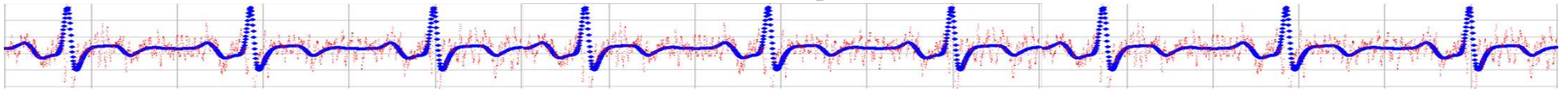# Empirical Research Methods in Information Science
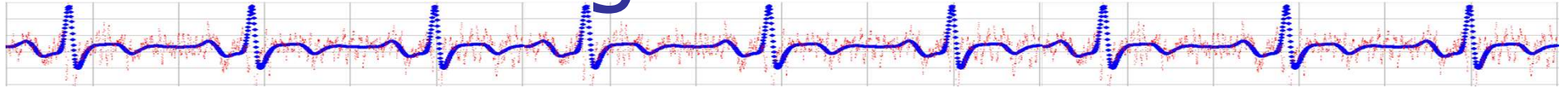
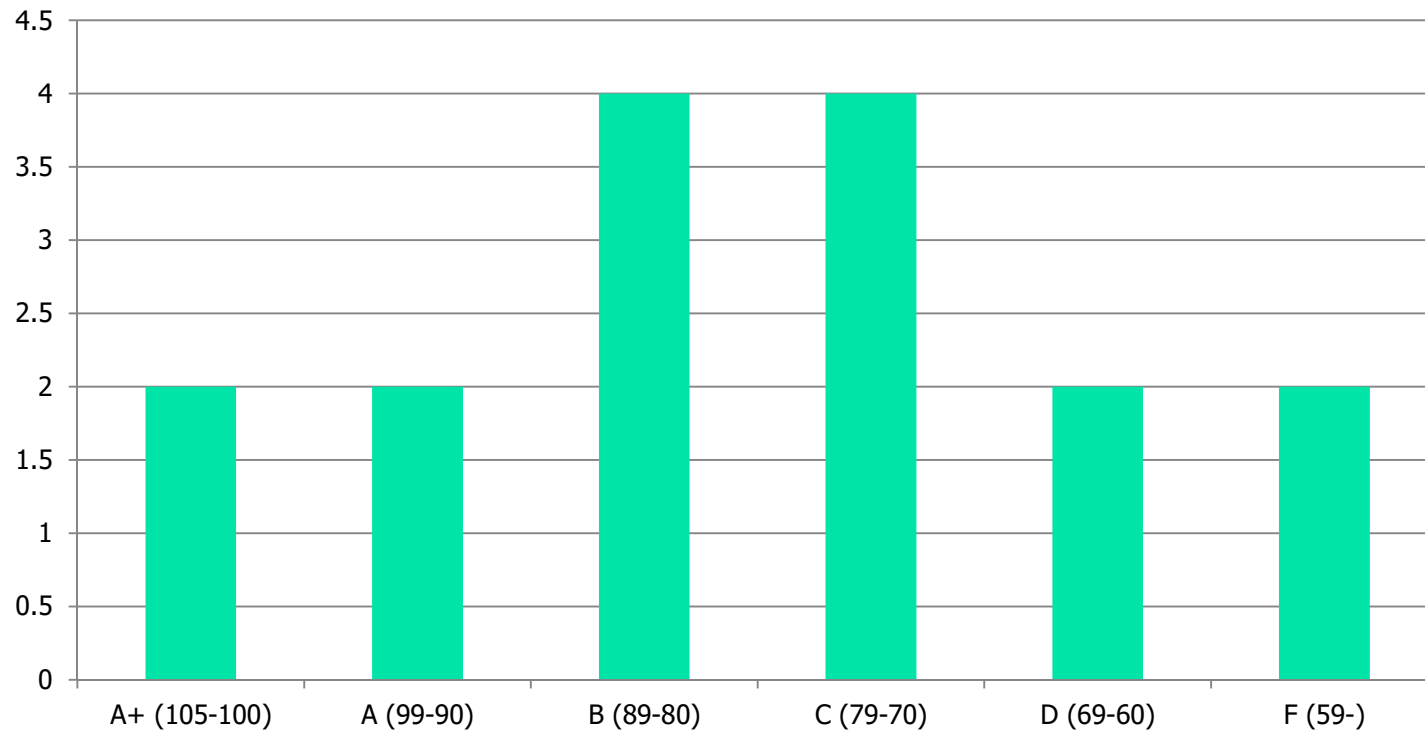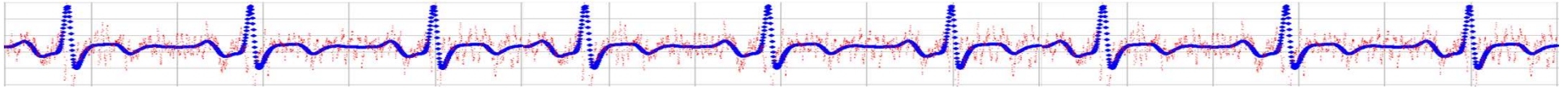# IS 4800 / CS6350

Lecture 16

Outline

- Reading assessment
- Midterm grades and discussion
- Reminder on project and expectations
- Reporting results
- Finding a research topic

Midterm grade distribution

**Midterm distribution
(with 13 point curve)**

Schedule for projects
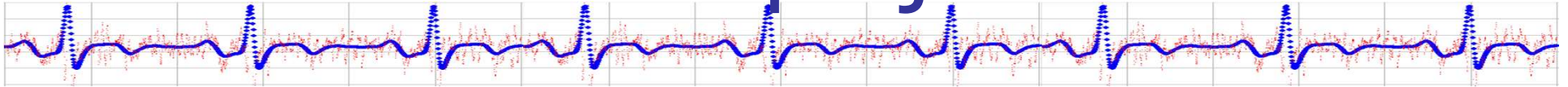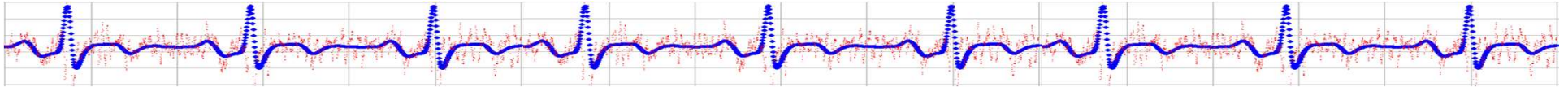
- Mon 3/25: T1 due/presentation
- Wed 3/27: T2 project description due
- Wed 4/3: T2 due/presentation
- Mon 4/8: T3 proposal due
- Wed 4/17: T3 due/presentation

Team project presentation
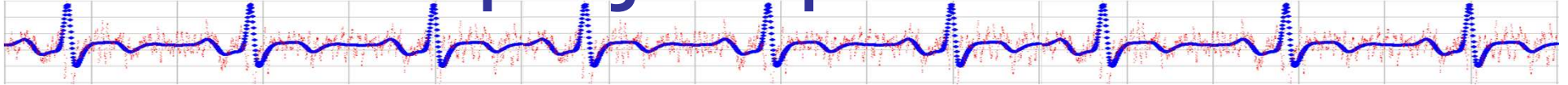
- 10 minute oral presentation (hard upper bound)

- 5 minutes critique (from all)

- Research question & motivation, hypotheses, study design, results, conclusions

- Visualization of data

Team evaluations

At completion of each project, I will ask you to evaluate your teammates

First team project
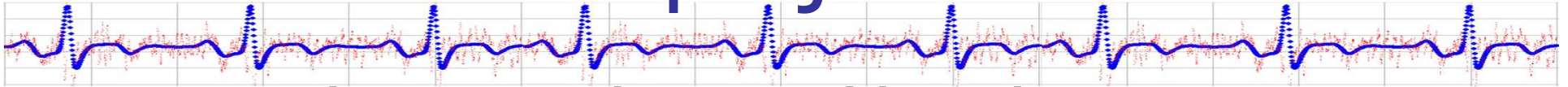
- Ethnographic and/or descriptive
- Priorities:
  - Research question & motivation
  - Methodology
  - Descriptive stats
  - Form of write-up & presentation
  - Conclusions/lessons learned

Team project bonus
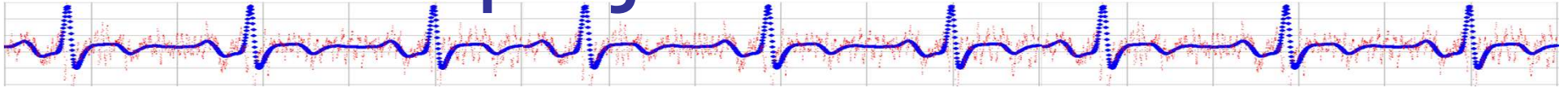
- Hand-in editing checklist (initialed by everyone on team) and get +10 point bonus

- https://docs.google.com/document/d/1s-Q_G2GaB3Ts1tv3nnK8IW5Ay-_P_VXBPnOifxQ9OeA/edit

(But if you say you checked each one and you clearly didn't, -10 point penalty!)

Presenting Research Results
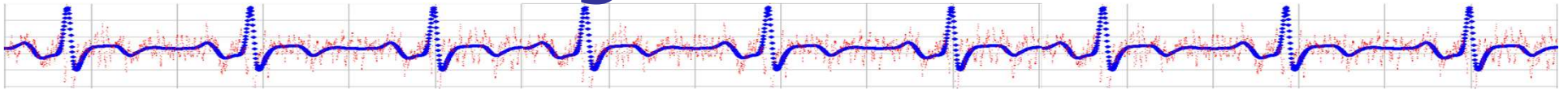
Written Study Reports
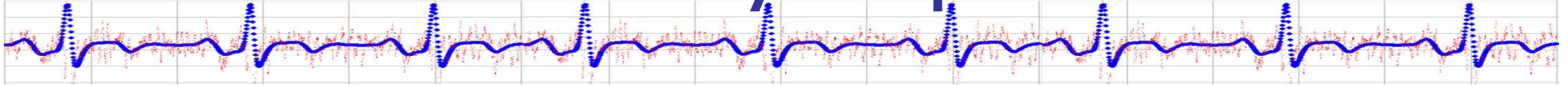
- Ignore most of the stuff on APA style in chapter (fonts, etc.)

- Information on <u>paper structure</u> very important & relevant

Written Study Reports
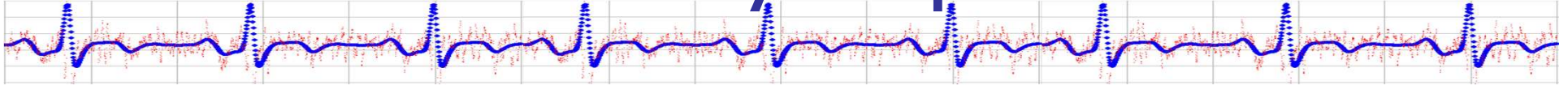
- Objectives (also critiques)
  - Describe what your study is about
  - Motivate your study
  - Assure reader you have conducted a sound study
    - Research Methods – often presented in small font
  - Present results in an objective manner
  - Discuss implications
  - Discuss future work

  - *Enable replication*

Typical Study vs. IS/CS/HCI Paper Structure
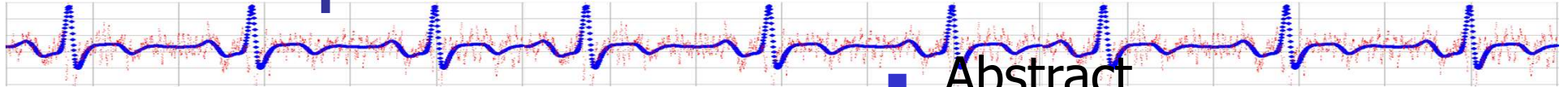
- Abstract
- Introduction
  - Motivation
  - Related work
  - Hypotheses
- Method
- Results
- Discussion
  - Limitations
  - Implications
  - Future work
- References

- Abstract
- Introduction
  - Motivation
- Related work
- System design
- Evaluation
  - Hypotheses
  - Method
  - Results
  - Discussion – summary, limitations
- Conclusion
  - Implications
  - Future work
- References

The Abstract

- Concise summary
- Abstract for an empirical study should include
    - Information on the problem under study
    - The nature of the subject sample
    - A description of methods, equipment, and procedures
    - A statement of the results
    - A statement of the conclusions drawn
- Often the last thing you write

14
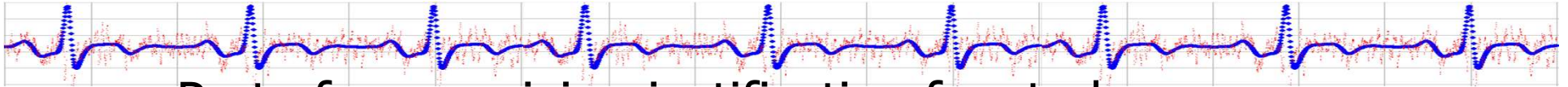
The Introduction

- Part of paper giving justification for study
- Usually has the following information
  - Introduction to the topic under study
  - Brief review of research and theory related to the topic
  - A statement of the problem to be addressed
  - A statement of the purpose of the research
  - A brief description of the research strategy
  - A description of predictions and hypotheses
- CS/IS papers often put Related Work as a separate section after Introduction
  - For each, *describe how your work is different*

Prior work
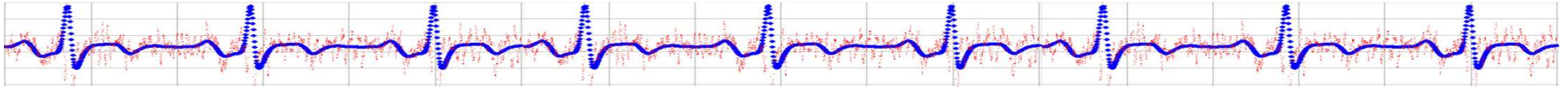
- Try to frame prior work in a positive light

- The natural tendency is to focus on the negative (e.g., X did Y but didn't do Z), but a better strategy is to focus on the positive (e.g., X enhanced the field by doing Y, setting up the opportunity to do Z in future work.)

16

Organization of the Introduction: General to Specific
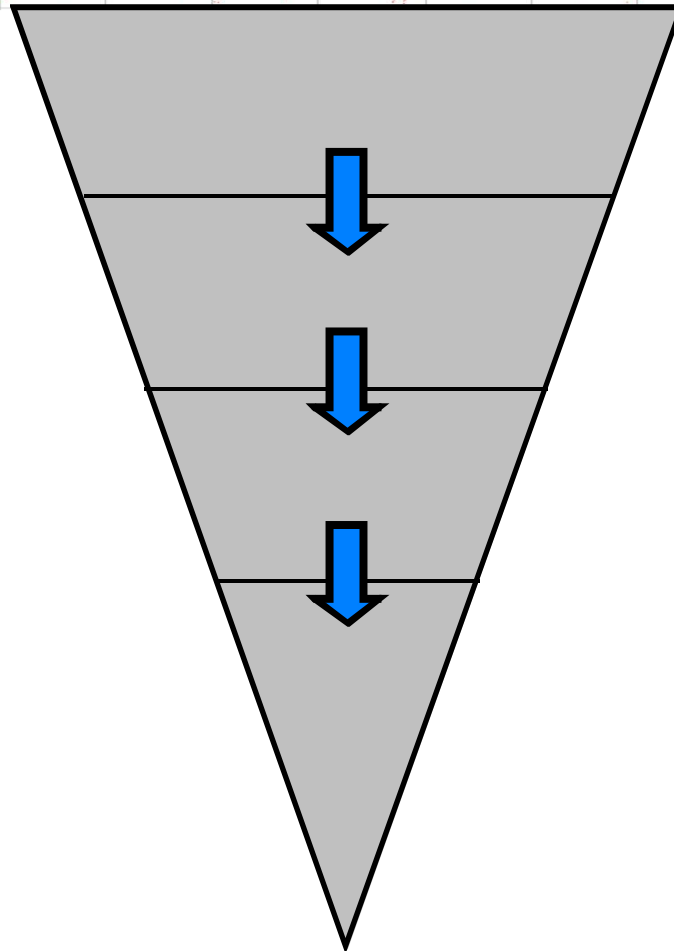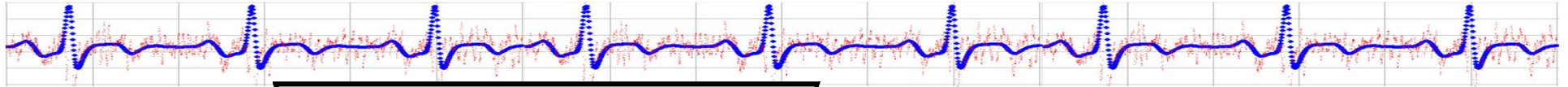
Present a general introduction to your topic

Review relevant literature

Link literature review to your hypotheses

State your hypotheses

The Method Section

- Includes information on exactly how a study was carried out
- Subsections
  - Participants or subjects
    - Describe in detail the participant or subject sample
    - Human participants go in a *Participants* subsection, and animal subjects in a *Subjects* subsection
  - Apparatus or materials
    - Describe in detail any equipment or materials used
    - Equipment is usually described in an *Apparatus* subsection and written materials in a *Materials* subsection

The Method Section

- Procedure
  - Describe
    - Exactly how the study was carried out
    - The conditions to which subjects were exposed or under which observed
    - The behaviors measured and how they were scored
    - When and where observations were made
    - Debriefing procedures
- Enough detail should be included in all sections so that the study could be replicated

The Results Section
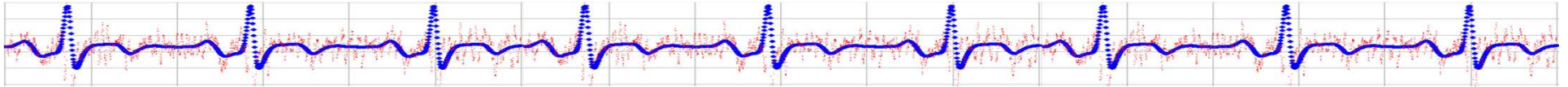
- Objective, dry, boring – *just the facts*
- All relevant data and analyses are reported in the results section
- Do not present raw data
- Data should be reported in summary form
  - Descriptive statistics (and assumptions)
  - Inferential statistics (and assumptions)
- Results of descriptive and inferential statistics must be presented in narrative format
- Describe the source of any unconventional statistical tests

Commonly Used Statistical Citations
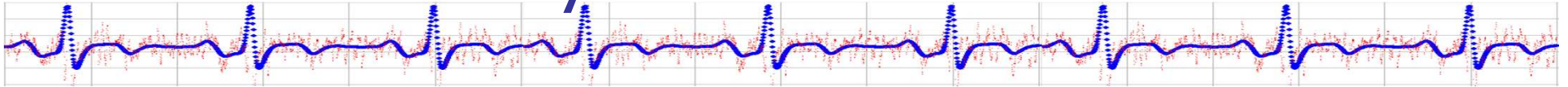
Statistical Test	Format
Analysis of variance	$F(1,85) = 5.96, p < .01$
Chi-square	$\chi^2(3) = 11.34, p < .01$
t test	$t(56) = 4.78, p < .01$
Pearson correlation coefficient	$r = -.87, p<.05$

Abbreviations for Statistical Notation

Abbreviation	Meaning
df	Degrees of freedom
F	F ratio
M	Arithmetic average (mean)
N	Number of subjects in entire sample
p	*p* value
SD	Standard deviation
t	*t* statistic
z	Results from a *z* test or *z* score
μ	Population mean (mu)
σ	Population stddev

The Discussion Section

- This is where you can take some liberties with describing what the results *mean*
- Results are interpreted, conclusions drawn, and findings are related to previous research
- Section begins with a brief restatement of hypotheses
- Next, indicate if hypotheses were confirmed
- The rest of the section is dedicated to integrating findings with previous research
- It is fine to speculate, but speculations should not stray far from the data

Organization of Discussion: Specific to General
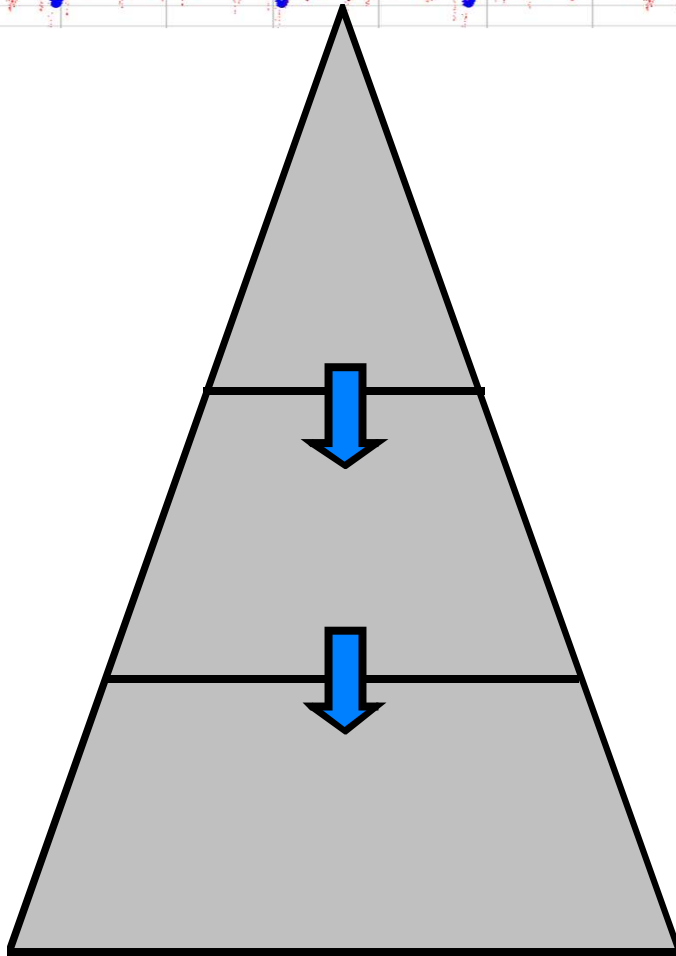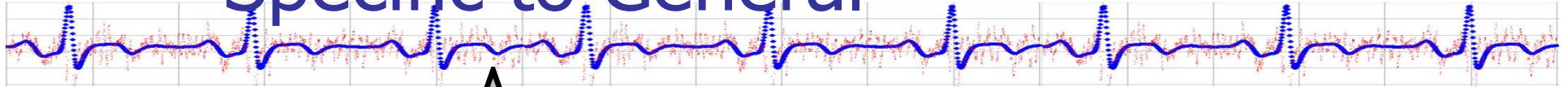
Restate your hypotheses or major finding

Tie your results with previous research and theory

State broad implications of your results, methodological implications, directions for future research

Example
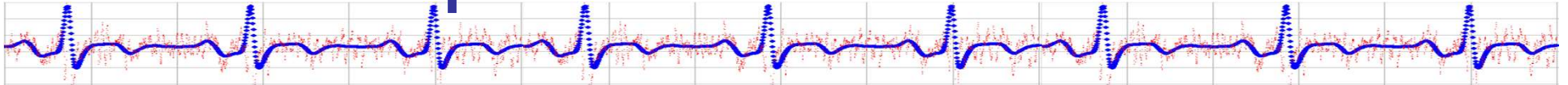
The Effects of Animated Characters on Anxiety, Task Performance, and Evaluations of User Interfaces

Raoul Rickenberg and Byron Reeves

Department of Communication
Stanford University
Stanford, CA 94305 USA
+1 650 725 3033
raoul@leland.stanford.edu
reeves@leland.stanford.edu

25

ABSTRACT

Animated characters are common in user interfaces, but important questions remain about whether characters work in all situations and for all users. This experiment tested the effects of different character presentations on user anxiety, task performance, and subjective evaluations of two commerce websites. There were three character conditions (no character, a character that ignored the user, and a character that closely monitored work on the website). Users were separated into two groups that had different attitudes about accepting help from others: people with control orientations that were *external* (users thought that other people controlled their success) and those with *internal* orientations (users thought they were in control). Results showed that the effects of monitoring and individual differences in thoughts about control worked as they do in real life. Users felt more anxious when characters monitored their website work and this effect was strongest for users with an external control orientation. Monitoring characters also decreased task performance, but increased trust in website content. Results are discussed in terms of design considerations that maximize the positive influence of animated agents.

INTRODUCTION

The history of ideas about animated characters in human-computer interaction is turning a corner. Initial debates concerned the presence of any character performing any kind of behavior. The questions were whether animated characters—as a general concept in interfaces—were good or bad, useful or useless. These debates rarely yielded an answer more satisfying than—"it depends." As has been the case with the introduction of all new media in the 20th century, the initial debate was framed too aggressively to

An elaboration of the conditions for animated characters to succeed is underway. There are several new studies that demonstrate the potential for animated characters to automate social interactions in ways that make computing more pleasing, productive, and easy. Research has focused on, for example, character appearance [19, 15], non-verbal behavior [6, 17], personality [12, 31], emotion [3, 4], and speech characteristics [20]. This research is important given the increasing use of animated characters in products and services ranging from search engines to shopping "bots" to virtual employees in commerce transactions.

Social facilitation

Sometimes it's nice to have company—a real person—when you work. Imagine, however, that you're working on a hard problem. Someone enters the room,

Locus of control

Now imagine two different people being monitored while they work on a complex task. The first person believes that she controls her own destiny and that other people have little to do with whether she fails or succeeds. The second person is convinced, however, that he is at the mercy of

EXPERIMENTAL METHODS

Subjects. Eighty-four people participated in the experiment (60% male and 40% female). An additional 20 people were used to pretest stimulus materials. All subjects were either undergraduate or graduate students recruited at Stanford University. All were experienced computer users (i.e., they knew how to word-process and manage a UNIX email account).

Experimental Design. The experiment was a between-subjects, full-factorial two-by-three design. The two factors were (1) the subjects' locus of control and (2) the monitoring activity of an animated character.

Locus of Control Pretest. Rotter's [27] Locus of Control Scale was used to determine the internal versus external orientations of potential subjects. This instrument consists of 23 forced-choice items that each present a pair of statements. In each pair, one statement expresses an internal viewpoint and the other an external viewpoint.

Scores on this scale can range from 0, indicating that no external statements are endorsed, to 23, indicating that all external statements are endorsed. The mean score on the pretest was 13.27 (SD = 3.8). Only the 42 subjects that scored lowest (internal) and highest (external) on the pretest were selected to participate. A two-tailed t-test indicated that the scores on the Locus of Control scale for these two groups was significantly different ($t(82)$ = 18.47, $p <$ .001).

Stimulus Material. The stimulus material consisted of two primary components: the animated characters that comprised the distinction between the Idle-Character and Monitoring-Character conditions and the web-based tasks that all subjects completed during the experiment.

The animated characters were specially developed for this experiment using Microsoft Agent software. The characters used in both the Idle-Character and the Monitoring-Character conditions were based upon Microsoft's "Genius" animations, so their physical features were identical (see Figure 1 for an illustration). The characters were approximately 1.5" tall (1152 x 870 resolution) and appeared in the lower-left corner of the Microsoft Explorer 4.5 browser that people used to view the web pages.

Apparatus. The computers used in the primary experiment were identical Hewlett Packard 440 mhz Kayak XWs with 21" color monitors (1152 x 870 resolution). These computers also had identical keyboards and mice, but were located in different experimental labs. The two labs were similar in terms of size and furnishings and use of the two labs was balanced across conditions.

Procedure. After arriving at a prescheduled time, people were brought to one of the labs in which the experiments were run and given a questionnaire. An experimenter then read an introductory script that was identical for all subjects before leaving the room.

People read specific instructions for each task in a frame that appeared at the bottom of their web-browser. When they found a solution to a task, they keyed their response into a text field in the frame and selected a *Submit* button. This brought up the instructions for the next task.

Anxiety Measure. Anxiety was assessed using a modified version of Speilberger, Gorsuch, & Lushene's [28] State Anxiety Scale. The items were answered on four-point Likert scales on a paper questionnaire. The following items exemplify those that appear on the questionnaire: *I felt calm; I felt secure; I felt strained;* etc.

Performance Measure. Performance was measured by adding the number or tasks completed correctly. The computers used in the experiment recorded people's answers for each task. Performance was calculated on the basis of logfiles compiled by these computers. Tasks included comparing the performance of various mutual funds, configuring computer hardware, etc.

Website Evaluations. Subjective evaluations of the two websites were measured with a sixteen-item questionnaire.

Manipulation check. The manipulation check was an index of three items answered by all people who were in one of the conditions that involved characters. People were asked whether the character seemed to be *watching* them, whether the character seemed to *record* their answers, and whether the character seemed to be *judging* them. These items were used to form a Monitoring Index that had a Cronbach's *alpha* of .70.

The manipulation was successful. A planned one-tailed *t*-test on the Monitoring Index showed that subjects in the Monitoring-Character condition reported a higher level of monitoring than subjects in the Idle-Character condition ($t(52) = 4.49$, p $<$ .001).

RESULTS

Full factorial ANOVAs were performed on all measures. A summary of these ANOVAs appears in Table 1. The planned tests of all hypotheses are discussed in detail below, as are results pertaining to the relationship of animated characters and locus of control to the evaluation of the websites.

Figure 2 shows the results for anxiety. One-tailed, *a priori* contrasts showed that people were more anxious when an idle character was present than when no character was present ($t(72) = 1.4$, $p = .08$). Also, people were more anxious if an animated character appeared to monitor them than if no character was present ($t(72) = 4.6$, $p < .001$). And people were more anxious if an animated character monitored them than if an idle character was present ($t(72) = 4.6$, $p < .001$). Users with an external locus of control were also more anxious when monitored by an animated character than were users with an internal locus of control ($t(72) = 4.6$, $p < .001$).

Anxiety

Internal Locus of Control
External Locus of Control

No Character   Idle   Monitoring***

Animated Character Behavior

36
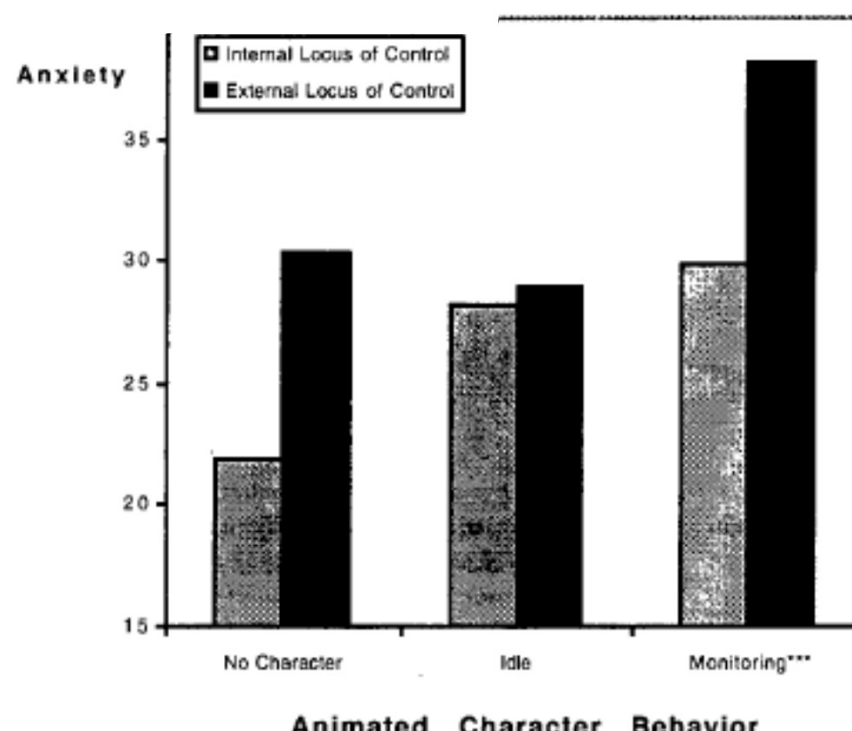
DISCUSSION

The perception of being monitored by an animated character has the same effects on Anxiety and Performance as being monitored by a human, either electronically or in person. When a character watches, users are more likely to feel anxious about their work and to perform less well. This anxiety is most pronounced among users who think that other people control their success.

At the most general level, these results suggests that decisions concerning the use of animated characters should address the details of execution and social presentation.' It is not sufficient–for celebration or condemnation–to focus on whether or not an animated character is present. Rather, the ultimate evaluation is similar to those for real people–it depends on what the character does, what it says, and how it presents itself. The effects of animated characters are not

Citations

- Liberally cite previous & related work.

- If you copy passages you *must* cite and, depending on length, format to indicate it is copied.

- Suggest using EndNote, BibTex or similar.

Endnote demo

Ethical Issues

- Report *all* of your findings (not just the ones you like)
- Adhere to your original plan
  - Report any deviations and why
  - Power analysis, statistics, measures
- Do not drop subjects or data points without rigorous justification
- If your hypothesis test was not significant *you cannot say anything about difference in means*
- If you did not do an experiment, attempting to control for extraneous variables, *you cannot mention or imply causality*

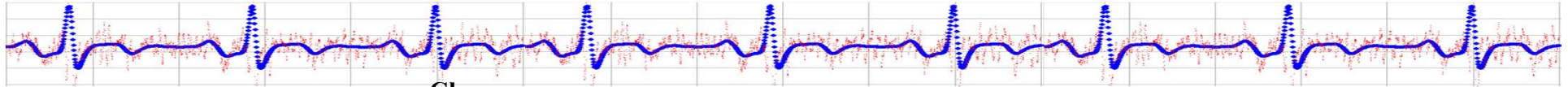# Oral Presentation of Study Results
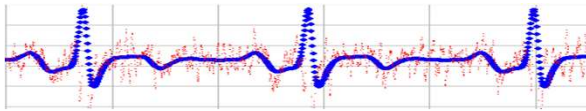
Oral Presentation

- Main concepts and ideas

- Do *not* go into great detail on experimental methods – just enough so people understand roughly what you did

- Focus on motivation, results, implications
 - If listener wants details they can read the paper or ask questions
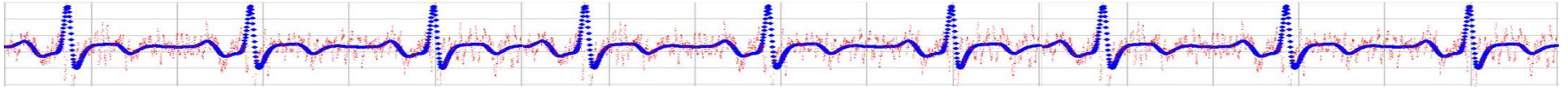
Oral Presentation
Don't do this…

Measure	Change From Day1	To Day2	ALL CONDS df	t	p	CONTROL df	t	p	NON-REL df	t	p	RELATIONL df	t	p
WAI/COMP	7	27	54	0.205	0.838				24	0.014	0.989	29	0.361	0.720
WAI/BOND	7	27	54	0.519	0.606				24	0.376	0.710	29	1.489	0.147
WAI/TASK	7	27	54	0.134	0.894				24	0.409	0.686	29	0.661	0.514
WAI/GOAL	7	27	54	0.155	0.877				24	0.081	0.936	29	0.329	0.745
CONTINUE LAURA	30	44	54	0.868	0.389				24	0.625	0.538	29	0.619	0.541
MIN/DAY	-6-0	22-30	81	1.470	0.145	26	1.274	0.214	24	0.124	0.903	29	1.104	0.279
	1-7	22-30	81	0.691	0.492	26	0.758	0.456	24	0.109	0.914	29	0.358	0.723
	22-30	38-44	81	3.626	**0.001**	26	2.480	**0.020**	24	1.959	0.062	29	1.804	0.082
DAY/WK>30MIN	-6-0	22-30	81	6.653	**0.000**	26	2.323	**0.028**	24	5.284	**0.000**	29	4.347	**0.000**
	1-7	22-30	81	6.272	**0.000**	26	2.401	**0.024**	24	3.818	**0.001**	29	4.597	**0.000**
	22-30	38-44	81	8.990	**0.000**	26	4.043	**0.000**	24	5.322	**0.000**	29	6.530	**0.000**
STEP/DAY	1-7	22-30	81	1.778	0.079	26	1.197	0.242	24	2.366	**0.026**	29	0.236	0.815
DAY/WK>10KSTEP	1-7	22-30	77	3.986	**0.000**	25	1.355	0.188	23	3.591	**0.002**	27	2.055	**0.050**
STAGE	Intake	30	81	6.988	**0.000**	26	3.403	**0.002**	24	4.000	**0.001**	29	4.738	**0.000**
	30	44	81	2.019	**0.047**	26	1.185	0.247	24	1.000	0.327	29	1.409	0.169
SELF-EFFICACY	1	29	81	4.782	**0.000**	26	0.872	0.391	24	3.314	**0.003**	29	4.750	**0.000**
	29	44	81	2.770	**0.007**	26	1.525	0.139	24	4.550	**0.000**	29	0.085	0.933
PROS	1	29	81	1.998	**0.049**	26	1.418	0.168	24	0.456	0.653	29	1.540	0.134
	29	44	81	0.393	0.695	26	1.147	0.262	24	0.225	0.824	29	0.308	0.760
CONS	1	29	81	0.902	0.370	26	1.124	0.271	24	0.499	0.622	29	0.823	0.417
	29	44	81	0.740	0.462	26	0.386	0.703	24	0.611	0.547	29	0.339	0.737
CONTINUE FT	30	44	81	1.520	0.133	26	1.442	0.161	24	1.163	0.256	29	0.000	1.000

43

Oral Presentation
Do use as many figures as possible
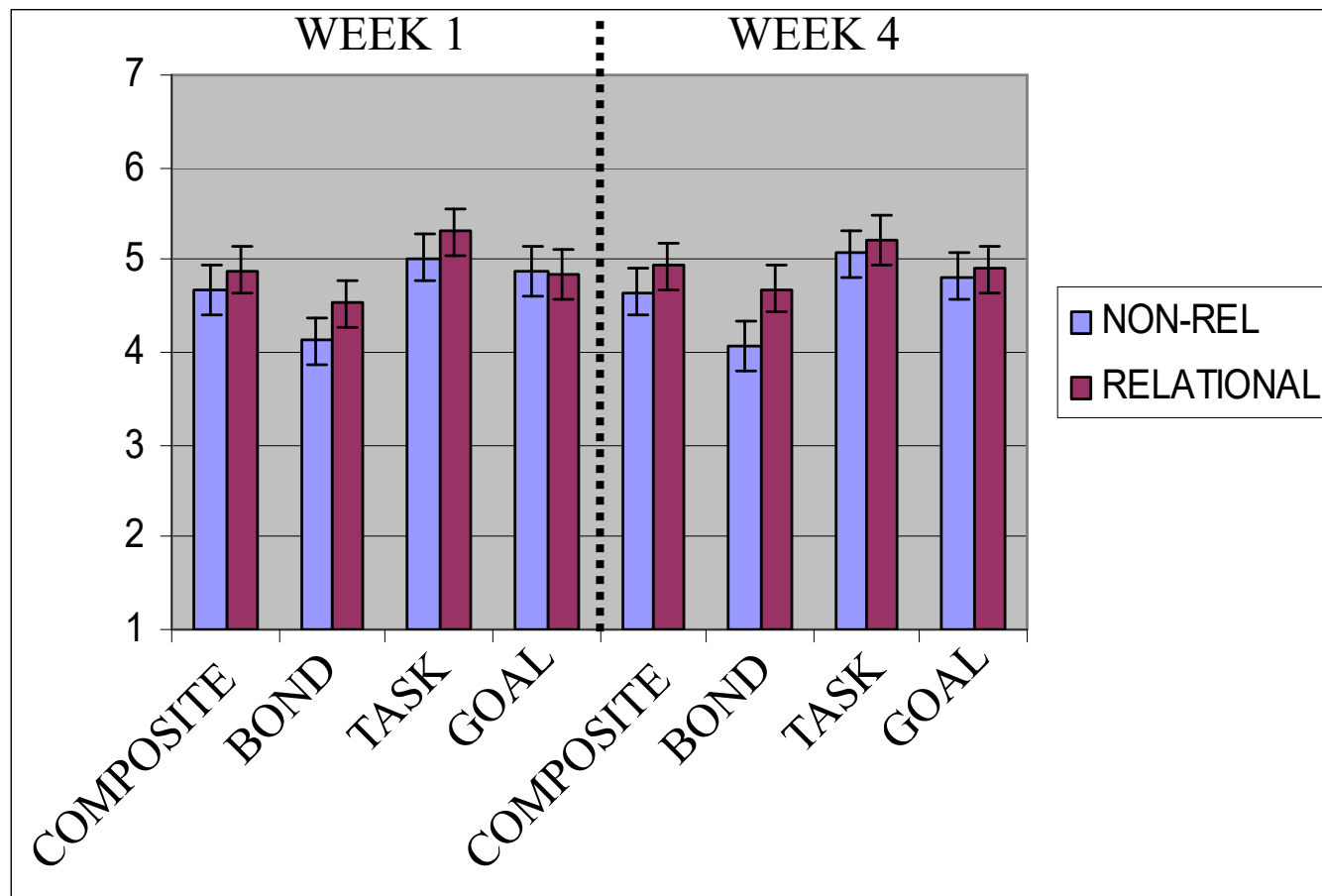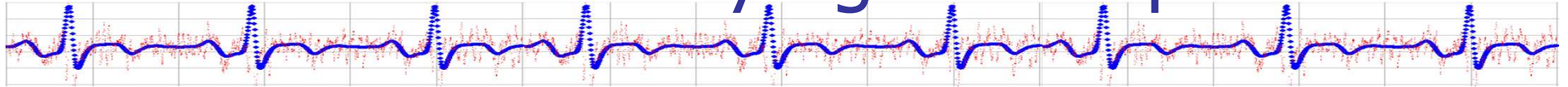
Oral Presentation Guide for Visuals

- Visuals should be *exhibits* that you talk about
  - Do not put lots of text on charts
  - Do not read your charts for your presentation
- Use ***relevant*** interactivity, video, images to keep your audience engaged

Do NOT…

- Put a lot of text on slides
- Use clip art or stock imagery
- Use complex images/tables without stepping the audience through them
- Include sloppy slides with typos
- Speak to the audience with your back to them

Common Questions
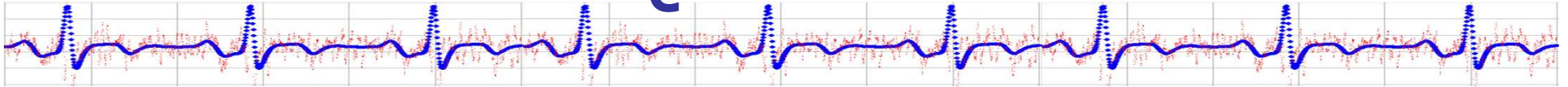
- How did you evaluate that?
- How did you measure that?
- How did you control for *extraneous variable X*?
- Why didn't you use statistic Y?
- Isn't that a biased sample?
- What was your control group?
- How did you do *study procedure Z?*

Outline for Descriptive Study Oral Presentation
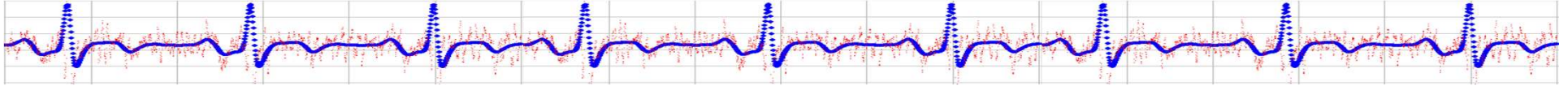
- **Motivation**

- **Research Questions**
  - Not hypotheses, unless meaningful Chi^2 GoF

- **Method**
  - Including exact text of any surveys

- **Results**
  - Demographics
  - Outcomes
  - Lots of visuals!

- **Conclusions / Discussion**

Tips
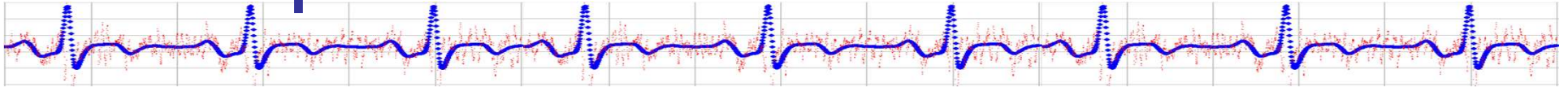
- Describe your sample
  - Minimal demographics – number of subjects, broken down by gender
  - Better: age, occupation, major, year
- Minimize text on your charts
- If you use a novel measure (e.g., new survey) you must give details on the measure
  - Actual questions asked
  - Any reliability/validity/psychometrics done
- If you do interviews, include actual quotes
- Build from data to conclusions
- Practice your timing/delivery with your project team