



# Modes of Participation, Identity Creation, and Group Maintenance in Different Technical Discourse Communities

Society for the Social Studies of Science  
Washington, DC (Crystal City)

stuart mawler  
2009/10/30



# What is this all about?

- ◆ Normatively identity-neutral interactions “hide” identity work
- ◆ Identity-oriented purposes are “hidden” in form and function of discourse events
- ◆ Community constructs combine
  - Discourse Communities
  - Epistemic Cultures
  - Communities of Practice
- ◆ Newsgroup analysis will illuminate these themes
  - Research still unfolding

# Newsgroup Backgrounder

- ◆ Newsgroups are subsections of the Usenet system
  - Created in 1979 at Duke University
- ◆ Users read and submit public messages (“posts”) to newsgroups
- ◆ Discussions are threaded
  - Largely driven by news reader client software
- ◆ Still available and see high-traffic use
  - Via portals like Google and Yahoo

# How was that again?

- ◆ Compare newsgroup discourse for C++ and Lisp
- ◆ Analyze all posts for 15 days: 7/1/2008-7/15/2008
  - comp.lang.c++ (1440 messages in 238 threads)
  - comp.lang.lisp (649 messages in 95 threads)
- ◆ Why these newsgroups?
  - Significantly different focus of each group
  - Both are high-traffic (in top ten within comp.lang.\* branch)
  - C++: Largest single subscriber base in comp.lang.\*

# What Do Newsgroups Do?

- ◆ “Official” or Normative purposes
  - Solve intractable technical questions
  - “Be there for the newbie”
- ◆ Identity-oriented purposes
  - Determine what counts as knowledge
  - Create and maintain group mythologies
  - Pass-on shared history
  - Validate and maintain practices
  - Establish individual credibility
  - Maintain and strengthen boundaries with other languages

# Technical Foundations... Mostly

- ◆ C++
  - High-level language, Object-oriented
  - General purpose
    - ◆ Commercial
  - Created by Bjarne Stroustrup at Bell Labs in 1979
  - ANSI/ISO standards maintained
    - ◆ Strongly “defended” or “protected”
- ◆ Lisp
  - Early high-level programming language
  - Central to classic AI research
    - ◆ Academic
  - Created by John McCarthy at MIT in 1958
  - Many sub-dialects, with two dominant
    - ◆ Common Lisp, created 1984
      - ANSI standard established 1994
    - ◆ Scheme, created 1970s
      - IEEE standard established 1990

# Newsgroup "Style"

- ◆ Both unmoderated – community and individual policing
- ◆ Both dominated by small group
  - C++ top 5 = 24%; Lisp top 5 = 26%
  - C++ dominated by one guy (James) at 10.83%
- ◆ `comp.lang.c++`
  - Standards for interaction published
  - Regular reference to "the standard"
  - Regular answer writers rarely start threads
- ◆ `comp.lang.lisp`
  - No official standard of interaction
  - Less policing of off-topic posts
  - References to standards more diffuse
  - More "professionals" initiate posts
- ◆ Differences are *qualitative*, rather than *quantitative*

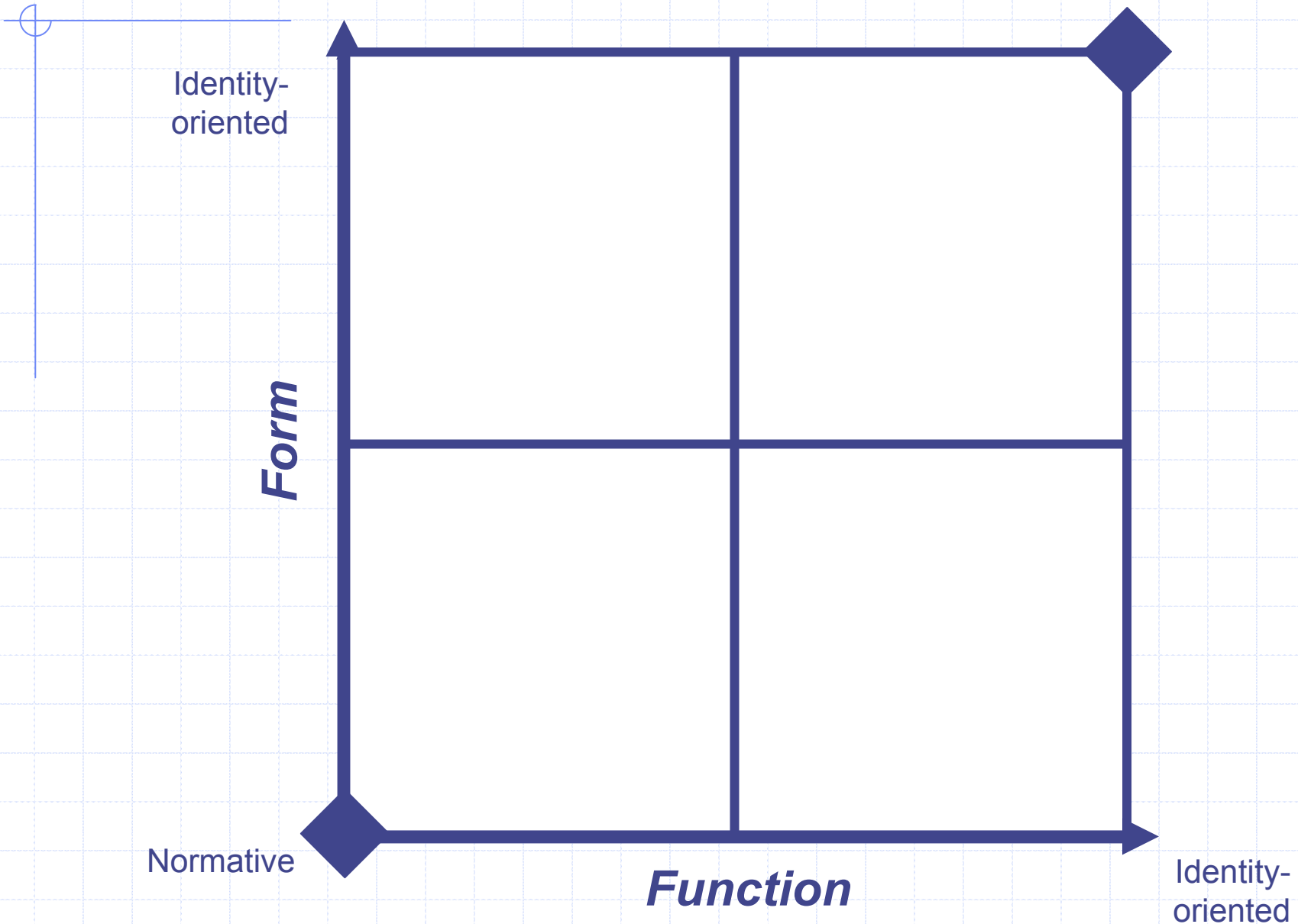
# Piles of Statistics

	<b>C++</b>		<b>Lisp</b>	
total posts	<b>1440</b>		<b>649</b>	
total threads	<b>238</b>		<b>95</b>	
avg posts / thread				
longest thread	46	3.19%	117	<b>18.03%</b>
top three threads	108	7.50%	225	<b>34.67%</b>
unique names	280	19.44%	134	20.65%
unique names which only responded	128	45.71%	92	<b>68.66%</b>
unique names which only originated	51	<b>18.21%</b>	11	8.21%
unique names which both orig & resp	98	<b>35.00%</b>	31	23.13%
top 5 unique name volume	347	24.10%	167	25.73%
top 5 unique name original posts	4	1.68%	12	<b>12.63%</b>
top 1 volume	156	10.83%	42	6.47%
top 1 original posts	1	0.42%	1	1.05%
top 10 participant volume	504	35.00%	261	40.22%
top 10 participant original posts	5	2.10%	15	<b>15.79%</b>
unique names with <=5 posts/name	225	80.36%	107	79.85%
total posts from unique names with <=5 posts	489	33.96%	201	30.97%
total single posts from unique names	91	6.32%	60	9.24%
total single posts from unique names as originators	48	<b>52.75%</b>	10	16.67%
total single posts from unique names as responders	43	47.25%	50	<b>83.33%</b>
total single posts from unique names to total threads	91	38.24%	60	<b>63.16%</b>

# Analyzing the texts

- ◆ From & Function on a spectrum
  - “Normative” to “Identity-oriented”
  - Differs from “discourse”, “style”, and “genre”
- ◆ Form
  - Modes of expression
    - ◆ Genre / generic structures (e.g., “article”)
    - ◆ Grammatical forms (e.g., “interrogative” or “third person”)
    - ◆ Literary forms (e.g., “irony” or “sarcasm”)
    - ◆ Cultural references (e.g., *Hitchhiker’s Guide to the Galaxy*)
- ◆ Function
  - Relevance to the community’s normative goals for the discourse event
- ◆ Re-emphasizes that differences are qualitative

# Visualizing Discourse Events



# Visualizing Discourse Events

```
Newsgroups: comp.lang.lisp
From: namekuseijin <namekuseijin@gmail.com>
Date: Fri, 1 Aug 2008 09:23:07 -0700 (PDT)
Local: Fri, Aug 1 2008 11:23 am
Subject: Re: The Fundamental Confusion of Xah
```

```
On 1 ago, 07:13, Jon Harrop <j...@ffconsultancy.com> wrote:
```

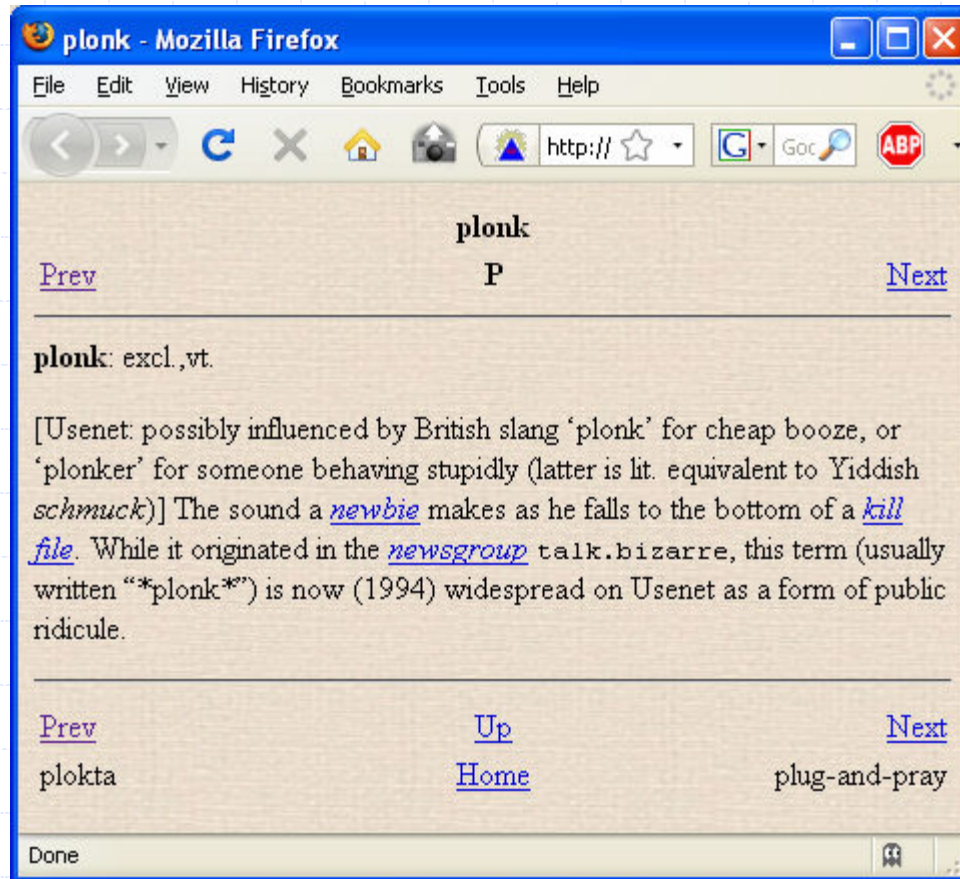
```
> What proportion of your code base is written in Java?
```

```
Hey, this is Java we're talking about! Even interfacing to it
probably demands a lot of glue code...
```

```
from: LSP0027.109
```



# Self-regulation by public ridicule



# A complete picture

- ◆ Understanding technical communities requires synthesis of three frameworks for community analysis
  - Programming newsgroups...
  - ...Are fundamentally about discourse (Discourse Community)
    - ◆ Traditionally, as a text-based community
    - ◆ Topically, since programming is a linguistic task
  - ...Create knowledge (Epistemic Culture)
    - ◆ Programming methods
    - ◆ History of programming
    - ◆ What counts as programming culture
  - ...Evaluate, establish, and maintain practices (Community of Practice)
    - ◆ Programming
    - ◆ Community interactions
      - For both newsgroups and physical-space