

Bridging Computer Science and Behavioral Science: Research Examples

Joseph A. Konstan
 Department of Computer Science and Engineering
 University of Minnesota
konstan@cs.umn.edu
<http://www.umn.edu/~konstan>

Two Research Projects

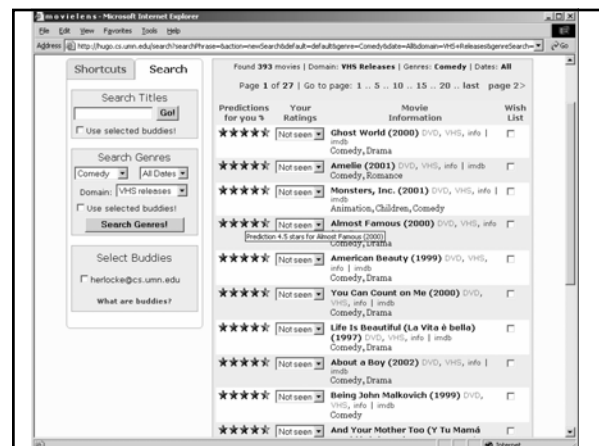
- Collaborative Research on Contribution to On-Line Communities
- MINTS – HIV-prevention study of Internet Latino Men

On-Line Communities Research

- Caveats
 - I'm describing it as a computer scientist
 - The reason for the collaboration is to have people who know more about social science than I do
 - Due to the size of this project, I'm not fully versed on the details of each component

The Collaboration

- University of Minnesota
 - Joseph Konstan, John Riedl, Loren Terveen
- Carnegie Mellon University
 - Bob Kraut, Sara Kiesler
- University of Michigan
 - Paul Resnick, Yan Chen
- Existing www.movielen.org website





Research Goals

- Build an understanding of how to elicit “optimal” participation in on-line communities
 - Test off-line theories in on-line context
 - Resolve cases where theories make conflicting predictions
- Deliver mid-level theory in the form of a “designer’s guide” for on-line communities

Types of Design Factors

- Visibility of Identity
- Status
 - Private or public
- Structure of Community
 - Number of members
 - Diversity
 - Self-Awareness
- Economic Structures

Types of Theory

- Social Psychology Theories
 - Identity leads to accountability
 - Affinity (liking) leads to participation
 - Uniqueness (need) leads to participation
- Economic Theories
 - Simple utility maximization
 - Inequality aversion

Usual Research Model

- Off-line analysis and experiments
 - Predict as well as possible based on historical data
- On-line user studies
 - Short-term
 - Long-term

An Initial Study

- A study of participation in discussions with two factors controlled
 - Similarity of tastes
 - Awareness of own uniqueness
- Results
 - Dissimilarity increased contribution
 - Awareness of own uniqueness increased contribution
 - Active discussants were not highly-active raters
 - Participants rated more than a control group

Motivational Follow-Up

- Class projects at CMU used an e-mail campaign to elicit ratings to discover:
 - Making users aware of their uniqueness increased rating
 - Giving users specific, achievable goals increased rating
 - But...
 - Reminding users of their self-benefit or benefit to others actually decreased the number of ratings!

Other Projects

- Member-Maintained Communities
 - What happens when you let the masses maintain the database?
- Social Preference
 - Using economic models to study user behavior
- Value of Information Analysis
 - Individual influence
 - Balancing community and individual value
- MovieLens MovieNight

Member-maintained community

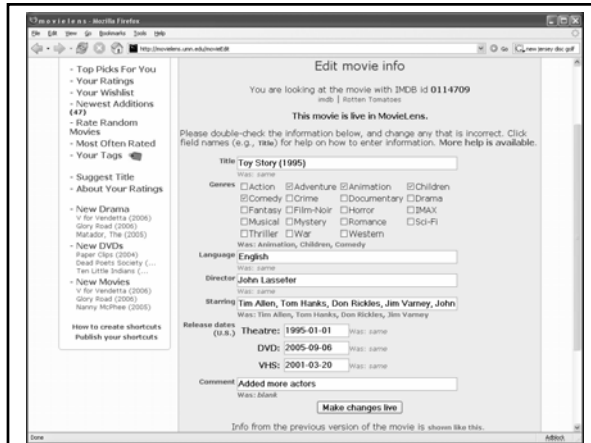
- Broad notion of online community
 - People interacting online
 - Directly or indirectly
- Member-maintained
 - Normally, few perform upkeep (Butler et al. 2005)
 - Why not many?

The virtues of the many

- Scale (Slashdot 2003)
- Speed (Viégas et al. 2004)
- Robustness against change
- Direction-setting (but, see Dibbel 1998 on LambdaMOO)

Two fundamental challenges

- Quality of contributions
 - Designing review mechanisms
 - Who can review, when?
- Quantity of contributions
 - Helping people find tasks
 - Intelligent task routing



Designing contribution review

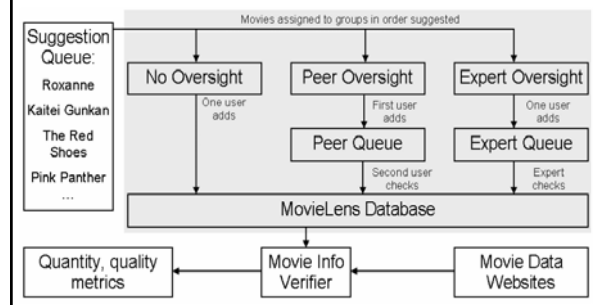
- Editorial oversight improves quality
 - Journalism
 - Peer review
- Wikipedia foregrounds questions
 - Who can review?
 - When to review?

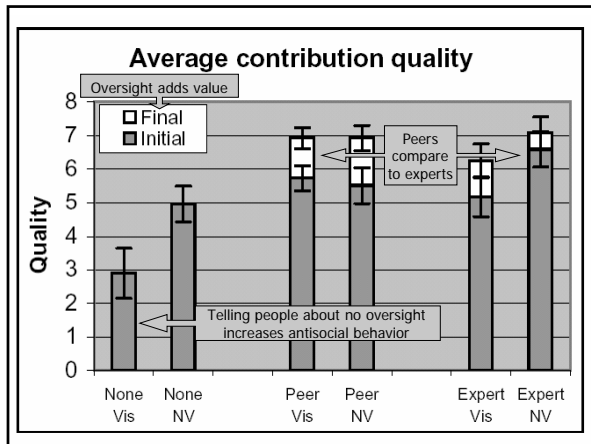
Who can review?

(Cosley et al., CHI05)

- Can peers do as well as experts?
- Task: add movies to MovieLens
- Hypothesis: Review increases motivation
- Design: 3 x 2
 - Level of review
 - Visibility of mechanism

Who can review overview

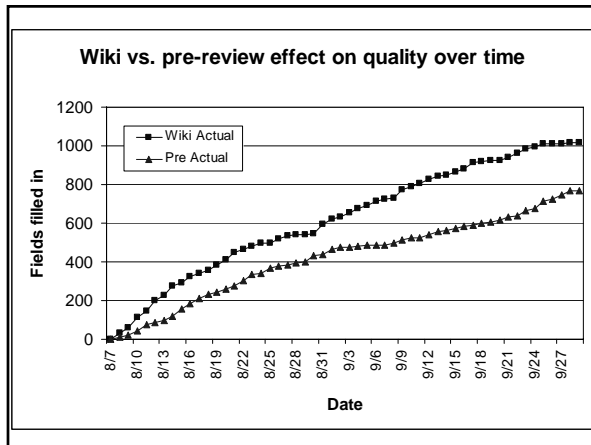




When to review

(Cosley et al., CHI 2006)

- Should contributions be visible before they are reviewed?
- Task: editing movies
 - Pre-review
 - Post-review ("Wiki-like")



Modeling review timing

- Multiple ways of knowing
- Increase generality
- Value simplicity (Axelrod 1985)
 - Many assumptions, clearly stated
 - Assumptions = design opportunities

Value changes over time

- DB value = sum of item values
- Value changes over time
- Some work helps, some hurts

$$V_{t+1} = V_t + G_t - B_t$$

Modeling Wiki-like's behavior

- DB value = sum of item values
- Value changes over time
- Some work helps, some hurts
- Tasks become harder to find as the database reaches its maximum value

$$V_{t+1} = V_t + \underbrace{(1 - P_t)}_{\text{Proportion of maximum DB value}} \gamma - P_t \beta$$

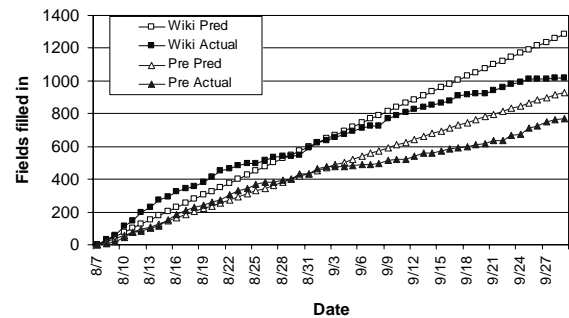
Pre-review imposes costs

- Some work is wasted
- Two people per completed contribution

$$V_{t+1} = V_t + \frac{1}{2} \frac{((1 - P_t)\gamma)^2 - (P_t\beta)^2}{(1 - P_t)\gamma + P_t\beta}$$

- Prediction: Wiki-like adds value faster

Predicted versus actual increases in quality



Equilibrium is the same

- Long-term, pre-review = Wiki-like
 - Again, provable
 - But, Wiki-like gets there faster
- Contributors determine value
 - Not surprising, but... surprising!
 - But, mechanism might affect *contributors*
- Needs to be validated

Economic Modeling of Contribution

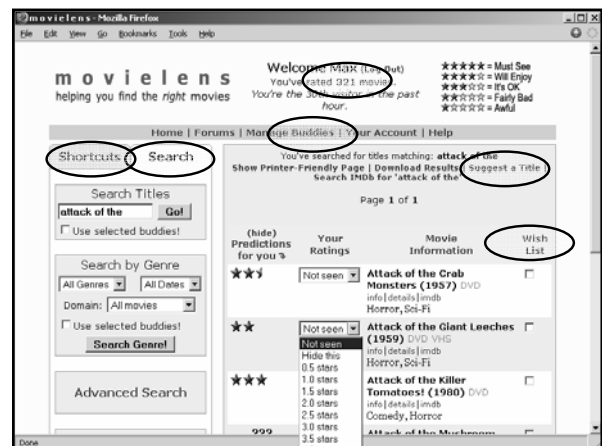
(Harper et al., UM 2005)

Social Preference Experiment

- Can we model user incentive for contributing to MovieLens?
- Can we use this model to create personalized messages to influence user behavior?
- Can we study which theories best predict behavior change?

Preliminary Model

- Benefit: prediction quality
 - Can be improved by rating
 - Diminishing returns
- Cost: time to rate
 - Increases over time (new movies harder to find)
- Rational agent will optimize



Improving the Model

- Check intuitions, learn motivations
 - Survey users
- Bottom-up, iterative approach:
 - Collect behavioral data
 - Hypothesize about effects
 - Build a new model
 - Lather, rinse, repeat

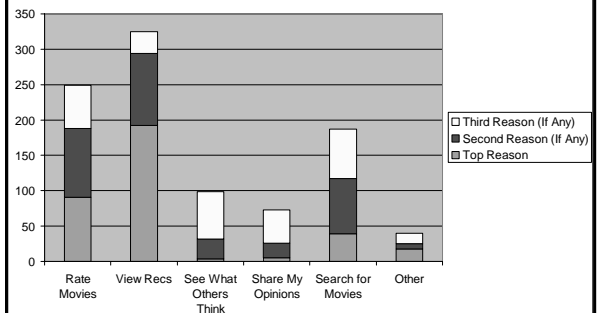
Improving the Model

- User study: entry criteria
 - 30 or more ratings
 - At least 3 separate logins
- 358 MovieLens users
 - mostly happy users
 - mostly power users...
 - >3/4 in the top quarter of users in terms of # ratings

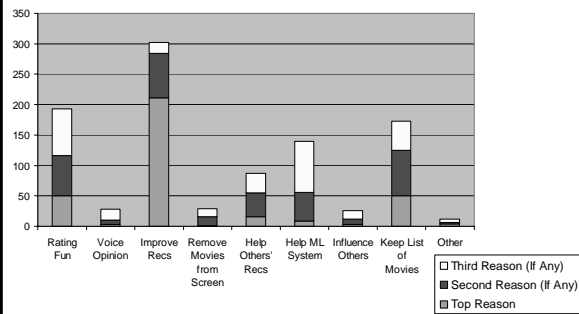
Improving the Model

- Survey data
 - Online survey
 - 7 multi-part questions
 - ~10 minutes
 - Examples:
 - "Rank your top 3 reasons to rate movies"
 - "How much would you pay to receive MovieLens recommendations?"
 - "Is rating movies fun?"
- Survey questions and results are at: <http://www.grouplens.org/data/mlsurvey0604.html>

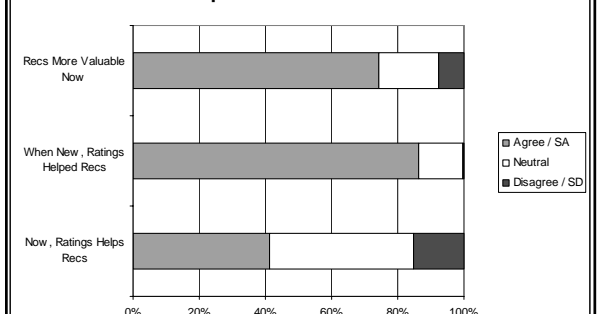
Top 3 Reasons for Using MovieLens



Top 3 Reasons to Rate Movies



Compare: New User vs Now



Survey Insights

- Some beliefs confirmed
 - Quality of recommendations matter
 - Recommendations get noticeably better over time
- Some new knowledge
 - Effects on others not too important
 - Rating things is fun

Improving the Model

- Behavioral Data
 - Examples:
 - "How many times have you logged in during the past 3 months?"
 - "How accurate have your recent predictions been?"
 - "How mainstream are your movie tastes?"

Calibrated Model

ratings
↓

$$\ln x_i = a_0 + a_1 \gamma_i + a_2 \beta_i + a_3 f_i + a_4 c_i + \vec{\lambda} \vec{Z} + \varepsilon_i.$$

↑ "Benefit" ↑ "Taste" ↑ "Fun" ↑ "Cost" ↑ Controls

Marginal Benefit (1/2)

$$\ln x_i = a_0 + \boxed{a_1 \gamma_i} + a_2 \beta_i + a_3 f_i + a_4 c_i + \vec{\lambda} \vec{Z} + \varepsilon_i.$$

- Benefit from recommendation quality
 - Survey: Oops! Users didn't like \$\$.
 - No useful direct measure
 - Estimate with other data, instead
- What's the scale? How do you use the data?
 - Well, we combined several variables...

Marginal Benefit (2/2)

$$\ln x_i = a_0 + \boxed{a_1 \gamma_i} + a_2 \beta_i + a_3 f_i + a_4 c_i + \vec{\lambda} \vec{Z} + \varepsilon_i.$$

- Survey
 - (+) Frequency, use ML to pick movies to watch
 - (+) Frequency, search for a movie
 - (-) Frequency, look only at first screen of recommendations
 - (+) Frequency, look at 5+ screens of recommendations
- Behavior
 - (+) # "hide this" ratings
 - (+) # saved searches (i.e. custom searches)

Results

- Regression Analysis
 - Adjusted $R^2 = .342$
 - $p = < .001$
- Behavior-only model possible

Next Steps

- Conducted an experiment evaluating the effect of conformity and inequality aversion prompts
- Designing personalized appeals to increase contribution

Reflections

- Jack Carroll has articulated the need for more theory in HCI
 - We think a large part of the problem is that theory from related fields isn't accessible to HCI practitioners
 - Resolve apparent conflicts
 - Develop design guidelines

MINTS – On-Line HIV Research

- Large NIH-funded project
 - Simon Rosser, PI, School of Public Health
 - Epidemiology
 - Co-Investigators from Medical School
 - Family Practice and Community Health
 - Program in Human Sexuality
 - Co-Investigators from Rhetoric
 - Co-Investigators from Education
 - Curriculum/Instruction
 - Co-Investigators from Computer Science
 - Outside experts (including community experts)

Background

- MINTS – Men's Internet Study
- Phase I – Assessment of MSM HIV risk for internet-using Latinos
 - Over 1000 valid subjects
- Phase II – Assessment of MSM HIV risk across US populations; develop and study Internet-based intervention
 - Approximately 3000 subjects

Six-Part Talk

- Design and Informed Consent
- Subject Data Privacy
- Data Integrity
 - the Evil Subject Naught
 - the Gang of 52
- "Content" Lessons from Phase I
- Lessons Learned about on-line Surveys
- A Glimpse at the MINTS-II Intervention

Design and Informed Consent

- Trust
- Ethical Behavior
- Attractiveness
- Bilingual Equivalency
 - Forward and back translated
 - Pilot tested for equivalency

Recruitment

We started by wanting to recruit 1,000 Internet-using Latino MSM Using banner advertisements at 3 layers of gay.com

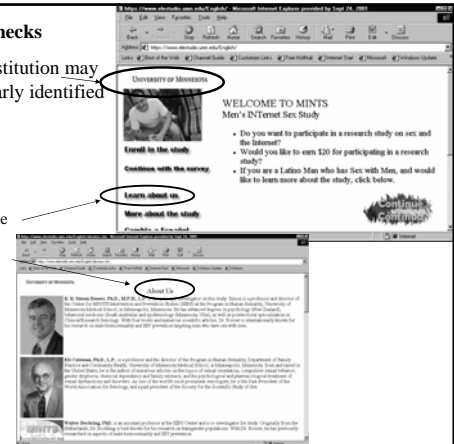


- "Bullet point language."
- e.g. 5 words to establish credibility.
- "Instant" decision making is more similar to market research than conventional studies

Credibility Checks

Sponsoring institution may need to be clearly identified and promoted.

Information on the researchers may require more researcher disclosure than in conventional studies



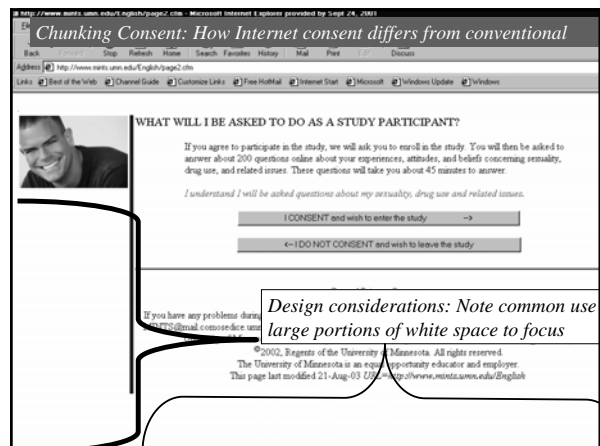
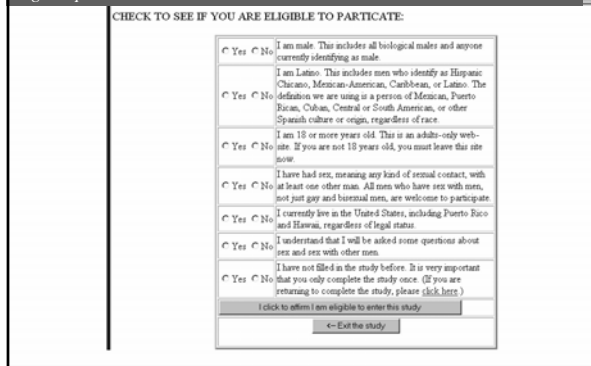
Informed Consent Online

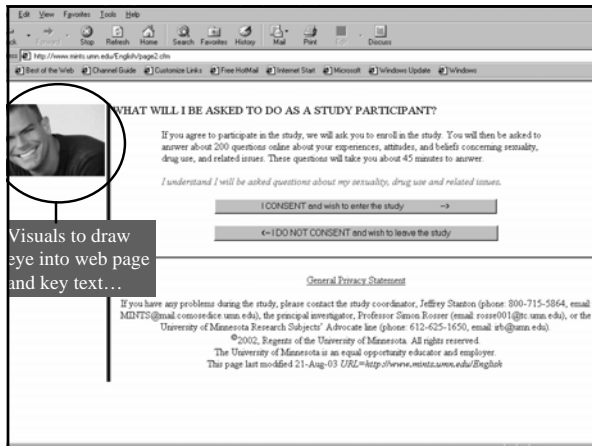
- Online reading different from print
 - Jumping around, hypertextual
 - Skimming, not reading
 - Bullets, not sentences
 - Nobody reads long consent forms
 - Harder to read on screen
- Solution: Chunked consent with active consent statements

We started with a welcome page which briefly summarized the points of key interest

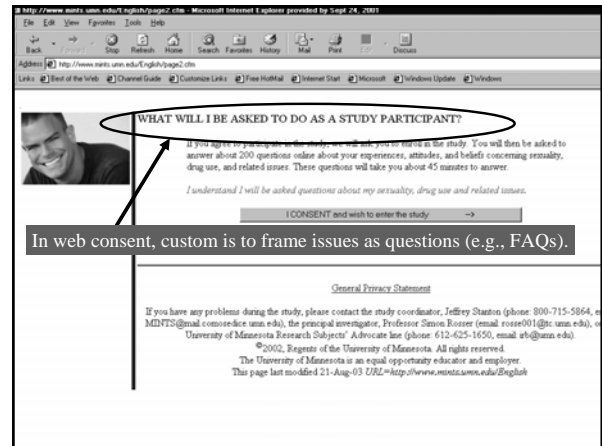


Internet time may be faster than real life time. For this reason, we brought our eligibility criteria forward so we did not waste non-eligible persons' time.

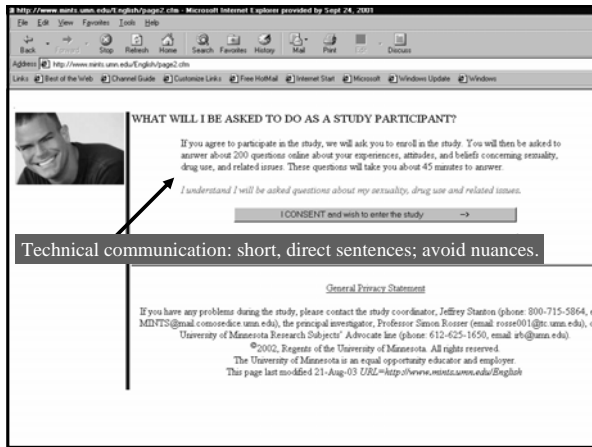




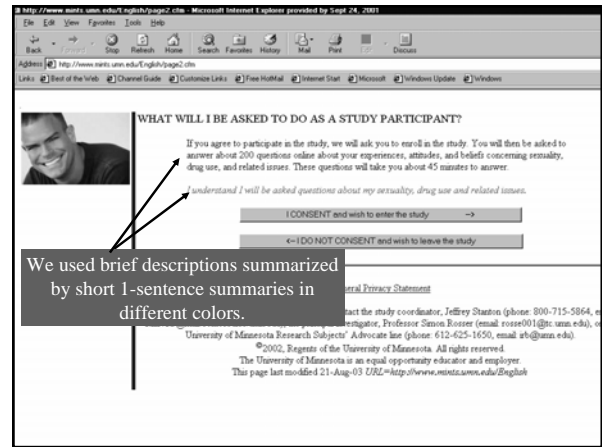
Visuals to draw eye into web page and key text...



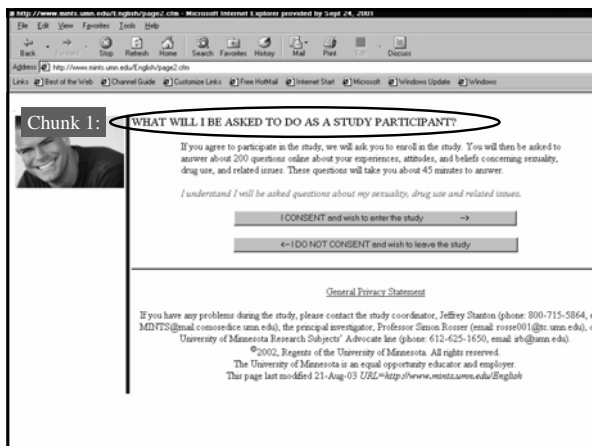
In web consent, custom is to frame issues as questions (e.g., FAQs).



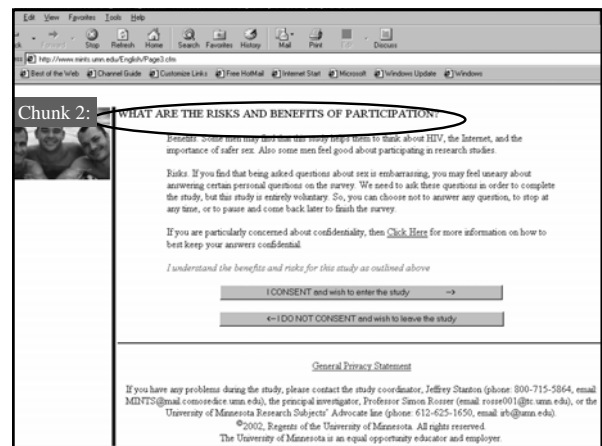
Technical communication: short, direct sentences; avoid nuances.



We used brief descriptions summarized by short 1-sentence summaries in different colors.



Chunk 1:



Chunk 2:

Microsoft Internet Explorer provided by Sept 24, 2003

Address: http://www.mnzt.unm.edu/English/sgprg4.cfm

Links: Best of the Web, Channel Guide, Custom Links, Free HotMail, Internet Star, Microsoft, Windows Update, Windows

Chunk 3: WILL I BE IDENTIFIED?

We want to assure you that we will not identify you. We do ask you for your zip code and telephone to make up a unique identifier that only you will know. This is necessary to register for the study. We will also ask you to provide an email address so we can contact you if there are any problems during the study. All email addresses will be deleted at the end of the study. Because we are asking sensitive questions, we have obtained a Certificate of Confidentiality from the National Institutes of Health. The researchers will use this certificate to challenge any demands for information that would identify you. For more information, press [here](#).

I understand that the study is confidential, and that only an email address will be used to contact me during the study.

I CONSENT and wish to enter the study →

← I DO NOT CONSENT and wish to leave the study

[General Privacy Statement](#)

If you have any problems during the study, please contact the study coordinator, Jeffrey Stanton (phone: 800-715-5864, email: MNZTS@mail.comconserve.unm.edu), the principal investigator, Professor Simon Rosser (email: rosse001@tc.unm.edu), or the University of Minnesota Research Subjects' Advocate line (phone: 612-625-1650, email: arb@umn.edu).

©2002, Regents of the University of Minnesota. All rights reserved.
The University of Minnesota is an equal opportunity educator and employer.
This page last modified 21-Aug-03 URL: http://www.mnzt.unm.edu/English

Microsoft Internet Explorer provided by Sept 24, 2003

Address: http://www.mnzt.unm.edu/English/confidentiality.cfm

Links: Best of the Web, Channel Guide, Custom Links, Free HotMail, Internet Star, Microsoft, Windows Update, Windows

ABOUT YOUR CONFIDENTIALITY

We are very concerned about the confidentiality of the information you provide us. As a government-funded study, the data we collect may be requested by others. Accordingly, we take the following precautions in collecting the data:

- We do not ask for an email address to aid communication. If you are worried that this may identify you, you can go to Hotmail, Yahoo or other web sites and create an email address just for this study. You may also opt not to provide an email address. All email addresses will be destroyed at the end of the study.
- We will not share, give or sell your email address to any other organization. It will be used solely for this study.
- We do not require you to provide any information that legally identifies you. At the end of the study, your unique identifier (based on your birthdate and zip code) will be destroyed.
- Any information you provide for purposes of receiving compensation will be stored separately from the rest of the study data and will be deleted after completion of the study.
- We will not secretly record or log any information about your visit to us. During data collection, we do collect IP addresses to ensure participants are valid. Email addresses are requested in case of problems with completing the survey. This information is deleted at the end of the study.
- We provide a secure web server to protect against intermediate sites intercepting your data.

In addition, to help us protect your privacy we have obtained a Certificate of Confidentiality from the National Institutes of Health. With this Certificate, we are forced to disclose information that may identify you, even by a court subpoena, in any federal, state, or local civil, criminal, administrative, legislative or other proceedings. We will use the certificate to resist any demands for information that would identify you, except as explained below:

- The Certificate cannot be used to resist a demand for information from personnel of the United States Government that is used for auditing or evaluation of this project.
- You should understand that a Certificate of Confidentiality does not prevent you from voluntarily releasing information about yourself or your involvement in this research. If you provide us with your written consent to receive research information, then we may not use the Certificate to withhold that information from you.
- You should also understand that a Certificate of Confidentiality does not prevent the researchers from releasing information about you to prevent serious harm to you or to someone else. However, it is highly unlikely that we would need to make such a disclosure, since we are not asking specific questions about such matters as child abuse, racial or homosexual interest or intentionally transmitting HIV.

Microsoft Internet Explorer provided by Sept 24, 2003

Address: http://www.mnzt.unm.edu/English/confidentiality.cfm

Links: Best of the Web, Channel Guide, Custom Links, Free HotMail, Internet Star, Microsoft, Windows Update, Windows

ABOUT YOUR CONFIDENTIALITY

We are very concerned about the confidentiality of the information you provide us. As a government-funded study, the data we collect may be requested by others. Accordingly, we take the following precautions in collecting the data:

- We do not ask for an email address to aid communication. If you are worried that this may identify you, you can go to Hotmail, Yahoo or other web sites and create an email address just for this study. You may also opt not to provide an email address. All email addresses will be destroyed at the end of the study.
- We will not share, give or sell your email address to any other organization. It will be used solely for this study.
- We do not require you to provide any information that legally identifies you. At the end of the study, your unique identifier (based on your birthdate and zip code) will be destroyed.
- Any information you provide for purposes of receiving compensation will be stored separately from the rest of the study data and will be deleted after completion of the study.
- We will not secretly record or log any information about your visit to us. During data collection, we do collect IP addresses to ensure participants are valid. Email addresses are requested in case of problems with completing the survey. This information is deleted at the end of the study.
- We provide a secure web server to protect against intermediate sites intercepting your data.

In addition, to help us protect your privacy we have obtained a Certificate of Confidentiality from the National Institutes of Health. With this Certificate, we are forced to disclose information that may identify you, even by a court subpoena, in any federal, state, or local civil, criminal, administrative, legislative or other proceedings. We will use the certificate to resist any demands for information that would identify you, except as explained below:

Section 1: how we will treat the personal information provided

- The Certificate cannot be used to resist a demand for information from personnel of the United States Government that is used for auditing or evaluation of this project.
- You should understand that a Certificate of Confidentiality does not prevent you from voluntarily releasing information about yourself or your involvement in this research. If you provide us with your written consent to receive research information, then we may not use the Certificate to withhold that information from you.
- You should also understand that a Certificate of Confidentiality does not prevent the researchers from releasing information about you to prevent serious harm to you or to someone else. However, it is highly unlikely that we would need to make such a disclosure, since we are not asking specific questions about such matters as child abuse, racial or homosexual interest or intentionally transmitting HIV.

Microsoft Internet Explorer provided by Sept 24, 2003

Address: http://www.mnzt.unm.edu/English/confidentiality.cfm

Links: Best of the Web, Channel Guide, Custom Links, Free HotMail, Internet Star, Microsoft, Windows Update, Windows

ABOUT YOUR CONFIDENTIALITY

We are very concerned about the confidentiality of the information you provide us. As a government-funded study, the data we collect may be requested by others. Accordingly, we take the following precautions in collecting the data:

- We do not ask for an email address to aid communication. If you are worried that this may identify you, you can go to Hotmail, Yahoo or other web sites and create an email address just for this study. You may also opt not to provide an email address. All email addresses will be destroyed at the end of the study.
- We will not share, give or sell your email address to any other organization. It will be used solely for this study.
- We do not require you to provide any information that legally identifies you. At the end of the study, your unique identifier (based on your birthdate and zip code) will be destroyed.
- Any information you provide for purposes of receiving compensation will be stored separately from the rest of the study data and will be deleted after completion of the study.
- We will not secretly record or log any information about your visit to us. During data collection, we do collect IP addresses to ensure participants are valid. Email addresses are requested in case of problems with completing the survey. This information is deleted at the end of the study.
- We provide a secure web server to protect against intermediate sites intercepting your data.

In addition, to help us protect your privacy we have obtained a Certificate of Confidentiality from the National Institutes of Health. With this Certificate, we are forced to disclose information that may identify you, even by a court subpoena, in any federal, state, or local civil, criminal, administrative, legislative or other proceedings. We will use the certificate to resist any demands for information that would identify you, except as explained below:

Section 2: Explaining a "Certificate of Confidentiality" and limits of confidentiality

- The Certificate cannot be used to resist a demand for information from personnel of the United States Government that is used for auditing or evaluation of this project.
- You should understand that a Certificate of Confidentiality does not prevent you from voluntarily releasing information about yourself or your involvement in this research. If you provide us with your written consent to receive research information, then we may not use the Certificate to withhold that information from you.
- You should also understand that a Certificate of Confidentiality does not prevent the researchers from releasing information about you to prevent serious harm to you or to someone else. However, it is highly unlikely that we would need to make such a disclosure, since we are not asking specific questions about such matters as child abuse, racial or homosexual interest or intentionally transmitting HIV.

Microsoft Internet Explorer provided by Sept 24, 2003

Address: http://www.mnzt.unm.edu/English/sgprg4.cfm

Links: Best of the Web, Channel Guide, Custom Links, Free HotMail, Internet Star, Microsoft, Windows Update, Windows

Section 3: Teaching enrollees steps they can take to increase their confidentiality ... from quitting web browsers to web anonymizers.

- You should also understand that a Certificate of Confidentiality does not prevent the researchers from releasing information about you to prevent serious harm to you or to someone else. However, it is highly unlikely that we would need to make such a disclosure, since we are not asking specific questions about such matters as child abuse, racial or homosexual interest or intentionally transmitting HIV.
- Except as listed above, we will not volunteer any personally identifiable information without your consent.

You are particularly concerned about privacy, we encourage you to consider the following other precautions:

- Make sure that nobody else can see your screen while you are answering the questions.
- Be sure to quit the web browser after finishing the study so the people cannot "go back" to the pages you completed.
- Verify that you have a secure connection to our site (look for the lock or key on your browser).
- If you are particularly concerned, you may want to connect to us through a web anonymizer service. You can find one by searching for "web anonymizer" on your favorite search engine. These services may even prevent your network provider from knowing that you connected to our survey.

[TURN TO PREVIOUS SCREEN AND THE RESEARCHERS HAVE THIS SITE](#)

[Links back directly](#)

[General Privacy Statement](#)

If you have any problems during the study, please contact the study coordinator, Jeffrey Stanton (phone: 800-715-5864, email: MNZTS@mail.comconserve.unm.edu), the principal investigator, Professor Simon Rosser (email: rosse001@tc.unm.edu), or the University of Minnesota Research Subjects' Advocate line (phone: 612-625-1650, email: arb@umn.edu).

©2002, Regents of the University of Minnesota. All rights reserved.
The University of Minnesota is an equal opportunity educator and employer.
This page last modified 21-Aug-03 URL: http://www.mnzt.unm.edu/English

Microsoft Internet Explorer provided by Sept 24, 2003

Address: http://www.mnzt.unm.edu/English/sgprg4.cfm

Links: Best of the Web, Channel Guide, Custom Links, Free HotMail, Internet Star, Microsoft, Windows Update, Windows

Chunk 5: IS THERE A CATCH TO THE STUDY?

No. We will not secretly keep your email address or send you any other information. Any information you provide for the purpose of receiving compensation will be stored separately from the rest of the study data, and will be destroyed at the end of the study. We will not leave co-headers (identifying files) on your system. We provide a secure web server to protect against intermediate sites intercepting your data. Please note that because you are answering questions via the Internet, your anonymity cannot be guaranteed.

I understand there are no secret catches to this study.

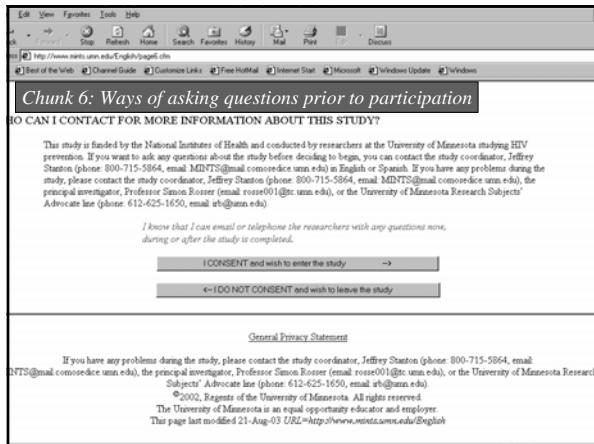
I CONSENT and wish to enter the study →

← I DO NOT CONSENT and wish to leave the study

[General Privacy Statement](#)

If you have any problems during the study, please contact the study coordinator, Jeffrey Stanton (phone: 800-715-5864, email: MNZTS@mail.comconserve.unm.edu), the principal investigator, Professor Simon Rosser (email: rosse001@tc.unm.edu), or the University of Minnesota Research Subjects' Advocate line (phone: 612-625-1650, email: arb@umn.edu).

©2002, Regents of the University of Minnesota. All rights reserved.
The University of Minnesota is an equal opportunity educator and employer.
This page last modified 21-Aug-03 URL: http://www.mnzt.unm.edu/English



Protecting Subject Data

- Context -- University of Minnesota
 - Extensive open records laws
- Context – Politically Sensitive Research
 - Need to protect against attack

Our Goals (1 of 2)

- Protect subjects
 - against knowledge of their participation
 - simply fact of participating reveals private data
 - protect against employer, family members, ...
- Protect integrity of data against attacks
 - hacker attacks on our systems
 - attacks on integrity of our data

Our Goals (2 of 2)

- Provide optional payment delivery
 - anonymous payment delivery
 - payment to charity
- Shield investigators from “discoverable knowledge” of subject identity
 - prevent subpoena-able knowledge

Key Trade-Offs

- Anonymity vs. Inducements
- Anonymity vs. Data Integrity
- Ease of Operation vs. Security
- Cost/Opportunity Cost

Pragmatic Design Principles (Selected Examples)

- Computer backups occur – avoid having sensitive data where an unexpected backup might archive it.
- Technology alone is insufficient – obtained certificate of confidentiality

Pragmatic Design Principles (More Selected Examples)

- Facilitate subjects in adding security. We cannot secure their machines, but we can advise them on doing so.
- Focus security on the weakest link. Don't oversecure the strong links!

Protection Against Monitoring

- Require SSL (encrypted communication)
- Advise subjects on how to clear responses from cache, on use of anonymizers, and on availability of free web-based e-mail.
- Several payment options:
 - check
 - PayPal
 - charity
 - no payment

When Subjects Have Questions

- Protocols for handling telephone and e-mail inquiries
 - deliberately avoid calling number identification
 - require enough data to prevent "spoofing"
 - collect no more data than necessary
- Staff training on handling sensitive communications

Sample Integrity

- Risks
 - Reaching a non-representative audience
 - Validate demographics (e.g., compare with census data)
 - Slow recruitment (too easy to get thousands of subjects before weekend)
 - Understand demographics
 - Cheating (multiple surveys)

MINTS-I Data

47,105,771 banner ads displayed (copy.com)	
29,024 clicks through advertising mails to the study site	
1742 enrollments - completion of eligibility screening and consent	
1152 completed surveys	
1026 surveys deemed valid and complete	
18 of these surveys are first completions by people who later completed one or more additional surveys	
110 repeat surveys (second or later complete survey from a single subject)	
65 surveys from "Subject Naught"	51 surveys from 17 other repeat IDs
5 surveys from participants deemed ineligible to participate	
3 evidence of non-Latinic identity	2 evidence of non-US residence
634 incomplete surveys, therefore rejected	
500 surveys flagged incomplete but otherwise valid (from a unique person)	
14 surveys deemed invalid and invalid (repeat submissions)	
58 surveys not started (did not answer any questions)	

What Happened?

- Significant duplicates submissions
 - Attempts to get compensation
- How were they detected?
 - Payment paper trail
 - IP address range
 - Time-to-completion
- Would it have mattered?
 - Yes – changed results!!!!

MINTS II – Aim I

- Still too early for final data, but ...
 - Over 150 subjects (out of 3000 completers) invalidated for evidence of being duplicates
 - The Gang of 52
 - A gut feeling (this is not the same as evidence, after all evidence can be wrong!) that duplicates increased when we increased compensation

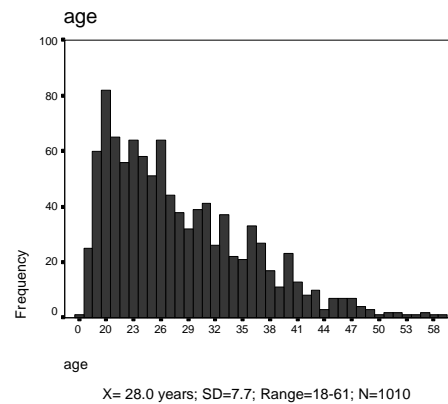
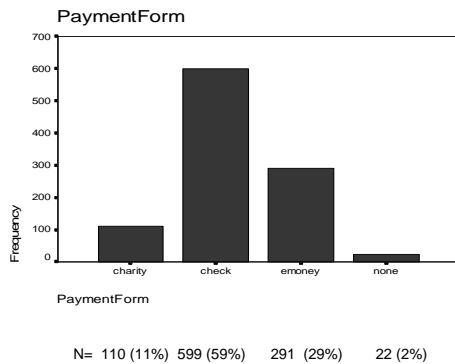
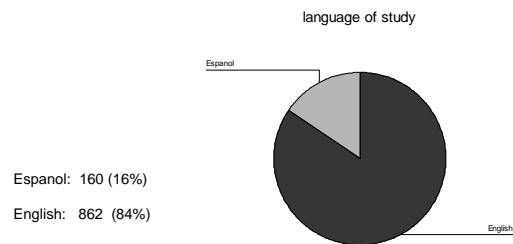
Validity Concerns

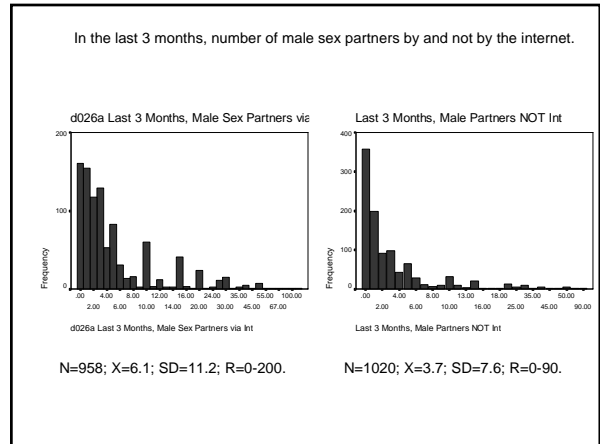
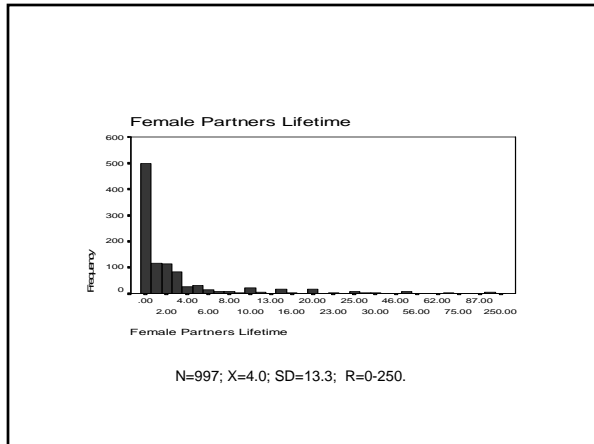
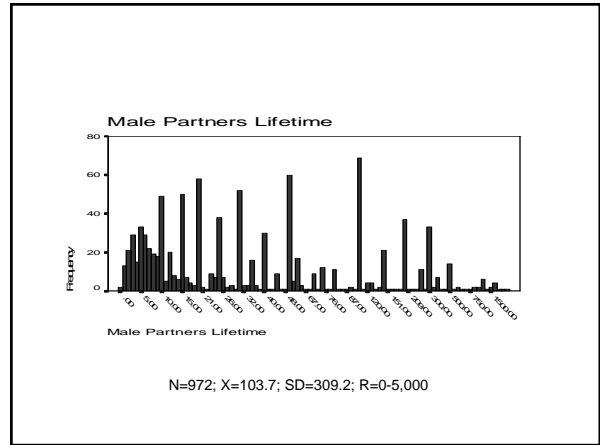
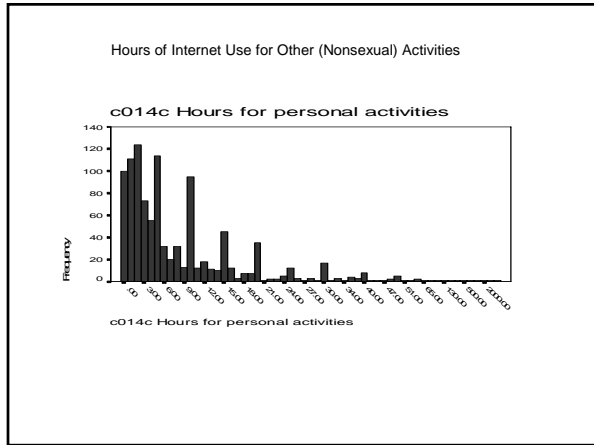
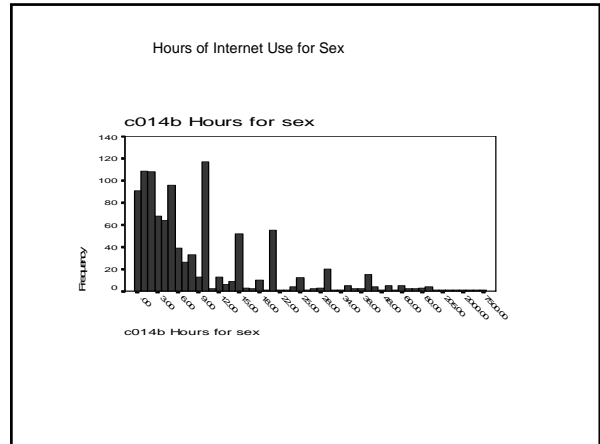
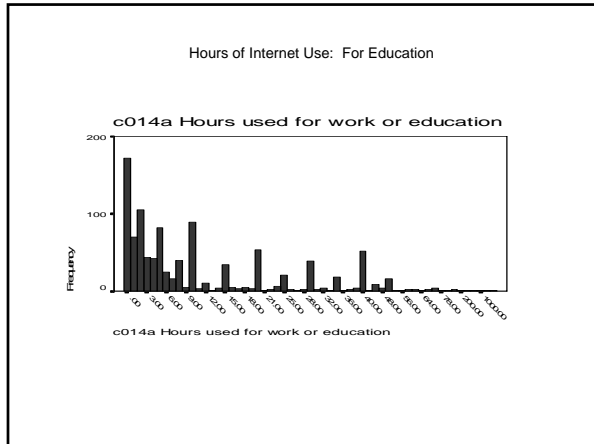
- Automated inspection shows numerous examples of inconsistent data
 - Age mismatch from enrollment to completion
 - Bogus data for some numeric fields (e.g., age = 4)
- Hypothesis (still being tested): attention and validity wane over a long survey
- Caveat – this is a very long internet survey
- Caveat #2 – users expect short!

Finally, Some Results

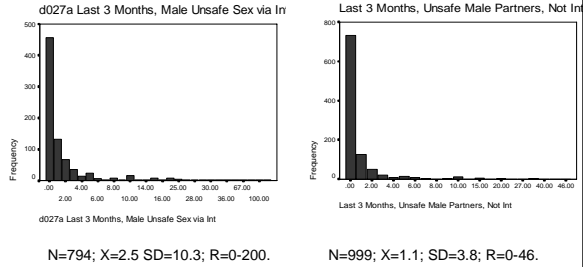
- Still in analysis phase, but ...

Language that Subjects Completed Study In

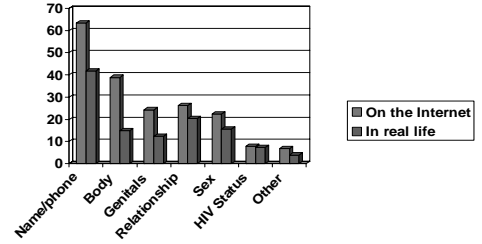




Comparison of number of male partners, unsafe sex met via the Internet and Met conventionally.



Lying on the Internet and in Real Life



Obvious, In Retrospect

- For subject #1, online is much harder than offline
 - Automating common sense
 - Tools still inadequate
 - Hard to change on the fly
 - Need for human monitoring
- Crossover somewhere in the thousands
 - Small marginal effort
 - Broader recruitment base

Thoughts Going Forward

- We really need to understand the nature and distribution of unreliability in internet studies
- When does larger sample size or reach outweigh reduced individual reliability
- I'd love to study incentives for attentive/reliable/accurate data

A Glimpse at the Future

- The MINTS-II Intervention
- co-Developed with experts in education, interactive media, sexology, public health
- Will be tested through a randomized clinical trial to assess reduction in sexual risk taking over 12 months
- Grounded in the Sexual Health Model, and based on a successful in-person workshop

Demo

Conclusions

- Challenging problems
- Need for Interdisciplinary Collaboration

Acknowledgements

- NSF ITR Grant: 03-24851
- NIH Grant #AG63688-01 (NIMH)

- and a whole bunch of collaborators

Bridging Computer Science and Behavioral Science: Research Examples

Joseph A. Konstan
Department of Computer Science and Engineering
University of Minnesota
konstan@cs.umn.edu
<http://www.umn.edu/~konstan>