Learning When to Quit: An Empirical Model of Experimentation in Standards Development

Bernhard Ganglmair   Timothy Simcoe   Emanuele Tarantino

NBER Summer Institute
July 17, 2019

Experimentation in Practice

"I haven't failed. I've just found 10,000 way that won't work."
- *Thomas Edison*

Experimentation in Practice

"I haven't failed. I've just found 10,000 way that won't work."
- *Thomas Edison*

CIHR_Wisdom
@CihrWisdom

How science works: fail fail fail fail fail fail fail fail fail succeed fail fail fail fail fail fail fail fail fail fail fail fail fail fail fail succeed fail fail fail fail fail fail fail fail fail fail fail fail fail fail fail fail fail fail fail fail fail succeed fail fail

03/29/18, 1:24 PM

1.1K RETWEETS **3.2K** LIKES

What We (Economists) Typically Observe

CIHR_Wisdom
@CihrWisdom

How science works: succeed
succeed
succeed

03/29/18, 1:24 PM

1.1K RETWEETS **3.2K** LIKES

This Paper

- Study learning in the *Internet Engineering Task Force* (IETF) – version-level data on 16k projects
- Engineers at IETF *develop* core internet protocols (TCP/IP, SIP, ...)
  Success = publication and failure = abandonment
- Learning about quality of idea/project by receiving "consensus" from community of peers

We Ask:

1. How do engineers come to project success or failure?
   - Anatomy of experimentation process
2. How fast do teams learn?
   - Dynamic discrete choice problem
   - Costly experimentation $\Rightarrow$ false negatives
   - Team experience, size, communication, demographics
3. Counterfactuals
   - Publication prize vs. cost subsidy
   - Costs of over/under confidence

Summary of Findings

1. Descriptive analysis of R&D in IETF:
   - Approx 3/4 projects abandoned, 40% with no revision
   - Content and impact increase with revisions
   - Consensus does not yield immediate publication
2. Parsimonious model of dynamic learning in R&D:
   - Pr[Good idea] = 59%; Pr[Consensus_t| good idea] = 17%
   - False negatives: 2/3 of "good" projects not published
   - Decreasing opportunity costs of revision
   - Experience, communication ⇒ faster learning

Contributions

- Simple DDC model of experimentation and learning
  - Erdem and Keane (1996); Crawford and Shum (2005); Dickstein (2014)
  - Pakes (1986); Nanda and Rhodes-Kropf (2015); Krieger (2017)
- Estimates of "rate of learning" and "revision costs" in consensus standardization
  - Rysman and Simcoe (2008); Fleming and Waguespack (2009); Simcoe (2012); Ganglmair and Tarantino (2014); Baron and Spulber (2019)
- Complement to "missing Einsteins" in patenting result
  - Bell, Chetty, Jaravel, Petkova and Van Reenen (2017)

Institution: The Internet Engineering Task Force (IETF)

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
- Largely decentralized platform, with exceptions: Area Directors expected to block projects that are in conflict with each other
- 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

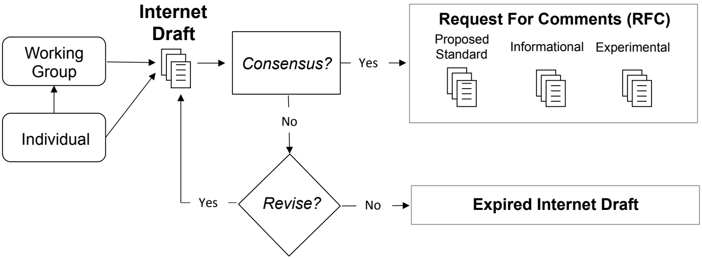
IETF Protocol Examples

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
TCP	Transmission Control Protocol	1981
IP	Internet Protocol	1981

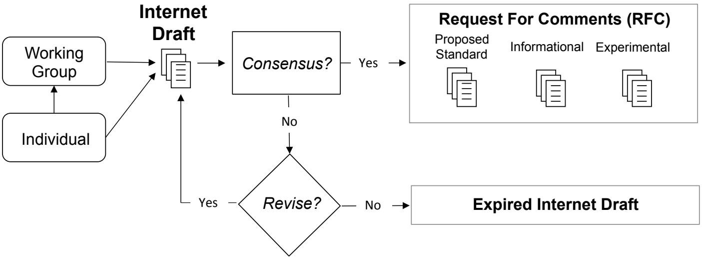
▸ Major Contributors

IETF Standardization



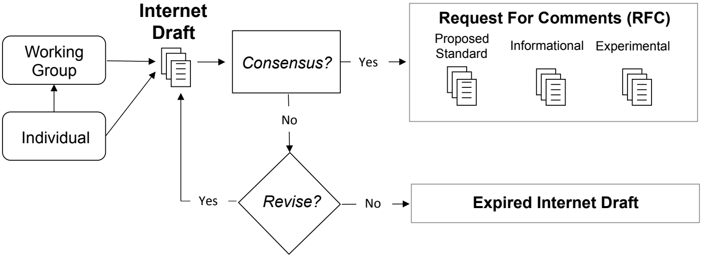
1. Identify problem and submit proposal (Internet Draft or ID)
   - Two types: Individual and Working Group
   - Two tracks: standards and non-standards
   - All projects posted to public repository

IETF Standardization



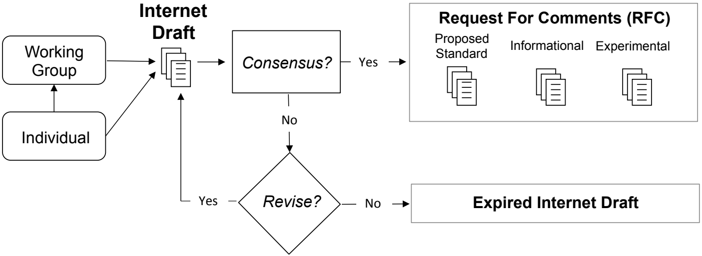
2. Community feedback via email and meetings

IETF Standardization



3. **Rough consensus** ⇒ ID published
   - Decision by WG Chair and IESG (*de facto* super-majority)
   - IETF guidelines: *"strongly held objections must be debated until most people are satisfied that these objections are wrong"*
   - Published ID's called Proposed Standards (or RFCs)

IETF Standardization



4. No consensus ⇒ sponsors have a choice
   - Revise ID → return to step (1) [submit revision]
   - Abandon ID → expires in 6 months

Data

```
HTTP Working Group                                          R. Fielding
INTERNET-DRAFT                                                UC Irvine
<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.

Internet-Drafts are draft documents valid for a maximum of 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 than as "work in progress".

To learn the current status of any Internet-Draft, please check the "1id-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 Coast).

Distribution of this document is unlimited. Please send comments to the HTTP working group at <http-wg@cuckoo.hpl.hp.com>. Discussions of the working group are archived at <URL:http://www.ics.uci.edu/pub/ietf/http/>. General discussions about HTTP and the applications which use HTTP should take place on the <www-talk@w3.org> 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 tasks, such as name servers and distributed object management systems, through extension of its request methods. A feature of HTTP is the typing and negotiation of data representation, allowing systems to be built independently of the data being transferred.

Fielding, et al [Page 1]

▸ URL

```
HTTP Working Group                                       R. Fielding
INTERNET-DRAFT                                             UC Irvine
<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.
```

▸ URL

Hypertext Transfer Protocol -- HTTP/1.1

RFC 2616

Document | IESG evaluation record | IESG writeups | Email expansions | History

Revision differences

From revision draft-ietf-http-v11-spec-rev-06 (1998-11-24)

To revision rfc2616 (1999-06-01)

Diff format Side-by-side | Before-after | Change bars | Wdiff | Submit

Document history

Date	Rev.	By	Action
1999-06-01	RFC	(System)	RFC published
1999-02-25	06	(System)	IESG has approved the document
1998-11-24	06	(System)	New version available: draft-ietf-http-v11-spec-rev-06.txt (diff from previous)
1998-09-30	05	(System)	Last call sent
1998-09-15	05	(System)	New version available: draft-ietf-http-v11-spec-rev-05.txt (diff from previous)
1998-08-14	04	(System)	New version available: draft-ietf-http-v11-spec-rev-04.txt (diff from previous)
1998-04-07	03	(System)	New version available: draft-ietf-http-v11-spec-rev-03.txt (diff from previous)
1998-02-24	02	(System)	New version available: draft-ietf-http-v11-spec-rev-02.txt (diff from previous)
1997-11-26	01	(System)	New version available: draft-ietf-http-v11-spec-rev-01.txt (diff from previous)
1997-10-16	00	(System)	New version available: draft-ietf-http-v11-spec-rev-00.txt

▸ URL

Main Variables

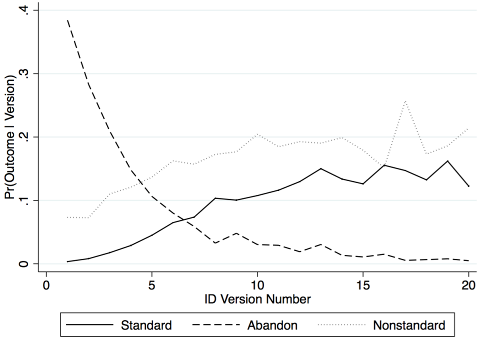
- Unit of observation: Internet Draft (i) Version (t)
- **Event**$_{it} \in \{$Revise, Abandon, Publish$\}$
- **Citations** $\Leftrightarrow$ E[π_i| Publish, t]
  - Count of non-patent prior art references to RFC
  - *Alternative: Count of RFC references to previous RFC*
- Author-team demographics
  - Cohort, team size (# authors), experience (max{previous IDs})
- Version feedback
  - Count of emails that specifically mention ID

Estimation Sample

IETF Submissions: 1996–2009

	Full Sample	Working Group	Stds-track Aband.	Stds-track Publ.	Nonstds-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

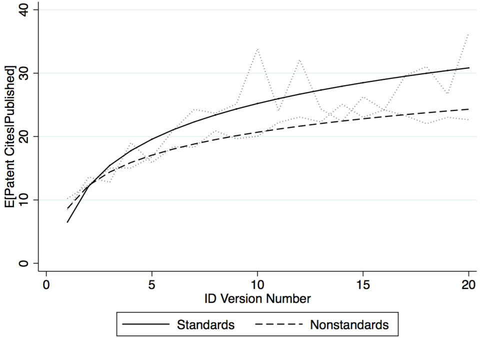
Fact #1: Hazard Rates



Pr[Event | survival to t]

- 40% of IDs never revised
- Increasing pub hazard
- Standards vs. nonstandards (“no learning” controls)
- $\Pr[\text{Publish}|t > 10] < 16\%$
  - Persistent revision ⇒ 2-phase model

Fact #2: Impact Increasing with Revisions



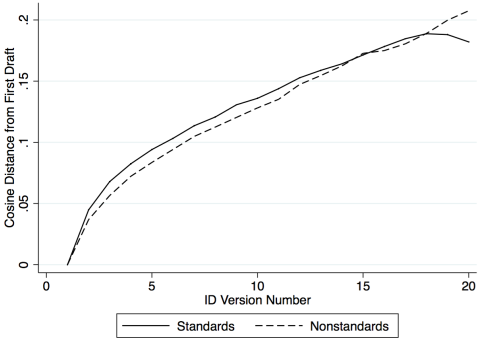
U.S. patent NPL cites proxy commercial impact

- Citations increase with number of revisions
- Lower for nonstandards-track RFCs than standards

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

Fact #3: Increasing Revision Distance

Revision ⇒ Substantive Change



- 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

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

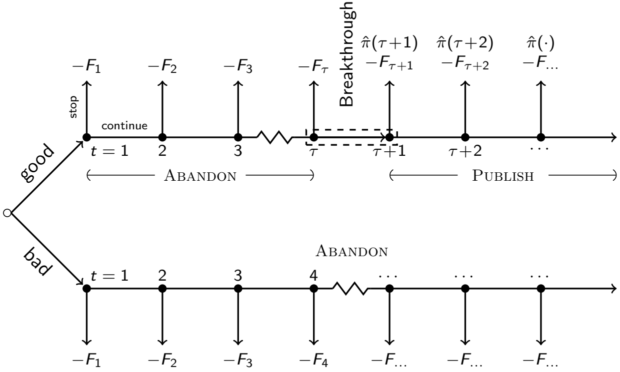
Model, Identification & Estimation

Empirical Model Overview

- Two-phases
  - Costly experiments to learn project quality (→**two-armed bandit**)
  - Further revisions to increase payoff (→**stopping problem**)
- Key Parameters
  - Project quality dist'n (p), Rate of learning (b), Revision costs (F)
- Identifying variation: hazard rates
- **Challenge**: Non-stationary DDC with unobserved state

Experimentation with Bayesian Learning

- Team gets project/idea with type $\theta \in \{\text{good,bad}\}$
  - Common prior: $Pr[\theta = \text{good}] = p$
- Team learns consensus ($\sigma_t = 1$) via "breakthrough"
  - $Pr[\sigma_{t+1} = 1 | \sigma_t = 0, \theta = \text{good}] = b$
  - $Pr[\sigma_{t+1} = 1 | \sigma_t = 0, \theta = \text{bad}] = 0$
- Beliefs: $\hat{p}(t) \equiv Pr[\theta = \text{good} | \sigma_t = 0] = \frac{p(1-b)^t}{(1-p)+p(1-b)^t}$
- Payoffs: $\pi(T, \sigma_T) = \sigma_T \hat{\pi}(T) - F_T$
  - $\hat{\pi}(t)$ increasing in t (from citations)
- Timing: 1) iid cost shock ε_t observed; 2) update beliefs $\hat{p}(t)$; 3) decide to **Revise** or **Stop**



$$F_t = \sum_{k=0}^{t-1} F(k, \sigma_k) + \varepsilon_k, \qquad F(0) = \varepsilon_0 = 0$$

Trade-Offs

- **Before consensus**: Exploration vs. exploitation
  - Beliefs: Increasing pessimism, $\hat{p}(t+1) < \hat{p}(t)$
  - Benefits: Increasing payoffs, $\hat{\pi}(t+1) > \hat{\pi}(t)$
  - Continue if and only if

$$b\hat{p}_t V_{t+1}^{\sigma=1} + (1 - b\hat{p}_t) V_{t+1}^{\sigma=0} \geq F_t + \varepsilon_t \qquad \Rightarrow \qquad \varepsilon_t \leq \bar{\varepsilon}_t^0$$

Trade-Offs

- **Before consensus**: Exploration vs. exploitation
  - Beliefs: Increasing pessimism, $\hat{p}(t+1) < \hat{p}(t)$
  - Benefits: Increasing payoffs, $\hat{\pi}(t+1) > \hat{\pi}(t)$
  - Continue if and only if

$$b\hat{p}_t V_{t+1}^{\sigma=1} + (1 - b\hat{p}_t) V_{t+1}^{\sigma=0} \geq F_t + \varepsilon_t \qquad \Rightarrow \qquad \varepsilon_t \leq \overline{\varepsilon}_t^0$$

- **After consensus**: Optimal stopping
  - Beliefs: Consensus achieved, so $\hat{p} = 1$
  - Tradeoff: Marginal costs and benefits of "polishing"
  - Continue if and only if

$$V_{t+1}^{\sigma_t=1} - [F_t + \varepsilon_t] \geq \hat{\pi}_t \qquad \Rightarrow \qquad \varepsilon_t \leq \overline{\varepsilon}_t^1$$

Estimation Assumptions

- Players know $\hat{\pi}(t) = E[Cites|t, \text{year} = 2000]$
  - Estimated in separate log-linear first stage
- Time horizon: $\overline{T} = 25$ (robust to higher values of $\overline{T}$)
- Quadratic revision costs: $F(t) = C_0 + C_1 t + C_2 t^2$
- Revision costs independent of breakthrough: $F(t, \sigma) \equiv F(t)$
  - Provides $\overline{T} - 1$ over-identifying restrictions
  - Can relax somewhat: $F(t, 1) = F(t, 0) + \kappa$
- $\varepsilon_t \sim \text{Logistic}(0, 1)$

Identification of b and p

- Probability of success **vs.** time to completion
  - Increase $p \Rightarrow$ patient to publish
  - Increase $b \Rightarrow$ faster screening
- Consider simulated outcomes at constant revision cost

	Baseline	Increase p	Increase b
Pr[Success]	47%	60%	53%
$E[T \mid \text{abandon}]$	9.1	11.9	7.2
$E[T \mid \text{publish}]$	11.5	11.7	11.4
p	0.59	0.69	0.59
b	0.17	0.17	0.29

Estimation Procedure

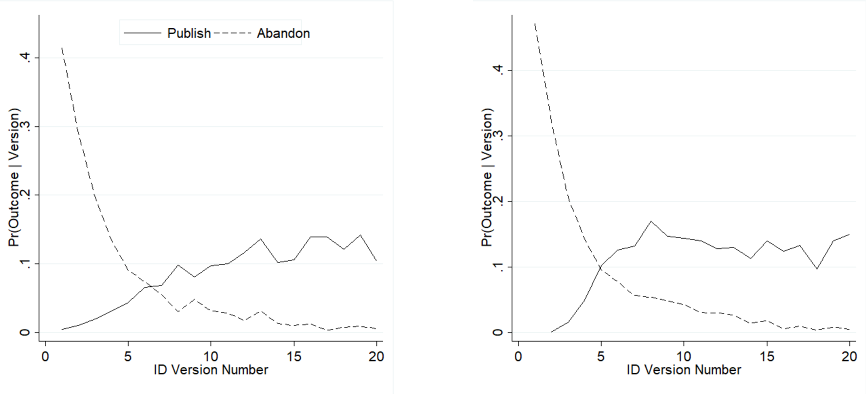
1. Iteratively search for values of $(b, p, F(t))$ that maximize the log-likelihood, where each iteration consists of two steps:
    1. Starting in period $\overline{T}$, recursively compute the sequence of cut-points $\{\bar{\varepsilon}_t^\sigma\}_{t=1}^{\overline{T}-1}$, along with the associated probabilities $G^\sigma(t)$ and continuation values $V(t, \sigma)$.
    2. Form $LL(b, p, F)$, retaining estimates of $F(t)$ if using nonstandards-track controls.
2. Bootstrap to obtain standard errors

Results

Baseline: Ex-Ante Identical Projects

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

Empirical and Simulated Hazard Rates



IETF Data **Simulated Data**

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: $F(t, \sigma_t = 1) = F(t, \sigma_t = 0) + \kappa$
4. Three-Step Estimation:
   - Using Assumption 2
5. Others: Include censored projects; RFC citations; $\overline{T} = 50$; only first project for each WG

▸ Table

Baseline Estimates: Summary

- Roughly 60% of IETF projects can generate consensus
- Robust to changes in assumptions and sampling
- Inform theory literature by estimating parameters driving Bayesian learning process in models of experimentation
- Around one-third of "publishable" ideas are published
  - Missing Knuth's? 

Team and Project Heterogeneity

Can add hetero in learning $b \equiv b(x)$, quality $p \equiv p(x)$, or both

Team and Project Heterogeneity

Can add hetero in learning $b \equiv b(x)$, quality $p \equiv p(x)$, or both

Specification	Communication	Experience		Commerciality	Team Size
(Explanatory variable x)	(Emails) (1)	(RFCs) (2)	(Projects) (3)	(Suit-to-Beard) (4)	(Authors) (5)
b (Category 1)	0.15	0.26	0.24	0.18	0.28
	[0.08, 0.21]	[0.20, 0.31]	[0.17, 0.30]	[0.10, 0.27]	[0.25, 0.32]
b (Category 2)	0.29	0.31	0.27	0.35	0.30
	[0.24, 0.34]	[0.28, 0.35]	[0.22, 0.32]	[0.33, 0.38]	[0.27, 0.34]
b (Category 3)	0.33	0.34	0.31	0.37	0.33
	[0.31, 0.37]	[0.32, 0.38]	[0.29, 0.35]	[0.35, 0.40]	[0.31, 0.37]
b (Category 4)	0.34	0.36	0.35	0.23	0.33
	[0.32, 0.39]	[0.34, 0.42]	[0.33, 0.39]	[0.13, 0.32]	[0.31, 0.38]
p	0.46	0.45	0.45	0.48	0.45
	[0.30, 0.59]	[0.29, 0.59]	[0.29, 0.59]	[0.35, 0.61]	[0.28, 0.59]
LL/Project	-2.343	-2.393	-2.392	-2.715	-2.400
Projects	14,444	13,922	13,922	10,710	13,922

Counterfactuals

Bottom Line

- Publication Prizes vs. Participation Subsidies
  - Subsidy ⇒ more "gross" innovation
  - Prize ⇒ higher *ex ante* project value
- Misaligned Priors
  - Do you want over or under-confident engineers?
  - Under-confidence is better b/c "dry wells" less costly

Conclusions

- Q: How do engineers learn from experimentation?
- **Descriptive analysis** of IETF protocol development
  - Fast failure, slow success & increasing payoffs $\Rightarrow$ two-stage model
- Parsimonious **structural model** of Bayesian learning
  - Finite horizon non-stationary DDC problem
  - Simple state-space $\Rightarrow$ fast & tractable
- Empirical Results
  - Good ideas ($p = 59\%$), slow learning ($b = 17\%$) & declining costs (**F**)
  - Model implies many "missing Knuth's"
  - Sanity check: WG projects learn twice as fast
  - Experience and communication $\Rightarrow$ faster learning
- Fascinating counterfactuals (read the paper!)

Thank you!

Please send comments or suggestions to

b.ganglmair@gmail.com
tsimcoe@bu.edu
or tarantino@uni-mannheim.de

Major Contributors

1992-1994		1992-2004	
1. Cisco	94	1. Cisco	1,787
2. Carnegie Mellon	51	2. Nortel	694
3. mtview.ca.us	48	3. Microsoft	581
4. IBM	44	4. Nokia	539
5. SNMP Research	38	5. Sun Microsystems	513
1995-1997			
1. Cisco	214	6. AT&T	513
2. IBM	140	7. IBM	490
3. Microsoft	140	8. Ericsson	398
4. Sun Microsystems	84	9. Lucent	343
5. USC (ISI)	79	10. Bell Labs	301
1998-2000			
1. Cisco	517	11. Alcatel	299
2. Nortel	321	12. Juniper Networks	260
3. AT&T	223	13. Intel	225
4. Microsoft	221	14. Columbia U.	220
5. Sun Microsystems	180	15. Siemens	200
2001-2004			
1. Cisco	962	16. Dynamicsoft	196
2. Nokia	404	17. USC (ISI)	195
3. Nortel	354	18. ACM	185
4. Ericsson	279	19. MIT	152
5. Sun Microsystems	234	20. NTT	149

Explanatory Variables

Explanatory variable *x* *For Team size: Counts in 4 categories*	N	Mean	SD	Min {1}	25th {2, 3}	50th {4, 5}	75th {6 – 72}	Max
Communication (Emails)					0.25	2.33	6.17	292
Full Sample	14,444	4.99	9.65	0				
WG Sample	3,168	5.20	8.08	0				
WG Sample (Exogenous)	3,086	4.66	3.35	0				
Experience (max RFCs)	13,922	3.82	8.13	0	0	0	4	68
Experience (max Projects)	13,922	14.80	26.45	0	1	5	17	254
Commerciality (Suit-to-Beard)	10,710	0.78	0.27	0.00	0.67	0.86	1.00	1.00
Team Size (Author Count)	13,922	2.25	1.85	5,935	5,750	1,708	529	

Back

Robustness Results

	Baseline	Area Specific	Calendar Time	Phase Specific (κ)	3-Step Estimator
Rate of Learning b	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 p	0.59	0.59	0.49	0.42	0.45
	[0.51,0.73]	(0.008)	(0.003)	(0.007)	[0.28,0.59]

Back