

Online Books: The Columbia Experience*

Paul Kantor, Tantalus Inc + Rutgers

Mary Summerfield, Columbia (*Consultant*)

Carol Mandel, Columbia (*New York University*)

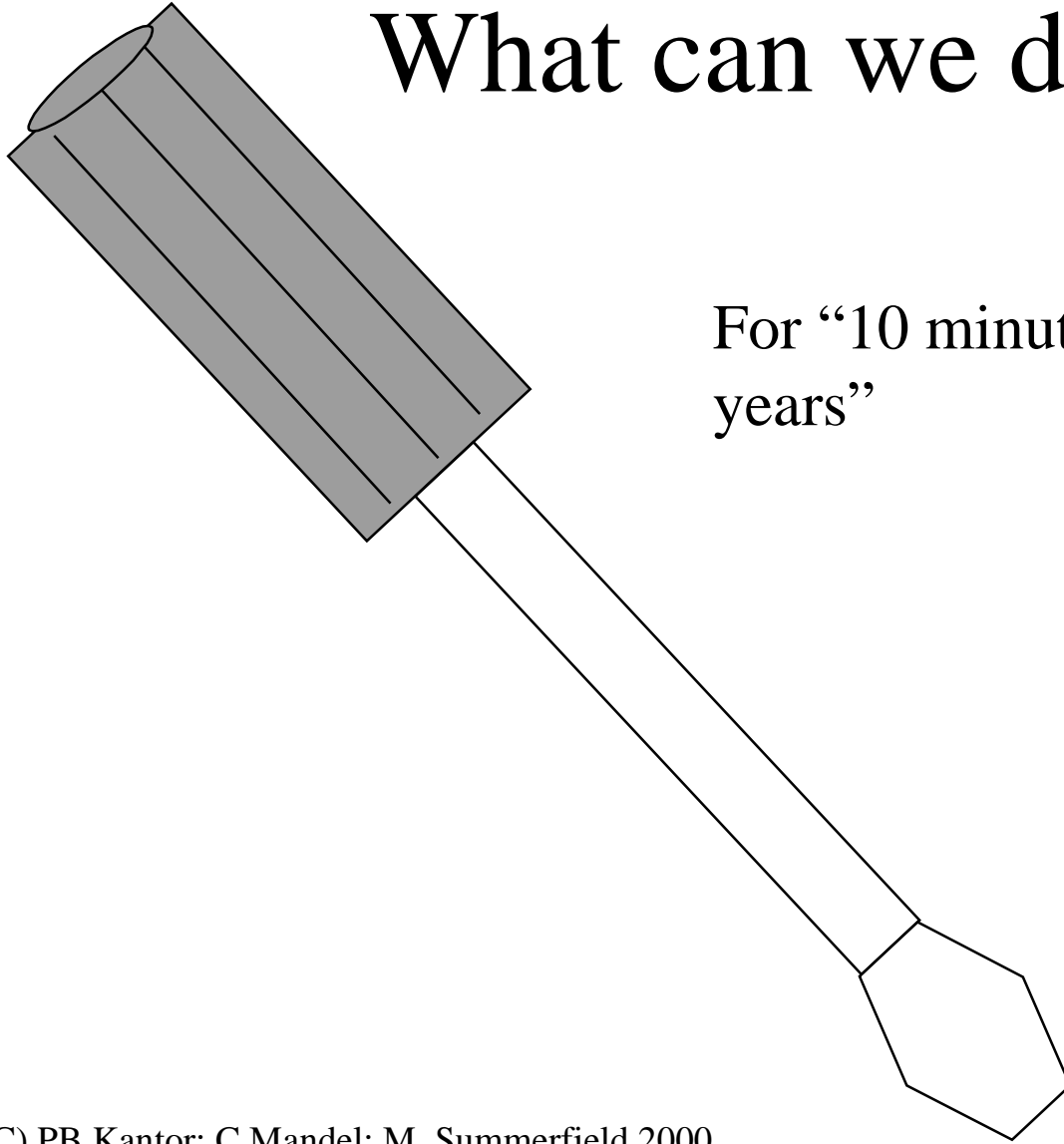
*Supported by the Andrew W. Mellon Foundation

Obligatory bad joke: A manager, engineer and computer scientist

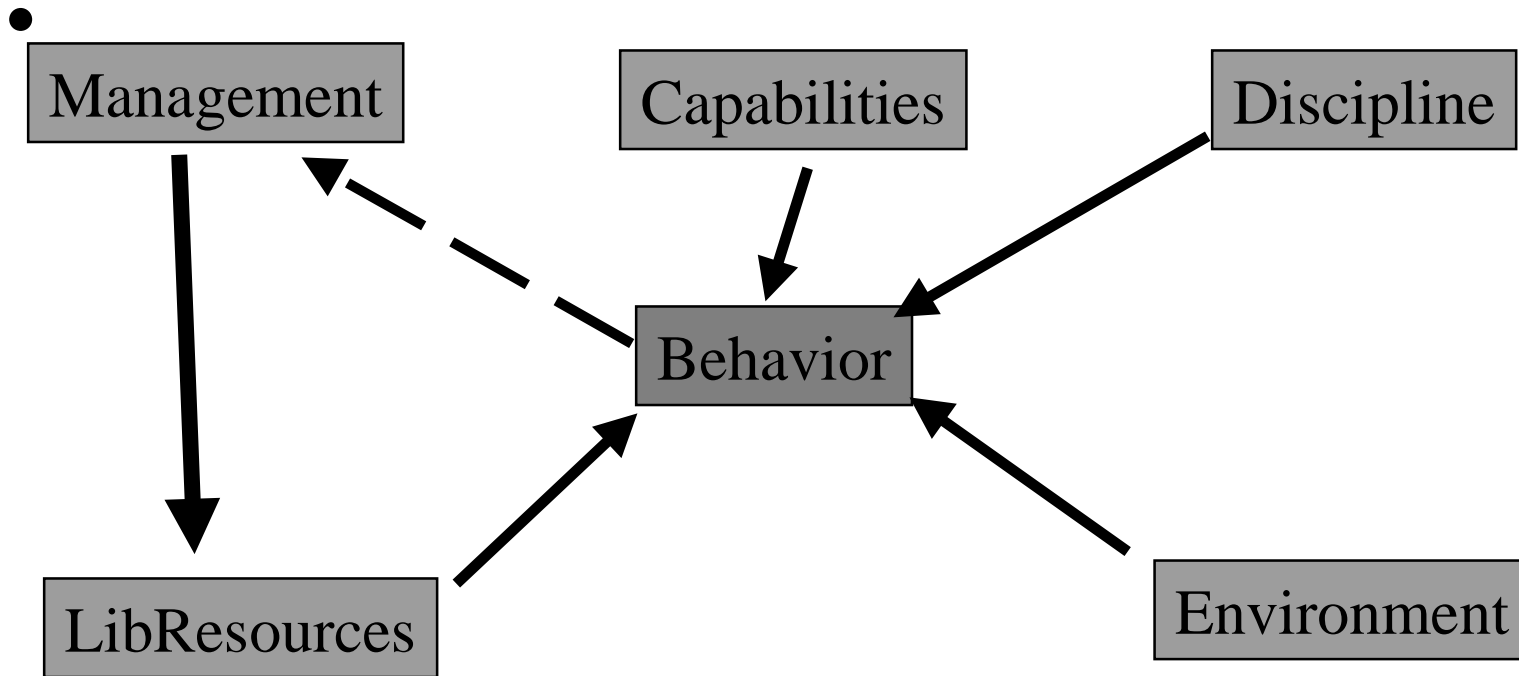
- Brakes fail - car stops just at edge of cliff
- Manager: *We'll form a focus team for a matrix review of vision and objectives*
- Engineer: *Let me have a screwdriver. I may be able to fix it in 10 minutes*
- Computer Scientist: *Let's push it back up to the top of the hill and see if they fail again.*

What can we do?

For “10 minutes” read “4
years”



The Variables



Overview

- 1995-1999
- Surveys and studies of users
- Preparation of books in HTML format
- Monitoring of National Environment
- Economic Perspective

Economic Perspective

- Actors weight the costs and benefits of alternatives A, A', A'', \dots
- Apply some personal “utility function” to those costs and benefits $u(c(A), b(A); me)$
- Chose the action with the largest personal utility.
- “Persons”: *students, faculty, staff, library, university, ...*

Forces Affecting Individuals

- Costs
 - Capital - Equipment needed
 - Capital - Skills needed
- Continuing
 - Costs of connection
 - Mental costs associated with use

Forces Affecting Individuals (2)

- Benefits
 - Ubiquity of access
 - Storage of important locations
 - Annotations
 - Search capability
 - Being up to date relative to peers
 - *plus Symbolic Utility (Nozick)*
- WOW!!

Forces Affecting Staff

- Costs
 - Learning curve
 - Continuous change
- Benefits
 - Better service to patrons
 - Ability to adapt materials
 - More portable skills (move to a .com)

Forces Affecting the Library

- Benefits
 - Competitiveness of the university
 - Professional goals of growth and service
- Costs
 - Equipment
 - Development of materials
 - Training

Columbia University

Online Books Evaluation Project

- Assumptions
- Environment
- Publishing costs
- Library costs
- Views of functions and design
- Use and user preferences
- Implications

Columbia University Online Books Evaluation Project

- Repackaged books for online delivery
- Studied their use
- Analyzed costs of development, delivery and use
- Related use and costs to context & potential

Publisher partners

- Columbia University Press
- Oxford University Press
- Garland Publishing
- Simon & Schuster Higher Education

Why put books online?

- Cheaper
 - to produce
 - to purchase
 - to acquire & maintain
- Increased functionality (e.g. searching, linking)
- Potential for enriched content
- Potential for expanded products (e.g., collections)
- 7 x 24 availability

Why not put books online?

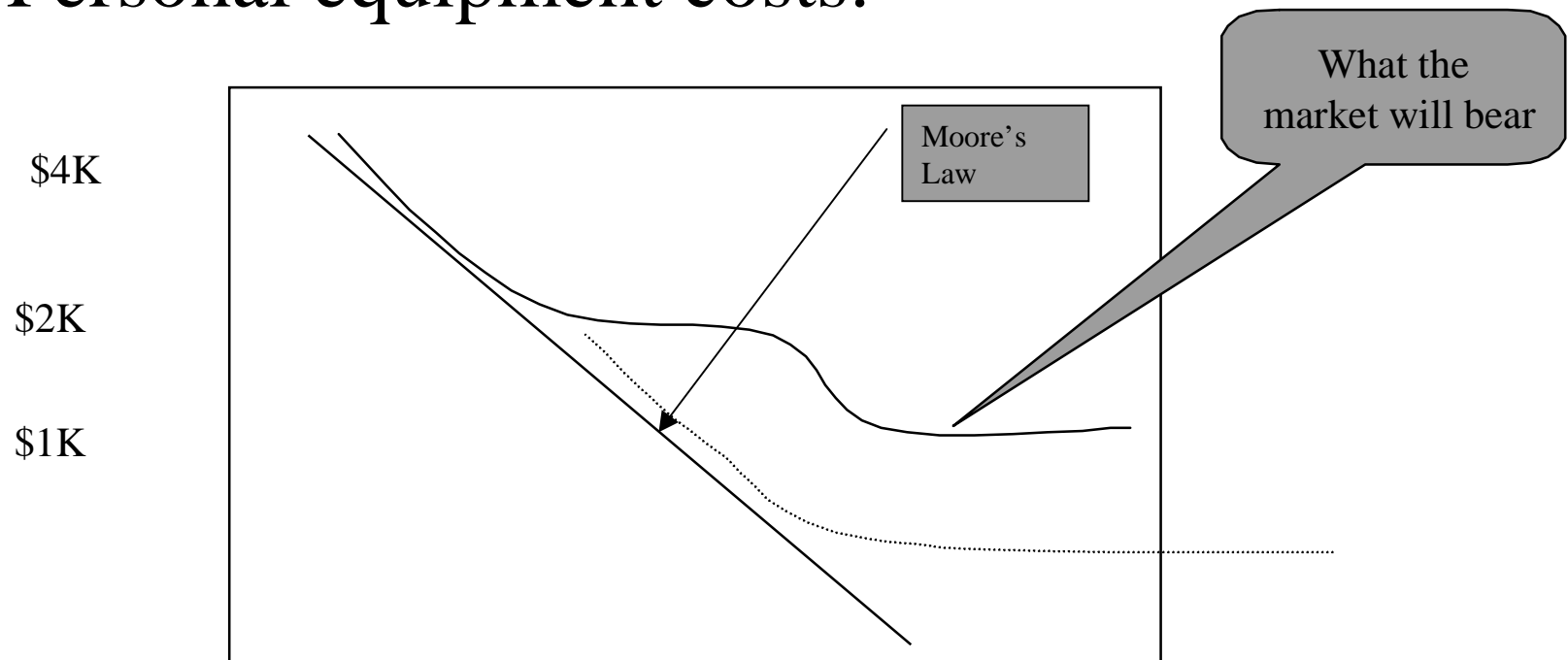
- No one wants to read online
- Difficult to use
- No feasible market model
- Access / connectivity inadequate
- Too costly to produce
- Authors will oppose

Environment for Online Books, 1995-99

- Improved price / power ratio for personal computers
- Penetration of Internet use to more than 50% of households by '99
- Half of all adults are Internet users (65 million in '98; 100 million in '99)
- No improvement in speed or ISP pricing since '97
- Hand-held bookreaders emerging in '98

Moore's Law, *Not!*

- Personal equipment costs:



Columbia Environment for Online Books, 1995-99

- Ethernet connectivity to all buildings and dorms
- By '97, 80% of students and faculty had adequate access to a networked computer
- By '97, most library users reported an average of 6 hours / week online

Columbia Environment for Online Books, 1995-99 (continued...)

- By spring '99, online full-text use is common, e.g., monthly JSTOR use = population
- Most online book use was from on-campus computers

OBEP Sources of Data

- Surveys
 - Online
 - Telephone
 - In-Class
- Focus Groups
 - Users
 - Potential Users
 - Librarians

OBEP Sources of Data (Focus here)

- Cost analyses
- Use data (Web, circ., etc.)
- Other studies, reports

Production Costs: In a print production environment, online is an additional cost

- Range *will be found in the full paper.*
- Low end cost requires a very standard electronic file

Sample e-book production costs

- Conversion: OCR, SGML
 - Conversion: ASCII to HTML
 - Conversion: PS to PDF
 - Conversion management
 - Books on server
- *will be found in the full paper.*

Potential savings for online only

Plant (typesetting) 10%

PPB (paper / print / bind) 15%

Plus, warehouse, shipping

Offsets to savings for online

- Mark up
- Marketing
- Customer service
- Continuing file maintenance

Lifecycle (30 yr.) costs are lower for
online books

Print

Online

Acq/Proc.

Storage/Maint.

Circulation

(incl. above)

Total (30 yrs.) *will be found
in the full*

paper.

Librarians' desired design features

- Searching across selected groups
- Stable, granular URL
- Bibliographic records
- Standard usage data
- Reliable migration to new platforms




Scholars' desired design features

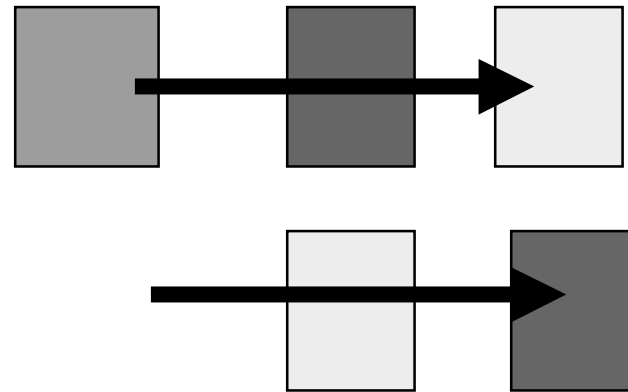
- Links from online catalog
- Flexible groupings for searching
- Comprehensive, linked TOC
- Browsable, linked thumbnail images
- Show two pages
- Show footnotes and text in parallel
- Pagination matching print

Scholars' desired design features, (continued...)

- Hyperlinked references
- Links to a dictionary
- Adjustable fonts, formats, etc.
- Annotation and highlighting capability
- Shared annotations

Users: Very Simplified View

- Technology 
- Behavior 
- Attitudes 



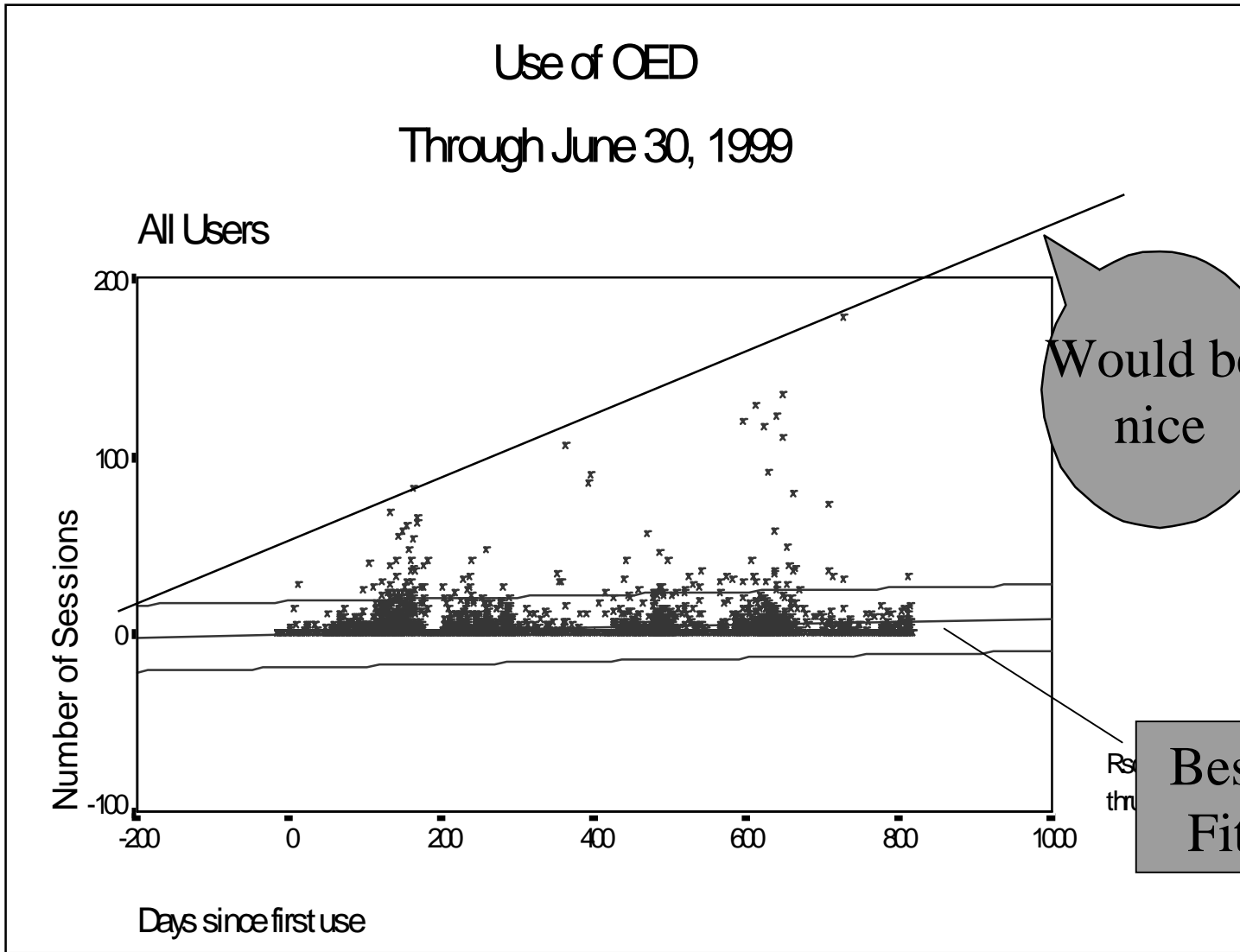
- *We expected picture (2): changes in attitudes would precede changes in behavior.*

Amount of Use

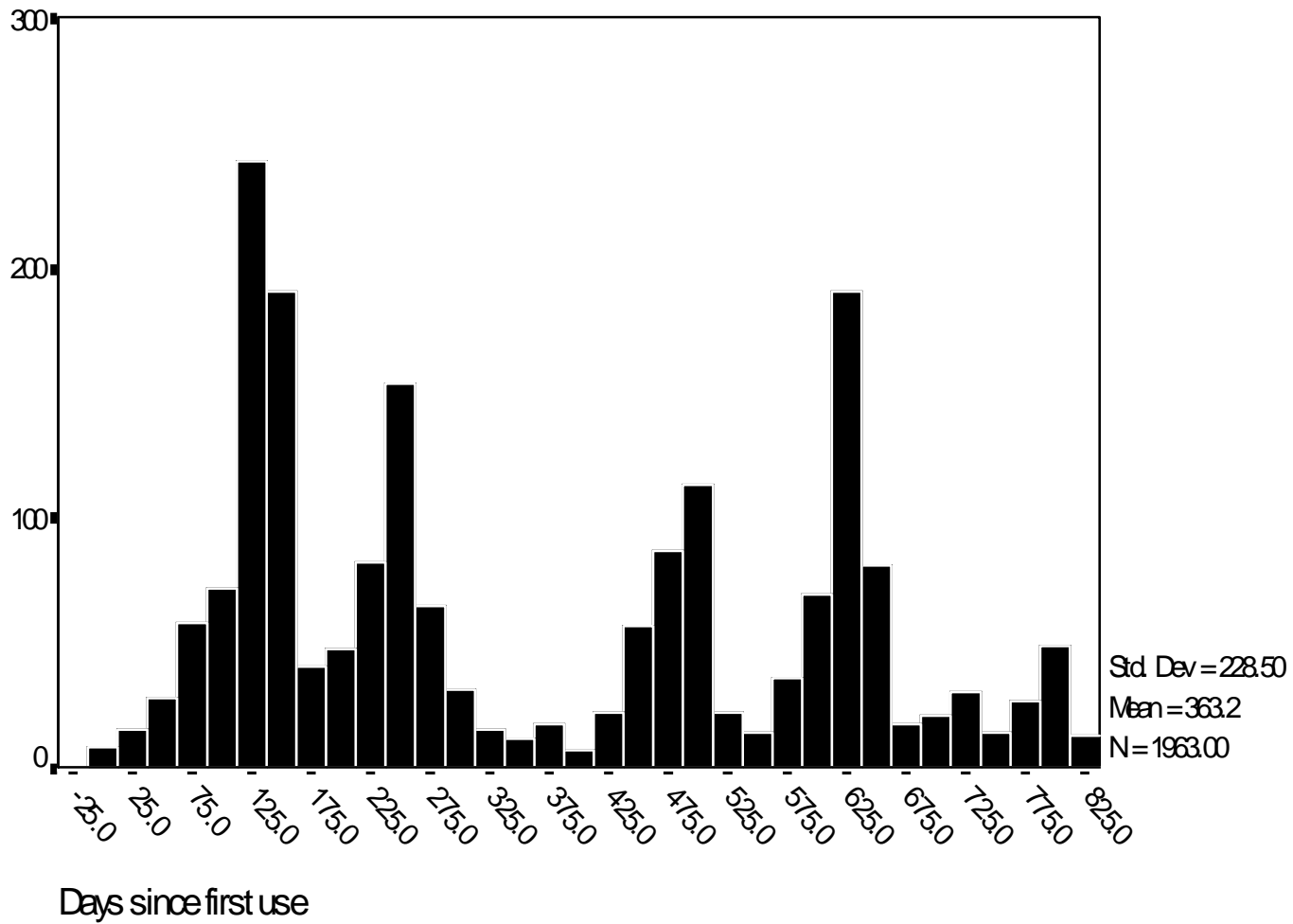
- 2 or 3 is the modal number of clicks, and above that the number of clicks a person made on the OED drops off exponentially.- chance to go on to two more clicks about 2/3rds at any given time.
- Similarly, time spend using OED online follows an exponential distribution.
- This is *not* a strong adoption pattern

Do users stay with the Online Version

- We can anonymously track individual users
- Plot how much they use the resource against how long it has been since they first used it
- With 100% adoption, this will be roughly linear.



Time since First Use: Users > 1 Session



Online books were used more than their
print counterparts

3 X as many readers per book online as
for paper version

(Note: Counted circulation but didn't count shelf
browsing)

Uses of online books (Qualitative)

- Browsing
- Grazing
- Citation checking
- Fact / quotation finding
- Reserve reading
- Determining need for paper copy
- Printing
- Reading

Patterns in online use (Quantitative)

- Linear: A,B,C,D,
- QuasiLinear: D,C,E,G (each only once)
- Uses of index:
 - I, D,C,E,G
 - I,A,I,C,I,G ... “Indo-Linear”?
- Hyperlinear: A,G,S,A,T,G,...
- Most use still Line, QL or IL

Paper Book in library not same as available

- In circulation
- N.O.S.
- Library not open
- OPAC \neq browsing

Scholars' Preferences for Book Access

	<i>Read Much</i>	<i>Read Little</i>
<i>Low cost book</i>	Buy	Online
<i>High cost book</i>	Borrow	Online

Users want online books for

- Convenient access
- Assured 7 x 24 availability
- Most book use needs
- Added functionality - annotate; hyperlink
example is NEC CiteSeer model

Users want paper books for reading

Libraries want online books for

- High demand (2nd + copy)
- Transient need (“rent-a-book”)
- Cost savings

Librarians concerned about

- Paying twice
- Uncertain preservation/migration
- Unwanted materials in bundled packages
(bundling general can increase both
consumer and producer benefit)

Options for library-oriented models, e.g.

- Online version for little or no additional cost
- Online collections priced attractively
- On-demand licensing
- On-demand print ordering

Characteristics of consumer-oriented models

- One-at-a-time use
- Proprietary readers
- Hand held devices
- Pricing unrelated to print purchases

Will models differ for:

- Textbooks
- Tradebooks
- Scholarly books
- Narrow interest (endangered?) scholarly books

Transition “compromises” include

- Print and online publication
- Backlist and frontlist
- Lead time for frontlist
- Limits on functionality for new titles

Knowledge Generation

- Use of online books can be tracked at a micro level, providing valuable information to authors and publishers.
- Scholarly authors are concerned about this data (the Dean needs to know)
- Adds one more factor in favor of online modality.

Some predictions

- Complex functionality will be reserved for huge sellers or subsidized projects
- “Endangered” monographs will be available from academic or society servers, LAPL, or authors. (i.e., not “published”)
- Many books will be electronic and print
- Netlibrary, or someone like it, (*not library experiments*) will define the product

Some predictions (continued...)

- No one will save money
 - Someone will make money (*lots of it??*)
 - Civilization as we know it will be
 - transformed beyond recognition
 - essentially unchanged
 - lost as media obsolesce
- } *Choose one*