not included to control for citations. Second, 396 full text articles could be downloaded from Scopus database due to license restrictions and only 138 articles make an explicit reference to the competing research tools. The majority of research articles is, therefore, classified as self-development. A multinomial treatment regression model simulating a maximal likelihood is used to correct for endogeneity. With a larger size of the sample, a matched control group could be built to gain higher accuracy.

Chapter 7

Conclusion

Scientific innovations need to get created and become widely adopted to fully exploit their potential as a dominant force in economic growth (Pejovich, 1996; Rogers, 2003; Rosenberg, 2004). Therefore, scientific research analyzes potential levers on the diffusion of scientific innovation with particular interest on institutions, e.g., settings of property rights. As institutional theory lacks in explaining the emergence and shaping of institutions, the institutional entrepreneur approach faces these limitations. According to the theory, institutional entrepreneurs combine logics from multiple fields and diffuse them in their social context to shape an emerging field.

While isolated quantitative analyses on the direct impact of institutions have been performed in prior research (Murray, 2006; Murray et al., 2007; Williams, 2013), a comprehensive end-to-end validation of the process from institutional entrepreneurs diffusing a logic to the indirect impact of selecting from competing open science and commercial research tools on knowledge diffusion was not yet done. In this thesis, three studies are performed to analyze and validate the end-to-end process in the research field of synthetic biology.

7.1 Summary of findings and practical implications

Study I aims to validate theories on institutional entrepreneurs, who are supposed to diffuse an institutional logic in order to shape an emerging organizational field. Prior research exists on activities of institutional entrepreneurs in emerging fields, diffusion of institutions and product innovation in industry (e.g. Van de Ven et al., 1993; Powell et al., 1996; Lenway et al., 2001; Garud et al., 2002; Baron et al., 1986; Guler et al., 2002; Ritvala et al., 2009). The context of scientific innovation was not analyzed and the diffusion of institutional logics by institutional entrepreneurs needed validation.

Results validate that institutional entrepreneurs use social cohesion to convince their context considering, both, co-authors and colleagues at affiliation. The high infectiousness of institutional entrepreneurs due to prior reputation increases social influence on socially proximate actors and thus enables them to compel others with the institutional logic. Characteristics of institutional entrepreneurs being not susceptible towards social influence are validated as well. This study extends the literature and provides a quantitative validation of theories on how institutional entrepreneurs drive diffusion of institutional logics in context of scientific innovation to shape an emerging field. The heterogeneous diffusion model is applied to scientific innovations and institutional logics in order to measure the impact of social influence mechanisms and validate theories on characteristics and strategies of institutional entrepreneurs. Transforming these results into a call for action, policy makers should be aware of institutional entrepreneurs in their seeding strategies. Identifying, empowering and directing

these key players enables the diffusion of an institutional logic and supports the emergence of a new field.

After validating antecedents of institutions, studies II and III aim to understand how institutions are crucial to exploit scientific innovations. Based on an established logic, institutions are shaped in the field. In the context of synthetic biology, open science initiatives are motivated by the engineering approach. To assess status quo, study II investigates the current impact of a central open science initiative, the BIOFAB project.

The study elaborates two advances for research. First, a new methodology to measure the re-usage of biological parts by matching DNA sequences with stored sequence data in patent applications is developed. Using the case of BIOFAB sequences, eleven times more re-usages can be identified than by counting references. Scholars can adopt this methodology to assess the diffusion of other query lists or validate prior research, which used, e.g., citation references to measure diffusion. Second, the current impact of open science in synthetic biology and BioBrick engineering, in particular, has priorly been assessed via a qualitative study (Kahl, 2015). The results of this survey are now confirmed by quantitatively analyzing the diffusion of open science parts in follow-on research. BIOFAB parts are diffused at a moderate level in patent applications. Based on the feedback of the company DNA2.0 which has implemented the sequences into their software, a more extensive diffusion than captured in patent applications must be assumed.

Practice can benefit from this study in two ways. First, a fast process to test for re-usage of parts was developed. Creators of biological parts can use the

protocol provided in A.2 to quantify the re-usage frequency. The BIOFAB team, in particular, can use the study results to get confirmation of their efforts and a potential feedback to their work. Second, the hypothesis concerning characterization quality as success factor is investigated. No definite effect of a higher characterization quality can be validated. Therefore, the community needs to reflect on potential reasons and to find solutions in case of further engagements. The influence model (Asavathiratham et al., 2001) is a framework to analyze change of behavior and can be applied to open science parts in synthetic biology to structure the process of problem solving. The recent release of the OpenMTA tries to consider the framework (Forsyth, 2017). The success of the OpenMTA should be monitored and the enumerated levers should be evaluated.

Study III extends the scope of study II and analyzes how selecting from competing open science and commercial research tools affect the knowledge diffusion of a paper. Prior research examined the effect of property right protection on the diffusion of a research tool in follow-on research regarding quantity and type of follow-on research. Scholars listed multiple positive factors of open science, e.g., to exploit the full potential of innovations (Bessen, 2004), and hypothesized that open science research tools might impact the diffusion of proprietary research tools (Murray et al., 2016). There was a lack of understanding whether selecting from competing open science and commercial research tools affects knowledge diffusion.

This study fills the gap by analyzing the diffusion of published studies considering the selection of the research tool using the case of Zinc Finger Nucleases. Results show that articles which use the open science research tool are predicted

to have higher citation rates than articles which pay for the commercial research tool. Therefore, this study follows Murray et al. (2016) evaluating effects of open science on follow-on research and validates that the usage of open science research tools is a factor that positively influences knowledge diffusion.

Accelerating research in the high-potential research area of synthetic biology is the primary goal of governments and industries (Peccoud, 2016; Raimbault et al., 2016; Singh, 2015). The co-existence of open science alternatives could negatively influence the economic success of commercial research tools. In the case of ZF, Sangamo did not start any lawsuit, because they predicted an increase of downstream applications with the co-existence of an open science alternative. These downstream applications would mostly require rights to use the commercial tool of Sangamo (Chandrasekharan et al., 2009). This hypothesis can be validated by exploring time series of research articles categorized by the selected research tool. No adverse effect can be observed due to the publication of an open science alternative.

7.2 Summary of limitations and future research opportunities

All three studies have limitations, both, concerning the datasets and the analysis process, and they have substantial opportunities for future research.

Study I has three limitations. First, the consideration of publication data only, second, a combination of causal mechanisms and third, limitations in data processing. The analysis is based on research articles accessible on Scopus.com.

Thus external actors such as mass media, consultants, funding agencies or professional communities and events like conferences are not considered. However research suggests that these external actors shape how actors interpret innovation (Wejnert, 2002; Strang et al., 1998). Second, even though causal mechanisms, e.g., information transmission, observation and learning are considered, the distinction and separation between them are not possible based on the existing dataset. Empirical analysis observing and interrogating actors might help to accomplish this step.

Third, data processing is limited by resources, e.g., the selection of articles with natural language processing can be improved with a more extensive training data set, and Scopus author identifier was used to disambiguate and collect all assigned articles. The error rate of Scopus author identifier algorithms can influence results.

Study II is as well limited in three dimensions. First, the analysis measures re-usages of biological parts in patent applications. Follow-on research, which is not secured in patent applications, is not considered in the dataset. Second, the query data consists of 478 BIOFAB parts and only a small set is re-used in patent applications. Thus, the two hypotheses are assessed by monitoring time-series data. A more significant quantity of query parts with a higher rate of diffusion is needed to establish statistically validated hypotheses. Third, the matching process is divided into an automated and a manual step. Thresholds in the computerized step are developed iteratively to gain the global optimum, but can still lead to false-negative results. In the manual review, characteristic phrases in the patent texts are used to validate re-usage. Prior training on how to read patents

given by experts and their sample checks increase the significance of results, but manual review can still be subject to inaccuracy.

The study III is limited by publication data and sample size. The analysis is based on research articles accessible on Scopus.com. Thus external actors such as mass media or professional communities and events like conferences are not considered, even though they can impact adoption processes (Wejnert, 2002; Strang et al., 1998). Scopus author identifier was used to disambiguate and collect all assigned articles. The error rate of Scopus author identifier algorithms can influence results. Also, variables considering quality of articles are not included to control for citations. 396 full text articles could be downloaded from Scopus database due to license restrictions. As only 138 research articles give explicit reference to the selected research tool, a majority of articles is categorized as self-development. A multinomial treatment regression model simulating maximal likelihood is used to correct for endogeneity. With a larger sample size a matched control group could be generated to reach higher accuracy.

Future research opportunities exist for all three studies. Study I validates social cohesion as a mechanism to diffuse institutional logics and a higher ability of institutional entrepreneurs to drive change due to their resources. The results are confirmed through multiple robustness checks. Future research opportunities are seen, both, in adapting the approach to alternative fields and in deep diving the present results. Scholars can use the large scale approach and validate results in alternative fields to confirm the universality of findings. Deep dives on the study results can be qualitative analyses expanding the work of Raimbault et al. (2016) and analyzing the dynamics of the core and peripheral communi-

ties building on Grodal (2017). In particular, case studies and interview series capturing perspectives of institutional entrepreneurs on their strategies to drive change and to influence their social context are essential to confirm and expand the results of this study.

Study II is a proof of concept to use sequence matching in patent applications to analyze the diffusion of scientific innovations in biotechnology. Using the case of BIOFAB parts, eleven times more re-usages could be identified than by counting references. Prior research using reference data can be verified with the new methodology. Also, a new research question is opened, i.e., success factors of open science technologies. A first predictor concerning the quality of parts could not be validated using the present dataset. Future research can re-analyze the effect using larger databases or investigate the effect of other success factors discussed in chapter 5.

Study III follows Murray et al. (2016) investigating the effect of open science upstream technologies on follow-on research and extends the research on predictors for knowledge diffusion. Identification of a used technology in follow-on research was performed through explicit references in full texts. Future research should back results using a larger database to create a matched control group.

Appendix A

Appendix

A.1 Validation Models for Chapter 4

To validate cox regression used in 4, calculation with varied threshold and alternative calculating methods are performed. Table A.1 visualizes six models calculated with 55% threshold on topic validation. Variations in regression results do not change statistical significances which are relevant to the predictions. Table A.2 contains calculation of the basic model using a logit and two parametric models, i.e. weibull and exponential distribution. Estimation results confirm the appropriate usage of cox regression in the main model.

Table A.1: Regression table with 55% threshold

	(1) SynBio Article	(2) SynBio Article	(3) SynBio Article	(4) SynBio Article	(5) SynBio Article	(6) SynBio Article
Africa	0	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)	(.)
Asia	0.605	0.496	0.501	0.494	0.501	0.496
	(1.81)	(1.49)	(1.50)	(1.48)	(1.50)	(1.49)
Australia	0.912**	0.764*	0.764*	0.763*	0.791*	0.788*
	(2.69)	(2.25)	(2.25)	(2.24)	(2.33)	(2.32)
Europe	1.062**	0.868**	0.872**	0.867**	0.880**	0.879**
	(3.18)	(2.60)	(2.61)	(2.60)	(2.64)	(2.63)
Latinamerica	0.0487	0.00871	0.00937	0.00783	0.0136	0.0110
	(0.14)	(0.03)	(0.03)	(0.02)	(0.04)	(0.03)
Middle-East	0.181	0.0664	0.0673	0.0657	0.0550	0.0461
	(0.53)	(0.20)	(0.20)	(0.19)	(0.16)	(0.14)
Northamerica	1.292***	1.029**	1.029**	1.030**	1.031**	1.026**
	(3.87)	(3.08)	(3.08)	(3.08)	(3.09)	(3.07)
Reputation	−0.658***	−0.677***	−0.679***	−0.677***	−0.802***	−0.833***
	(−83.47)	(−86.62)	(−86.71)	(−86.62)	(−84.62)	(−67.80)
Tenure	−0.0442***	−0.0485***	−0.0483***	−0.0485***	−0.0462***	−0.0477***
	(−29.62)	(−32.14)	(−32.04)	(−32.14)	(−30.49)	(−31.64)
Constraint	−0.0583	0.103**	0.105**	0.103**	0.121***	0.116***
	(−1.69)	(3.02)	(3.09)	(3.02)	(3.56)	(3.42)
Knowledge Diversity	0.302***	0.228***	0.230***	0.227***	0.257***	0.237***
	(6.55)	(4.93)	(4.99)	(4.93)	(5.56)	(5.14)
Affinity towards SynBio	1.769***	1.516***	1.530***	1.516***	1.563***	1.548***
	(16.69)	(14.19)	(14.31)	(14.19)	(14.65)	(14.49)
Reputation of Affiliation	−0.0296***	−0.0727***	−0.0726***	−0.0729***	−0.0713***	−0.0683***
	(−10.26)	(−22.82)	(−22.78)	(−22.72)	(−22.42)	(−21.37)
Affinity of Affiliation towards SynBio	−0.208***	−0.382***	−0.381***	−0.382***	−0.383***	−0.375***
	(−15.33)	(−22.24)	(−22.13)	(−22.22)	(−22.23)	(−21.74)
Structural Equivalence	0.0832***	0.0426***	0.0406***	0.0425***	0.0509***	0.0352***
	(32.04)	(7.02)	(6.69)	(7.01)	(8.95)	(5.68)
Knowledge Equivalence	0.00490***	0.00422***	0.00421***	0.00421***	0.00446***	0.00418***
	(5.98)	(5.17)	(5.16)	(5.16)	(5.47)	(5.12)
Missing Data Control	−0.306***	−0.187***	−0.192***	−0.186***	−0.263***	−0.209***
	(−12.39)	(−7.45)	(−7.66)	(−7.43)	(−10.41)	(−8.33)
Co-Authors in SynBio		0.343***	0.195**	0.343***	−0.325***	0.300***
		(8.97)	(3.28)	(8.96)	(−6.66)	(7.79)
Reputation of Co-Authors in SynBio		0.0762***	0.0670***	0.0764***	0.129***	0.0811***
		(9.47)	(7.91)	(9.48)	(15.50)	(10.07)
Authors in SynBio at Affiliation		0.0489***	0.0498***	0.0545**	0.0459***	−0.0242
		(4.09)	(4.17)	(2.97)	(3.84)	(−1.90)
Reputation of Authors in SynBio at Affiliation		0.113***	0.112***	0.114***	0.116***	0.118***
		(16.37)	(16.27)	(14.95)	(16.74)	(17.07)
Co-Authors in SynBio × Reputation of Co-Authors in SynBio			0.0327***			
			(3.34)			
Authors in SynBio at Affiliation × Reputation of Authors in SynBio at Affiliation				−0.00153		
				(−0.40)		
Co-Authors in SynBio × Reputation					0.242***	
					(28.91)	
Authors in SynBio at Affiliation × Reputation						0.0765***
						(18.08)
Observations	153812	153812	153812	153812	153812	153812

t statistics in parentheses

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

Table A.2: Regression table for logit and parametric models

	(1) SynBio Article	(2) SynBio Article	(3) SynBio Article
main			
Africa	0	0	0
	(.)	(.)	(.)
Asia	−0.516*	0.0372	−0.190
	(−2.56)	(0.39)	(−1.15)
Australia	−0.165	−0.133	0.0415
	(−0.79)	(−1.34)	(0.24)
Europe	−0.0247	−0.183	0.151
	(−0.12)	(−1.91)	(0.92)
Latinamerica	−0.788***	0.303**	−0.642***
	(−3.77)	(3.06)	(−3.73)
Middle-East	−0.666**	0.283**	−0.635***
	(−3.21)	(2.86)	(−3.72)
Northamerica	−0.0360	−0.337***	0.356*
	(−0.18)	(−3.51)	(2.16)
Reputation	−0.765***	0.296***	−0.685***
	(−117.94)	(108.82)	(−119.95)
Tenure	−0.0472***	0.0339***	−0.0469***
	(−35.02)	(54.90)	(−43.31)
Constraint	−0.0449	0.102***	−0.0676**
	(−1.46)	(6.96)	(−2.69)
Knowledge Diversity	0.528***	−0.242***	0.306***
	(13.06)	(−12.88)	(9.41)
Affinity towards SynBio	2.452***	−0.659***	0.979***
	(14.29)	(−9.43)	(10.39)
Reputation of Affiliation	−0.0774***	0.0520***	−0.0763***
	(−25.76)	(38.48)	(−32.76)
Affinity of Affiliation towards SynBio	−0.233***	0.261***	−0.406***
	(−15.75)	(37.58)	(−32.91)
Structural Equivalence	0.162***	−0.0591***	0.0693***
	(6.96)	(−12.43)	(20.77)
Knowledge Equivalence	0.00512***	−0.0152***	0.0217***
	(6.64)	(−37.03)	(29.27)
Missing Data Control	−0.609***	0.254***	−0.339***
	(−26.87)	(24.32)	(−18.88)
Co-Authors in SynBio	0.364***	−0.134***	0.226***
	(13.11)	(−11.24)	(10.95)
Reputation of Co-Authors in SynBio	0.0846***	−0.0230***	0.0469***
	(12.27)	(−7.58)	(8.95)
Authors in SynBio at Affiliation	0.0892***	−0.0458***	0.0748***
	(9.61)	(−10.71)	(10.09)
Reputation of Authors in SynBio at Affiliation	0.163***	−0.0477***	0.108***
	(25.96)	(−16.34)	(21.69)
Constant	−0.553**	3.332***	−6.802***
	(−2.66)	(33.49)	(−39.13)
ln_sig			
Constant		−0.152***	
		(−32.32)	
ln_p			
Constant			0.754***
			(139.64)
Observations	153727	153727	153727

t statistics in parentheses

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

A.2 Code to run BLAST query for BIOFAB parts matching

Build database with nucleotides in USPTO patent applications downloaded from PatentLens:

```
makeblastdb -in USPTO_Applications.fa -parse_seqids
-dbtype nucl -max_file_sz 10GB
```

Run Blastn Standalone with specified parameters:

```
blast -db USPTO_Applications.fa
-query BIOFABDB.fasta
-outfmt ''6 qseqid sseqid pident length mismatch gapopen qstart qend
 sstart send evalue bitscore qcovs''
-evalue ''1e-6''
-task blastn
-max_target_seqs 5000000
-best_hit_overhang 0.1
-best_hit_score_edge 0.1
-num_threads 4
-out USPTOblastn.txt
```

Table A.3: Overview BIOFAB parts

Number	Type	Sequence Name	Category
1	promoter	apFAB29	priorly used
2	promoter	apFAB30	new created
3	promoter	apFAB31	new created
4	promoter	apFAB32	new created
5	promoter	apFAB33	new created
6	promoter	apFAB34	new created
7	promoter	apFAB35	priorly used
8	promoter	apFAB36	new created
9	promoter	apFAB37	new created
10	promoter	apFAB38	new created
11	promoter	apFAB39	new created
12	promoter	apFAB40	new created
13	promoter	apFAB41	new created
14	promoter	apFAB42	new created
15	promoter	apFAB43	new created
16	promoter	apFAB44	new created
17	promoter	apFAB45	new created
18	promoter	apFAB46	new created
19	promoter	apFAB47	new created
20	promoter	apFAB48	new created
21	promoter	apFAB49	new created
22	promoter	apFAB50	new created
23	promoter	apFAB51	new created
24	promoter	apFAB52	new created
25	promoter	apFAB53	new created
26	promoter	apFAB54	new created
27	promoter	apFAB55	new created
28	promoter	apFAB56	new created
29	promoter	apFAB57	new created
30	promoter	apFAB58	new created
31	promoter	apFAB59	new created
32	promoter	apFAB60	priorly used
33	promoter	apFAB61	new created
34	promoter	apFAB62	new created
35	promoter	apFAB63	new created
36	promoter	apFAB64	new created
37	promoter	apFAB65	new created
38	promoter	apFAB66	new created
39	promoter	apFAB67	new created
40	promoter	apFAB68	new created

Overview BIOFAB parts (2)

Number	Type	Sequence Name	Category
41	promoter	apFAB69	new created
42	promoter	apFAB70	new created
43	promoter	apFAB71	new created
44	promoter	apFAB72	new created
45	promoter	apFAB73	new created
46	promoter	apFAB74	new created
47	promoter	apFAB75	new created
48	promoter	apFAB76	new created
49	promoter	apFAB77	new created
50	promoter	apFAB78	new created
51	promoter	apFAB79	new created
52	promoter	apFAB80	new created
53	promoter	apFAB81	new created
54	promoter	apFAB82	new created
55	promoter	apFAB83	new created
56	promoter	apFAB84	new created
57	promoter	apFAB85	new created
58	promoter	apFAB86	new created
59	promoter	apFAB87	new created
60	promoter	apFAB88	new created
61	promoter	apFAB89	new created
62	promoter	apFAB90	new created
63	promoter	apFAB91	priorly used
64	promoter	apFAB92	new created
65	promoter	apFAB93	new created
66	promoter	apFAB94	new created
67	promoter	apFAB95	new created
68	promoter	apFAB96	new created
69	promoter	apFAB97	new created
70	promoter	apFAB98	new created
71	promoter	apFAB99	new created
72	promoter	apFAB100	new created
73	promoter	apFAB101	new created
74	promoter	apFAB102	new created
75	promoter	apFAB103	new created
76	promoter	apFAB104	new created
77	promoter	apFAB105	new created
78	promoter	apFAB106	new created
79	promoter	apFAB107	new created
80	promoter	apFAB108	new created

Overview BIOFAB parts (3)

Number	Type	Sequence Name	Category
81	promoter	apFAB110	new created
82	promoter	apFAB111	new created
83	promoter	apFAB112	new created
84	promoter	apFAB113	new created
85	promoter	apFAB114	new created
86	promoter	apFAB115	new created
87	promoter	apFAB117	new created
88	promoter	apFAB118	new created
89	promoter	apFAB119	new created
90	promoter	apFAB120	new created
91	promoter	apFAB121	new created
92	promoter	apFAB122	new created
93	promoter	apFAB123	new created
94	promoter	apFAB124	new created
95	promoter	apFAB125	new created
96	promoter	apFAB126	new created
97	promoter	apFAB127	new created
98	promoter	apFAB128	new created
99	promoter	apFAB129	new created
100	promoter	apFAB130	new created
101	promoter	apFAB131	new created
102	promoter	apFAB133	new created
103	promoter	apFAB134	new created
104	promoter	apFAB136	new created
105	promoter	apFAB137	new created
106	promoter	apFAB138	new created
107	promoter	apFAB139	new created
108	promoter	apFAB140	new created
109	promoter	apFAB141	new created
110	promoter	apFAB142	new created
111	promoter	apFAB143	new created
112	promoter	apFAB144	new created
113	promoter	apFAB145	new created
114	promoter	apFAB146	new created
115	promoter	apFAB147	new created
116	promoter	apFAB148	new created
117	promoter	apFAB149	new created
118	promoter	apFAB150	new created
119	promoter	apFAB151	new created
120	promoter	apFAB152	new created

Overview BIOFAB parts (4)

Number	Type	Sequence Name	Category
121	promoter	apFAB156	new created
122	promoter	apFAB157	new created
123	promoter	apFAB159	new created
124	promoter	apFAB160	new created
125	promoter	apFAB161	new created
126	promoter	apFAB162	new created
127	promoter	apFAB164	new created
128	promoter	apFAB166	new created
129	promoter	apFAB167	new created
130	promoter	apFAB168	new created
131	promoter	apFAB177	new created
132	promoter	apFAB180	new created
133	promoter	apFAB181	new created
134	promoter	apFAB182	new created
135	promoter	apFAB183	new created
136	promoter	apFAB184	new created
137	promoter	apFAB186	new created
138	promoter	apFAB187	new created
139	promoter	apFAB188	new created
140	promoter	apFAB189	new created
141	promoter	apFAB190	new created
142	promoter	apFAB192	new created
143	promoter	apFAB193	new created
144	promoter	apFAB194	new created
145	promoter	apFAB195	new created
146	promoter	apFAB197	new created
147	promoter	apFAB199	new created
148	promoter	apFAB200	new created
149	promoter	apFAB201	new created
150	promoter	apFAB202	new created
151	promoter	apFAB203	new created
152	promoter	apFAB204	new created
153	promoter	apFAB205	new created
154	promoter	apFAB206	new created
155	promoter	apFAB207	new created
156	promoter	apFAB208	new created
157	promoter	apFAB209	new created
158	promoter	apFAB210	new created
159	promoter	apFAB211	new created
160	promoter	apFAB212	new created

Overview BIOFAB parts (5)

Number	Type	Sequence Name	Category
161	promoter	apFAB213	new created
162	promoter	apFAB215	new created
163	promoter	apFAB216	new created
164	promoter	apFAB217	new created
165	promoter	apFAB220	new created
166	promoter	apFAB221	new created
167	promoter	apFAB224	new created
168	promoter	apFAB225	new created
169	promoter	apFAB226	new created
170	promoter	apFAB227	new created
171	promoter	apFAB228	new created
172	promoter	apFAB229	new created
173	promoter	apFAB230	new created
174	promoter	apFAB231	new created
175	promoter	apFAB241	new created
176	promoter	apFAB251	new created
177	promoter	apFAB252	new created
178	promoter	apFAB253	new created
179	promoter	apFAB254	new created
180	promoter	apFAB255	priorly used
181	promoter	apFAB256	new created
182	promoter	apFAB257	new created
183	promoter	apFAB258	new created
184	promoter	apFAB259	new created
185	promoter	apFAB260	new created
186	promoter	apFAB261	new created
187	promoter	apFAB262	new created
188	promoter	apFAB263	new created
189	promoter	apFAB264	new created
190	promoter	apFAB265	new created
191	promoter	apFAB266	new created
192	promoter	apFAB267	new created
193	promoter	apFAB268	new created
194	promoter	apFAB270	new created
195	promoter	apFAB271	new created
196	promoter	apFAB272	new created
197	promoter	apFAB273	new created
198	promoter	apFAB274	new created
199	promoter	apFAB276	new created
200	promoter	apFAB278	new created

Overview BIOFAB parts (6)

Number	Type	Sequence Name	Category
201	promoter	apFAB279	new created
202	promoter	apFAB280	new created
203	promoter	apFAB281	new created
204	promoter	apFAB282	new created
205	promoter	apFAB284	new created
206	promoter	apFAB285	new created
207	promoter	apFAB286	new created
208	promoter	apFAB287	new created
209	promoter	apFAB293	new created
210	promoter	apFAB294	new created
211	promoter	apFAB295	new created
212	promoter	apFAB296	new created
213	promoter	apFAB297	new created
214	promoter	apFAB298	new created
215	promoter	apFAB299	new created
216	promoter	apFAB300	new created
217	promoter	apFAB301	new created
218	promoter	apFAB302	new created
219	promoter	apFAB303	new created
220	promoter	apFAB304	new created
221	promoter	apFAB305	new created
222	promoter	apFAB306	new created
223	promoter	apFAB307	new created
224	promoter	apFAB308	new created
225	promoter	apFAB309	new created
226	promoter	apFAB310	new created
227	promoter	apFAB311	new created
228	promoter	apFAB312	new created
229	promoter	apFAB313	new created
230	promoter	apFAB314	new created
231	promoter	apFAB315	new created
232	promoter	apFAB316	new created
233	promoter	apFAB317	new created
234	promoter	apFAB318	priorly used
235	promoter	apFAB319	new created
236	promoter	apFAB321	new created
237	promoter	apFAB322	new created
238	promoter	apFAB323	new created
239	promoter	apFAB324	new created
240	promoter	apFAB325	new created

Overview BIOFAB parts (7)

Number	Type	Sequence Name	Category
241	promoter	apFAB326	new created
242	promoter	apFAB327	new created
243	promoter	apFAB329	new created
244	promoter	apFAB331	new created
245	promoter	apFAB332	new created
246	promoter	apFAB333	new created
247	promoter	apFAB334	new created
248	promoter	apFAB335	new created
249	promoter	apFAB337	new created
250	promoter	apFAB338	new created
251	promoter	apFAB339	new created
252	promoter	apFAB340	new created
253	promoter	apFAB341	new created
254	promoter	apFAB342	new created
255	promoter	apFAB343	new created
256	promoter	apFAB345	new created
257	promoter	apFAB346	new created
258	promoter	apFAB347	new created
259	BCD	apFAB535	new created
260	BCD	apFAB536	new created
261	BCD	apFAB537	new created
262	BCD	apFAB538	new created
263	BCD	apFAB539	new created
264	BCD	apFAB540	new created
265	BCD	apFAB541	new created
266	BCD	apFAB542	new created
267	BCD	apFAB543	new created
268	BCD	apFAB544	new created
269	BCD	apFAB545	new created
270	BCD	apFAB546	new created
271	BCD	apFAB547	new created
272	BCD	apFAB548	new created
273	BCD	apFAB549	new created
274	BCD	apFAB550	new created
275	BCD	apFAB551	new created
276	BCD	apFAB552	new created
277	BCD	apFAB553	new created
278	BCD	apFAB554	new created
279	BCD	apFAB555	new created
280	BCD	apFAB556	new created

Overview BIOFAB parts (8)

Number	Type	Sequence Name	Category
281	BCD	apFAB557	new created
282	BCD	apFAB558	new created
283	BCD	apFAB559	new created
284	BCD	apFAB560	new created
285	BCD	apFAB561	new created
286	BCD	apFAB562	new created
287	BCD	apFAB563	new created
288	BCD	apFAB564	new created
289	BCD	apFAB565	new created
290	BCD	apFAB566	new created
291	BCD	apFAB567	new created
292	BCD	apFAB568	new created
293	BCD	apFAB569	new created
294	BCD	apFAB570	new created
295	BCD	apFAB571	new created
296	BCD	apFAB572	new created
297	BCD	apFAB573	new created
298	BCD	apFAB574	new created
299	BCD	apFAB575	new created
300	BCD	apFAB576	new created
301	BCD	apFAB577	new created
302	BCD	apFAB578	new created
303	BCD	apFAB579	new created
304	BCD	apFAB580	new created
305	BCD	apFAB581	new created
306	BCD	apFAB582	new created
307	BCD	apFAB583	new created
308	BCD	apFAB584	new created
309	BCD	apFAB585	new created
310	BCD	apFAB586	new created
311	BCD	apFAB587	new created
312	BCD	apFAB588	new created
313	BCD	apFAB589	new created
314	BCD	apFAB590	new created
315	BCD	apFAB591	new created
316	BCD	apFAB592	new created
317	BCD	apFAB593	new created
318	BCD	apFAB594	new created
319	BCD	apFAB595	new created
320	BCD	apFAB596	new created

Overview BIOFAB parts (9)

Number	Type	Sequence Name	Category
321	BCD	apFAB597	new created
322	BCD	apFAB598	new created
323	BCD	apFAB599	new created
324	BCD	apFAB600	new created
325	BCD	apFAB601	new created
326	BCD	apFAB602	new created
327	BCD	apFAB603	new created
328	BCD	apFAB604	new created
329	BCD	apFAB605	new created
330	BCD	apFAB606	new created
331	BCD	apFAB607	new created
332	BCD	apFAB608	new created
333	BCD	apFAB609	new created
334	BCD	apFAB610	new created
335	BCD	apFAB611	new created
336	BCD	apFAB612	new created
337	BCD	apFAB613	new created
338	BCD	apFAB614	new created
339	BCD	apFAB615	new created
340	BCD	apFAB616	new created
341	BCD	apFAB617	new created
342	BCD	apFAB618	new created
343	BCD	apFAB619	new created
344	BCD	apFAB620	new created
345	BCD	apFAB621	new created
346	BCD	apFAB622	new created
347	BCD	apFAB623	new created
348	BCD	apFAB624	new created
349	BCD	apFAB625	new created
350	BCD	apFAB626	new created
351	BCD	apFAB627	new created
352	BCD	apFAB628	new created
353	BCD	apFAB629	new created
354	BCD	apFAB630	new created
355	BCD	apFAB631	new created
356	BCD	apFAB632	new created
357	BCD	apFAB633	new created
358	BCD	apFAB634	new created
359	BCD	apFAB635	new created
360	BCD	apFAB636	new created

Overview BIOFAB parts (10)

Number	Type	Sequence Name	Category
361	BCD	apFAB637	new created
362	BCD	apFAB638	new created
363	BCD	apFAB639	new created
364	BCD	apFAB640	new created
365	BCD	apFAB641	new created
366	BCD	apFAB642	new created
367	BCD	apFAB643	new created
368	BCD	apFAB644	new created
369	BCD	apFAB645	new created
370	BCD	apFAB646	new created
371	BCD	apFAB647	new created
372	BCD	apFAB648	new created
373	BCD	apFAB649	new created
374	BCD	apFAB650	new created
375	BCD	apFAB651	new created
376	BCD	apFAB652	new created
377	BCD	apFAB653	new created
378	BCD	apFAB654	new created
379	BCD	apFAB655	new created
380	BCD	apFAB656	new created
381	BCD	apFAB657	new created
382	BCD	apFAB658	new created
383	BCD	apFAB659	new created
384	BCD	apFAB660	new created
385	BCD	apFAB661	new created
386	BCD	apFAB662	new created
387	BCD	apFAB663	new created
388	BCD	apFAB664	new created
389	BCD	apFAB665	new created
390	BCD	apFAB666	new created
391	BCD	apFAB667	new created
392	BCD	apFAB668	new created
393	BCD	apFAB669	new created
394	BCD	apFAB670	new created
395	BCD	apFAB671	new created
396	BCD	apFAB672	new created
397	BCD	apFAB673	new created
398	BCD	apFAB674	new created
399	BCD	apFAB675	new created
400	BCD	apFAB676	new created

Overview BIOFAB parts (11)

Number	Type	Sequence Name	Category
401	BCD	apFAB677	new created
402	BCD	apFAB678	new created
403	BCD	apFAB679	new created
404	BCD	apFAB680	new created
405	terminator	amyA	priorly used
406	terminator	amyA_(L1)	new created
407	terminator	amyA_(L2)	new created
408	terminator	amyA_(S)	new created
409	terminator (LET)	attCaadA7	new created
410	d_terminator	attCaadA7_l_s80GC0.3	new created
411	terminator	BBa_B1002	priorly used
412	terminator (LET)	BBa_B1002_(L)	new created
413	terminator (LET)	BBa_B1002_(LS)	new created
414	terminator	BBa_B1006	priorly used
415	terminator (LET)	BBa_B1006_(L)	new created
416	terminator	BBa_B1006_(S1)	new created
417	terminator (LFFT)	BBa_B1006_(S2)	new created
418	terminator (LFFT)	BBa_B1006_noH	new created
419	terminator	BBa_B1006_U10	new created
420	terminator	BBa_B1006_U4	new created
421	terminator (LET)	BBa_B1006_U4_(L)	new created
422	terminator	BBa_B1006_U6	new created
423	terminator_mutant	BBa_B1006_T_noSL	new created
424	terminator	crp	priorly used
425	terminator	crp_[min]	priorly used
426	d_terminator	crp_T_linker_his_T	new created
427	d_terminator	EOU_double_term	priorly used
428	terminator (LFFT/ET)	his	priorly used
429	terminator	his_[min]	new created
430	terminator	his_[min]_(S)	new created
431	terminator (LFFT)	his_var	new created
432	terminator (LFFT)	his_var_(L)	new created
433	d_terminator	his_T_min_l_crp_T_min	new created
434	terminator	ilvGEDA	priorly used
435	terminator	ilvGEDA_[min]	priorly used
436	terminator (LFFT)	lambda_tR2	priorly used
437	terminator (LET)	lambda_tR2_(L)	new created
438	terminator (LFFT)	lambda_tR2_[min]	new created
439	terminator (LFFT)	lambda_tR2.1	priorly used
440	terminator_mutant	lambda_tR2_S	priorly used

Overview BIOFAB parts (12)

Number	Type	Sequence Name	Category
441	terminator_mutant	lambdatR2_S_min	priorly used
442	terminator	M13_central	priorly used
443	terminator	M13_central_[min]	priorly used
444	d_terminator	M13_central_T_l_rrnD_T1	new created
445	terminator (ET)	RNAI	priorly used
446	terminator (ET)	rnpB_T1	priorly used
447	terminator (LFFT)	rnpB_T1_[min]	priorly used
448	terminator	rpoC	priorly used
449	terminator (LFFT)	rpoC_[min]	priorly used
450	terminator (ET)	rrnA_T1	priorly used
451	terminator (ET)	rrnB_T1	priorly used
452	terminator	rrnB_T1_[min]	priorly used
453	terminator (ET)	rrnD_T1	priorly used
454	terminator	rrnD_T1_[min]	priorly used
455	spacer	spacer_rd1.1	not considered
456	spacer	spacer_rd1.2	not considered
457	spacer	spacer_rd2	not considered
458	spacer	spacer_small	not considered
459	spacer	spacer40GC0.32	not considered
460	spacer	spacer50GC0.23	not considered
461	spacer	spacer80GC0.3	not considered
462	d_terminator	s80GC0.3_l_attCaadA7	new created
463	spacer	spacer80GC0.5	not considered
464	spacer	spacer80GC0.7	not considered
465	terminator	T21	new created
466	terminator	T3_early	new created
467	terminator	T7_early	priorly used
468	terminator	tetAC	priorly used
469	terminator (ET)	tonB	priorly used
470	terminator (LET)	tonB_[min]	priorly used
471	d_terminator	tonB_T_min_l_his_T_min	new created
472	terminator	trp	priorly used
473	terminator	trp_(L)	priorly used
474	terminator	trp_[min]	priorly used
475	terminator	trp_1419	priorly used
476	terminator	trp_L126	priorly used
477	terminator	trp_L126_(L)	priorly used
478	terminator_mutant	trp_att_L126_LST	new created

A.3 Variation in citation rates per author per technology

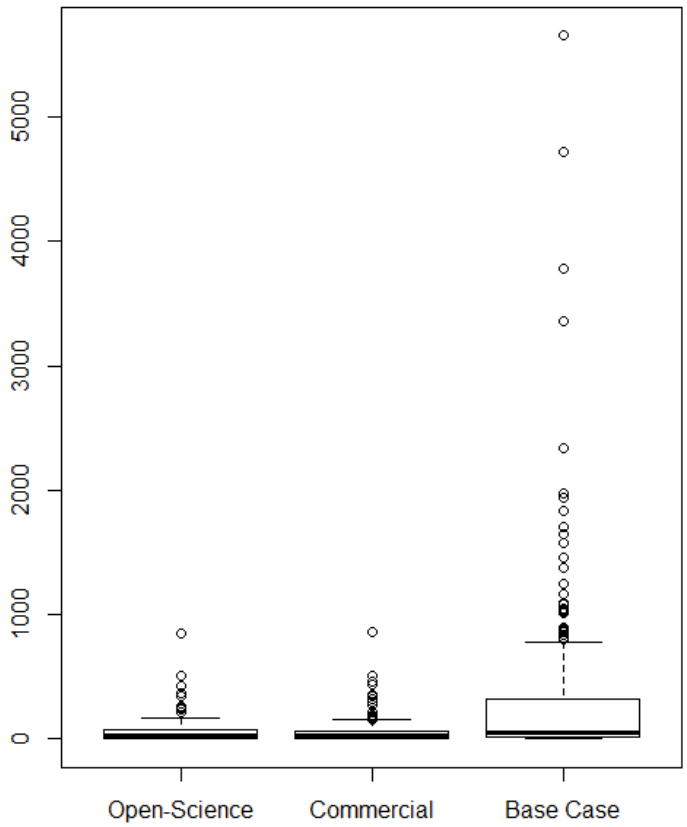


Figure A.1: Box-plot on total citations per author categorized by used research tool

A.4 Validation of instrument variables for multinomial treatment regression

The complete procedure in stata to validate choice of instrument variables for multinomial treatment regression calculation in chapter 6.

```
. ivreg2 citationslog age agesquared pagecount titlelength numberreferences countauthors groupsumpriorcitslog priorpubli
> cationaffid (sangamo consortium  = grouppriorsangamo_au grouppriorconsortium_au group_priorcoauthor_sang group_priorco
> author_con group_prioraff_sang group_prioraff_con  ), first

First-stage regressions
```

First-stage regression of sangamo:

Statistics consistent for homoskedasticity only
Number of obs = **396**

sangamo	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
grouppriorsangamo_au	**-.2030559**	.0751231	**-2.70**	**0.007**	**-.3507636**	**-.0553481**
grouppriorconsortium_au	**-.1632926**	.0740317	**-2.21**	**0.028**	**-.3088544**	**-.0177307**
group_priorcoauthor_sang	**.1631886**	.0558317	**2.92**	**0.004**	**.0534118**	**.2729655**
group_priorcoauthor_con	**-.1576612**	.0534149	**-2.95**	**0.003**	**-.2626861**	**-.0526363**
group_prioraff_sang	**.0655625**	.0588604	**1.11**	**0.266**	**-.0501695**	**.1812944**
group_prioraff_con	**-.0542975**	.0661437	**-0.82**	**0.412**	**-.18435**	**.0757549**
age	**-.0311091**	.0227898	**-1.37**	**0.173**	**-.0759187**	**.0137005**
agesquared	**-.0005193**	.0019325	**-0.27**	**0.788**	**-.0043191**	**.0032804**
pagecount	**.0111231**	.007585	**1.47**	**0.143**	**-.0037906**	**.0260367**
titlelength	**.0007526**	.0006701	**1.12**	**0.262**	**-.0005649**	**.0020702**
numberreferences	**-.0040068**	.0014422	**-2.78**	**0.006**	**-.0068425**	**-.0011711**
countauthors	**.0086274**	.0053339	**1.62**	**0.107**	**-.0018601**	**.019115**
groupsumpriorcitslog	**-.0076705**	.0187919	**-0.41**	**0.683**	**-.0446194**	**.0292783**
priorpublicationaffid	**-6.37e-07**	**4.81e-07**	**-1.33**	**0.186**	**-1.58e-06**	**3.08e-07**
_cons	**.4752816**	.1807811	**2.63**	**0.009**	**.1198279**	**.8307352**

F test of excluded instruments:
  F(6, 381) = **6.81**
  Prob > F = **0.0000**
Sanderson-Windmeijer multivariate F test of excluded instruments:
  F(5, 381) = **6.72**
  Prob > F = **0.0000**

First-stage regression of consortium:

Statistics consistent for homoskedasticity only
Number of obs = **396**

consortium	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
grouppriorsangamo_au	**-.1073152**	.0574652	**-1.87**	**0.063**	**-.2203037**	**.0056734**
grouppriorconsortium_au	**.1487506**	.0566303	**2.63**	**0.009**	**.0374036**	**.2600977**
group_priorcoauthor_sang	**.009238**	.0427083	**0.22**	**0.829**	**-.0747355**	**.0932115**
group_priorcoauthor_con	**-.0029692**	.0408596	**-0.07**	**0.942**	**-.0833077**	**.0773692**
group_prioraff_sang	**-.0562256**	.0450251	**-1.25**	**0.213**	**-.1447544**	**.0323031**
group_prioraff_con	**.0435718**	.0505964	**0.86**	**0.390**	**-.0559114**	**.143055**
age	**.0418985**	.017433	**2.40**	**0.017**	**.0076215**	**.0761755**
agesquared	**-.0039034**	.0014783	**-2.64**	**0.009**	**-.00681**	**-.0009968**
pagecount	**.0019056**	.0058021	**0.33**	**0.743**	**-.0095025**	**.0133138**
titlelength	**.0000688**	.0005126	**0.13**	**0.893**	**-.0009391**	**.0010766**
numberreferences	**.0013748**	.0011032	**1.25**	**0.213**	**-.0007944**	**.003544**
countauthors	**.0079265**	.0040802	**1.94**	**0.053**	**-.0000959**	**.0159489**
groupsumpriorcitslog	**-.0114084**	.0143748	**-0.79**	**0.428**	**-.0396723**	**.0168555**
priorpublicationaffid	**7.98e-07**	**3.68e-07**	**2.17**	**0.031**	**7.52e-08**	**1.52e-06**
_cons	**-.0355587**	.1382879	**-0.26**	**0.797**	**-.3074618**	**.2363444**

F test of excluded instruments:
  F(6, 381) = **5.20**
  Prob > F = **0.0000**
Sanderson-Windmeijer multivariate F test of excluded instruments:
  F(5, 381) = **5.35**
  Prob > F = **0.0001**

Summary results for first-stage regressions

		(Underid)		(Weak id)
Variable	F(**6**, **381**) P-val	SW Chi-sq(**5**) P-val		SW F(**5**, **381**)
sangamo	**6.81** **0.0000**	**34.91** **0.0000**		**6.72**
consortium	**5.20** **0.0000**	**27.81** **0.0000**		**5.35**

Stock-Yogo weak ID F test critical values for single endogenous regressor:
 5% maximal IV relative bias **19.28**
 10% maximal IV relative bias **11.12**
 20% maximal IV relative bias **6.76**
 30% maximal IV relative bias **5.15**
 10% maximal IV size **29.18**
 15% maximal IV size **16.23**
 20% maximal IV size **11.72**
 25% maximal IV size **9.38**
Source: Stock-Yogo (2005). Reproduced by permission.
NB: Critical values are for Sanderson-Windmeijer F statistic.

Underidentification test
Ho: matrix of reduced form coefficients has rank=K1-1 (underidentified)
Ha: matrix has rank=K1 (identified)
Anderson canon. corr. LM statistic Chi-sq(**5**)=**25.00** P-val=**0.0001**

Weak identification test
Ho: equation is weakly identified
Cragg-Donald Wald F statistic **4.28**

Stock-Yogo weak ID test critical values for K1=2 and L1=6:
 5% maximal IV relative bias **15.72**
 10% maximal IV relative bias **9.48**
 20% maximal IV relative bias **6.08**
 30% maximal IV relative bias **4.78**
 10% maximal IV size **21.68**
 15% maximal IV size **12.33**
 20% maximal IV size **9.10**
 25% maximal IV size **7.42**
Source: Stock-Yogo (2005). Reproduced by permission.

```
Weak-instrument-robust inference
Tests of joint significance of endogenous regressors B1 in main equation
Ho: B1=0 and orthogonality conditions are valid
Anderson-Rubin Wald test          F(6,381)=      2.41     P-val=0.0266
Anderson-Rubin Wald test          Chi-sq(6)=     15.05    P-val=0.0199
Stock-Wright LM S statistic       Chi-sq(6)=     14.49    P-val=0.0246

Number of observations              N  =         396
Number of regressors                K  =          11
Number of endogenous regressors     K1 =           2
Number of instruments               L  =          15
Number of excluded instruments      L1 =           6

IV (2SLS) estimation
--------------------

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

                                                    Number of obs =       396
                                                    F( 10,   385) =     35.01
                                                    Prob > F      =    0.0000
Total (centered) SS     =  812.0918471              Centered R2   =    0.4119
Total (uncentered) SS   =  4272.147555              Uncentered R2 =    0.8882
Residual SS             =  477.5726302              Root MSE      =     1.098

------------------------------------------------------------------------------
 citationslog |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
      sangamo |   -1.07371   .4801999    -2.24   0.025    -2.014885   -.1325359
   consortium |  -.9952059   .7187315    -1.38   0.166    -2.403894    .4134818
          age |   .7862856   .0726349    10.83   0.000     .6439239    .9286473
   agesquared |  -.0507171   .0061245    -8.28   0.000     -.062721   -.0387132
    pagecount |   .0326484   .0224403     1.45   0.146    -.0113338    .0766306
  titlelength |  -.0042354    .001955    -2.17   0.030     -.008067   -.0004037
numberreferences |   .0027354   .0044675     0.61   0.540    -.0060207    .0114915
  countauthors |   .0736074   .0136739     5.38   0.000     .0468071    .1004077
groupsumpriorcitslog |   .1353309   .0518452     2.61   0.009     .0337162    .2369456
priorpublicationaffid |   2.20e-06   1.58e-06     1.40   0.163    -8.90e-07   5.29e-06
         _cons |  -.6561967   .5620203    -1.17   0.243    -1.757736    .4453429
------------------------------------------------------------------------------

Underidentification test (Anderson canon. corr. LM statistic):          24.998
                                                    Chi-sq(5) P-val =    0.0001
------------------------------------------------------------------------------
Weak identification test (Cragg-Donald Wald F statistic):                4.279
Stock-Yogo weak ID test critical values:  5% maximal IV relative bias    15.72
                                         10% maximal IV relative bias     9.48
                                         20% maximal IV relative bias     6.08
                                         30% maximal IV relative bias     4.78
                                         10% maximal IV size             21.68
                                         15% maximal IV size             12.33
                                         20% maximal IV size              9.10
                                         25% maximal IV size              7.42
Source: Stock-Yogo (2005).  Reproduced by permission.
------------------------------------------------------------------------------
Sargan statistic (overidentification test of all instruments):           6.340
                                                    Chi-sq(4) P-val =    0.1752
------------------------------------------------------------------------------
Instrumented:         sangamo consortium
Included instruments: age agesquared pagecount titlelength numberreferences
                      countauthors groupsumpriorcitslog priorpublicationaffid
Excluded instruments: grouppriorsangamo_au grouppriorconsortium_au
                      group_priorcoauthor_sang group_priorcoauthor_con
                      group_prioraff_sang group_prioraff_con
------------------------------------------------------------------------------

. ivhettest
IV heteroskedasticity test(s) using levels of IVs only
Ho: Disturbance is homoskedastic
    Pagan-Hall general test statistic   :  21.768  Chi-sq(14) P-value = 0.0835

. ivreg2 citationslog age agesquared pagecount titlelength numberreferences countauthors groupsumpriorcitslog priorpubli
> cationaffid (sangamo consortium  = grouppriorsangamo_au grouppriorconsortium_au group_priorcoauthor_sang group_priorco
> author_con group_prioraff_sang group_prioraff_con  ), first liml

First-stage regressions
-----------------------

First-stage regression of sangamo:

Statistics consistent for homoskedasticity only
Number of obs =                   396
------------------------------------------------------------------------------
     sangamo |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 grouppriorsangamo_au |  -.2030559   .0751231    -2.70   0.007    -.3507636   -.0553481
grouppriorconsortium_au |  -.1632926   .0740317    -2.21   0.028    -.3088544   -.0177307
group_priorcoauthor_sang |   .1631886   .0558317     2.92   0.004     .0534118    .2729655
group_priorcoauthor_con |  -.1576612   .0534149    -2.95   0.003    -.2626861   -.0526363
group_prioraff_sang |   .0655625   .0588604     1.11   0.266    -.0501695    .1812944
group_prioraff_con |  -.0542975   .0661437    -0.82   0.412      -.18435    .0757549
          age |  -.0311091   .0227898    -1.37   0.173    -.0759187    .0137005
   agesquared |  -.0005193   .0019325    -0.27   0.788    -.0043191    .0032804
    pagecount |   .0111231    .007585     1.47   0.143    -.0037906    .0260367
  titlelength |   .0007526   .0006701     1.12   0.262    -.0005649    .0020702
numberreferences |  -.0040068   .0014422    -2.78   0.006    -.0068425   -.0011711
  countauthors |   .0086274   .0053339     1.62   0.107    -.0018601     .019115
groupsumpriorcitslog |  -.0076705   .0187919    -0.41   0.683    -.0446194    .0292783
priorpublicationaffid |  -6.37e-07   4.81e-07    -1.33   0.186    -1.58e-06   3.08e-07
         _cons |   .4752816   .1807811     2.63   0.009     .1198279    .8307352
------------------------------------------------------------------------------

F test of excluded instruments:
  F(  6,   381) =     6.81
  Prob > F      =   0.0000
Sanderson-Windmeijer multivariate F test of excluded instruments:
  F(  5,   381) =     6.72
  Prob > F      =   0.0000
```

First-stage regression of consortium:

Statistics consistent for homoskedasticity only
Number of obs = **396**

consortium	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
grouppriorsangamo_au	**-.1073152**	**.0574652**	**-1.87**	**0.063**	**-.2203037**	**.0056734**
grouppriorconsortium_au	**.1487506**	**.0566303**	**2.63**	**0.009**	**.0374036**	**.2600977**
group_priorcoauthor_sang	**.009238**	**.0427083**	**0.22**	**0.829**	**-.0747355**	**.0932115**
group_priorcoauthor_con	**-.0029692**	**.0408596**	**-0.07**	**0.942**	**-.0833077**	**.0773692**
group_prioraff_sang	**-.0562256**	**.0450251**	**-1.25**	**0.213**	**-.1447544**	**.0323031**
group_prioraff_con	**.0435718**	**.0505964**	**0.86**	**0.390**	**-.0559114**	**.143055**
age	**.0418985**	**.017433**	**2.40**	**0.017**	**.0076215**	**.0761755**
agesquared	**-.0039034**	**.0014783**	**-2.64**	**0.009**	**-.00681**	**-.0009968**
pagecount	**.0019056**	**.0058021**	**0.33**	**0.743**	**-.0095025**	**.0133138**
titlelength	**.0000688**	**.0005126**	**0.13**	**0.893**	**-.0009391**	**.0010766**
numberreferences	**.0013748**	**.0011032**	**1.25**	**0.213**	**-.0007944**	**.003544**
countauthors	**.0079265**	**.0040802**	**1.94**	**0.053**	**-.0000959**	**.0159489**
groupsumpriorcitslog	**-.0114084**	**.0143748**	**-0.79**	**0.428**	**-.0396723**	**.0168555**
priorpublicationaffid	**7.98e-07**	**3.68e-07**	**2.17**	**0.031**	**7.52e-08**	**1.52e-06**
_cons	**-.0355587**	**.1382879**	**-0.26**	**0.797**	**-.3074618**	**.2363444**

F test of excluded instruments:
F(6, 381) = **5.20**
Prob > F = **0.0000**
Sanderson-Windmeijer multivariate F test of excluded instruments:
F(5, 381) = **5.35**
Prob > F = **0.0001**

Summary results for first-stage regressions

Variable	F(6, 381)	P-val	(Underid) SW Chi-sq(5)	P-val	(Weak id) SW F(5, 381)
sangamo	**6.81**	**0.0000**	**34.91**	**0.0000**	**6.72**
consortium	**5.20**	**0.0000**	**27.81**	**0.0000**	**5.35**

Stock-Yogo weak ID F test critical values for single endogenous regressor:
10% maximal LIML size **4.45**
15% maximal LIML size **3.34**
20% maximal LIML size **2.87**
25% maximal LIML size **2.61**
Source: Stock-Yogo (2005). Reproduced by permission.
NB: Critical values are for i.i.d. errors only.

Underidentification test
Ho: matrix of reduced form coefficients has rank=K1-1 (underidentified)
Ha: matrix has rank=K1 (identified)
Anderson canon. corr. LM statistic Chi-sq(**5**)=**25.00** P-val=**0.0001**

Weak identification test
Ho: equation is weakly identified
Cragg-Donald Wald F statistic **4.28**

Stock-Yogo weak ID test critical values for K1=2 and L1=6:
10% maximal LIML size **4.06**
15% maximal LIML size **2.95**
20% maximal LIML size **2.63**
25% maximal LIML size **2.46**
Source: Stock-Yogo (2005). Reproduced by permission.

Weak-instrument-robust inference
Tests of joint significance of endogenous regressors B1 in main equation
Ho: B1=0 and orthogonality conditions are valid
Anderson-Rubin Wald test F(**6,381**)= **2.41** P-val=**0.0266**
Anderson-Rubin Wald test Chi-sq(**6**)= **15.05** P-val=**0.0199**
Stock-Wright LM S statistic Chi-sq(**6**)= **14.49** P-val=**0.0246**

Number of observations N = **396**
Number of regressors K = **11**
Number of endogenous regressors K1 = **2**
Number of instruments L = **15**
Number of excluded instruments L1 = **6**

LIML estimation
k =1.01543
lambda =1.01543

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

Number of obs = **396**
F(10, 385) = **31.33**
Prob > F = **0.0000**
Total (centered) SS = **812.0918471** Centered R2 = **0.3399**
Total (uncentered) SS = **4272.147555** Uncentered R2 = **0.8745**
Residual SS = **536.0490798** Root MSE = **1.163**

citationslog	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sangamo	**-1.357323**	**.5610373**	**-2.42**	**0.016**	**-2.456936**	**-.2577101**
consortium	**-1.328763**	**.8625438**	**-1.54**	**0.123**	**-3.019318**	**.3617915**
age	**.7850874**	**.0787019**	**9.98**	**0.000**	**.6308345**	**.9393403**
agesquared	**-.0516125**	**.0066888**	**-7.72**	**0.000**	**-.0647224**	**-.0385026**
pagecount	**.03689**	**.024037**	**1.53**	**0.125**	**-.0102216**	**.0840016**
titlelength	**-.0039144**	**.0020884**	**-1.87**	**0.061**	**-.0080077**	**.0001788**
numberreferences	**.0020367**	**.0047966**	**0.42**	**0.671**	**-.0073643**	**.0114378**
countauthors	**.0757777**	**.0146024**	**5.19**	**0.000**	**.0471574**	**.1043979**
groupsumpriorcitslog	**.126868**	**.0555463**	**2.28**	**0.022**	**.0179992**	**.2357368**
priorpublicationaffid	**2.31e-06**	**1.73e-06**	**1.34**	**0.180**	**-1.07e-06**	**5.70e-06**
_cons	**-.505993**	**.6085835**	**-0.83**	**0.406**	**-1.698795**	**.6868087**

Underidentification test (Anderson canon. corr. LM statistic): **24.998**
Chi-sq(**5**) P-val = **0.0001**

Weak identification test (Cragg-Donald Wald F statistic): **4.279**

```
Stock-Yogo weak ID test critical values: 10% maximal LIML size            4.06
                                         15% maximal LIML size            2.95
                                         20% maximal LIML size            2.63
                                         25% maximal LIML size            2.46
Source: Stock-Yogo (2005).  Reproduced by permission.
------------------------------------------------------------------------------
Sargan statistic (overidentification test of all instruments):           6.018
                                                     Chi-sq(4) P-val =    0.1978
------------------------------------------------------------------------------
Anderson-Rubin statistic (overidentification test of all instruments):   6.064
                                                     Chi-sq(4) P-val =    0.1944
------------------------------------------------------------------------------
Instrumented:         sangamo consortium
Included instruments: age agesquared pagecount titlelength numberreferences
                      countauthors groupsumpriorcitslog priorpublicationaffid
Excluded instruments: groupprior sangamo_au groupprior consortium_au
                      group_priorcoauthor_sang group_priorcoauthor_con
                      group_prioraff_sang group_prioraff_con
------------------------------------------------------------------------------
```

```
. condivreg citationslog age agesquared pagecount titlelength numberreferences countauthors groupsumpriorcitslog priorpu
> blicationaffid ( technologythree  = groupprior sangamo_au groupprior consortium_au group_priorcoauthor_sang group_priorc
> oauthor_con group_prioraff_sang group_prioraff_con ), liml interval

Instrumental variables (LIML) regression

First-stage results                                   Number of obs =     396
                                                      Wald chi2( 9) =  324.94
F(  6,   381) =    4.28                               Prob > w      =  0.0000
Prob > F      =  0.0003                               R-squared     =  0.3479
R-squared     =  0.1419                               Adj R-squared =  0.3327
Adj R-squared =  0.1103                               Root MSE      =   1.171

------------------------------------------------------------------------------
 citationslog |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
technologythree |   .9402081   .4187902     2.25   0.025     .1168126    1.763604
          age |    .839001   .0675867    12.41   0.000     .7061167    .9718852
    agesquared |  -.0551643   .0062865    -8.78   0.000    -.0675244   -.0428041
    pagecount |   .0332847   .0236651     1.41   0.160    -.0132439    .0798134
  titlelength |  -.0041995   .0020595    -2.04   0.042    -.0082487   -.0001502
numberreferences |  .0045144   .0043253     1.04   0.297    -.0039898    .0130186
 countauthors |   .0743559   .0144588     5.14   0.000      .045928    .1027837
groupsumpriorcitslog |  .1228792   .0554528     2.22   0.027     .0138519    .2319065
priorpublicationaffid |  3.54e-06   1.50e-06     2.36   0.019     5.93e-07    6.48e-06
        _cons |  -3.561585   1.164582    -3.06   0.002    -5.851304   -1.271866
------------------------------------------------------------------------------
Instrumented:  technologythree
Instruments:   age agesquared pagecount titlelength numberreferences
               countauthors groupsumpriorcitslog priorpublicationaffid
               groupprior sangamo_au groupprior consortium_au
               group_priorcoauthor_sang group_priorcoauthor_con
               group_prioraff_sang group_prioraff_con
Confidence set and p-value for technologyt~e are based on normal approximation
------------------------------------------------------------------------------
```

```
------------------------------------------------------------------------------
             Coverage-corrected confidence interval and p-value
                  for Ho: _b[technologythree] = 0
              LIML estimate of _b[technologythree] = .9402081

------------------------------------------------------------------------------
Test                          Confidence Interval                   p-value
------------------------------------------------------------------------------
Conditional LR              [ .1665058,  2.373742]                    0.0168
------------------------------------------------------------------------------
```

Table A.4: Regression table for 2nd step robustness check with SJR

	(1) Citations
Open Science Research Tool	-0.306^{***} (-5.99)
Commercial Research Tool	-0.750^{***} (-18.39)
SJR	0.377^{***} (34.78)
Age of article	0.748^{***} (65.85)
Age squared of article	-0.0450^{***} (-50.18)
Page count	0.0467^{***} (8.63)
Title length	-0.00673^{***} (-20.64)
Number of references	-0.000434 (-0.70)
Count of authors	0.0529^{***} (22.10)
Authors' prior citations	0.0902^{***} (11.02)
Constant	-0.449^{***} (-5.66)

t statistics in parentheses

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

Bibliography

Abbott, Andrew (1988). *The system of professions: An essay on the division of labor*. Chicago, IL: University of Chicago Press.

Abramovitz, Moses (1956). "Resource and output trends in the United States since 1870". In: *NBER Chapters*, pp. 1–23.

Addgene (2017). *Addgene mission statement*. URL: https://www.addgene.org/mission/ (visited on 08/19/2017).

Alberts, Bruce, Dennis Bray, Karen Hopkin, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter (2013). *Essential cell biology*. London: Garland Science.

Aldrich, Howard E and C Marlene Fiol (1994). "Fools rush in? The institutional context of industry creation". In: *Academy of Management Review* 19.4, pp. 645–670.

Allison, Paul D (1995). *Survival analysis using SAS: A practical guide*. New York: SAS Institute.

Alston, Lee J, Gary D Libecap, and Bernardo Mueller (1999). *Titles, conflict, and land use: The development of property rights and land reform on the Brazilian amazon frontier*. Ann Arbor, MI: University of Michigan Press.

Alston, Lee J, Gary D Libecap, and Robert Schneider (1996). "The determinants and impact of property rights: Land titles on the Brazilian frontier". In: *The Journal of Law, Economics, and Organization* 12.1, pp. 25–61.

Altshuler, Jill S, Erin Balogh, Anna D Barker, Stephen L Eck, Stephen H Friend, Geoffrey S Ginsburg, Roy S Herbst, Sharyl J Nass, Christopher M Streeter, and John A Wagner (2010). "Opening up to precompetitive collaboration". In: *Science translational medicine* 2.52, p. 52cm26.

Anand, Narasimhan and Mary R Watson (2004). "Tournament rituals in the evolution of fields: The case of the Grammy Awards". In: *Academy of Management Journal* 47.1, pp. 59–80.

Anderson, Terry L and Peter J Hill (1975). "The evolution of property rights: A study of the American West". In: *The Journal of Law and Economics* 18.1, pp. 163–179.

Andrews, A B (1990). "Interdisciplinary and interorganizational collaboration". In: *Encyclopedia of Social Work*. Ed. by L. Ginsberg. Silver Spring, MD: National Association of Social Workers Press, pp. 1479–1489.

Ardal, Christine, Annette Alstadsæter, and John-Arne Røttingen (Dec. 2011). "Common characteristics of open source software development and applicability for drug discovery: A systematic review". In: *Health Research Policy and Systems* 9.1, p. 36.

Argyres, Nicholas S and Julia Porter Liebeskind (1998). "Privatizing the intellectual commons: Universities and the commercialization of biotechnology". In: *Journal of Economic Behavior and Organization* 35.4, pp. 427–454.

Arrow, Kenneth (1962). "Economic welfare and the allocation of resources for invention". In: *The Rate and Direction of Inventive Activity* 1.1, S. 609–626.

Asavathiratham, Chalee, Sandip Roy, Bernard Lesieutre, and George Verghese (2001). "The influence model". In: *IEEE Control Systems* 21.6, pp. 52–64.

Ashforth, Blake E and Barrie W Gibbs (1990). "The double-edge of organizational legitimation". In: *Organization science* 1.2, pp. 177–194.

Bailey, Claire, Heather Metcalf, and Brian Crook (2012). *Synthetic biology: A review of the technology, and current and future needs from the regulatory framework in Great Britain. Research Report 944.* Tech. rep. Health and Safety Executive.

Bailey, Norman T J (1975). *The mathematical theory of infectious diseases and its applications*. High Wycombe: Charles Griffin & Company Ltd.

Baker, Monya (2014). "Repositories share key research tools". In: *Nature* 505.7483, p. 272.

Bandura, Albert, C Barr Taylor, S Lloyd Williams, Ivan N Mefford, and Jack D Barchas (1985). "Catecholamine secretion as a function of perceived coping self-efficacy". In: *Journal of Consulting and Clinical Psychology* 53.3, p. 406.

Baron, James N, Frank R Dobbin, and P Devereaux Jennings (1986). "War and peace: The evolution of modern personnel administration in US industry". In: *American journal of Sociology* 92.2, pp. 350–383.

Bascle, G. (2008). "Controlling for endogeneity with instrumental variables in strategic management research". In: *Strategic Organization* 6.3, pp. 285–327.

Battilana, Julie (2006). "Agency and institutions: The enabling role of individuals' social position". In: *Organization* 13.5, pp. 653–676.

Baxter, John D, James L Roberts, Peter H Seeburg, Howard M Goodman, and John Shine (Mar. 1982). *Adrenocorticotropin-lipotropin precursor gene*.

Becker, Gary S (1974). "On the relevance of the new economics of the family". In: *The American Economic Review* 64.2, pp. 317–319.

Beckert, Jens (1999). "Agency, entrepreneurs, and institutional change. The role of strategic choice and institutionalized practices in organizations". In: *Organization Studies* 20.5, pp. 777–799.

Benford, Robert D and David A Snow (2000). "Framing processes and social movements: An overview and assessment". In: *Annual Review of Sociology* 26.1, pp. 611–639.

Benkler, Yochai (Aug. 2004). "Intellectual property: Commons-based strategies and the problems of patents". In: *Science* 305.5687, pp. 1110–1111.

Berger, Peter and Thomas Luckmann (1967). *The social construction of reality: A treatise on the sociology of education*. Garden City, NY.

Bessen, James (2004). "Holdup and licensing of cumulative innovations with private information". In: *Economics Letters* 82.3, pp. 321–326.

Bibikova, Marina, Mary Golic, Kent G. Golic, and Dana Carroll (2002). "Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases". In: *Genetics* 161.3, pp. 1169–1175.

Binswanger, Hans P and Klaus Deininger (1997). "Explaining agricultural and agrarian policies in developing countries".

BioBricks Foundation (2016). *BBF Allstars*. URL: http://2016.igem.org/Team:BioBricks (visited on 08/20/2016).

— (2017a). *Public Domain Chronicle*. URL: https://publicdomainchronicle.org/ (visited on 10/25/2017).

— (2017b). *SB X.0 Series*. URL: www.sb7.info (visited on 08/28/2017).

Boeke, Jef D, George Church, Andrew Hessel, Nancy J Kelley, Adam P Arkin, Yizhi Cai, Rob Carlson, Aravinda Chakravarti, Virginia W Cornish, and Liam Holt (2016). "The genome project-write". In: *Science* 353.6295, pp. 126–127.

Bohannon, John (2016). "Who's downloading pirated papers? Everyone". In: *Science* 352.6285, pp. 508–512.

Bornmann, Lutz, Rüdiger Mutz, Christoph Neuhaus, and Hans Dieter Daniel (2008). "Citation counts for research evaluation: Standards of good practice for analyzing bibliometric data and presenting and interpreting results". In: *Ethics in Science and Environmental Politics* 8.1, pp. 93–102.

Bothner, Matthew S. (May 2003). "Competition and social influence: The diffusion of the sixth-generation processor in the global computer industry". In: *American Journal of Sociology* 108.6, pp. 1175–1210.

Bourdieu, Pierre and Loïc J D Wacquant (1992). *An invitation to reflexive sociology*. Chicago, IL: University of Chicago press.

Boxenbaum, Eva and Julie Battilana (2005). "Importation as innovation: Transposing managerial practices across fields". In: *Strategic Organization* 3.4, pp. 355–383.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin (2009). "Identification of peer effects through social networks". In: *Journal of Econometrics* 150.1, pp. 41–55.

Breiger, Ronald L and John W Mohr (2004). "Institutional logics from the aggregation of organizational networks: Operational procedures for the analysis of counted data". In: *Computational & Mathematical Organization Theory* 10.1, pp. 17–43.

Brent, Roger (2004). "A partnership between biology and engineering". In: *Nature Biotechnology* 22.10, pp. 1211–1214.

Breschi, Stefano, Francesco Lissoni, and Fabio Montobbio (2007). "The scientific productivity of academic inventors: New evidence from Italian data". In: *Economics of Innovation and New Technology* 16.2, pp. 101–118.

Breslow, Norman (1974). "Covariance analysis of censored survival data". In: *Biometrics* 30.1, pp. 89–99.

Brickson, Shelley (2000). "The impact of identity orientation on individual and organizational outcomes in demographically diverse settings". In: *Academy of Management Review* 25.1, pp. 82–101.

Bromley, Daniel W (1989). *Economic interests and institutions: The conceptual foundations of public policy*. New York and Oxford: Basil Blackwell.

Buhr, Sarah (2016). *OpenTrons aims to be the 'PC' of biotech labs*. URL: https://techcrunch.com/2016/02/01/opentrons-aims-to-be-the-pc-of-biotech-labs/ (visited on 08/20/2017).

Burt, Ronald S. (1982). *Toward a structural theory of action: Network models of social structure, perception, and action*. Academic Press, p. 381.

— (1987). "Social contagion and innovation: Cohesion versus structural equivalence". In: *American Journal of Sociology* 92.6, pp. 1287–1335.

— (2004). "Structural holes and good ideas". In: *American Journal of Sociology* 110.2, pp. 349–399.

Calvert, Jane (2012). "Ownership and sharing in synthetic biology: A 'diverse ecology' of the open and the proprietary?" In: *BioSocieties* 7.2, pp. 169–187.

Calvert, Jane and P.-B. Joly (2011). "How did the gene become a chemical compound? The ontology of the gene and the patenting of DNA". In: *Social Science Information* 50.2, pp. 157–177.

Cambia (2006). *Lens.org - Solving the problem of problem solving*. URL: https://www.lens.org/lens/ (visited on 08/22/2017).

— (2009). *PatSeq Finder*. URL: https://www.lens.org/lens/bio/patseqfinder (visited on 10/01/2016).

Cameron, A C and Pravin K. Trivedi (2005). *Microeconometrics: Methods and applications*. Cambridge University Press.

Campbell, Eric G, Brian R Clarridge, Lauren Birenbaum, Stephen Hilgartner, and David Blumenthal (2002). "Evidence from a national survey". In: *American Medical Association* 287.4, pp. 473–480.

Campbell, John L (2004). *Institutional change and globalization*. Princeton, NJ: Princeton University Press.

Casini, Arturo, Marko Storch, Geoffrey S Baldwin, and Tom Ellis (2015). "Bricks and blueprints: Methods and standards for DNA assembly". In: *Nature Reviews Molecular Cell Biology* 16.9, pp. 568–576.

Cerulo, Karen A (1997). "Identity construction: New issues, new directions". In: *Annual Review of Sociology* 23.1, pp. 385–409.

Chandras, Christina, Thomas Weaver, Michael Zouberakis, Damian Smedley, Klaus Schughart, Nadia Rosenthal, John M Hancock, George Kollias, Paul N Schofield, and Vassilis Aidinis (2009). "Models for financial sustainability

of biological databases and resources". In: *Database: The Journal of Biological Databases and Curation* 2009, bap017.

Chandrasekharan, Subhashini, Sapna Kumar, Cory M Valley, and Arti K Rai (Feb. 2009). "Proprietary science, open science and the role of patent disclosure: the case of zinc-finger proteins". In: *Nature Biotechnology* 27.2, pp. 140–144.

Coase, Ronald H (1937). "The nature of the firm". In: *Economica* 4.16, pp. 386–405.

Cole, Jonathan R (2000). "A short history of the use of citations as a measure of the impact of scientific and scholarly work". In: *The Web of Knowledge: A Festschri in Honor of Eugene Garfield* 1.1, pp. 281–300.

Coleman, James Samuel, Elihu Katz, and Herbert Menzel (1966). *Medical innovation: A diffusion study*. New York: Bobbs-Merrill.

Colomy, Paul (1998). "Neofunctionalism and neoinstitutionalism: Human agency and interest in institutional change". In: *Sociological Forum* 13.2, pp. 265–300.

Cook-Deegan, Robert and Christopher Heaney (Sept. 2010). "Patents in genomics and human genetics". In: *Annual Review of Genomics and Human Genetics* 11.1, pp. 383–425.

Cornish, William R, Mike Adcock, and Margaret Llewelyn (2003). *Intellectual property rights (IPRs) and genetics-a study into the impact and management of intellectual property rights within the healthcare sector*. Cambridge: Public Health Genetics Unit, Cambridge Genetics Knowledge Park.

Cox, David R (1972). "Regression models and life tables (with discussion)". In: *Journal of the Royal Statistical Society* 34, pp. 187–220.

— (1975). "Partial likelihood". In: *Biometrika* 62.2, pp. 269–276.

Creed, W E Douglas, Maureen A Scully, and John R Austin (2002). "Clothes make the person? The tailoring of legitimating accounts and the social construction of identity". In: *Organization Science* 13.5, pp. 475–496.

Crick, Francis (Aug. 1970). "Central dogma of molecular biology". In: *Nature* 227.5258, pp. 561–563.

Crouzier, Thomas (2014). "Researchers in the cloud". In: *Trends in Biochemical Sciences* 39.8, pp. 344–346.

Cyert, Richard M and James G March (1963). *A behavioral theory of the firm*. 2nd. Englewood Cliffs, NJ: Wiley-Blackwell.

Czarnitzki, Dirk, Wolfgang Glänzel, and Katrin Hussinger (2007). "Patent and publication activities of German professors: An empirical assessment of their co-activity". In: *Research Evaluation* 16.4, pp. 311–319.

— (2009). "Heterogeneity of patenting activity and its implications for scientific research". In: *Research Policy* 38.1, pp. 26–34.

D'Espaux, Leo, Amit Ghosh, Weerawat Runguphan, Maren Wehrs, Feng Xu, Oliver Konzock, Ishaan Dev, Melissa Nhan, Jennifer Gin, and Amanda Reider Apel (2017). "Engineering high-level production of fatty alcohols by Saccharomyces cerevisiae from lignocellulosic feedstocks". In: *Metabolic Engineering* Jul.42, pp. 115–125.

Dasgupta, Partha and Paul A David (1994). "Toward a new economics of science". In: *Research Policy* 23.5, pp. 487–521.

David, Paul A (2000). "The digital technology boomerang: New intellectual property rights threaten global 'open science'". Stanford.

— (2001). "Will building good fences really make good neighbors in science". In: *Report to European Commission (DG-Research) STRATA-ETAN workshop on IPR aspects of internal collaborations*.

— (2004). "Can 'open science' be protected from the evolving regime of IPR protections?" In: *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, pp. 9–34.

Davis, Gerald F and Henrich R Greve (1997). "Corporate elite networks and governance changes in the 1980s". In: *American Journal of Sociology* 103.1, pp. 1–37.

Davis, Philip M, Bruce V Lewenstein, Daniel H Simon, James G Booth, and Mathew J L Connolly (2008). "Open access publishing, article downloads, and citations: randomised controlled trial". In: *British Medical Journal* 337.1, a568.

Deb, Partha and Pravin K. Trivedi (2006a). "Maximum simulated likelihood estimation of a negative binomial regression model with multinomial endogenous treatment". In: *Stata Journal* 6.2, pp. 246–255.

— (2006b). "Specification and simulated likelihood estimation of a non-normal treatment-outcome model with selection: Application to health care utilization". In: *Econometrics Journal* 9.2, pp. 307–331.

Déjean, Frédérique, Jean-Pascal Gond, and Bernard Leca (2004). "Measuring the unmeasured: An institutional entrepreneur strategy in an emerging industry". In: *Human Relations* 57.6, pp. 741–764.

Delmer, Deborah P, Carol Nottenburg, Greg D Graff, and Alan B Bennett (2003). "Intellectual property resources for international development in agriculture". In: *Plant Physiology* 133.December, pp. 1666–1670.

Demil, Benoît and Jabril Bensédrine (2005). "Processes of legitimization and pressure toward regulation: Corporate conformity and strategic behavior". In: *International Studies of Management & Organization* 35.2, pp. 56–77.

Demsetz, Harold (1967). "Toward a theory of property rights". In: *American Economic Review* 57.2, pp. 347–359.

Dennis, Carina (2004). "Biologists launch 'open-source movement'". In: *Nature* 431.7008, p. 494.

Dew, Nicholas (2006). "Institutional entrepreneurship: A Coasian perspective". In: *The International Journal of Entrepreneurship and Innovation* 7.1, pp. 13–22.

Diamond Jr, Arthur M (1986). "What is a citation worth?" In: *Journal of Human Resources* 21.2, pp. 200–215.

DiMaggio, Paul J (1988). "Interest and agency in institutional theory". In: *Institutional patterns and organizations: Culture and environment*. Ed. by Lynne G Zucker. Cambridge, MA: Ballinger, pp. 3–21.

— (1995). "Comments on 'What theory is not'". In: *Administrative Science Quarterly* 40.3, pp. 391–397.

— (1997). "Culture and cognition". In: *Annual Review of Sociology* 23.1, pp. 263–287.

DiMaggio, Paul J and Walter W. Powell (Apr. 1983). "The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields". In: *American Sociological Review* 48.2, p. 147.

— (1991). "Introduction to the new institutionalism in organization studies". In: *The New Institutionalism in Organizational Analysis*. Ed. by Walter W Powell and Paul J DiMaggio. University of Chicago Press, pp. 1–38.

Dixon, Nancy M (1999). *The organizational learning cycle: How we can learn collectively*. Aldershot: Gower Publishing, Ltd.

Dogterom, Marileen (2017). *Dutch researchers join forces to build synthetic cell*. URL: https://www.tudelft.nl/en/2017/tu-delft/dutch-researchers-join-forces-to-build-synthetic-cell/.

Dorado, Silvia (2005). "Institutional entrepreneurship, partaking, and convening". In: *Organization Studies* 26.3, pp. 385–414.

Dosi, Giovanni (2000). *Innovation, organization and economic dynamics: Selected essays*. Cheltenham, UK, Northampton, MA: Edward Elgar Publishing.

Douglas, Mary (1986). *How institutions think*. Syracuse: Syracuse University Press.

Dupré, John (2004). "Understanding contemporary genomics". In: *Perspectives on Science* 12.3, pp. 320–338.

Durand, Rodolphe and Jean McGuire (2005). "Legitimating agencies in the face of selection: The case of AACSB". In: *Organization Studies* 26.2, pp. 165–196.

Dymond, Jessica S, Sarah M Richardson, Candice E Coombes, Timothy Babatz, Heloise Muller, Narayana Annaluru, William J Blake, Joy W Schwerzmann, Junbiao Dai, and Derek L Lindstrom (2011). "Synthetic chromosome arms function in yeast and generate phenotypic diversity by design". In: *Nature* 477.7365, pp. 471–476.

Edelman, Lauren B and Mark C Suchman (1997). "The legal environments of organizations". In: *Annual Review of Sociology* 23.1, pp. 479–515.

Efron, Bradley (Sept. 1977). "The efficiency of Cox's likelihood function for censored data". In: *Journal of the American Statistical Association* 72.359, pp. 557–565.

Eisenhardt, Kathleen M, Nathan R Furr, and Christopher B Bingham (2007). "Small worlds and regional innovation". In: *Organization Science* 18.6, pp. 938–954.

Endy, Drew (2005). "Foundations for engineering biology". In: *Nature* 438.7067, pp. 449–453.

— (2009). *Open biotechnology and the BioBrick public agreement*. URL: http://www.openwetware.org/images/f/fd/Why%7B%5C_%7Dthe%7B%5C_%7DBPAv1.pdf (visited on 08/13/2017).

— (2010). *Overview and Context of the Science and Technology of Synthetic Biology. Presentation to the Presidential Commission for the Study of Bioethical Issues, 8 July*. URL: http://www.bioethics.gov/transcripts/%20synthetic-biology/070810/overview-and-contextof-the-science-and-technology.html (visited on 08/13/2017).

Enyeart, Peter J and Andrew D Ellington (2011). "Synthetic biology: A yeast for all reasons". In: *Nature* 477.7365, pp. 413–414.

Esquire (2008). *The 75 most influential people of the 21st century*. URL: http://www.esquire.com/news-politics/g82/most-influential-21st-century-1008/?.

Eysenbach, Gunther (2006). "Citation advantage of open access articles". In: *PLoS biology* 4.5, e157.

Fabrizio, Kira R and Alberto Di Minin (2008). "Commercializing the laboratory: Faculty patenting and the open science environment". In: *Research Policy* 37.5, pp. 914–931.

Falagas, Matthew E., Angeliki Zarkali, Drosos E. Karageorgopoulos, Vangelis Bardakas, and Michael N. Mavros (2013). "The impact of article length on the number of future vitations: A bibliometric analysis of general medicine journals". In: *PLoS ONE* 8.2, e49476.

Fichman, Robert G (2004). "Real options and IT platform adoption: Implications for theory and practice". In: *Information systems research* 15.2, pp. 132–154.

Finbow, Robert (1993). "Ideology and institutions in North America". In: *Canadian Journal of Political Science/Revue canadienne de science politique* 26.4, pp. 671–697.

Fine, Gary Alan (2001). *Difficult reputations: Collective memories of the evil, inept, and controversial.* Chicago, IL: University of Chicago Press.

Fligstein, Neil (1985). "The spread of the multidivisional form among large firms, 1919-1979". In: *American Sociological Review* 50.3, p. 377.

— (1997). "Social skill and institutional theory". In: *American Behavioral Scientist* 40.4, pp. 397–405.

— (2001a). "Institutional entrepreneurs and cultural frames-The case of the European Union's single market program". In: *European Societies* 3.3, pp. 261–287.

— (2001b). "Social skill and the theory of fields". In: *Sociological theory* 19.2, pp. 105–125.

Fligstein, Neil and Iona Mara-Drita (1996). "How to make a market: Reflections on the attempt to create a single market in the European Union". In: *American Journal of Sociology* 102.1, pp. 1–33.

Forsyth, Lynne (June 2017). *Twist Bioscience and BioBricks Foundation announce unique partnership*. URL: https://synbiobeta.com/twist-bioscience-and-biobricks-foundation-announce-unique-partnership/ (visited on 08/30/2017).

Fosfuri, A and Alfonso Gambardella (2001). *Markets for technology: The economics of innovation and corporate strategy*. Cambridge, MA: MIT Press.

Foucault, Michel (1972). *The archeology of knowledge*. New York: Pantheon.

Freeman, Chris (1994). "The economics of technical change: Critical survey". In: *Cambridge Journal of Economics* 18.5, pp. 463–514.

Freeman, Christopher (1989). *Technology policy and economic performance*. London: Pinter Publishers Great Britain.

Friedland, Roger and Robert R Alford (1991). "Bringing society back in: Symbols, practices and institutional contradictions". In: *The New Institutionalism in Organizational Analysis*. Ed. by Walter W Powell and Paul J DiMaggio. Chicago, IL: University of Chicago Press, pp. 232–263.

Fu, Fengli, Jeffry D Sander, Morgan L. Maeder, Stacey Thibodeau-Beganny, J. Keith Joung, Drena Dobbs, Leslie Miller, and Daniel F. Voytas (2009). "Zinc finger database (ZiFDB): A repository for information on C2H2 zinc fingers and engineered zinc-finger arrays". In: *Nucleic Acids Research* 37.SUPPL. 1, pp. 279–283.

Fu, Fengli and Daniel F. Voytas (2013). "Zinc finger database (ZiFDB) v2.0: A comprehensive database of C 2H2 zinc fingers and engineered zinc finger arrays". In: *Nucleic Acids Research* 41.D1, pp. 452–455.

Furman, Jeffrey L and Scott Stern (Aug. 2011). "Climbing atop the shoulders of giants: The impact of institutions on cumulative research". In: *American Economic Review* 101.5, pp. 1933–1963.

Galanie, Stephanie, Kate Thodey, Isis J Trenchard, Maria Filsinger Interrante, and Christina D Smolke (2015). "Complete biosynthesis of opioids in yeast". In: *Science* 349.6252, pp. 1095–1100.

Galaskiewicz, Joseph and Ronald S. Burt (Mar. 1991). "Interorganization contagion in corporate philanthropy". In: *Administrative Science Quarterly* 36.1, p. 88.

Gans, Joshua S and Scott Stern (2000). "Incumbency and R&D incentives: Licensing the gale of creative destruction". In: *Journal of Economics and Management Strategy* 9.4, pp. 485–511.

Garud, Raghu, Sanjay Jain, and Arun Kumaraswamy (2002). "Institutional entrepreneurship in the sponsorship of common technological standards: The case of Sun Microsystems and Java". In: *Academy of Management Journal* 45.1, pp. 196–214.

Garud, Raghu and Peter Karnøe (2003). "Bricolage versus breakthrough: Distributed and embedded agency in technology entrepreneurship". In: *Research policy* 32.2, pp. 277–300.

Gaule, Patrick and Nicolas Maystre (2011). "Getting cited: Does open access help?" In: *Research Policy* 40.10, pp. 1332–1338.

Genspace (2009). *Genspace mission statement*. URL: https://www.genspace.org/mission/ (visited on 08/31/2017).

George, Elizabeth, Prithviraj Chattopadhyay, Sim B Sitkin, and Jeff Barden (2006). "Cognitive underpinnings of institutional persistence and change: A framing perspective". In: *Academy of Management Review* 31.2, pp. 347–365.

Gewin, Virginia (2013). "'Open-hardware' pioneers push for low-cost lab kit". In: *Nature* 495.7438, pp. 13–14.

Gibson, Daniel G, John I Glass, Carole Lartigue, Vladimir N Noskov, Ray-Yuan Chuang, Mikkel A Algire, Gwynedd A Benders, Michael G Montague, Li Ma, and Monzia M Moodie (2010). "Creation of a bacterial cell controlled by a chemically synthesized genome". In: *Science* 329.5987, pp. 52–56.

Giddens, Anthony (1984). *Tlie constitution of society*. Berkeley, CA: University of California Press.

Gilbert, Richard and Carl Shapiro (1990). "Optimal patent length and breadth". In: *RAND Journal of Economics* 21.1, pp. 106–112.

Glynn, Mary Ann (2008). "Beyond constraint: How institutions enable identities". In: *The Sage Handbook of Organizational Institutionalism*. Ed. by Royston Greenwood, Christine Oliver, and Roy Suddaby. London: SAGE Publications, pp. 413–430.

Goldberg, Victor P (1974). "Institutional change and the quasi-invisible hand". In: *The Journal of Law and Economics* 17.2, pp. 461–492.

Gourieroux, Christian and Alain Monfort (1996). *Simulation-based econometric methods*. Oxford university press.

Granovetter, Mark (1985). "Economic action and social structure: The problem of embeddedness". In: *American Journal of Sociology* 91.3, pp. 481–510.

Gray, Barbara (1985). "Conditions facilitating interorganizational collaboration". In: *Human Relations* 38.10, pp. 911–936.

Gray, Barbara, Jill M Purdy, and Shahzad Ansari (2015). "From interactions to institutions: Microprocesses of framing and mechanisms for the structuring of institutional fields". In: *Academy of Management Review* 40.1, pp. 115–143.

Green, Jerry R and Suzanne Scotchmer (1995). "On the division of profit in sequential innovation". In: *RAND Journal of Economics* 26.1, pp. 20–33.

Greenwood, Royston and Christopher R Hinings (1996). "Understanding radical organizational change: Bringing together the old and the new institutionalism". In: *Academy of Management Review* 21.4, pp. 1022–1054.

Greenwood, Royston and Roy Suddaby (2006). "Institutional entrepreneurship in mature fields: The big five accounting firms". In: *Academy of Management Journal* 49.1, pp. 27–48.

Greenwood, Royston, Roy Suddaby, and Christopher R Hinings (2002). "Theorizing change: The role of professional associations in the transformation of institutionalized fields". In: *Academy of Management Journal* 45.1, pp. 58–80.

Greve, Henrich R (1998). "Performance aspirations, and risky change organizational change". In: *Administrative Science Quarterly* 43.1, pp. 58–86.

— (2005). "Interorganizational learning and heterogeneous social structure". In: *Organization Studies* 26.7, pp. 1025–1047.

Grodal, Stine (2017). "Field expansion and contraction: How communities shape social and symbolic boundaries". In: *Administrative Science Quarterly*, pp. 1–36.

Guler, Isin, Mauro F Guillén, and John Muir Macpherson (2002). "Global competition, institutions, and the diffusion of organizational practices: The international spread of ISO 9000 quality certificates". In: *Administrative science quarterly* 47.2, pp. 207–232.

Habibzadeh, Farrokh and Mahboobeh Yadollahie (2010). "Are shorter article titles more attractive for citations? Crosssectional study of 22 scientific journals". In: *Croatian medical journal* 51.2, pp. 165–170.

Hall, Bronwyn H (2004). "Innovation and diffusion". NBER Working Paper 10212.

Hardy, Cynthia (1994). "Underorganized interorganizational domains: The case of refugee systems". In: *The Journal of Applied Behavioral Science* 30.3, pp. 278–296.

Hardy, Cynthia and Steve Maguire (2008). "Institutional entrepreneurship". In: *The Sage Handbook of Organizational Institutionalism*. Ed. by Royston Greenwood, Christine Oliver, and Roy Suddaby. London: SAGE Publications, pp. 198–217.

Hardy, Cynthia and Nelson Phillips (1999). "No joking matter: Discursive struggle in the Canadian refugee system". In: *Organization Studies* 20.1, pp. 1–24.

Hargadon, Andrew B (1998). "Firms as knowledge brokers: Lessons in pursuing continuous innovation". In: *California Management Review* 40.3, pp. 209–227.

Hargadon, Andrew B and Yellowlees Douglas (2001). "When innovations meet institutions: Edison and the design of the electric light". In: *Administrative Science Quarterly* 46.3, pp. 476–501.

Harris, Ron (1997). "Political economy, interest groups, legal institutions, and the repeal of the Bubble Act in 1825". In: *The Economic History Review* 50.4, pp. 675–696.

Haselhoff, Jim (2010). "Designer life: Scotland's next industrial revolution?" In: *Panel discussion at Edinburgh Science Festival*.

Haunschild, Pamela R (1994). "How much is that company worth?: Interorganizational relationships, uncertainty, and acquisition premiums". In: *Administrative Science Quarterly* 39.3, pp. 391–411.

Haveman, Heather A and Hayagreeva Rao (1997). "Structuring a theory of moral sentiments: Institutional and organizational coevolution in the early thrift industry". In: *American Journal of Sociology* 102.6, pp. 1606–1651.

Heller, M A, Rebecca S. Eisenberg, G. Hardin, Harold Demsetz, Robert P Merges, Richard R Nelson, D. Lehman, R. Lempert, and R. Nisbett (1998). "Can patents deter innovation? The anticommons in biomedical research." In: *Science* 280.5364, pp. 698–701.

Hellman, T (2007). "Bridging the science to market gap: The role of patents". In: *Journal of Economic Behavior and Organization* 63.4, pp. 624–647.

Henderson, Rebecca M and Adam B Jaffe (1995). "The Bayh-Dole Act and Trends in University Patenting: 1965 - 1988". In: *Proceedings of the Conference on University Goals, Institutional Mechanisms and the "Industrial Transferability" of Research*. Discussion paper series. Stanford Center for Economic Policy Research.

Henkel, Joachim and Stephen M Maurer (June 2007). "The economics of synthetic biology". In: *Molecular Systems Biology* 3.1, p. 117. URL: http://msb.embopress.org/cgi/doi/10.1038/msb4100161.

Hensmans, Manuel (2003). "Social movement organizations: A metaphor for strategic actors in institutional fields". In: *Organization Studies* 24.3, pp. 355–381.

Herscovitch, Melanie, Eric Perkins, Andy Baltus, and Melina Fan (Apr. 2012). "Addgene provides an open forum for plasmid sharing". In: *Nature Biotechnology* 30.4, pp. 316–317.

Higgs, Robert (1996). "Legally induced technical regress in the Washington salmon fishery". In: *Empirical Studies in Institutional Change*. Ed. by Lee J. Alston,

Thrainn Eggertsson, and Douglass C. North. Cambridge: Cambridge University Press, pp. 247–279.

Hirsch, Paul M (1997). "Review essay: Sociology without social structure: Neoinstitutional theory meets brave new world". In: *American Journal of Sociology* 102.6, pp. 1702–1723.

Hoffman, Andrew J (1999). "Institutional evolution and change: Environmentalism and the US chemical industry". In: *Academy of Management Journal* 42.4, pp. 351–371.

Hoffman, Andrew J and Marc Ventresca (2002). *Organizations, policy and the natural environment: Institutional and strategic perspectives*. Stanford: Stanford University Press.

Holm, Petter (1995). "The dynamics of institutionalization: Transformation processes in Norwegian fisheries". In: *Administrative Science Quarterly* 40.3, pp. 398–422.

Homans, George C (1961). *Human behavior: Its elementary forms*. New York: Harcourt, Brace.

Hughes, Sally Smith (2001). "Making dollars out of DNA: The first major patent in biotechnology and the commercialization of molecular biology, 1974-1980". In: *Isis* 92.3, pp. 541–575.

Hutchison, Clyde A, Ray-Yuan Chuang, Vladimir N Noskov, Nacyra Assad-Garcia, Thomas J Deerinck, Mark H Ellisman, John Gill, Krishna Kannan, Bogumil J Karas, and Li Ma (2016). "Design and synthesis of a minimal bacterial genome". In: *Science* 351.6280, aad6253.

iGEM Foundation (2003). *BioBricks Scars*. URL: http://parts.igem.org/partsdb/scars.cgi (visited on 08/28/2017).

— (2017a). *iGEM Alumni*. URL: http://igem.org/Alumni/About.

— (2017b). *iGEM Competitions Archive*. URL: http://igem.org/Previous%7B%5C_%7DiGEM%7B%5C_%7DCompetitions (visited on 07/07/2017).

— (2017c). *iGEM Jugding Criteria*. URL: http://2017.igem.org/Judging (visited on 08/29/2017).

— (2017d). *iGEM Startups*. URL: http://igem.org/IGEM%7B%5C_%7DStartups (visited on 07/07/2017).

Jackall, Robert (1988). *Moral mazes*. New York: Oxford University Press.

Jacques, Thomas S and Neil J Sebire (2010). "The impact of article titles on citation hits: An analysis of general and specialist medical journals." In: *JRSM short reports* 1.1, p. 2.

Jefferson, Osmat A, Deniz Köllhofer, Thomas H Ehrich, and Richard A Jefferson (2015). "The ownership question of plant gene and genome intellectual properties". In: *Nature Biotechnology* 33.11, pp. 1138–1143.

Jefferson, Richard A (2006). "Science as social enterprise: The CAMBIA BiOS initiative". In: *Innovations: Technology, governance, globalization* 1.4, pp. 13–44.

Johnson, Rick (2009). "Synthetic biology: Challenges of ownership, access & rights". In: *Presentation at Symposium on Opportunities and Challenges in the Emerging Field of Synthetic Biology. National Academies' Keck Center*, pp. 9–10.

Joung, J. Keith, Daniel F. Voytas, and Joanne Kamens (2015). "Accelerating research through reagent repositories: The genome editing example." In: *Genome biology* 16.1, p. 255.

Juhas, Mario (2015). "On the road to synthetic life: The minimal cell and genome-scale engineering". In: *Critical Reviews in Biotechnology* 8551.August, pp. 1–8.

Juhas, Mario, Peter W. Davenport, James R. Brown, Orr Yarkoni, and James W. Ajioka (2013). "The Cambridge BioDesign techevent - synthetic biology, a new 'age of wonder'?" In: *Biotechnology Journal* 8.7, pp. 761–763.

Kahl, Linda J (Jan. 2015). "Realizing positive network effects in synthetic biology". In: *Journal of Responsible Innovation* 2.1, pp. 137–139.

— (2017). *OpenMTA (Draft 11 June 2017)*. URL: https://dspace.mit.edu/handle/1721.1/111012 (visited on 09/01/2017).

Kahl, Linda J and Drew Endy (2013). "A survey of enabling technologies in synthetic biology". In: *Journal of Biological Engineering* 7.1, p. 13.

Kaiser, Jocelyn (Dec. 2005). "Gene therapy: Putting the fingers on gene repair". In: *Science* 310.5756, pp. 1894–1896.

— (2008). *DNA 'Scissors' go open source*. URL: http://www.sciencemag.org/news/2008/07/dna-scissors-go-open-source (visited on 08/26/2017).

— (2012). "The visionary". In: *Science* 335.6069, pp. 651–653.

Kamens, Joanne (2015). "The Addgene repository: An international nonprofit plasmid and data resource". In: *Nucleic Acids Research* 43.D1, pp. D1152–D1157.

Kane, Rosalie A (1975). "The interprofessional team as a small group". In: *Social Work in Health Care* 1.1, pp. 19–32.

Keasling, Jay D (2013). *Why synthetic biology is the field of the future*. URL: http://www.pbs.org/wgbh/nova/next/tech/why-synthetic-biology-is-the-field-of-the-future/ (visited on 08/13/2017).

Kelavkar, Uddhav P (Nov. 2001). "Human Genome Project". In: *Encyclopedia of Life Sciences*. Chichester: John Wiley & Sons, Ltd.

Keller, Evelyn Fox (2000). *The century of the gene*. Cambridge, MA: Harvard University Press.

Kelman, Herbert C (2006). "Interests, relationships, identities: Three central issues for individuals and groups in negotiating their social environment". In: *Annual Review of Psychology* 57.1, pp. 1–26.

Kim, Y G, J Cha, and Srinivasan Chandrasegaran (1996). "Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain". In: *Proceedings of the National Academy of Sciences* 93.3, pp. 1156–1160.

Kitch, Edmund W (1977). "The nature and function of the patent system". In: *The Journal of Law and Economics* 20.2, pp. 265–290.

Kitchener, Martin (2002). "Mobilizing the logic of managerialism in professional fields: The case of academic health centre mergers". In: *Organization Studies* 23.3, pp. 391–420.

Klein, John P. and Melvin L. Moeschberger (Oct. 2003). *Survival analysis: Techniques for censored and truncated data*. New York: Springer, p. 536.

Klemperer, Paul (1990). "How broad should the scope of patent protection be?" In: *RAND Journal of Economics* 21.1, pp. 113–130.

Klug, Aaron and Daniela Rhodes (1987). "'Zinc fingers': A novel protein motif for nucleic acid recognition". In: *Trends in Biochemical Sciences* 12.C, pp. 464–469.

Knight, Frank H (1921). *Risk, uncertainty and profit*. New York: Hart, Schaffner and Marx.

Kono, Clifford, Donald Palmer, Roger Friedland, and Matthew Zafonte (1998). "Lost in space: The geography of corporate interlocking directorates". In: *American Journal of Sociology* 103.4, pp. 863–911.

Kraatz, Matthew S and James H Moore (2002). "Executive migration and institutional change". In: *Academy of Management Journal* 45.1, pp. 120–143.

Krimsky, Sheldon (2004). "Science in the private interest: Has the lure of profits corrupted biomedical research?" In: *The Quarterly Review of Biology* 79.3, pp. 341–343.

Krueger, Anne (Feb. 1988). *The political economy of controls: American sugar*. Tech. rep. Cambridge, MA: National Bureau of Economic Research. URL: http://www.nber.org/papers/w2504.pdf.

Kumar, Sapna and Arti K Rai (2007). "Synthetic biology: The intellectual property puzzle". In: *Texas Law Review* 85.1, pp. 1745–1768.

La Croix, Sumner J and James Roumasset (1990). "The evolution of private property in nineteenth-century Hawaii". In: *The Journal of Economic History* 50.4, pp. 829–852.

La Paillasse (2011). *La Paillasse*. URL: https://lapaillasse.org/ (visited on 08/31/2017).

Lagerkvist, U L F (1978). ""Two out of three": An alternative method for codon reading". In: *Proceedings of the National Academy of Sciences* 75.4, pp. 1759–1762.

Laurance, William F., D. Carolina Useche, Susan G. Laurance, and Corey J. a. Bradshaw (2013). "Predicting publication success for biologists". In: *BioScience* 63.10, pp. 817–823.

Lawrence, Thomas B (1999). "Institutional strategy". In: *Journal of Management* 25.2, pp. 161–187.

Lawrence, Thomas B, Cynthia Hardy, and Nelson Phillips (2002). "Institutional effects of interorganizational collaboration: The emergence of proto-institutions". In: *Academy of Management Journal* 45.1, pp. 281–290.

Lawrence, Thomas B and Nelson Phillips (2004). "From Moby Dick to Free Willy: Macro-cultural discourse and institutional entrepreneurship in emerging institutional fields". In: *Organization* 11.5, pp. 689–711.

Lawrence, Thomas B and Roy Suddaby (2006). "Institutions and institutional work". In: *The Sage Handbook of Organization Studies*. Ed. by Stewart R Clegg, Cynthia Hardy, Tom Lawrence, and Walter R Nord. Vol. 2. New York: SAGE Publications, pp. 215–254.

Leblebici, Huseyin, Gerald R Salancik, Anne Copay, and Tom King (Sept. 1991). "Institutional change and the transformation of interorganizational fields: An organizational history of the US radio broadcasting industry". In: *Administrative Science Quarterly* 36.3, pp. 333–363.

Leca, Bernard and Philippe Naccache (2006). "A critical realist approach to institutional entrepreneurship". In: *Organization* 13.5, pp. 627–651.

Lederberg, Joshua (1952). "Cell genetics and hereditary symbiosis". In: *Physiological Reviews* 32.4, pp. 403–430.

Lee, Kyungmook and Johannes M Pennings (2002). "Mimicry and the market: Adoption of a new organizational form". In: *Academy of Management Journal* 45.1, pp. 144–162.

Leenders, Roger (2002). "Modeling social influence through network autocorrelation: Constructing the weight matrix". In: *Social Networks* 24.1, pp. 21–47.

Lenway, Stefanie Ann, Jeffrey A Hart, and Thomas P Murtha (Nov. 2001). *Managing new industry creation: Global knowledge formation and entrepreneurship in high technology*. Stanford: Stanford University Press, p. 288.

Lessig, Lawrence (2002). *The future of ideas: The fate of the commons in a connected world*. New York: Random House Inc.

Liddicoat, Johnathon E and Kathleen Liddell (2016). *Open innovation with large bioresources: Goals, challenges and proposals*. Tech. rep. 2. Cambridge, UK: University of Cambridge, Centre for Science and Policy.

Lindenberg, Siegwart (1998). "The cognitive turn in institutional analysis: Beyond NIE and NIS?" In: *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft* 154.4, pp. 716–727.

Loecher, Markus (2016). "RgoogleMaps: Overlays on static maps". In: *R package version 1.4.1*.

Lokker, Cynthia, K Ann McKibbon, R James McKinlay, Nancy L Wilczynski, and R Brian Haynes (2008). "Prediction of citation counts for clinical articles at two years using data available within three weeks of publication: Retrospective cohort study". In: *British Medical Journal* 336.7645, pp. 655–657.

Looy, Bart Van, Koenraad Debackere, and Petra Andries (2003). "Policies to stimulate regional innovation capabilities via university–industry collaboration: An analysis and an assessment". In: *R&D Management* 33.2, pp. 209–229.

Lounsbury, Michael (2002). "Institutional transformation and status mobility: The professionalization of the field of finance". In: *Academy of Management Journal* 45.1, pp. 255–266.

— (2007). "A tale of two cities: Competing logics and practice variation in the professionalizing of mutual funds". In: *Academy of Management Journal* 50.2, pp. 289–307.

Lounsbury, Michael, Marc Ventresca, and Paul M Hirsch (2003). "Social movements, field frames and industry emergence: A cultural–political perspective on US recycling". In: *Socio-Economic Review* 1.1, pp. 71–104.

Luisi, Pier Luigi, Francesca Ferri, and Pasquale Stano (Jan. 2006). "Approaches to semi-synthetic minimal cells: A review". In: *Naturwissenschaften* 93.1, pp. 1–13.

Lundvall, Bengt-Ake (1992). *National innovation system: Towards a theory of innovation and interactive learning*. London, UK: Pinter.

Maeder, Morgan L. et al. (2008). "Rapid 'open-source' engineering of customized zinc-finger nucleases for highly efficient gene modification". In: *Molecular Cell* 31.2, pp. 294–301.

Maguire, Steve and Cynthia Hardy (2006). "The emergence of new global institutions: A discursive perspective". In: *Organization Studies* 27.1, pp. 7–29.

Maguire, Steve, Cynthia Hardy, and Thomas B Lawrence (2004). "Institutional entrepreneurship in emerging fields: HIV/AIDS treatment advocacy in Canada". In: *Academy of Management Journal* 47.5, pp. 657–679.

Mann, Ronald J and Thomas W Sager (2007). "Patents, venture capital, and software startups". In: *Research Policy* 36.2, pp. 193–208.

Mansfield, Edwin (1995). "Academic Research Underlying Industrial Innovations: Sources, Characteristics, and Financing". In: *The Review of Economics and Statistics* 77.1, p. 55.

Mansfield, Edwin and Jeong-Yeon Lee (1996). "The modern university: Contributor to industrial innovation and recipient of industrial R&D support". In: *Research Policy* 25.7, pp. 1047–1058.

Manski, Charles F. (1993). "Identification of endogenous social effects: The reflection problem". In: *The Review of Economic Studies* 60.3, pp. 531–542.

March, James G and Johan P Olsen (1989). *Rediscovering institutions: The organizational basis of politics*. New York: Free Press.

Marsden, Peter V.; and Noah E. Friedkin (Aug. 1993). "Network Studies of Social Influence". In: *Sociological Methods & Research* 22.1, pp. 127–151.

Maurer, Stephen M (2009). "Before it's too late. Why synthetic biologists need an open-parts collaboration and how to build one". In: *EMBO reports* 10.8, pp. 806–9.

McGinnis, Scott and Thomas L. Madden (2004). "BLAST: At the core of a powerful and diverse set of sequence analysis tools". In: *Nucleic Acids Research* 32.suppl_2, W20–W25.

McGuire, Deborah B (1999). "Building and maintaining an interdisciplinary research team". In: *Alzheimer Disease & Associated Disorders* 13.Suppl 1, S17–21.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook (2001). "Birds of a feather: Homophily in social networks". In: *Annual Review of Sociology* 27.1, pp. 415–444.

Medin, Douglas L (1989). "Concepts and conceptual structure". In: *American Psychologist* 44.12, p. 1469.

Meindl, James R, Charles Stubbart, and Joseph F Porac (1994). "Cognition within and between organizations: Five key questions". In: *Organization Science* 5.3, pp. 289–293.

Merges, Robert P and Richard R Nelson (1990). "On the complex economics of patent scope". In: *Columbia Law Review* 90.4, pp. 839–916.

— (1994). "On limiting or encouraging rivalry in technical progress: The effect of patent scope decisions". In: *Journal of Economic Behavior and Organization* 25.1, pp. 1–24.

Merton, Robert K (1973). "The normative structure of science". In: *The Sociology of Science: Theoretical and Empirical Investigations*. Ed. by Robert K Merton. Chicago, IL: University of Chicago Press, pp. 267–280.

Meyer, John W and Brian Rowan (1977). "Institutionalized organizations: Formal structure as myth and ceremony". In: *American Journal of Sociology* 83.2, pp. 340–363.

Meyer, John W and W Richard Scott (1992). *Organizational environments: Ritual and rationality*. Beverly Hills: Sage Publications, Inc.

Meyer, Renate E and Gerhard Hammerschmid (2006). "Changing institutional logics and executive identities: A managerial challenge to public administration in Austria". In: *American Behavioral Scientist* 49.7, pp. 1000–1014.

Misangyi, Vilmos F, Gary R Weaver, and Heather Elms (2008). "Ending corruption: The interplay among institutional logics, resources, and institutional entrepreneurs". In: *Academy of Management Review* 33.3, pp. 750–770.

Mitchell, Rudolph, Yehudit Judy Dori, and Natalie H. Kuldell (2011). "Experiential engineering through iGEM - An undergraduate summer competition in

synthetic biology". In: *Journal of Science Education and Technology* 20.2, pp. 156–160.

Mizruchi, Mark S and Lisa C Fein (1999). "The social construction of organizational knowledge: A study of the uses of coercive, mimetic, and normative isomorphism". In: *Administrative Science Quarterly* 44.4, pp. 653–683.

Mohr, John W and Vincent Duquenne (1997). "The duality of culture and practice: Poverty relief in New York City, 1888 - 1917". In: *Theory and society* 26.2-3, pp. 305–356.

Monteiro, Pedro and Davide Nicolini (2015). "Recovering materiality in institutional work: Prizes as an assemblage of human and material entities". In: *Journal of Management Inquiry* 24.1, pp. 61–81.

Montiel, Ivan and Bryan W Husted (2009). "The adoption of voluntary environmental management programs in Mexico: First movers as institutional entrepreneurs". In: *Journal of Business Ethics* 88.1, pp. 349–363.

Moore, Finola E, Deepak Reyon, Jeffry D Sander, Sarah A Martinez, Jessica S Blackburn, Cyd Khayter, Cherie L Ramirez, J. Keith Joung, and David M Langenau (2012). "Improved somatic mutagenesis in zebrafish using transcription activator-like effector nucleases (TALENs)". In: *PLoS ONE* 7.5, e37877.

Mosher, David (2010). *DIY biotech hacker space opens in NYC*. URL: https://www.wired.com/2010/12/genspace-diy-science-laboratory/.

Moss, Lenny (2004). *What genes can't do*. Cambridge, MA: MIT Press.

Mowery, David C and Richard R Nelson (1999). *Sources of industrial leadership: Studies of seven industries*. Cambridge, UK: Cambridge University Press.

Mowery, David C, Richard R Nelson, Bhaven N Sampat, and Arvids A Ziedonis (2001). "The growth of patenting and licensing by US universities: An assessment of the effects of the Bayh–Dole act of 1980". In: *Research policy* 30.1, pp. 99–119.

Murray, Fiona (2006). "The Oncomouse that roared: Resistance and accommodation to patenting in academic science". Sloan School of Management Working Paper.

Murray, Fiona, Philippe Aghion, Mathias Dewatripont, Julian Kolev, and Scott Stern (2016). "Of mice and academics: Examining the effect of openness on innovation". In: *American Economic Journal: Economic Policy* 8.1, pp. 212–252.

Murray, Fiona and Scott Stern (2007). "Do formal intellectual property rights hinder the free flow of scientific knowledge? An empirical test of the anti-

commons hypothesis". In: *Journal of Economic Behavior and Organization* 63.4, pp. 648–687.

Mutalik, Vivek K, Joao C Guimaraes, Guillaume Cambray, Colin Lam, Marc Juul Christoffersen, Quynh-Anh Mai, Andrew B Tran, Morgan Paull, Jay D Keasling, and Adam P Arkin (2013). "Precise and reliable gene expression via standard transcription and translation initiation elements". In: *Nature Methods* 10.4, pp. 354–360.

Mutch, Alistair (2007). "Reflexivity and the institutional entrepreneur: A historical exploration". In: *Organization Studies* 28.7, pp. 1123–1140.

Mutch, Alistair, Rick Delbridge, and Marc Ventresca (Sept. 2006). "Situating organizational action: The relational sociology of organizations". In: *Organization* 13.5, pp. 607–625.

Nawy, Tal (2011). "Yeast 2.0". In: *Nature Methods* 8.11, p. 895.

Nelson, Bryn (2014). "Synthetic biology: Cultural divide." In: *Nature* 509.7499, pp. 152–154.

Nelson, Richard R (1959). "The simple economics of basic scientific research". In: *Journal of Political Economy* 67.3, pp. 297–306.

— (1993). *National innovation systems: A comparative analysis*. New York and London: Oxford university press.

Nelson, Richard R and Nathan Rosenberg (1993). "Technical innovation and national systems". In: *National innovation systems: A comparative analysis*. Ed. by Richard R Nelson. Oxford, UK: Oxford University Press, pp. 3–21.

Nieminen, Pentti, James Carpenter, Gerta Rucker, and Martin Schumacher (2006). "The relationship between quality of research and citation frequency". In: *BMC Medical Research Methodology* 6.3, p. 42.

Nirenberg, M., P. Leder, M. Bernfield, R. Brimacombe, J. Trupin, F. Rottman, and C. O'Neal (May 1965). "RNA codewords and protein synthesis, VII. On the general nature of the RNA code". In: *Proceedings of the National Academy of Sciences* 53.5, pp. 1161–1168.

North, Douglass C (1990). *Institutions, institutional change and economic performance*. Cambridge, UK, New York: Cambridge University Press.

North, Douglass C and Robert Paul Thomas (1970). "An economic theory of the growth of the western world". In: *The Economic History Review* 23.1, pp. 1–17.

O'Brien, By Daniel P. and W. Murray Spruill (Apr. 2013). "Does inherency have a place in determinations of obviousness?" In: *Biotechnology Law Report* 32.1-2, pp. 3–16.

Oakes, Leslie S, Barbara Townley, and David J Cooper (1998). "Business planning as pedagogy: Language and control in a changing institutional field". In: *Administrative Science Quarterly* 43.1, pp. 257–292.

Ocasio, William (1997). "Towards an attention-based view of the firm". In: *Strategic Management Journal* 18.S1, pp. 187–206.

OECD (2007). "Growth: Rationale for an innovation strategy". Luxembourg.

Ogilvie, Sheilagh (2007). "'Whatever is, is right'? Economic institutions in pre-industrial Europe". In: *The Economic History Review* 60.4, pp. 649–684.

OpenPlant (2016). *Progress Report*. URL: https://www.openplant.org/s/OpenPlant%7B%5C_%7Dmanual%7B%5C_%7D2016-tal8.pdf (visited on 08/13/2017).

OpenTrons (2011). *OpenTrons Protocol*. URL: https://protocols.opentrons.com/ (visited on 08/22/2017).

Orru, Marco, Nicole Woolsey Biggart, and Gary G Hamilton (1991). "Organizational isomorphism in East Asia: Broadening the new institutionalism". In: *The New Institutionalism in Organizational Analysis*. Ed. by Walter W. Powell and Paul J DiMaggio. Chicago, IL: University of Chicago Press, pp. 361–389.

Overwalle, Geertrui Van (2009). *Gene patents and collaborative licensing models: Patent pools, clearinghouses, open source models, and liability regimes*. Cambridge, UK, New York: Cambridge University Press.

Oye, Kenneth A and Rachel Wellhausen (2009). "The intellectual commons and property in synthetic biology". In: *Synthetic Biology: The technoscience and its societal consequences*. Ed. by Markus Schmidt, Alexander Kelle, Agomoni Ganguli-Mitra, and Huib Vriend. Dordrecht: Springer Netherlands, pp. 121–140.

Pacheco, D. F., J. G. York, Thomas J Dean, and S. D. Sarasvathy (2010). "The co-evolution of institutional entrepreneurship: A tale of two theories". In: *Journal of Management* 36.4, pp. 974–1010.

Pagan, Adrian Rodney and Anthony David Hall (1983). "Diagnostic tests as residual analysis". In: *Econometric Reviews* 2.2, pp. 159–218.

Parsons, Jill Petraglia and Clifford S Duke (2013). "Strategies for developing and innovating living stocks collections: An ESA workshop report". In: *The Bulletin of the Ecological Society of America* 94.1, pp. 118–129.

Parsons, Talcott (1956). "Suggestions for a sociological approach to the theory of organizations". In: *Administrative Science Quarterly* 1.1, pp. 63–85.

Pearson, Helen (Sept. 2008). "Protein engineering: The fate of fingers". In: *Nature* 455.7210, pp. 160–164.

Peccoud, Jean (2016). "Synthetic biology: Fostering the cyber-biological revolution". In: *Synthetic Biology* 1.1, ysw001.

Pejovich, Svetozar (June 1996). "Property rights and technological innovation". In: *Social Philosophy and Policy* 13.02, p. 168.

Pénin, Julien and Jean Pierre Wack (2008). "Research tool patents and free-libre biotechnology: A suggested unified framework". In: *Research Policy* 37.10, pp. 1909–1921.

Peplow, Mark (Feb. 2013). *Malaria drug made in yeast causes market ferment*.

Perkel, Jeffrey M (2012). "Genome engineering: Writing a better genome". In: *Biotechniques* 53.4, pp. 213–215.

Perkmann, Markus and Andre André Spicer (2007). "Healing the scars of history: Projects, skills and field strategies in institutional entrepreneurship". In: *Organization Studies* 28.7, pp. 1101–1122.

Petersen, A. M., S. Fortunato, R. K. Pan, K. Kaski, O. Penner, A. Rungi, M. Riccaboni, H. E. Stanley, and F. Pammolli (Oct. 2014). "Reputation and impact in academic careers". In: *Proceedings of the National Academy of Sciences* 111.43, pp. 15316–15321.

Pfeffer, Jeffrey and Gerald R Salancik (1978). *The external control of organizations: A resource dependence approach*. New York: Harper and Row Publishers.

Piwowar, Heather A., Roger S. Day, and Douglas B. Fridsma (2007). "Sharing detailed research data is associated with increased citation rate". In: *PLoS ONE* 2.3, e308.

Polletta, Francesca and James M. Jasper (2001). "Collective identity and social movements". In: *Annual Review of Sociology* 27.1, pp. 283–305.

Porteus, Matthew H and David Baltimore (2003). "Chimeric nucleases stimulate gene targeting in human cells". In: *Science* 300.5620, p. 763.

Powell, Walter W, Kenneth W Koput, and Laurel Smith-Doerr (1996). "Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology". In: *Administrative science quarterly*, pp. 116–145.

Raffaelli, Ryan and Mary Ann Glynn (2014). "Turnkey or tailored? Relational pluralism, institutional complexity, and the organizational adoption of more or less customized practices". In: *Academy of Management Journal* 57.2, pp. 541–562.

Rai, Arti K (2005). "Open and collaborative research: A new model for biomedicine". In: *Intellectual Property Rights in Frontier Industries: Software and Biotechnology*. Ed. by Robert W. Hahn. Ann Arbor, MI: University of Michigan, pp. 131–158.

Rai, Arti K and James Boyle (2007). "Synthetic biology: Caught between property rights, the public domain, and the commons". In: *PLoS Biology* 5.3, pp. 389–393.

Raimbault, Benjamin, Jean Philippe Cointet, and Pierre Benot Joly (2016). "Mapping the emergence of synthetic biology". In: *PLoS ONE* 11.9, pp. 1–19.

Ramirez, Cherie L et al. (May 2008). "Unexpected failure rates for modular assembly of engineered zinc fingers". In: *Nature Methods* 5.5, pp. 374–375.

Rao, Hayagreeva (1994). "The social construction of reputation: Certification contests, legitimation, and the survival of organizations in the American automobile industry: 1895–1912". In: *Strategic Management Journal* 15.S1, pp. 29–44.

— (1998). "Caveat emptor: The construction of nonprofit consumer watchdog organizations". In: *American Journal of Sociology* 103.4, pp. 912–961.

Rao, Hayagreeva, Philippe Monin, and Rodolphe Durand (2003). "Institutional change in Toque Ville: Nouvelle cuisine as an identity movement in French gastronomy". In: *American Journal of Sociology* 108.4, pp. 795–843.

— (2005). "Border crossing: Bricolage and the erosion of categorical boundaries in French gastronomy". In: *American Sociological Review* 70.6, pp. 968–991.

Richardson, Sarah M, Leslie A Mitchell, Giovanni Stracquadanio, Kun Yang, Jessica S Dymond, James E DiCarlo, Dongwon Lee, Cheng Lai Victor Huang, Srinivasan Chandrasegaran, and Yizhi Cai (2017). "Design of a synthetic yeast genome". In: *Science* 355.6329, pp. 1040–1044.

Ritvala, Tiina and Nina Granqvist (2009). "Institutional entrepreneurs and local embedding of global scientific ideas - The case of preventing heart disease in Finland". In: *Scandinavian Journal of Management* 25.2, pp. 133–145.

Rogers, Everett M (2003). *Diffusion of innovations*. New York: Free Press.

Rogers, Everett M and D Lawrence Kincaid (1981). *Communication networks: Toward a new paradigm for research*. New York: Free Press.

Rosch, Eleanor (1975). "Cognitive representations of semantic categories". In: *Journal of Experimental Psychology* 104.3, p. 192.

Rosenberg, Nathan (2004). "Innovation and economic growth. OECD Working Paper". URL: https://www.oecd.org/cfe/tourism/34267902.pdf.

Ruef, Martin (1999). "Social ontology and the dynamics of organizational forms: Creating market actors in the healthcare field, 1966-1994". In: *Social Forces* 77.4, pp. 1403–1432.

Ruttan, Vernon W. (Dec. 2006). "Social science knowledge and induced institutional innovation: An institutional design perspective". In: *Journal of Institutional Economics* 2.03, p. 249.

Sage (2009). *Synapse*. URL: https://www.synapse.org/ (visited on 08/22/2017).

Sampson, R. J. (Aug. 1997). "Neighborhoods and violent crime: A multilevel study of collective efficacy". In: *Science* 277.5328, pp. 918–924.

Sander, Jeffry D, Peter Zaback, J. Keith Joung, Daniel F. Voytas, and Drena Dobbs (2007). "Zinc finger targeter (ZiFiT): An engineered zinc finger/target site design tool". In: *Nucleic Acids Research* 35.SUPPL.2, pp. 599–605.

Santolini, Marc (2017). *iGEMeta Analysis*. URL: http://2017.igem.org/Resources/iGEMeta (visited on 09/15/2017).

SCHER, SCENIHR, and SCCS (2014). *Opinion on synthetic biology and definition*. Tech. rep. September. Scientific Committees, pp. 1–62.

Scotchmer, Suzanne (1991). "Standing on the shoulders of giants: Cumulative research and the patent law". In: *Journal of Economic Perspectives* 5.1, pp. 29–41.

— (2004). *Innovation and incentives*. Cambridge, MA: MIT press.

Scott, Christopher Thomas (2005). "The zinc finger nuclease monopoly". In: *Nature Biotechnology* 23.8, pp. 915–918.

Scott, W Richard (1991). "Unpacking institutional arguments". In: *The New Institutionalism in Organizational Analysis*. Ed. by Walter P Powell and Paul J DiMaggio. Chicago, IL: University of Chicago Press, pp. 164–182.

— (1995). *Institutions and organizations. Foundations for organizational science*. Thousand Oaks: SAGE Publications.

— (2000). *Institutional change and healthcare organizations: From professional dominance to managed care*. Chicago, IL: University of Chicago Press.

Scott, W Richard (2001). *Institutions and organizations*. Thousand Oaks: SAGE Publications.

Searle, John R (1995). "The construction of social reality". In: *Sociological Inquiry* 44.4, pp. 257–268.

Seglen, Per O (1997). "Why the impact factor of journals should not be used for evaluating research". In: *British Medical Journal* 314.7079, p. 498.

Seiler, Catherine Y, Jin G Park, Amit Sharma, Preston Hunter, Padmini Surapaneni, Casey Sedillo, James Field, Rhys Algar, Andrea Price, and Jason Steel (2013). "DNASU plasmid and PSI: Biology-Materials repositories. Resources to accelerate biological research". In: *Nucleic Acids Research* 42.D1, pp. D1253–D1260.

Selivanov, Dmitriy (2016). *text2vec: Modern text mining framework for R*. URL: https://cran.r-project.org/package=text2vec.

Selznick, Philip (1949). *TVA and the grass roots: A study in the sociology of formal organization*. 3rd ed. Berkeley, CA: University of California Press.

— (1957). *Leadership in administration: A sociological interpretation*. Berkeley, CA: University of California Press.

Seo, Myeong-Gu and W E Douglas Creed (2002). "Institutional contradictions, praxis, and institutional change: A dialectical perspective". In: *Academy of Management Review* 27.2, pp. 222–247.

Setubal, Carlos and Joao Meidanis (1997). *Introduction to computational molecular biology*. Boston: PWS Publishing.

Sewell Jr, William H (1992). "A theory of structure: Duality, agency, and transformation". In: *American Journal of Sociology* 98.1, pp. 1–29.

Shannon, C.E. (Jan. 1949). "Communication in the presence of noise". In: *Proceedings of the IRE* 37.1, pp. 10–21.

Shetty, Reshma P, Drew Endy, and Thomas F Knight (2008). "Engineering BioBrick vectors from BioBrick parts." In: *Journal of Biological Engineering* 2.1, p. 5.

Shipilov, Andrew, Henrich Greve, and Timothy Rowley (2010). "When do interlocks matter? Institutional logics and the diffusion of multiple corporate governance practices". In: *Academy of Management Journal* 53.4, pp. 846–864.

Sine, Wesley D and Brandon H Lee (2009). "Tilting at windmills? The environmental movement and the emergence of the US wind energy sector". In: *Administrative Science Quarterly* 54.1, pp. 123–155.

Singer, Judith D. and John B. Willett (May 2003). *Applied longitudinal data analysis*. Cambridge, UK: Oxford University Press.

Singh, Harpreet, David Kryscynski, Xinxin Li, and Ram Gopal (Aug. 2016). "Pipes, pools, and filters: How collaboration networks affect innovative performance". In: *Strategic Management Journal* 37.8, pp. 1649–1666.

Singh, Jitendra V (1986). "Performance, slack, and risk taking in organizational decision making". In: *Academy of Management Journal* 29.3, pp. 562–585.

Singh, Param Vir and Corey Phelps (2013). "Networks, social influence, and the choice among competing innovations: Insights from open source software licenses". In: *Management Science* 24.3, pp. 539–560.

Singh, Ranjan (2015). *Synthetic biology market by products*. Tech. rep. Portland, OR: Allied Market Research.

Smolke, Christina D (2009). "Building outside of the box: iGEM and the BioBricks Foundation". In: *Nature Biotechnology* 27.12, pp. 1099–1102.

Snow, David A and Robert D Benford (1992). "Master frames and cycles of protest". In: *Frontiers in social movement theory*. Ed. by Aldon D Morris and Carol McClurg Mueller. London, New Haven: Yale University Press, pp. 133–155.

Snow, David A, E Burke Rochford Jr, Steven K Worden, and Robert D Benford (1986). "Frame alignment processes, micromobilization, and movement participation". In: *American Sociological Review* 51.4, pp. 464–481.

Stephan, Paula E, Shiferaw Gurmu, Albert J Sumell, and Grant Black (2007). "Who's patenting in the university? Evidence from the survey of doctorate recipients". In: *Economics of Innovation and New Technology* 16.2, pp. 71–99.

Stock, James H and Motohiro Yogo (2002). "Testing for weak instruments in linear IV regression". NBER Working Paper 0284. Cambridge, MA.

Stokes, Donald E (2011). *Pasteur's quadrant: Basic science and technological innovation*. Washington, D.C.: Brookings Institution Press.

Strang, David and Sarah A Soule (1998). "Diffusion in organizations and social movements: From hybrid corn to poison pills". In: *Annual Review of Sociology* 24.1, pp. 265–290.

Strang, David and Nancy Brandon Tuma (1993). "Spatial and temporal heterogeneity in diffusion". In: *American Journal of Sociology* 99.3, pp. 614–639.

Straus, Joseph, Henrik Holzapfel, and Matthias Lindenmeir (2002). *Empirical survey on genetic inventions and patent law*. Tech. rep. Muenchen: Federal Ministry of Education and Research.

Suchman, Mark C (1995). "Managing legitimacy: Strategic and institutional approaches". In: *Academy of Management Review* 20.3, pp. 571–610.

Suddaby, Roy and Royston Greenwood (2005). "Rhetorical strategies of legitimacy". In: *Administrative Science Quarterly* 50.1, pp. 35–67.

SynBERC (2006). *Vision and history of synberc*. URL: http://www.synberc.org/about (visited on 08/13/2017).

Synthetic Yeast 2.0 (2017). *Collaborators in Synthetic Yeast 2.0 Project*. URL: http://syntheticyeast.org/collaborators/ (visited on 08/31/2017).

Tajfel, Henri and John C Turner (1979). "An integrative theory of intergroup conflict". In: *The Social Psychology of Intergroup Relations* 33.47, p. 74.

Talukdar, Mrinal (2014). *Dictionary of computer and information technology*. New Delhi: Prabhat Prakashan.

Tan, Elaine S (2005). "Ideology, interest groups, and institutional change: The case of the British prohibition of wages in kind". In: *Journal of Institutional Economics* 1.2, pp. 175–191.

Thornton, Patricia H (2002). "The rise of the corporation in a craft industry: Conflict and conformity in institutional logics". In: *Academy of Management Journal* 45.1, pp. 81–101.

— (2004). *Markets from culture: Institutional logics and organizational decisions in higher education publishing*. Stanford: Stanford University Press.

Thornton, Patricia H and William Ocasio (1999). "Institutional logics and the historical contingency of power in organizations: Executive succession in the higher education publishing industry, 1958–1990". In: *American Journal of Sociology* 105.3, pp. 801–843.

— (2008). "Institutional Logics". In: *The Sage Handbook of Organizational Institutionalism*. Ed. by Royston Greenwood, Christine Oliver, and Roy Suddaby. London: SAGE Publications, pp. 99–129.

Thursby, Marie and Richard Jensen (2001). "Proofs and prototypes for sale: The licensing of university inventions". In: *American Economic Review* 91.1, pp. 240–259.

Thursby, Marie, Jerry Thursby, and Swasti Gupta-Mukherjee (2007). "Are there real effects of licensing on academic research? A life cycle view". In: *Journal of Economic Behavior and Organization* 63.4, pp. 577–598.

Torrance, Andrew W (2010). "Synthesizing law for synthetic biology". In: *Minnesota Journal of Law, Science & Technology* 11.2, p. 629.

— (2011). "DNA copyright". In: *Valparaiso University Law Review* 46.1, p. 1.

Torrance, Andrew W and Linda J Kahl (2013). "Bringing standards to life: Synthetic biology standards and intellectual property". In: *Santa Clara High Technology Law Journal* 30.2, pp. 199–230.

Townley, Barbara (2002). "The role of competing rationalities in institutional change". In: *Academy of Management Journal* 45.1, pp. 163–179.

Train, Kenneth E. (2003). *Discrete choice methods with simulation*. Cambridge, UK: Cambridge University Press.

Trajtenberg, Manuel, Rebecca Henderson, and Adam B Jaffe (1997). "University versus corporate patents: A window on the basicness of invention". In: *Economics of Innovation and New Technology* 5.1, pp. 19–50.

Trist, Eric (1983). "Referent organizations and the development of inter-organizational domains". In: *Human Relations* 36.3, pp. 269–284.

Tyler, Tom R (1999). "Why people cooperate with organizations: An identity-based perspective." In: *Research in Organizational Behavior* 21.1, pp. 201–246.

Van de Ven, Andrew H and Raghu Garud (1993). "Innovation and industry development: The case ofcochlear implants". In: *Research on technological innovation, management and Policy*. Ed. by R. A. Burgelman and R. S. Rosenbloom. Vol. 5. Greenwich, Connecticut: JAI Press Inc., pp. 1–46.

Van Overwalle, Geertrui (2010). "Turning patent swords into shares". In: *Science* 330.6011, pp. 1630–1631.

Wade, Nicholas (Dec. 2009). *In new way to edit DNA, hope for treating disease*. New York.

Walker, Joan L., Emily Ehlers, Ipsita Banerjee, and Elenna R. Dugundji (May 2011). "Correcting for endogeneity in behavioral choice models with social

influence variables". In: *Transportation Research Part A: Policy and Practice* 45.4, pp. 362–374.

Walsh, John P, Ashish Arora, and Wesley M Cohen (Feb. 2003). "Working through the patent problem". In: *Science* 299.5609, pp. 1021–1021.

Walsh, John P, Charlene Cho, and Wesley M Cohen (2005). "View from the bench: Patents and material transfers". In: *Science* 309.5743, pp. 2002–2003.

Walsh, John P, Wesley M Cohen, and Charlene Cho (2007). "Where excludability matters: Material versus intellectual property in academic biomedical research". In: *Research Policy* 36.8, pp. 1184–1203.

Warren, Roland L (1967). "The interorganizational field as a focus for investigation". In: *Administrative Science Quarterly* 12.3, pp. 396–419.

Weinberger, Armin, Karsten Stegmann, and Frank Fischer (2007). "Knowledge convergence in collaborative learning: Concepts and assessment". In: *Learning and Instruction* 17.4, pp. 416–426.

Wejnert, Barbara (2002). "Integrating models of diffusion of innovations: A conceptual framework". In: *Annual Review of Sociology* 28.1, pp. 297–326.

Wellhausen, Rachel and Gautam Mukunda (2009). "Aspects of the political economy of development and synthetic biology". In: *Systems and synthetic biology* 3.1-4, p. 115.

White, Harrison C (1992). *Identity and control: A structural theory of social action*. Princeton, NJ: Princeton University Press.

Wijen, Frank and Shahzad Ansari (2007). "Overcoming inaction through collective institutional entrepreneurship: Insights from regime theory". In: *Organization Studies* 28.7, pp. 1079–1100.

Williams, Heidi (2013). "Intellectual property rights and innovation: Evidence from the human genome". In: *Journal of Political Economy* 121.1, pp. 1–27.

Williamson, Oliver E (1985). *The economic institutions of capitalism: Firms, markets, relational contracting*. New York: Free Press.

Wilson Center (2015). *U.S. trends in synthetic biology research funding*. Tech. rep. September. Washington, D.C.: Wilson Center, pp. 1–9.

Wooldridge, Jeffrey M (2013). *Introductory Econometrics: A modern approach*. Vol. 5. Mason, OH: South-Western Cengage Learning.

Wooten, Melissa (2006). "The evolution of the black higher education field, 1854 - 1996". PhD thesis. Ann Arbor, MI.

Wright, David A., Jeffrey A. Townsend, Ronnie Joe Winfrey, Phillip A. Irwin, Jyothi Rajagopal, Patricia M. Lonosky, Bradford D. Hall, Michael D. Jondle, and Daniel F. Voytas (2005). "High-frequency homologous recombination in plants mediated by zinc-finger nucleases". In: *Plant Journal* 44.4, pp. 693–705.

Xi, Sigma and The Scientific Research Society (1988). *Removing the boundaries: Perspectives on cross-disciplinary research: Final report on an inquiry into cross-disciplinary science*. New Haven, Conn.: Sigma Xi, The Scientific Research Society.

Y-Combinator (2017). *OpenTrons at Y-Combinator*. URL: http://blog.ycombinator.com/opentrons-w16-does-lab-grunt-work-so-scientists-can-focus-on-bigger-things/ (visited on 08/20/2017).

Yegros-Yegros, Alfredo, Ismael Rafols, and Pablo D'Este (2015). "Does interdisciplinary research lead to higher citation impact? The different effect of proximal and distal interdisciplinarity". In: *PLoS ONE* 10.8, e0135095.

Zajac, Edward J and James D Westphal (2004). "The social construction of market value: Institutionalization and learning perspectives on stock market reactions". In: *American Sociological Review* 69.3, pp. 433–457.

Zerbe, Richard O and C Leigh Anderson (2001). "Culture and fairness in the development of institutions in the California gold fields". In: *The Journal of Economic History* 61.1, pp. 114–143.

Zilber, Tammar B (2002). "Institutionalization as an interplay between actions, meanings, and actors: The case of a rape crisis center in Israel". In: *Academy of Management Journal* 45.1, pp. 234–254.

Zinc Finger Consortium (2016). *Zinc finger consortium reagents on Addgene*. URL: https://www.addgene.org/zfc/ (visited on 10/01/2016).

Zucker, Lynne G (1977). "The role of institutionalization in cultural persistence". In: *American Sociological Review* 42.5, pp. 726–743.