# Heuristic Evaluation (Pinelle)

- Heuristic evaluation is a method of qualitative evaluation of software.

449: A Design Space for Evaluation
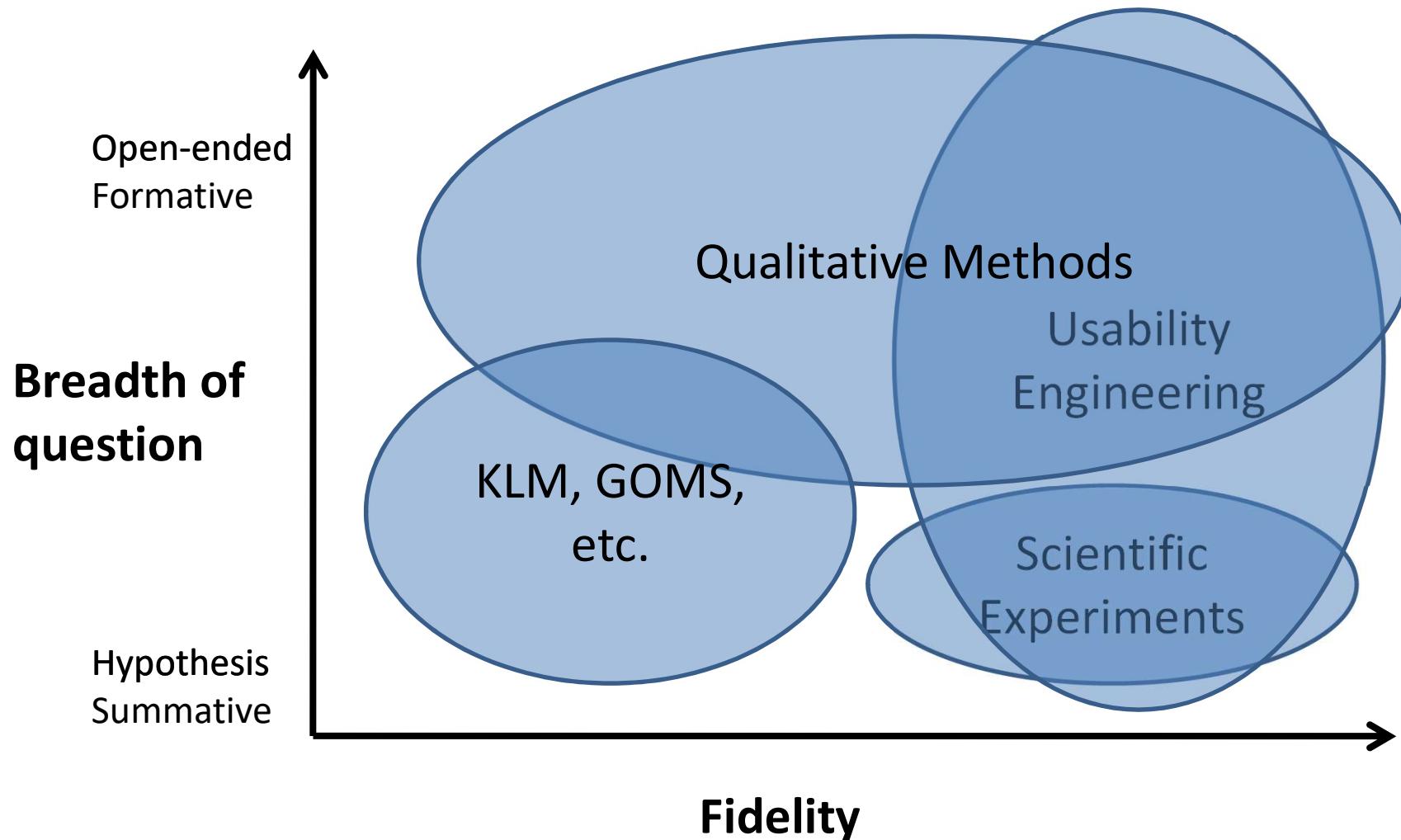
Qualitative Evaluation

- Constructivist claims
- Very common in design
  - Can be used either during design or after design complete
  - Can also be used before design to understand world
- Broad categories
  - Walkthroughs/thinkalouds
  - Interpretive
  - Predictive

Interpretive Evaluation

- Need real-world data of application use
- Need knowledge of users in evaluation
- Techniques (will revisit after talking about data collection)
  - Contextual Inquiry
    - Similar to for user understanding, but applied to final product
  - Cooperative and Participative evaluation
    - Cooperative evaluation allows users to walkthrough selected tasks, verbalize problems
    - Participative evaluation also encourages users to select tasks
  - Ethnographic methods
    - Intensive observation, in-depth interviews, participation in activities, etc. to evaluate
    - Master-apprentice is one restricted example of evaluation that can yield ethnographic data

Predictive Evaluation

- Avoid extensive user testing by predicting usability
- Includes
  - Person down the hall testing
  - Usage modeling
  - **Inspection methods**

Inspection methods

- Inspect aspects of technology
- Specialists who know both technology and user are used
- Emphasis on dialog between user and system
- Include usage simulations, heuristic evaluation, walkthroughs, and other forms of discount evaluation
  - Also includes standards inspection
    - Test compliance with standards
  - Consistency inspection
    - Test a suite for similarity

Aside: Discount Evaluation
(UW Research)

- Adam Fourney and Mike Terry
  - Mine Google suggest
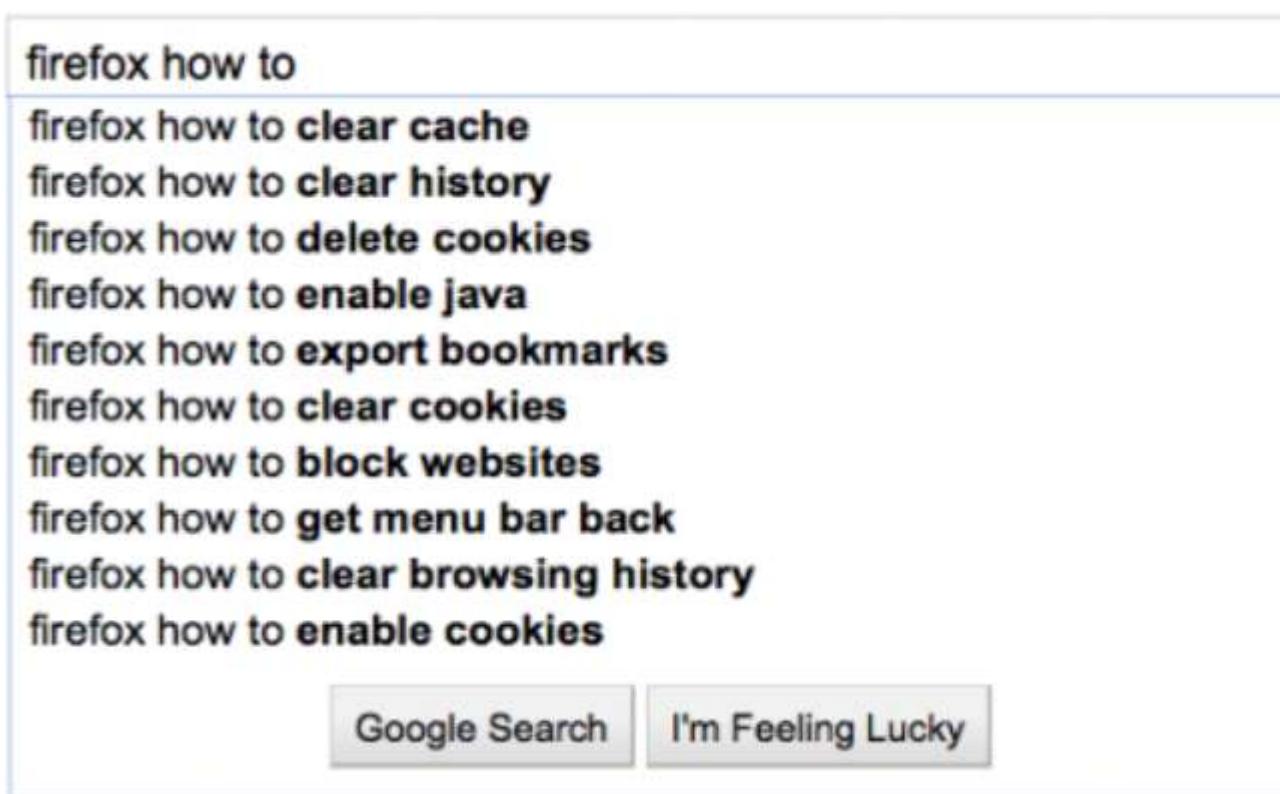
firefox how to

firefox how to **clear cache**
firefox how to **clear history**
firefox how to **delete cookies**
firefox how to **enable java**
firefox how to **export bookmarks**
firefox how to **clear cookies**
firefox how to **block websites**
firefox how to **get menu bar back**
firefox how to **clear browsing history**
firefox how to **enable cookies**

Google Search   I'm Feeling Lucky

Inspection Methods: Heuristic evaluation

- Set of high level heuristics guide expert evaluation
  - High-level heuristics are a set of key usability issues of concern
- Guidelines are often quite generic
  - Simple natural dialog
  - Speaks users' language
  - Minimizes memory load
  - Consistent
  - Gives feedback
  - Has clearly marked exits
  - Has shortcuts
  - Provides good error messages
  - Prevents errors

Process

- Each review does two passes
  - Inspects flow from screen to screen
  - Inspects each screen against heuristics
- Sessions typically one to two hours
- Evaluators aggregate and list problems

Heuristic Evaluation of Games

- Goal is to come up with heuristics so designers, companies, etc. can do a form of *predictive evaluation,* ***heuristic evaluation.***
  - Goal of paper is to create heuristics.
- To do this, a 3 stage process
  - Researchers individually identify problems based on 108 reviews, resulting in 50 problem categories (by summing problems from each researcher)
  - Researchers collaborate to eliminate 8 problem categories as not salient, then categorize the remaining 42, yielding 12 usability problems common in contemporary games.
  - Researcher invert the categories to create heuristics.

Inspection Methods: Heuristic evaluation

- Set of high level heuristics guide expert evaluation
  - High-level heuristics are a set of key usability issues of concern
- Guidelines are often quite generic
  - Simple natural dialog
  - Speaks users' language
  - Minimizes memory load
  - Consistent
  - Gives feedback
  - Has clearly marked exits
  - Has shortcuts
  - Provides good error messages
  - Prevents errors

Pinelle et al. Game Heuristics

1. Consistent response to actions
2. Customize video, audio, difficulty, speed
3. Predictable or reasonable NPCs
4. Clear, unobstructed views
5. Skip non-playable or repeated content
6. Intuitive and customizable input mappings
7. Controls with appropriate sensitivity and responsiveness
8. Game status information
9. Provide instructions/training and help
10. Easy to interpret representations that minimize micromanagement.

Two Considerations

- Methodology
  - Was the method well-explained, reasonable
  - Could you replicate what they did?
- Utility
  - Are these useful

Problem category	Key issues	Example	Total
1. Unpredictable / inconsistent response to user's actions	poor hit detection, poor in-game physics, inconsistent response to input	"You'll often find yourself swinging away and watching your blade go right through your foes with no effect whatsoever." [21]	42
2. Does not allow enough customization	does not allow user to change video and audio settings, difficulty, or game speed	"Worst of all, there is no way to adjust the game's speed. You're stuck constantly veering between way-too-fast combat and way-too-slow travel to engage enemies and explore maps." [34]	11
3. Artificial intelligence problems	problems with pathfinding, problems with computer controlled teammates	"Your own teammates are a mixed bag in terms of intelligence, as they're sometimes good about getting behind cover and laying down covering fire, but they also do crazy things, like rushing forward into the open or constantly getting in your line of fire." [27]	22
4. Mismatch between camera/view and action	bad camera angle, view is obstructed, view does not adjust to user's action quickly enough	"If you leave the camera alone you'll either be running around blind or the camera will get stuck on something and just twitch randomly for a while." [22]	23
5. Does not let user skip non-playable content	cannot skip video and audio clips, frequently repeated sequences	"Making matters even worse, you can't skip over any of this blah-blah-blah wordiness and often have to exhaust all of the dialogue options to open up new discussion topics or the ability to solve a problem." [35]	6
6. Clumsy input scheme	bad input mappings, limited device support, limited control customization	"You can't type your name in—you must instead select letters with your mouse. And you can't use the keyboard to navigate." [14]	21
7. Difficult to control actions in the game	oversensitive controls, unnatural controls, unresponsive controls	"Chopper controls run the gamut from very touchy to absurdly touchy. For instance, it's almost impossible to keep an Osprey's rotors pointed toward the sky." [36]	31
8. Does not provide enough information on game status	does not provide adequate information on character, game world, or enemies. visual indicators, icons, and maps are inadequate.	"The map function reveals everything about a particular location except where you are, making it useless until you do some heavy exploring to get your bearings." [31]	28
9. Does not provide adequate training and help	does not provide default and recommended choices; does not provide suggestions and help; does not provide adequate documentation, instructions, tutorials, and training missions.	"A tutorial and a better online manual are sorely needed. There is a really nifty feature where you can ask specific questions of your wrestlers and booking team, but it's just not enough." [32]	36
10. Command sequences are too complex	learning curve is too steep; requires too much micromanagement; command sequences are complex, lengthy, and awkward, making the game difficult to play	"Not only does the main map have four modes and many subscreens, but nearly all the political activities take place on separate screens designed more to resemble scenes in Roman life than to maintain any continuity to the interface." [28]	20
11. Visual representations are difficult to interpret	bad visualization of information, too much screen clutter, too many characters or game elements on the screen at the same time, difficult to visually distinguish interactive content from non-interactive content	"Buildings are completely barren. And the satellite-styled overhead map is tough to read, because it consists of mashes of green terrain and colored blobs representing enemy soldiers and objectives." [33]	16
12. Response to user's action not timely enough	slow response time interferes with user's ability to interact with the game successfully	"Often, though, you'll need to hit the key before the indicator appears, as the game takes a moment to register your keystroke." [8]	29
		Combined Total	285

iTunes Paper (Voida)
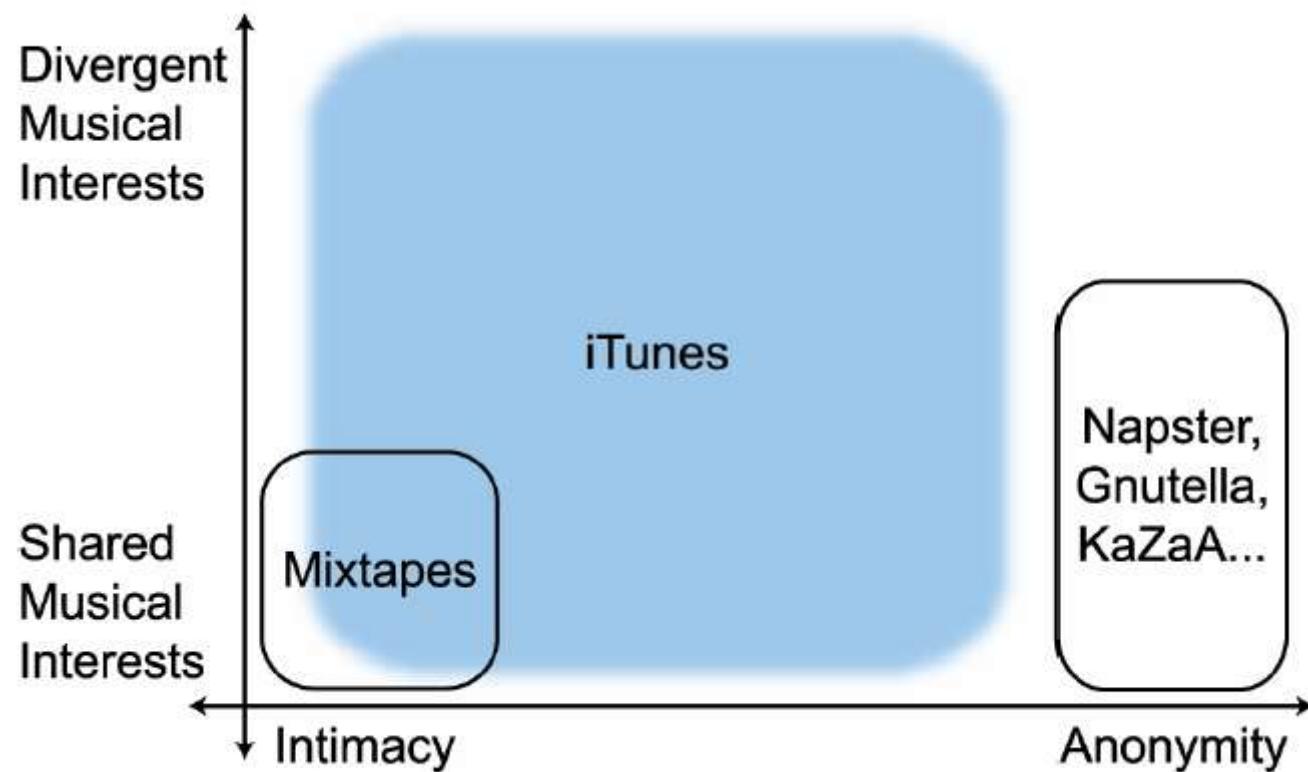
Method: 2 Paragraphs

We conducted 13 semi-structured interviews of iTunes users. The interviews lasted approximately 45 minutes each and were held in the participants' offices. To the extent possible, the interviews focused on specific examples of social aspects of iTunes use. For example, we asked participants to tell us about the last time they discovered a new music library in iTunes. The 13 participants were all employees of a mid-sized (~175 employees) corporation. Ten of the participants were researchers in various technical disciplines; three of the participants were administrative support staff.

The network topology of this company consisted of four wired subnets. Three of the subnets were defined by the physical layout of the building – floor 1, floor 2, and floor 3. The fourth subnet was used by the members of a department within that corporation. Theoretically, then, our participants belonged to four different groups of iTunes users; participants were able to view and share the music only of those members of their subnet group. In reality, we interviewed between two and eight members of each of three subnet groups, ranging in size from 3 to 12 known members. One last participant did not share his music library; if he had tried, he would have belonged to the third floor subnet group which had no other members [Table 1].

Analytical Approach

Analytical Approach

- Privacy Personas: Clustering Users via Attitudes and Behaviors toward Security Practices
  - https://dl.acm.org/citation.cfm?id=2858214
- Thoughts?

Contributions

- Results:
  - Adoption/Critical mass – ethos of sharing
  - Impression management
    - Concern about what your music says about you
    - Judgments about what others' music says about them
  - Dynamics of system
    - At work versus not, people leaving company
- Design space issues:
  - Gray area between intimacy and anonymity
  - Additional motivation to create sharing

Meta-Level Comments: Qualitative CHI Paper

- Common to see themes (3 or 4)
  - Get to this by iterating on data
    - Open coding
    - Axial coding to aggregate themes
- Common to see "Implications for Design"
  - Here inserted into themes
    - Sort of a "why should we care" section

Contrasting Papers

- Quantitative
  - 5 different mode switching techniques
- Qualitative
  - How people think about and perform sharing in work environments

Appendix – An Interview Question snapshot used by the authors

- What convinced you to initiate iTunes sharing on your subnet?
- Did you have any privacy concerns in deciding to share your music?
- How do you feel about the arrival of new collections on the network?
- How do you feel when a music library has disappeared from the network?
- How do you feel when you close your iTunes connection?
- What kind of identity do you portray though your music library?
- Have you tried to portray an identity through your own music library?
- Does your music library project an image of you to others sharing your music?
- Do you have any musical expertise that you would share through your library?
- Have you noticed other people changing the names of their libraries?
- How is your music library representative of yourself?
- How does others' music libraries affect your impression of them, if at all?
- How do you feel about users obscuring their own names?
- Would you like to be able to access libraries outside of your subnet?
- Has iTunes music sharing allowed your community to become more intimate?
- How do you feel when you have to cut someone off from your music without the ability to warn them?
- What kind of improvements can you imagine for the iTunes music-sharing feature?

Taken from http://ccrma.stanford.edu/~sonian/220D/

Impression Management and Access Control

- – *I just went through it and said, "Eh, I wonder what kind of image this is, you know, giving me," right? I just went through it to see if there was not like stuff that would be like, I don't know, annoying; that I would not like people to know that I had (P11).*
- – *When the sharing happened...I had not ripped everything from my CD collection....It was fairly heavily skewed toward the classical and soundtrack part of my collection...the order in which I'd popped the CDs in. And I remember thinking about this and was like, "Gee, that's not very cool...." So when we started sharing, I started reripping things, adding stuff to my collection....I added more to kind of rebalance it and cover a wider breadth of genres that I had in my collection (P11).*

- Another participant had not given the contents of his music library the same degree of scrutiny:
  - – *I mean if people are looking at my playlist to get a picture of the kind of music I like and don't like, you know. Or to get a little insight into what I'm about, it'd be kind of inaccurate 'cuz there's, you know, there's Justin Timberlake and there's another couple of artists on here that...Michael McDonald, you know. Some of this stuff I would not, you know, want to be like kind of associated with it....I guess part of it is it wouldn't be bad if, you know, people thought I was kind of hip and current with my music instead of like an old fuddy duddy with music.*

Impression Management and Access Control

- Another participant used his own national identity to give his library… *…a particular focus on all of the German bands actually that I have, because…if I have something to offer on the network, I'd like to be able to give, you know, albums and artists that other people don't have (P11).*

- These participants described their expertise as being in an area they felt that, at best, others would not "relate to" and, at worst, would be a "horrible experience":

  - *I have a lot of Hindi music that is stuff that I listen and I don't expect other people to relate to. So that is not there (P4).*

  - *I don't want to bother sharing all of my stupid band clips 'cuz that would probably be a pretty horrible experience (P12).*

Impressions of Others

- For the potential listening audience, these carefully crafted views into others' music libraries constituted "little windows into what they are about" (P1). In some cases, participants would browse through the list of genres represented in others' libraries to come to the conclusion that someone is "eclectic" or "easy because he has only one genre" (P11). One participant (P1) drew his impressions not so much from the musical content of others' libraries as from characteristics of the custom playlists that some users generated from their content.
    - People can give names to their collections that are not necessarily obvious. So the first few times that SmallieBiggs here appeared on my list, I was really curious who the heck is SmallieBiggs?... So that was, you know, enjoyable detective work (P11).
    - I wish I could find out who these people are. That's one thing that would be cool. I mean its kind of a small group. (P10)

Impression of Others (Conclusion)

- Despite the close examination of others' libraries, participants seldom felt that these musical impressions significantly changed their view of a coworker. Rather, they felt it mostly "serves to reinforce impressions I've already got" (P12). Occasionally, however, a participant admitted that knowledge of others' musical tastes impacted his opinion of them:
  - *"[P6] I have learned is a big fan of whatever current pop is which I suppose to some degree lowers my estimation of him but not by too much" (P12).*
- The more significant and longer-lasting impact of these musical impressions seems to be the binary judgment that frequently gets made:
  - *So when there is someone new, I spend a fair amount of time listening to what they have and then...binary process, either I just decide well there is nothing in there for me or I really like it and will come back to it. (P11).*
- In other words, the first examination of another person's library seems to have a strong influence on whether the visitor will ever return to that library.