# Learning When to Quit: An Empirical Model of Experimentation

Bernhard Ganglmair Timothy Simcoe Emanuele Tarantino

January 2018

Introduction

## This Paper

### Learning in R&D

How do scientists/engineers learn whether or not they have a good idea?

- Study dynamic learning in R&D decision making at the project level
- Learning about the quality (or type: good or bad) of a project that materializes via "consensus" by a community of peers
- Parsimonious structural model that combines Bayesian learning with ongoing improvements to project
- Estimate on project-level data for successful and abandoned projects from the *Internet Engineering Task Force* (IETF)

### How do Researchers and Engineers Use Time?

Ideally, focus on valuable projects and quickly cut loose projects without potential. But how to know which is which? (Allen, 1966)

- Researchers' actions (i.e., employment of time) depend on beliefs and the (institutional) environment
- This paper:
  - 1. Estimate key parameters in researchers' R&D decisions (beliefs, opportunity costs) based on observed actions and outcomes
  - 2. Counterfactual: Are cost subsidies or publication prizes the better policy?
  - 3. Counterfactual: What is the effect/cost of over-confidence?
  - 4. Heterogeneity: How do communication and prior experience relate to learning?

Introduction

# Summary of Findings

- 1. Estimates for "rate of learning," prior beliefs about quality, and opportunity costs of revisions
  - Three out of five (59%) projects submitted to IETF are publishable (can generate value)
  - Less than 1/3 (31%) of these projects are published (successful)
  - Realization of value for 17% of good projects per revision
  - Decreasing opportunity costs of revising a project's draft
- 2. Both *subsidies* and *prizes* raise projects' *gross* value, only the latter increase *net* value (i.e., considering private revision costs)
- 3. Over-confidence is more costly than pessimism (misaligned priors)
- 4. More communication and experience of authors result in faster learning; higher commerciality of projects slows the rate of learning

Introduction

### Related Literature

- Dynamic decisions with learning
  - Erdem and Keane (1996); Crawford and Shum (2005); Dickstein (2014)
  - Pakes (1986); Nanda and Rhodes-Kropf (2015); Krieger (2017)
- Consensus standardization (IETF)
  - Rysman and Simcoe (2008); Fleming and Waguespack (2009); Simcoe (2012); Ganglmair and Tarantino (2014)
- Prizes v. Subsidies
  - Murray, Stern, Campbell, and MacKormack (2012); Galasso, Mitchell, and Virag (2016); Hall and Van Reenen (2000)
- Optimism in innovation
  - Galasso and Simcoe (2011); Hirshleifer, Low, and Hong Teoh (2012)

Ganglmair, Simcoe, and Tarantino

Learning When to Quit

### Internet Engineering Task Force Institutional Background

- IETF is the main forum for internet protocol development
- Anyone can participate. In practice, corporate, academic and individual engineers, and computer scientists
- Main motivation: Advance technology
- Transparent process  $\Rightarrow$  rich data
  - Repository with every version of every project (success and failure)
  - E-mail server where project-related communication occurs
  - Tri-annual meetings, held around the world

# **IETF** Protocol Examples

Important Standards

	Description	Year
RTP	Real-time Transport Protocol	2003
SIP	Session Initiation Protocol	2002
HTTP	Hypertext Transfer Protocol	1999
IPV6	Internet Protocol, Version 6 (IPv6)	1998
DHCP	Dynamic Host Configuration Protocol	1997
POP3	Post Office Protocol – Version 3	1996
NAT	Network Address Translator	1994
FTP	File Transfer Protocol	1985
ТСР	Transmission Control Protocol	1981
IP	Internet Protocol	1981

Major Contributors

# **IETF** Standardization

Process Overview: Data-Generating Process

- 1. Identify problem and submit proposal (Internet Draft or ID)
  - Two types: Individual and Working Group
  - All projects posted to public repository
- 2. Community feedback via email and meetings
- 3. Rough consensus  $\Rightarrow$  ID published
  - Decision by WG Chair and IESG (de facto super-majority)
  - Published ID's called Proposed Standards (or RFCs)
- 4. No consensus  $\Rightarrow$  sponsors have a choice
  - Revise ID  $\rightarrow$  return to step (1) [submit revision]
  - Abandon ID  $\rightarrow$  expires in 6 months

# Standards and Nonstandards

Parallel Publication Tracks

- Standards
  - Normative document ⇒ formal IETF endorsement
  - Objective: promote commercial implementation
  - Example: "Domain Host Control Protocol" RFC2131
- Nonstandards
  - Informative document
  - Two types: Informational & Experimental

- Identical review and publication process
  - Nonstandards have less uncertainty → "no learning" baseline

### **IETF** Standardization

Possible Outcomes



Minimum Control     Minimum Control	<pre>Intermet-main Grant-left-http-vil-spec-rev-00.txt Grant-left-http-vil-spec-rev-00.txt The Grant- ere of the Grant State The State of the Grant State The State of the State of State of State The State of the State of State of State State of the State of State of State of State State of State of State of State of State of State of State State of State of State of State of State of State of State State of State of</pre>		
<pre>HTTERT.TEMATY Control is appeared to be the set of a set of a</pre>	<pre>Intermet-main Grant-left-http-vil-spec-rev-00.txt Grant-left-http-vil-spec-rev-00.txt The Grant- ere of the Grant State The State of the Grant State The State of the State of State of State The State of the State of State of State State of the State of State of State of State State of State of State of State of State of State of State State of State of State of State of State of State of State State of State of</pre>		
<text><text><section-header><text><text><text><text><text><text></text></text></text></text></text></text></section-header></text></text>	<text><text><text><text><text><text><text><text><text></text></text></text></text></text></text></text></text></text>	INTERNET-DEAFT UC Irvine <draft-ietf-http-vll-spec-rev-00.txt> J. Gettys J. G. Mogul DEC B. Frystyk</draft-ietf-http-vll-spec-rev-00.txt>	
<ul> <li>Actua of this Mana</li> <li>This document is an internet-Draft. Internet-Drafts are different to the document and the source of particular internet-oration of the document as internet-oration.</li> <li>Actua of the document will document as internet-oration of the document as internet-oration of the document and the documen</li></ul>	<ul> <li>Actua of this Memo</li> <li>This document is an internet-braft. Internet-brafts are without an experiment of the second second second second property of the second second second second second second second second second sec</li></ul>	MIT/LCS	
The information is an income-badi, income-badis and the information of the polynomic polynomic polynomic polynomic (IET?), its area, and its working groups. Note that other (IET?), its area, and its working documents and intermet- badis. Intermet-badis are draft documents wild for a maximum of six months and may be updated, replaced, or made about by Intermet-brates are ference material of to cite them other than an 'work in progress'. To learn the current status of any Intermet-brate, please check the 'lid-abstratis, itt' listing groups are intermet-brate in through the status of any Intermet-brate (IS Bast Coast), or ftp.isl.edu (IS West Coast). Distribution of this document is unlimited. Please send sechet the The document is unlimited. Please send sechet at UBT working group are etter are and the place on the vow-talked.org mailing list. Abstract The hypertext Transfer Protocol (HTTP) is an application- level protocol for distributed, collaborative, hyperseci- ciented protocol which can be used for many task, such as a may are an an and distributed of data spresention. All distributed of data spresention.	This decount is an intermet-Draft, Intermet-Drafts are priving social in the Theorem Information State Free (IETF), its areas, and its working groups. Note that other (IETF), its areas, and its working documents as Intermet- Drafts. Intermet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or made obsolute by Intermet-Drafts are started material or to clut them other than as 'work in progress'. To learn the current status of any Intermet-Draft, please check the 'lid-sbuttactic stat' listing routined in the control of the document is unlimited. Please send check the 'lid-sbuttactic stat' listing routined (05 West Coast). Distribution of this document is unlimited. Please send comments to the MTT working group at chttp: decound in the TT working group at chttp: decound in the place on the wow-talked.org mailing list. Abstract The hypertext Transfer Protocol (MTTP) is an application- level protocol for distributed collaborative, hypersedia oriented of used for used for any state, such as a may any state of the document is under the place in a schirod at cullistic place on the wow-talked.org mailing list.	Hypertext Transfer Protocol HTTP/1.1	
working documents of the Internet Engineering Task Force (ITTT), is array, and its working documents as Internet- Darfs. Internet-Internet are distributed working documents as Internet- Darfs. Internet-Internet are distributed on the Internet of the Internet-Internet are distributed on the Internet than an 'work in progress'. To learn the current status of any Internet-Darfs, please check the 'lid-abstract.utt' listing contained in the Internet-Internet Sadow Directories on fit, iso, as (Africa), de Internet-Internet (Of Bast Coat), or fit, islewh (Of West Coat). Distribution of this document is unlimited. Please send commont is to the INTY working group as chttp- wereokon bak.hg.comb. Discusions of the working proop sci fit. Astract The hypertext Transfer Protocol (HTTP) is an application- level protocol for distributed, collaborative, Sympewsed coinsted protocol which can be used for many task, such as new servers and distributed object samagement systems and and megiciains of data septemention, allowing systems to be built independently of the data being transferred.	working documents of the Internet Engineering Task Force (ITTT), its area, and its working groups, how that other bards. Internet-Drafts are draft documents are internet- bards. Internet-Drafts are draft documents valid for a maximum of six month and may be updated, replaced, or made obsolute by other document at any time. It in langyrografts to use than as 'work in progress'. To learn the current status of any Internet-Draft, plases check the 'lid-abstracts.txt' listing contained in the internet-Drafts Baddor Directorise and the loss and Article documents.txt of the any Internet-Draft, plases check the 'lid-abstracts.txt' listing contained in the internet-Drafts Baddor Directorise and the loss and Article documents to the any Internet-Draft, plases check the 'lid-abstracts.txt' listing contained in the internet-Drafts Baddor Directorise and the loss and Article documents to the HT working proop at <htp: webcoko.bgl.hg.com. Discussions of the working proop arc essential the HTT working proop at the the internet-Draft and the applications which use HTT wholed take place on the <a href="http://www.baddits.type://wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww&lt;/td&gt;&lt;td&gt;Status of this Memo&lt;/td&gt;&lt;/tr&gt;&lt;tr&gt;&lt;td&gt;six moths and may be updated, replaced, or made obsolute by&lt;br&gt;other documents any time. It in inappropriate to use&lt;br&gt;that as " in="" progress".<br="" work="">To learn the current status of any to cita the other intermet-Tork Bindsor Directories on fip.inc.co. Intermet-Tork Bindsor Directories on fip.inc.co. Advisor of the status of the status of the status desintermic-ore (UE Bast Coast), or fip.inc.do (US West Coast). Distribution of this document is unlimited. Please send comments to the HTT working proop at chttp- weferbook hellsp.come. Discussions of the basis of the status wefferbook hellsp.come. Discussions of the basis of the status wefferbook hellsp.come. Discussions of the basis of the status "Wefferbook hellsp.come. Discussions of the basis of the basis description of the document is unlimited. Please send comments to the HTT working prime at the application the use HTT working that he applications while the list. The Hypertext Transfer Protocol (HTTP) is an application- lawi protocol for distributed, collaborative, Hypermetry ciented protocol which can be used for many tasks, such as new servers and distributed object management systems, allowing aystems to be built independently of the data being transferrod.</a></htp: 	six months and may be updated, replaced, or made obsolute by other documents at my time. It is inappropriate to use than as "work in progress". To learn the current status of any interest-braft, plases check the "lid-abstracts tat" listing contained in the intermet-fract Badev Directories on the listics can [Africa], ds.internic.net (US Bast Coast), or ftp.isl.edu (US West Coast). Distribution of this document is unlimited. Plases emd comments to the HTT working group at other, wyferwid hyl hip-tere. Jubersions directories on the listic transmission about HTT and the applications which use HTT should take place on the cow-talktwiors mailing list. Abstract The Hypertext Transfer Protocol (HTTP) is an application- level protocol for distributed, collaborative, hypermedia contents of the used for any tasks, such as news envers and distributed object management system, allocing systems to be built independently of the data being transfer	working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-
check the 'lid-abstract.tt' lising contained in the intermeth-frait Badow Directories on the licics an [Africa), intermeth-frait Badow Directories on the licics an [Africa), ds.intermic.ene(US East Coast), or ftp.isl.edu(US Went Coast). Distribution of this document is unlimited. Please send commant to the HTT working group at chttp- deneral directories about TTT and the applications which the sender of the sender of the sender of the sender archived at CMLintspi//wow.ics.usi.edu/pub/ietf/http/. deneral directories about TTT and the applications which list. Abstract The Mypertext Transfer Frotocol (MTTP) is an application- level protocol for distributed, collaborative, Mypermedia cointed protocol which can be used for many task, such as name servers and distributed object management systems. allowing systems to be built independently of the data being transferred.	check the "lid-abstracts.tx" liking contained in the intermet infat Bindow Directions on Typilo.co.al direction of the state of the state of the state of the ds.intermic.net(OS East Coast), or fop.isi.edu (OS West Coast). Distribution of this document is unlimited. Please send commont to the HTT working group at citty- ds.intermicent of the state of the state of the state service of the state of the state of the state of the service of the state of the state of the state of the service of the state of the state of the state of the service of the state of the state of the state of the service of the state of the state of the state of the state of the state of the state of the state of the state of the list of the state of the state of the state of the state of the list of the state of the s	six months and may be updated, replaced, or made obsolete by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other	
comments to the HTFP working group at chttp- wijerucko.phi.ph.com. Discussions of the working group, are wijerucko.phi.ph.com. Discussions of the working troup, are descrid discussions about HTFP and the applications which use HTTP should take place on the vow-taikWJ.orgp mailing Abstract The Hypersex Transfer Protocol (HTTP) is an epication- level protocol for distributed, collaborative, hypermedia information systems. It is a generic, stateless, object- oriented protocol Which can be used for many takek, such as through extension of its request methods. A feature of HTTP is the typing and negotiation of data represention, allowing systems to be built independently of the data being transferred.	comments to the HTTP working group at thtp- wiferucto.phi.ph.come. Discoursions of the working group, are wiferucto.phi.ph.come. Discoursions of the working group. General discoussions about HTTP and the applications which use HTTP should take place on the «now-takkWolorey mailing list. Abstract The Hypertext Transfer Protocol (HTTP) is an application- level protocol for distributed collaborative, hypermedia information systems. It is a generic, stateless, object- oriented protocol which can be used for many takes, such as through extension of its request scholes, a feature of HTTP is the typing and megotiation of data representation, allowing systems to be built independently of the data being transferred.	check the "lid-abstracts.txt" listing contained in the Internet-Drafts Shadow Directories on ftp.is.co.za (Africa), nic.nordu.net (Europe), munnari.oz.au (Pacific Rim), ds.internic.net (US East Coast), or ftp.isi.edu (US West	
The Hypertext Transfer Protocol (HTF) is an application- text of a second of a distributed of allaberting hypermedi- information systems. It is a generic, stateless, object- oriented protocol which can be used for many tasks, such as new servers and distributed object anagement systems, and and allowing the second of the systems, mer- it be typical megodiation of data representation, allowing systems to be built independently of the data being transferred.	The hypertest Transfer Protocol (NTT) is an application- new of percoad for distribution of bilaberative, hyperaedla information systems. It is a generic, statedes, object- oriented protocol which can be used for many tasks, such as news servers and distributed object management systems, mern is the typical and negotiation of data representation, allowing systems to be built independently of the data being transferred.	comments to the HTTP working group at <http- wg@cuckoo.hpl.hp.com&gt;. Discussions of the working group are archived at <urlihttp: http="" ietf="" pub="" www.ics.uci.edu=""></urlihttp:>. General discussions about HTTP and the applications which use HTTP should take place on the <www-talk@w3.org> mailing</www-talk@w3.org></http- 	
level protocol for distributed, collaborative, hypermedia information systems. It is a generic, stateless, object- oriented protocol which can be used for many tasks, such as name servers and distributed object management systems, through extension of its request methods. A feature of HTPP allowing systems to be built independently of the data being transferred.	level protocol for distributed, collaborative, hypermedia information systems. It is a generic, stateless, object- oriented protocol which can be used for many tasks, such as name servers and distributed object management systems, through extension of its request methods. A feature of HTP allowing systems to be built independently of the data being transferred.	Abstract	
Fielding, et al [Page 1]	Fielding, et al [Page 1]	level protocol for distributed, collaborative, hypermedia information systems. It is a generic, stateless, object- oriented protocol which can be used for many taaks, such as mase servers and distributed object nangement systems, much and the system server and the server and the server and the lowing systems to be built independently of the data being	
		Fielding, et al [Page 1]	



HTTP Working Group R. Fielding INTERNET-DRAFT UC Trvine <draft-ietf-http-v11-spec-rev-00.txt> J. Gettys J. C. Mogul DEC H. Frystyk T. Berners-Lee MIT/LCS Expires January 30, 1998 July 30, 1997 Hypertext Transfer Protocol -- HTTP/1.1 Status of this Memo This document is an Internet-Draft. Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.



RFC 2616					
Document II	ESG evaluati	on record II	ESG writeups Email expansions History		
Revision	differe	nces			
From revision draft-ietf-http-v11-spec-rev-06 (1998-11-24)					
To revision rfc2616 (1999-			-06-01) -		
Diff format	1	Side-by-side	Before-after Change bars Wdiff Submit		
Documer	nt histo	ory			
	Rev. +		* Action * RFC published		
Date +	Rev. +	By ¢			
Date +	Rev. +	By + (System)	RFC published		
Date + 1999-06-01 1999-02-25	Rev. RFC 06	By + (System) (System)	RFC published IESG has approved the document		
Date         +           1999-06-01         1           1999-02-25         1           1998-11-24         1	Rev.         #           RFC         06           06         06	By (System) (System) (System)	RFC published IESG has approved the document New version available: draft-ietf-http-v11-spec-rev-06.txt (diff from previous)		
Date         +           1999-06-01         1           1999-02-25         1           1998-11-24         1           1998-09-30         1	Rev.         +           RFC         -           06         -           05         -	By • (System) (System) (System) (System)	RFC published IESG has approved the document New version available: draft-ietf-http-v11-spec-rev-06.txt (diff from previous) Last call sent		
Date         *           1999-06-01         1           1999-02-25         1           1998-11-24         1           1998-09-30         1           1998-09-15         1	Rev.         +           RFC         -           06         -           06         -           05         -	By • (System) (System) (System) (System) (System)	RFC published IESG has approved the document New version available: draft-ietf-http-v11-spec-rev-06.txt (diff from previous) Last call sent New version available: draft-ietf-http-v11-spec-rev-05.txt (diff from previous)		
Date         +           1999-06-01         1           1999-02-25         1           1998-11-24         1           1998-09-30         1           1998-09-15         1           1998-09-14         1	Rev.         +           RFC            06            06            05            05            04	By ¢ (System) (System) (System) (System) (System) (System)	RFC published IESG has approved the document New version available: draft-ietf-http-v11-spec-rev-06.txt (diff from previous) Last call sent New version available: draft-ietf-http-v11-spec-rev-05.txt (diff from previous) New version available: draft-ietf-http-v11-spec-rev-04.txt (diff from previous)		
Date         •           1999-06-01         1           1999-02-25         1           1998-01-24         1           1998-09-30         1           1998-09-15         1           1998-08-14         1           1998-04-07         1	Rev.         +           RFC         0           06         0           05         0           05         0           05         0           05         0           05         0           05         0           05         0           05         0           05         0           05         0	By         +           (System)         -           (System)         -	RFC published IESG has approved the document New version available: draft-ietf-http-v11-spec-rev-06.txt (diff from previous) Last call sent New version available: draft-ietf-http-v11-spec-rev-05.txt (diff from previous) New version available: draft-ietf-http-v11-spec-rev-04.txt (diff from previous) New version available: draft-ietf-http-v11-spec-rev-05.txt (diff from previous)		



### Main Variables

- Decisions: **Event** in  $t \in \{\text{Continue}, \text{Abandon}, \text{Publish}\}$
- **Citations**  $\Leftrightarrow$  E[ $\pi$ |Publish, Version]
  - Count of non-patent prior art references to RFC
  - Alternative: Count of RFC references to previous RFC
- Demographics
  - Cohort, team size (authors), experience (max{previous IDs})
- Feedback
  - Count of emails that specifically mention ID
- Revision size
  - Cosine distance for document-pair word vectors

### **Estimation Sample**

IETF Submissions: 1996-2009

	Full	Working	Stds-track		Nonstds-
	Sample	Group	Aband.	Publ.	track
WG (%)	24.44	100.00	14.11	65.25	46.39
Team Size (Author Count)	2.28	2.45	2.22	2.43	2.48
Experience (max Projects)	15.01	15.69	13.50	21.87	16.84
Versions	3.55	5.60	2.09	9.33	6.71
Communication (Emails)	21.20	33.78	9.87	69.19	41.03
Published RFC (%)	23.97	56.10	0.00	100.00	100.00
Citations	2.99	8.30	0.76	12.23	7.19
N (Projects)	16,091	3,932	12,234	2,210	1,647
N (Versions)	57,179	22,025	25,511	20,622	11,046

- Author-teams on individual and WG submissions have similar size and experience; but strong correlation between experience and publication outcome
- WG proposals account for 25% of all IDs, but have higher publication rates
- Published IDs go through more revisions than abandoned IDs
- Patent citations increase with publication as RFCs; standards-track proposals are cited more

### Mean Distance to Original Proposal



- Textual distance of a version *T* from initial version *t* = 1
- Proposals change throughout the revision process
- Plot very similar for standards-track and nonstandards-track proposals

n

$$\mathsf{dist}(T,1) \equiv 1 - \frac{x_T \cdot x_1}{||x_T|| \ ||x_1||} \qquad \mathsf{with} \ ||x_i|| = \sqrt{\sum_{l=1}^n x_{i,l}^2}.$$

Ganglmair, Simcoe, and Tarantino

Learning When to Quit

### Hazard Rates



- Probability of exit, conditional on survival to t
- 40% of proposals never revised
- Some immediate publication of nonstandards (no learning)
- Publication hazard (for standards-track) always below 16%: Authors continue to revise ⇒ 2-phase model

### Citations Increase with Revisions



- Citations from U.S. patents (NPL) as proxy for commercial impact
- Citations increase with number of revisions
- Lower for nonstandards-track RFCs than standards

 $Cites_i = \alpha_y + (\beta_1 + \beta_2 Nonstandard_i) + \log(Versions_i) + \varepsilon_i$ 

Ganglmair, Simcoe, and Tarantino

Learning When to Quit

## Implications

- Without learning, linear hazard rates of abandonment ...
- ... but data suggests process is not *purely* about learning
  - Distance measure  $\rightarrow$  cumulative improvement
- Data suggest two phases: before and after consensus
  - Nonstandards also take time to publish: not all are immediately published
  - Decreasing hazard rates of abandonment, increasing hazard rates of publication
- Empirical goals:
  - Recover primitives: opportunity costs, learning parameter, prior belief of project quality
  - Replicate hazard rates

### Model & Estimation

### **Bayesian Learning Model**

- Model of R&D in the spirit of Roberts and Weitzman (1981)
- Researchers develop a project; endowed with an initial version/idea of unknown quality (good/bad)
- Only good projects generate value; good type must be learned (via breakthrough  $\sim$  consensus)
- Two-phase model:
  - Costly experiments to learn the project's type ( $\rightarrow$ two-armed bandit)
  - Once type is learned in a period  $\tau$ , further project revisions to increase value ( $\rightarrow$ stopping problem)
- In each period, after observing a cost shock  $\varepsilon_t \sim G(\cdot)$ , the team decides to continue or stop

### Bayesian Learning Model



Ganglmair, Simcoe, and Tarantino

Learning When to Quit

### Stylized Model: Two-Period Example Illustrate: and Motivate Empirical Identification

- p: prior belief/probability that project is good
- *b*: probability that researcher learns good type (breakthrough) after posting a revision (= 0 if bad project)
- $\hat{\pi}_t$ : payoffs when published in t;  $\hat{\pi}_1 < \hat{\pi}_2$
- $(A_t, S_t)$ : share of abandoned/published in t:  $A_1 + A_2 + S_1 + S_2 = 1$
- Two-period model timing:
  - t = 0: researcher endowed with initial proposal (good with p)
  - t = 1: researcher observes  $\varepsilon_1$ ; decides to *continue* or *stop* (allowing for learning with b)
  - t = 2: if good type revealed, project is success (post-breakthrough), otherwise failure (pre-breakthrough)





Beliefs in t = 1 are updated using Bayes' rule:

$$\hat{p}^{\mathsf{pre}} = rac{p(1-b)}{1-pb},$$

for pre-breakthrough, and  $\hat{p}^{\text{post}} = 1$  for post-breakthrough.

Ganglmair, Simcoe, and Tarantino

Learning When to Quit

### Trade-Offs

- Before breakthrough (not yet learned good type):
  - Beliefs: Failed experiments  $\Rightarrow$  researchers become more pessimistic

$$\hat{p}^{\mathsf{pre}} < p$$

• Benefits: More revisions  $\Rightarrow$  larger potential publication payoff

$$\hat{\pi}_2 > \hat{\pi}_1 > 0$$

- Tradeoff: Exploitation vs. exploration
- Continue if and only if

$$b\hat{p}^{\mathsf{pre}}\hat{\pi}_2 - [F^{\mathsf{pre}}(1) + \varepsilon_1] \geq 0 \qquad 
ightarrow \qquad G(b\hat{p}^{\mathsf{pre}}\hat{\pi}_2 - F^{\mathsf{pre}}(1))$$

### Trade-Offs

### • After breakthrough (learned good type):

- Beliefs: Good type has been learned so that  $\hat{p}^{\text{post}} = 1$
- Tradeoff: Simple comparison of marginal costs and benefits
- Continue if and only if

$$\hat{\pi}_2 - \left[ \mathsf{F}^{\mathsf{post}}(1) + arepsilon_1 
ight] \geq \hat{\pi}_1 \qquad 
ightarrow \qquad \mathsf{G}(\hat{\pi}_2 - \hat{\pi}_1 - \mathsf{F}^{\mathsf{post}}(1))$$

### From Model to Identification

• Use conditional choice probabilities together with data on payoffs and distribution of the structural error terms to identify cost functions:

$$egin{aligned} \mathsf{Pr}(\mathsf{stop}|\mathsf{post}) &= rac{\mathcal{S}_1}{\mathcal{N}_1^\mathsf{post}} = 1 - \mathcal{G}(\hat{\pi}_2 - \hat{\pi}_1 - \mathcal{F}^\mathsf{post}(1)) \ \mathsf{Pr}(\mathsf{stop}|\mathsf{pre}) &= rac{\mathcal{A}_1}{\mathcal{N}_1^\mathsf{pre}} = 1 - \mathcal{G}(b\hat{p}^\mathsf{pre}\hat{\pi}_2 - \mathcal{F}^\mathsf{pre}(1)) \end{aligned}$$

where

$$\begin{split} N_1^{\text{post}} &= S_1 + S_2 - A_2 \frac{b\hat{p}^{\text{pre}}}{1 - b\hat{p}^{\text{pre}}} \\ N_1^{\text{pre}} &= A_1 + A_2 \frac{1}{1 - b\hat{p}^{\text{pre}}} \end{split}$$

are the number of projects in the post- and pre-breakthrough phases.

Ganglmair, Simcoe, and Tarantino

Learning When to Quit

# From Model to Identification

### Theorem

For any  $b \ge \underline{b}$ , there exists a unique p(b) and F(b) such that for all t, the probability of publication (abandonment) in period t equals  $S_t$  ( $A_t$ ).

- For any admissible *b*, we can choose other structural parameters to rationalize the data
- For point identification, we need more information
- We consider the cost functions!

### From Model to Identification

### Assumption

Revisions costs are independent of the breakthrough phase:

$$F^{post}(1) = F^{pre}(1) = F(1)$$

Then:

### Theorem

If  $G(\cdot)$  is the logistic cdf and given Assumption 1, there exists a unique solution so long as  $p \ge 1 - A_1^2$ .

- We establish this for a general  $\overline{T}$  period model
- Point identification of b, p, and costs F(t)

### Assumption

Nonstandards-track projects have the same cost function F(t) as standards-track projects, but are always published (i.e., p = b = 1).

- Second identification strategy, exploiting nonstandards-track RFCs as unique institutional feature of the IETF
- Reasonable assumption: the two tracks are very similar in terms of content and publication process, but differ in commercial implication
- Under this Assumption 2, costs are identified by payoffs and distribution of stopping times for nonstandards-track RFCs, while b and p identified by the payoffs and stopping times for the standards-track RFCs

### From Two-Periods Model to ML Estimation

- Time horizon:  $\overline{T} = 25$  (robust to higher values of  $\overline{T}$ )
- Quadratic incremental costs:

$$F(t) = C_0 + C_1 t + C_2 t^2$$

•  $\varepsilon_t \sim \text{Logistic}(0,1)$  with CDF

$$G(\varepsilon) = rac{1}{1 + \exp(-\varepsilon)}.$$

### **Estimation Procedure**

- 1. Values  $\hat{\pi}(t)$  are predicted citations of projects in U.S. patents
- 2. If Assumption 2: Estimate costs for sample of nonstandards-track projects, based on  $\hat{\pi}(t)$  for nonstandards-track from Step 1.
- Iteratively search for values of (b, p, F) that maximize the log-likelihood, retaining estimates of F from Step 2 if using Assumption 2.

### Results
## Baseline: Ex-Ante Identical Projects

	Full Sample	WG Sample
Rate of Learning ( <i>b</i> )	0.17 [0.16,0.20]	0.34 [0.32,0.37]
Quality Prior ( <i>p</i> )	0.59 [0.51,0.73]	0.73 [0.64,0.83]
Costs $F(1)$	2.35 [1.95,3.25]	3.74 [2.90,5.15]
Costs $F(10)$	1.25 [1.12,1.81]	1.14 [1.01,1.72]
Costs <i>F</i> (20)	0.51 [0.48,0.65]	0.66 [0.59,0.86]
Projects: Standards track	14,444	3,168

Ganglmair, Simcoe, and Tarantino

## Empirical and Simulated Hazard Rates



## Robustness to Alternative Assumptions

- 1. Area Specific Citations:
  - Allow for citations (Step 1) to vary with observed technology areas
- 2. Calendar Time:
  - Duration of projects measured in quarters rather than versions
- 3. Phase-Specific Costs:
  - Relax Assumption 1:  $F^{\text{post}}(t) = F^{\text{pre}}(t) + \kappa$
- 4. Three-Step Estimation:
  - Including Assumption 2
- 5. Others: Include censored projects; RFC citations;  $\overline{T} = 50$ ; only first project for each WG

Ganglmair, Simcoe, and Tarantino

Learning When to Quit

January 2018 35 / 48

## **Robustness Results**

	Baseline	Area Specific	Calendar Time	Phase Specific ( $\kappa$ )	3-Step Estimator
Rate of Learning <i>b</i>	0.17	0.10	0.20	0.26	0.30
	[0.16,0.20]	(0.003)	(0.003)	(0.006)	[0.27,0.34]
Quality Prior <i>p</i>	0.59	0.59	0.49	0.42	0.45
	[0.51,0.73]	(0.008)	(0.003)	(0.007)	[0.28,0.59]

## Baseline Estimates: Summary

- Roughly 60% of projects submitted to IETF can potentially generate value
- One third of these is published (successful)
- Arrival rate of breakthrough/consensus is 17% per revision
- Our parsimonious model is robust to changes in estimation assumptions and sampling
- Inform theory literature by estimating parameters driving Bayesian learning process in models of experimentation

### Counterfactuals

Counterfactuals

## **Counterfactual Simulations**

#### • R&D Cost Subsidies and Publication Prizes

- Which of these generates more value?
- Over-confidence and Pessimism
  - What are the costs of misaligned priors?
  - Is over-confidence good for R&D?

## Prizes and Subsidies

- **Prize:** award  $\delta_p \ge 0$  "citations" to published RFCs
  - akin to a publication prize, or higher patent values
- **Subsidy:** reduce "citation cost" by  $\delta_s \ge 0$  for *all* revisions
  - akin to an R&D tax credit: not conditional on success of the investment
- Keeping the overall budget (in terms of cost of the regime) constant:
  - How do these policies affect authors' incentives?
  - Which policy generates larger values (to authors/organization)?

## Incentive Effects

- Differential effects on exploitation-exploration trade-off
- **Prize:** effect weakens in t
  - After consensus there is no effect, hence no effect on projects that are successful in the baseline without a prize.

incremental increase of  $\hat{\pi} = \hat{\pi}(t+1) + \delta_{p} - [\hat{\pi}(t) + \delta_{p}]$ 

- Before consensus, the relative effect is decreasing with *t* because authors update beliefs and expect to receive prize with lower probability
- Subsidy: effect increases in t
  - Paid per revision, both before and after consensus
  - Constant in absolute terms but increasing in relative terms as incremental opportunity costs are decreasing

# Citations (Gross Value)



- Subsidies with stronger effect on citations than prize
- Subsidies with effect both on intensive and extensive margin
- Both policies with elasticity larger than one

#### Net Project Value



- Total value of projects, accounting for authors' costs of submitting (more) revisions
- Only a (modest) prize can increase value

## Optimism and Pessimism

- What are the costs of misaligned priors about project quality?
- Suppose p is the true (estimated) prior quality, and authors have subjective priors p'.
  - True prior determines the selection of project types (good and bad)
  - Subjective priors determine the authors' decisions to continue
- We refer to p' > p as "over-confidence" and and p' < p as "pessimism"
- Take (p, b) as given, and consider an author-team with prior p'; simulate standards-track projects for p' ∈ (0, 1)

### **Project Value**



- Both over-confidence and pessimism are distortive
- Costs of over-confidence in terms of total value are higher
- Over-confident researchers over-invest in unpromising lines of research ("dry wells")

Counterfactuals Optimism and Pessimism

#### Project Duration



- More over-confident author-teams continue longer
- For higher values of p', this effect is driven by bad projects

Ganglmair, Simcoe, and Tarantino

#### **Results: Project Heterogeneity**

Results: Project Heterogeneity

## Exploit Observed Project Heterogeneity

- 1. Communication: positive effect on rate of learning
- 2. Experience: positive effect on rate of learning
- 3. Commerciality: suit-to-beard ratio with negative effect on rate of learning
- 4. Team Size: positive effect on rate of learning

Conclusions

# Summary of Results

- Parsimonious model of dynamic learning in R&D decision-making at the project level
- Data from the IETF (with institutional features for identification)
- Learning (b) "about use of time:"
  - Higher rate of learning: bad projects are abandoned faster, yielding higher-value process
  - More communication, experience, commerciality, and team size induce a higher rate of learning
- Prizes yield better outcome than subsidies (~R&D tax credit) when accounting for private costs of overdeveloping bad ideas
- Over-confidence is more costly than pessimism

Ganglmair, Simcoe, and Tarantino

Conclusions

# What's Next?

- What is in a patent?
  - Combine data on patent declarations to the IETF to study how patent disclosure changes project content
- Network structure
  - Consider researchers in a network: Is there a relationship between nature of network ties, author affiliation/centrality/experience and outcomes?
- Attention
  - Use information on emails to study causal relationship between communication and outcomes

#### Thank you!

Please send comments or suggestions to

ganglmair@utdallas.edu tsimcoe@bu.edu or tarantino@uni-mannheim.de

## Major Contributors

#### Count of Project Submissions

<u>1992-1994</u> 1. Cisco 2. Carnegie Mellon 3. mtview.ca.us 4. IBM 5. SNMP Research	94 51 48 44 38	<u>1992-2004</u> 1. Cisco 2. Nortel 3. Microsoft 4. Nokia 5. Sun Microsystems	1,787 694 581 539 513
<u>1995-1997</u> 1. Cisco 2. IBM 3. Microsoft 4. Sun Microsystems 5. USC (ISI)	214 140 140 84 79	<ol> <li>AT&amp;T</li> <li>IBM</li> <li>Ericsson</li> <li>Lucent</li> <li>Bell Labs</li> </ol>	513 490 398 343 301
<u>1998-2000</u> 1. Cisco 2. Nortel 3. AT&T 4. Microsoft 5. Sun Microsystems	517 321 223 221 180	<ol> <li>Alcatel</li> <li>Juniper Networks</li> <li>Intel</li> <li>Columbia U.</li> <li>Siemens</li> </ol>	299 260 225 220 200
2001-2004 1. Cisco 2. Nokia 3. Nortel 4. Ericsson 5. Sun Microsystems	962 404 354 279 234	<ol> <li>Dynamicsoft</li> <li>USC (ISI)</li> <li>ACM</li> <li>MIT</li> <li>NTT</li> </ol>	196 195 185 152 149

### Dynamic Host Configuration Protocol RFC 2131

Abstract

The Dynamic Host Configuration Protocol (DHCP) provides a framework for passing configuration information to hosts on a TCPIP network. DHCP is based on the Bootstrap Protocol (BOOTP) [7], adding the capability of automatic allocation of reusable network addresses and additional configuration options [19]. DHCP captures the behavior of BOOTP relay agents [7, 21], and DHCP participants can interoperate with BOOTP participants [9].



#### Known TCP Implementation Problems RFC 2525

#### Introduction

This memo catalogs a number of known TCP implementation problems. The goal in doing so is to improve conditions in the existing Internet by enhancing the quality of current TCP/IP implementations. It is hoped that both performance and correctness issues can be resolved by making implementors aware of the problems and their solutions. In the long term, it is hoped that this will provide a reduction in unnecessary traffic on the network, the rate of connection failures due to protocol errors, and load on network servers due to time spent processing both unsuccessful connections and retransmitted data. This will help to ensure the stability of the global Internet.

