07. Security Warnings; Designing and Analyzing Quantitative Studies

Blase Ur

April 23rd, 2019

CMSC 23210 / 33210

THE UNIVERSITY OF CHICAGO

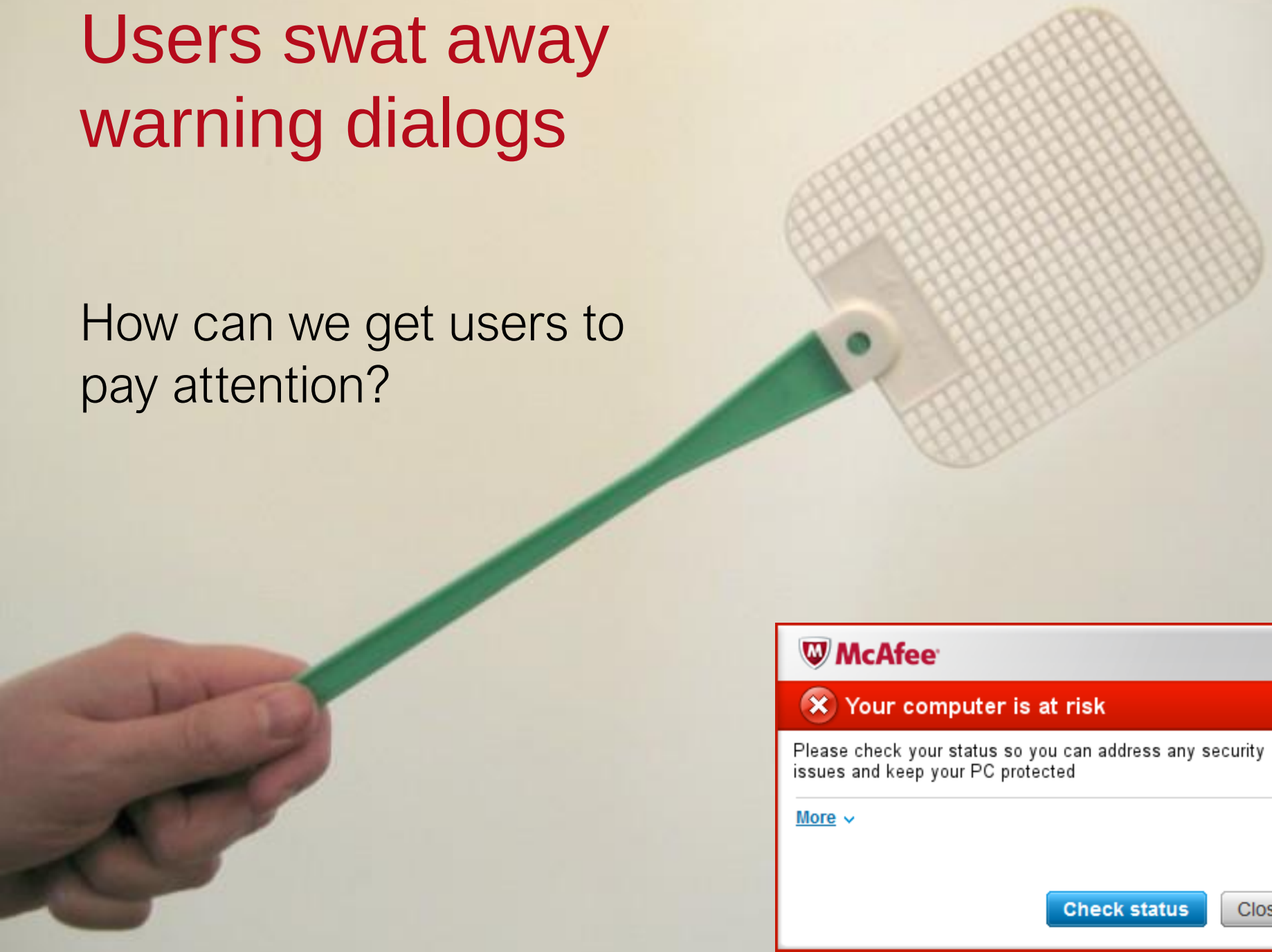
Security, Usability, & Privacy Education & Research

Security Warnings


Security Error: Domain Name Mismatch
Something happened and you need to click OK to get on with doing things.
Certificate mismatch security identification administrator communication intercept liliputian snotweasel foxtrot omegaforce.
Technical Crap
Cancel
OK

Users swat away warning dialogs

How can we get users to pay attention?



NEAT and SPRUCE (from Microsoft)

Rob Reeder, Ellen Cram Kowalczyk, and Adam Shostack. Poster: Helping engineers design NEAT security warnings. SOUPS 2011.

http://cups.cs.cmu.edu/soups/2011/posters/soups_posters-Reeder.pdf

- NEAT – 4 questions to ask when you design a security or privacy UX
- SPRUCE – 6 elements to include in a security or privacy UX
  - Good advice, but sometimes it may be better to keep it short and simple rather than include all 6 elements

Microsoft

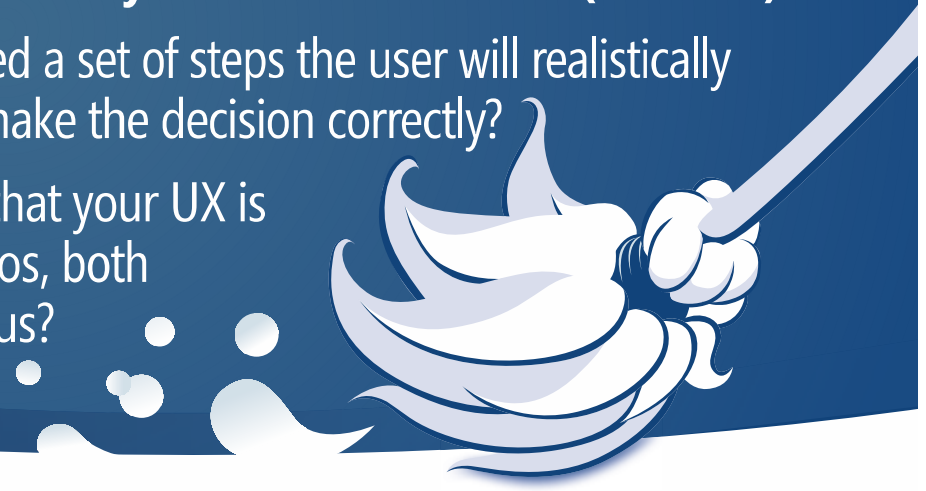
Ask yourself: Is your security or privacy UX:

NECESSARY? Can you change the architecture to eliminate or defer this user decision?

EXPLAINED? Does your UX present all the information the user needs to make this decision? **Have you followed SPRUCE? (see back)**

ACTIONABLE? Have you determined a set of steps the user will realistically be able to take to make the decision correctly?

TESTED? Have you checked that your UX is NEAT for all scenarios, both benign and malicious?



NEAT

When you involve the user in a NEAT security or privacy decision, explain the decision using these 6 elements:

SOURCE: State who or what is asking the user to make a decision

PROCESS: Give the user actionable steps to follow to make a good decision

RISK: Explain what bad thing could happen if the user makes the wrong decision

UNIQUE KNOWLEDGE user has: Tell the user what information they bring to the decision

CHOICES: List available options and clearly recommend one

EVIDENCE: Highlight information the user should factor in or exclude in making the decision

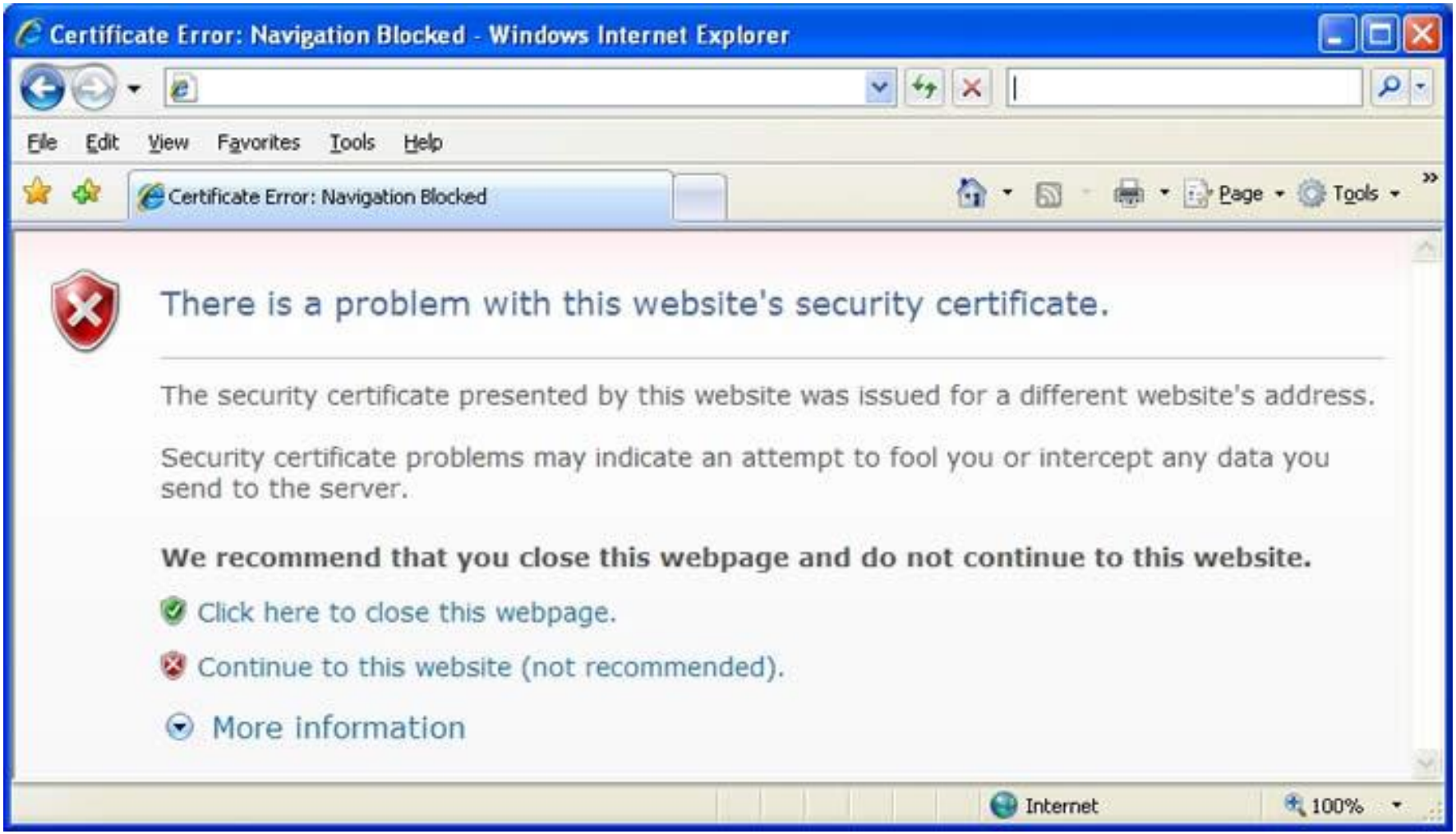
SPRUCE

For more info, contact **neatux@microsoft.com**

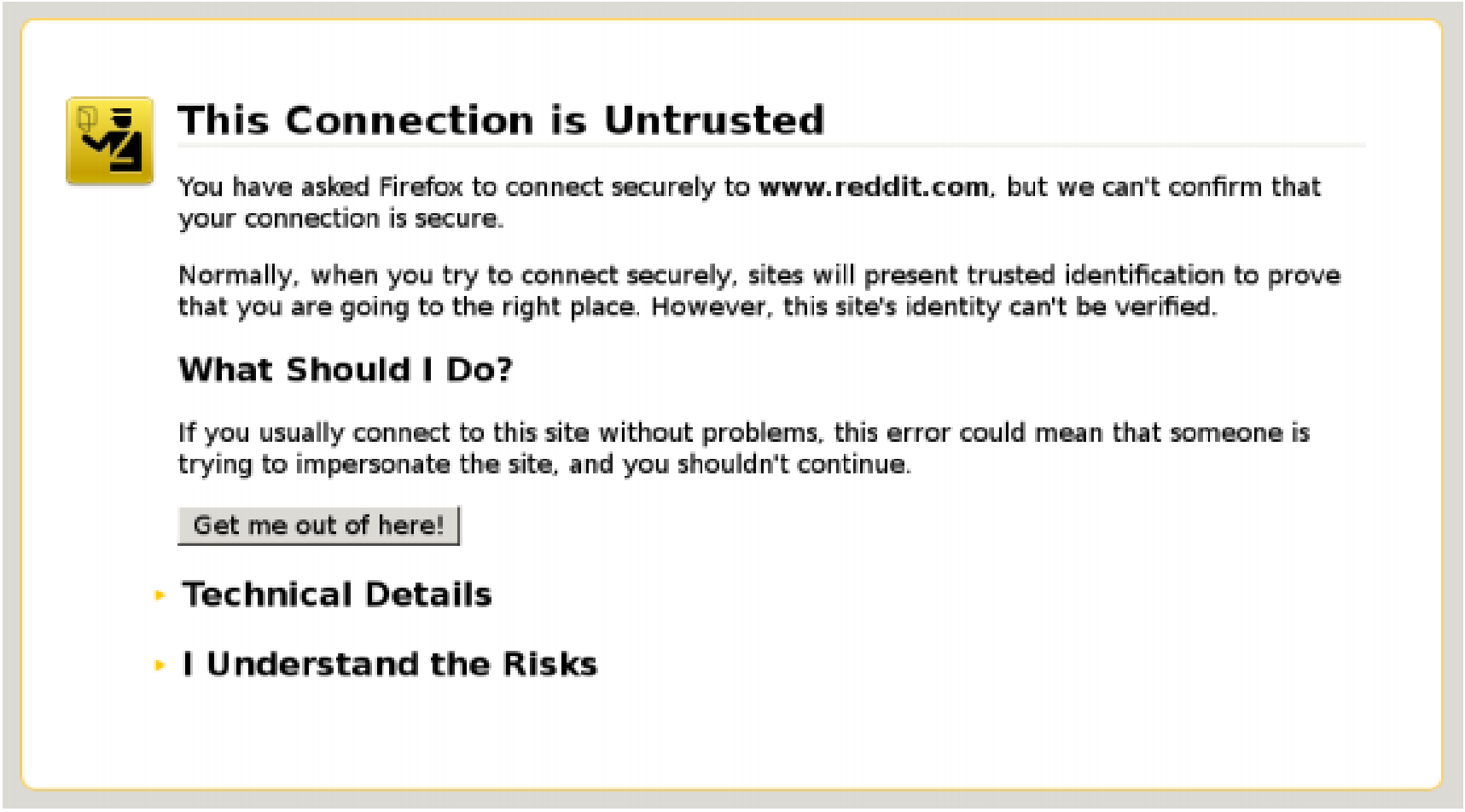
Alice in Warningland

Old Warning (IE 6)

Slightly Newer Warning (IE 7)

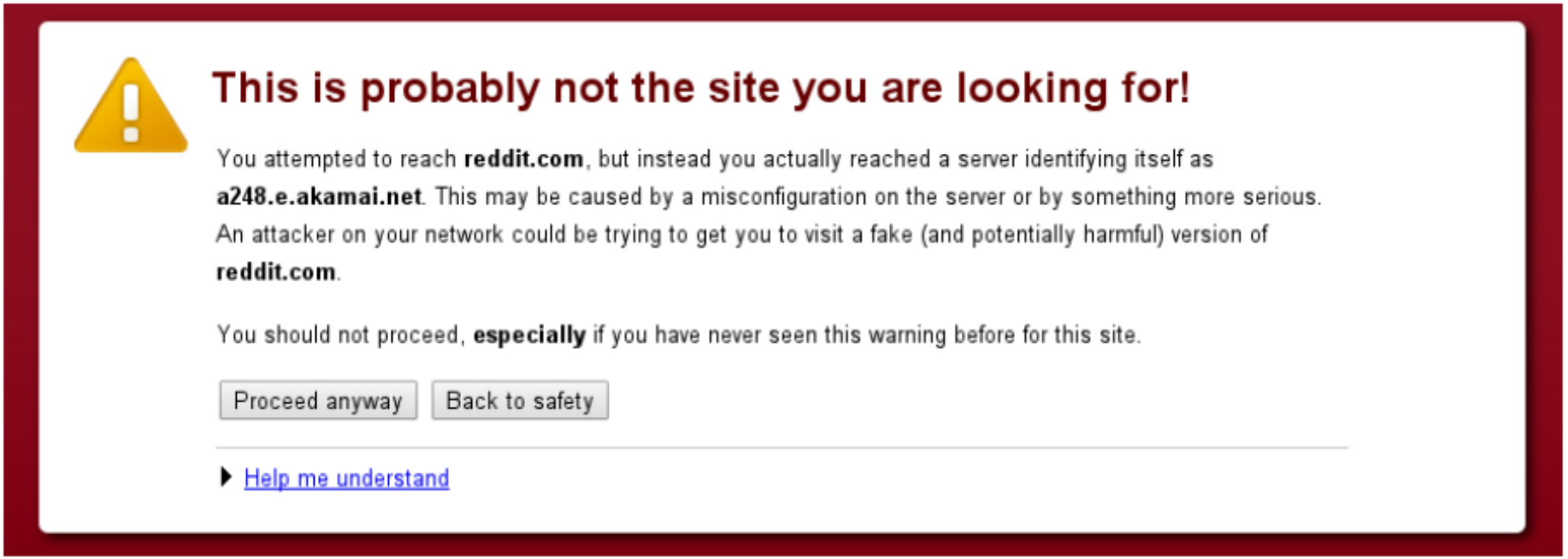


Newer Warning (Firefox)



Newer Warning (Firefox): Step 2

Newer Warning (Chrome)



Alice in Warningland takeaways

- Field study: correlation or causation?
- Is clicking through an SSL warnings always wrong?
  - Technically skilled users (e.g., Linux users) ignored warnings more often
- Comparison with lab studies
- Prior lab study using eye-tracking software

More takeaways

- Passive warnings vs. interstitial warnings
- Consent and ethics
- Sampling bias
- Dealing with noisy data
- Certificate pinning
- HSTS (HTTP Strict Transport Security)
  - Prevents protocol downgrade attacks

How do you know when you are actually at risk?

Some hazards are ALWAYS dangerous

Some hazards are context dependent

Computer security dialogs are context-dependent

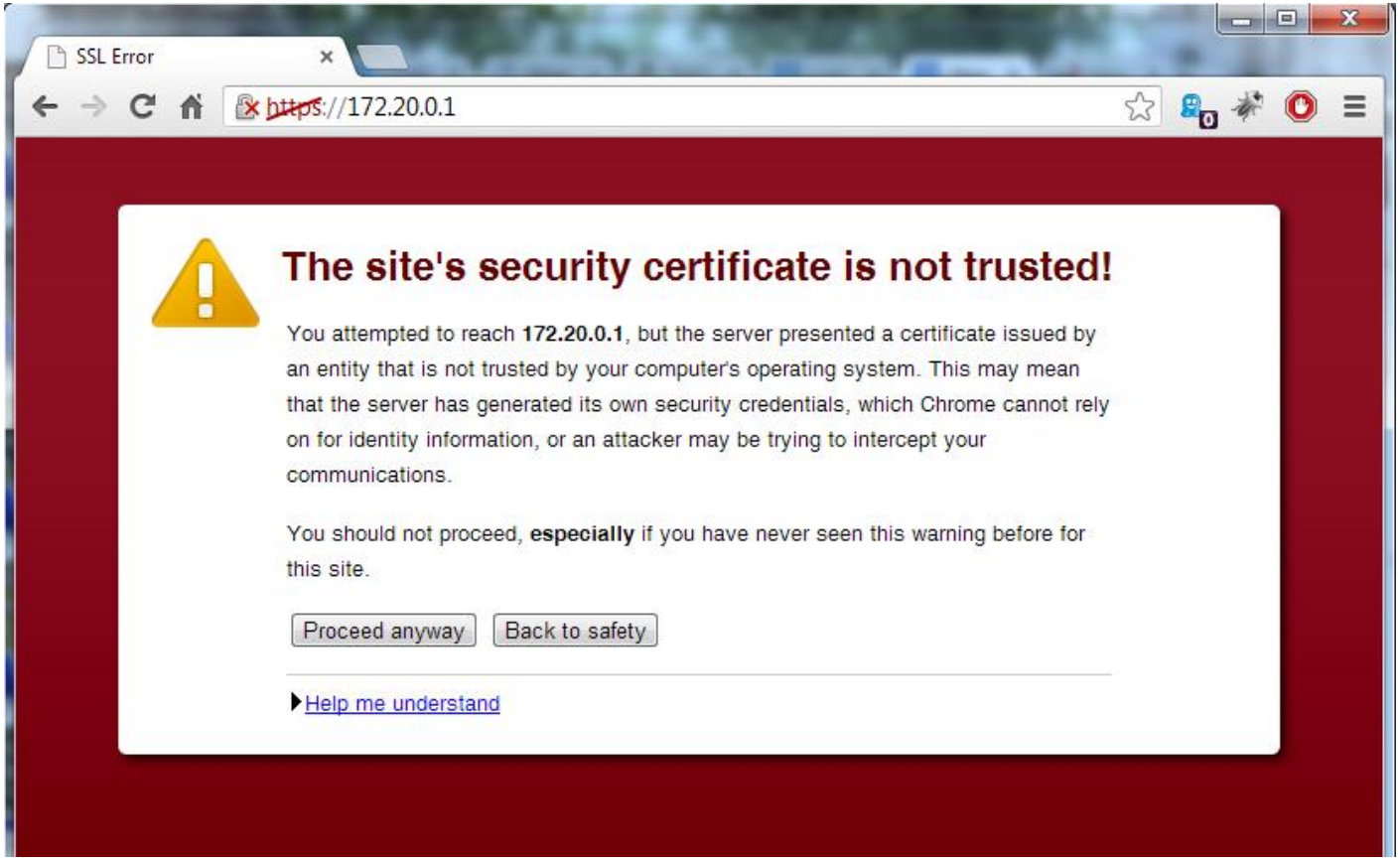
- Security warning dialogs more like warnings on wine than warnings on poison
- Software developers place burden of assessing risk on users

A good warning helps users determine whether they are at risk

- Stops users from doing something dangerous in risky context
- Doesn’t interfere with non-risky contexts
- Need to test warnings in both contexts

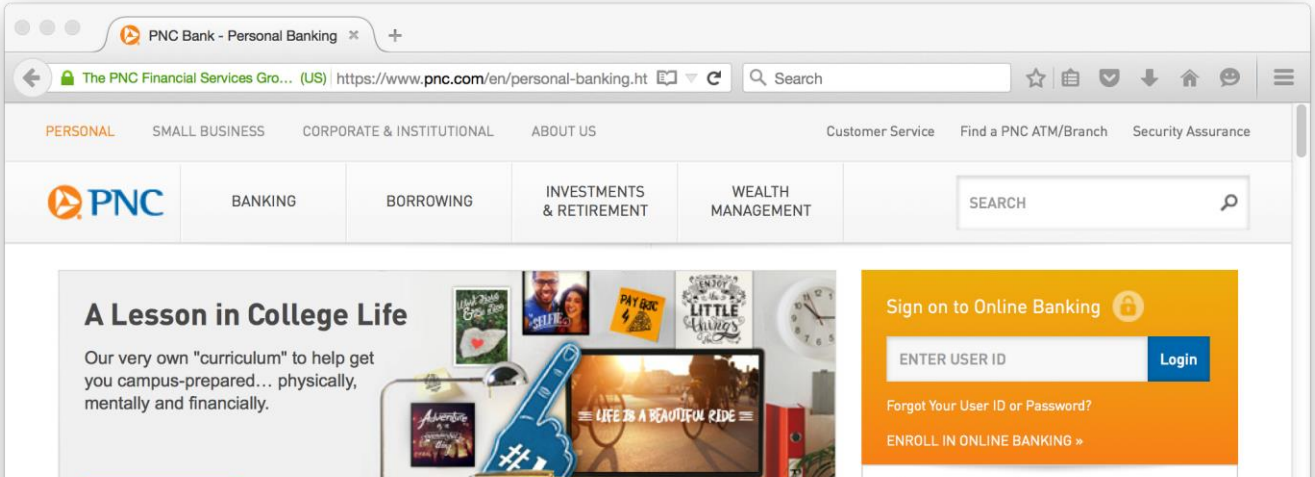
Non-risky context

- Encounter self-signed certificate (familiar experience for developers)



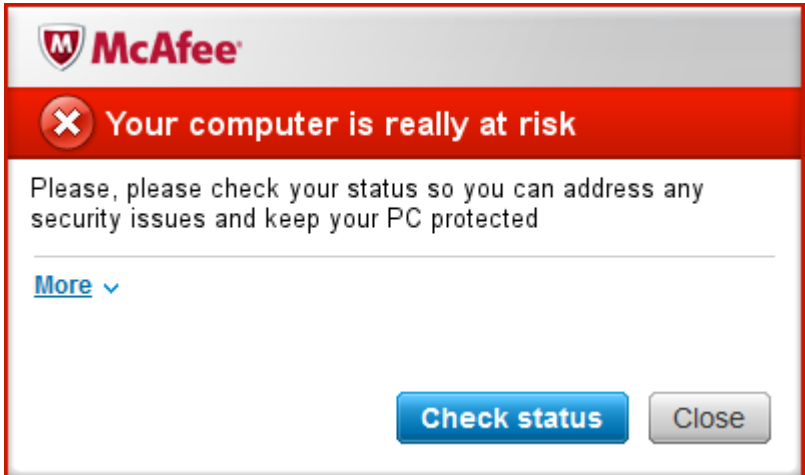
Prior study idea: Create risky context

- Put users in situation where they have something they care about at risk
  - Come to our lab and check bank account balance online
- Make users think they are actually at risk
  - Use web proxy to do man-in-the-middle attack



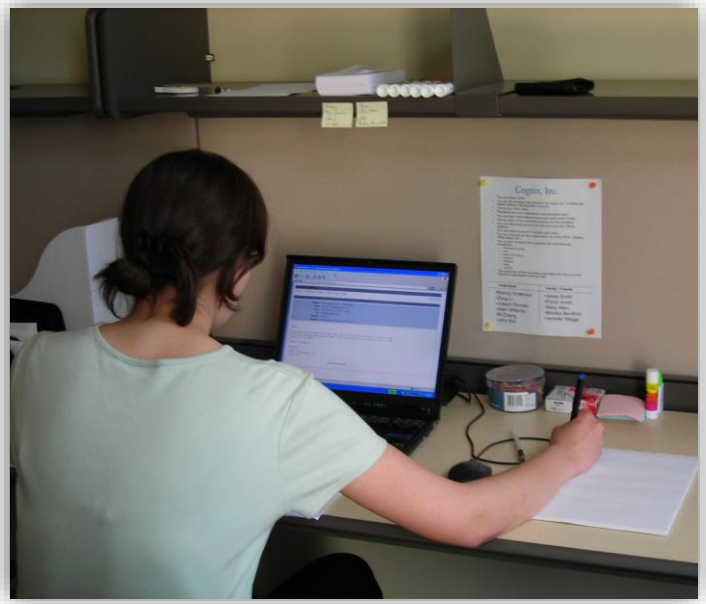
Revised plan for former study

- Remove root certificate from browser
- Web site certificates can’t be verified
- Visits to secure sites will trigger warnings



Lab study challenges

- Participants may feel safe
- They may think they have to do everything we tell them
- Their priority may be to finish study fast and get paid

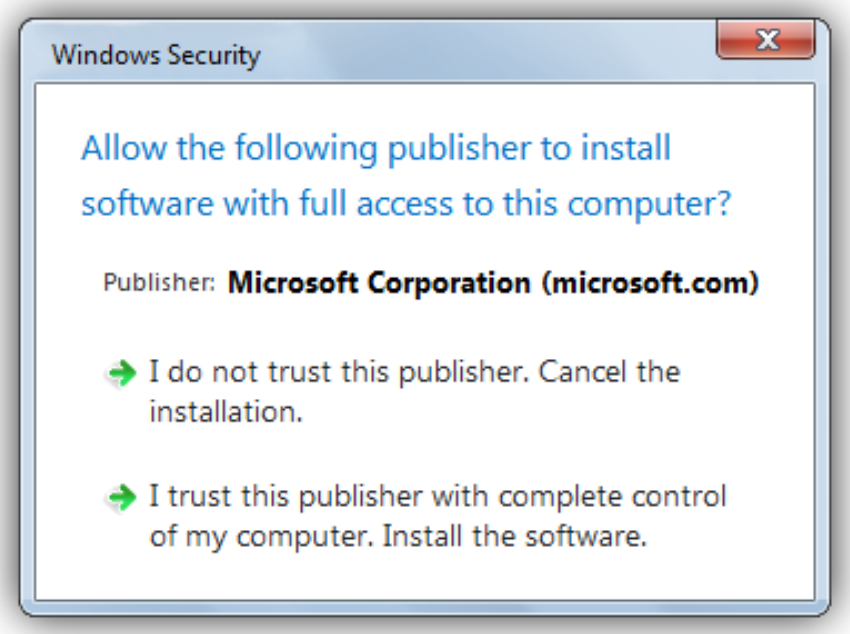


Security-decision UI study

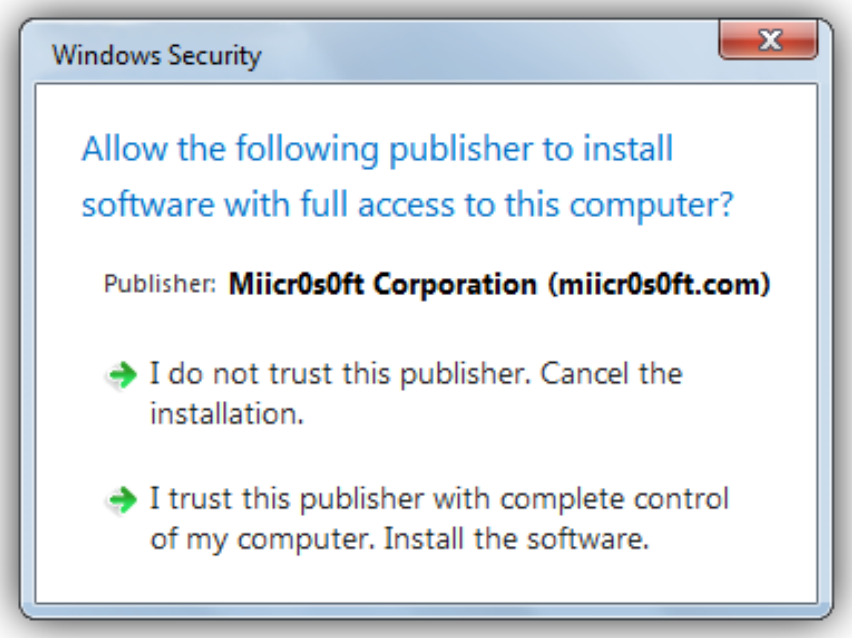
- How can we focus users' attention on key information they need to make informed decisions?

C. Bravo-Lillo, L.F. Cranor, J. Downs, S. Komanduri, R.W. Reeder, S. Schechter, and M. Sleeper. **Your Attention Please: Designing security-decision UIs to make genuine risks harder to ignore**. SOUPS 2013

Can you spot the suspicious software?



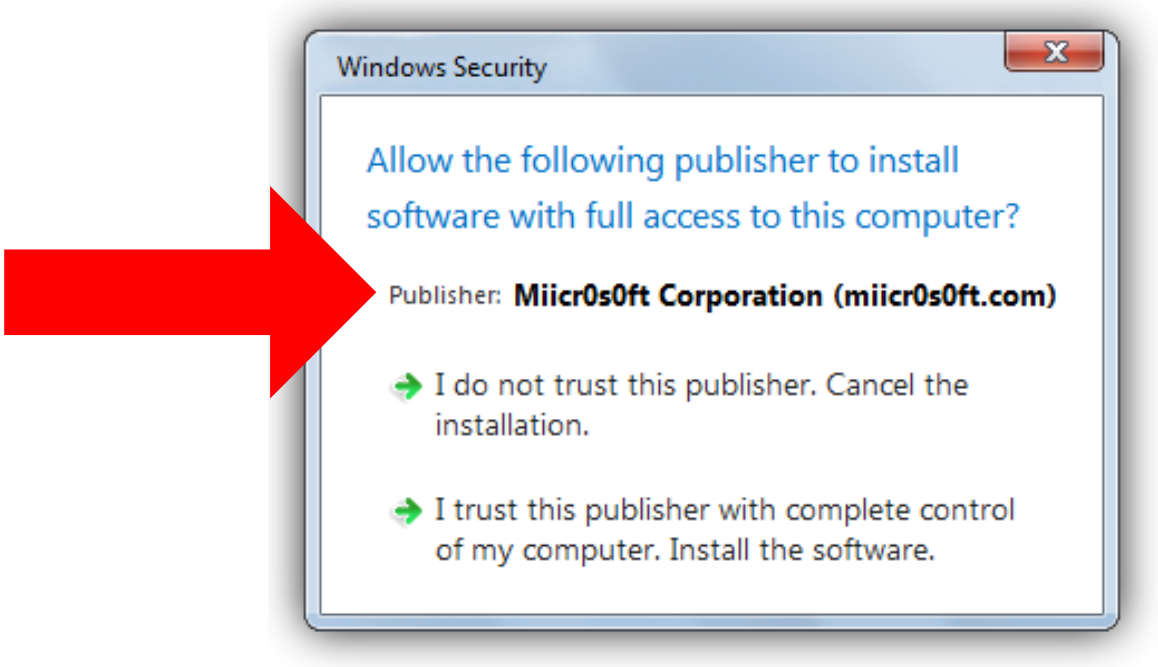
benign



suspicious

Key question: Do you trust publisher?

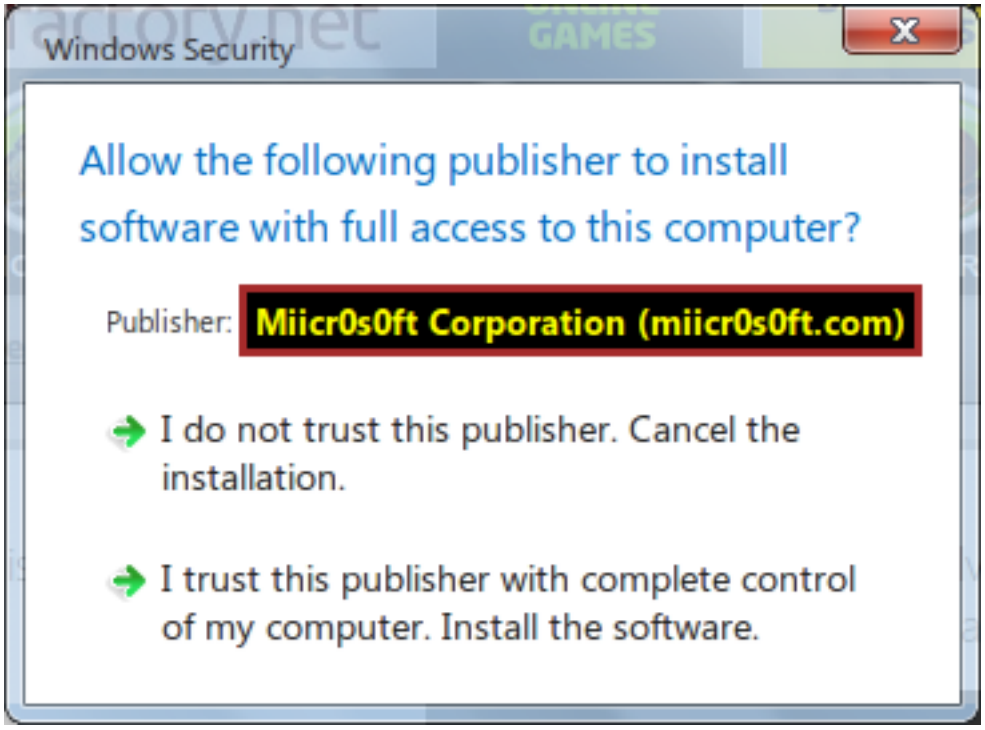
Name of publisher is critical information in trust decision



How can we get users to notice suspicious publishers?

- Use **attractors** to draw attention to publisher name
- Force delay before users can install
- Force interaction before users can install
- Force users to read publisher name

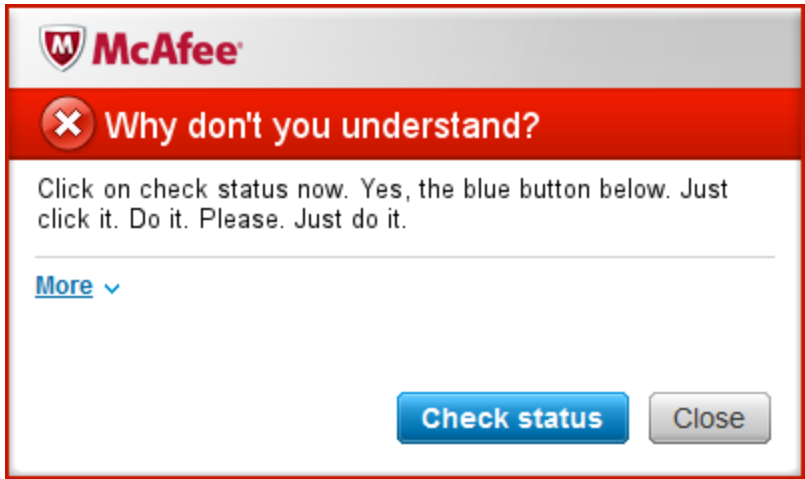
ANSI standard warning colors



Obstruct install button until user types publisher name

Do any of these work?

- Do attractors and other techniques prevent suspicious installs without preventing benign installs?
- How much do attractors delay benign installs?

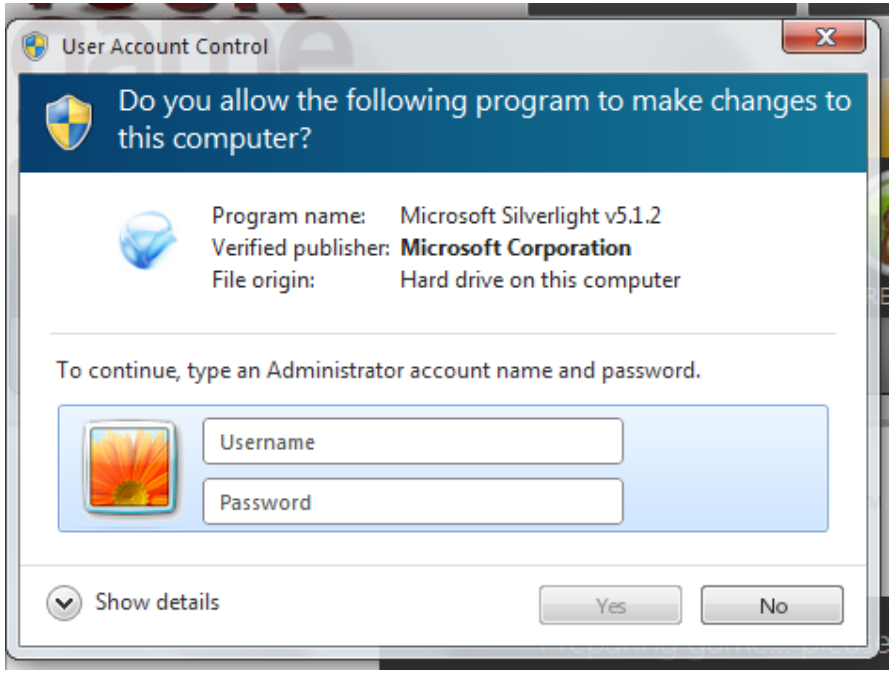


Methodology requirements

- Massive, inexpensive, quick
- Remote observation/recording of behavior
- Participants should feel safety/risk and behave as they would in real life
- But should not actually be at increased risk through participation in experiment

Use Mturk game ruse

- Ruse previously developed for study of whether users would fall for fake OS password dialogs



Operating System Framed in Case of Mistaken Identity: Measuring the success of web-based spoofing attacks on OS password-entry

Online games evaluation survey

Online games evaluation survey

Carnegie Mellon U

Purpose of the study

This survey is part of a research study conducted by Dr. Julie Downs at Carnegie Mellon University. The purpose of this study is to evaluate online games according to criteria that will be explained in the next pages. You will be asked to go to websites, play a game for 2 to 3 minutes, then return to this survey to give us your opinion on each. The whole survey should take you between 15 and 20 minutes in total.

Participants requirements

Participation in this study is limited to individuals age 18 and older. **You have to physically be in the United States of America to be eligible to participate in this study, and not having taken before any early version of the same survey**.

Risks, benefits, and compensation

The risks and discomfort associated with participation in this study are no greater than those ordinarily encountered in daily life or during other online activities. There may be no personal benefit from your participation in the study but the knowledge received may be of value to humanity. You will receive $1.00 as a compensation for participation in this study. There will be no cost to you if you participate in this study.

The data captured for the research does not include any personally identifiable information about you. We will collect your IP address only to check whether you qualify for the study.

Confidentiality

By participating in this research, you understand and agree that Carnegie Mellon may be required to disclose your consent form, data and other personally identifiable information as required by law, regulation, subpoena or court order. Otherwise, your confidentiality will be maintained in the

Amazon Mechanical Turk | Carnegie Mellon University

saucers.cups.cs.cmu.edu/yacot/mnt/wtk/survey/index.php?t=1&i=A2NUXAJFPAX4Z2

This is a test version of the CMU Online Games Evaluation Study. You are currently using Microsoft Windows 7.

Online games evaluation survey

Instructions to evaluate the game:

1. Click on the
   game to loa
2. When the g
3. Return to this survey to answer the questions below.

Assigned game #1: Mars Buggy Online
http://www.gametop.com/online-free-games/mars-buggy-online/?i=A2NUXAJFPAX4Z2

Attention: The website whose URL appears above is external to this study. Our researchers **do not** control its content.

1. Were you able

- Yes
- No (you w

Next

Assigned game #1: Mars Buggy Online

Attention: The website whose URL appears above is external to this study. Our researchers do not control its contents

Carnegie Mellon ... | sc08 - Paint | EN | 9:10 PM 10/9/2012

need to be rescued.

Play this free online game today and bring your crew back to earth.

Do you like this game? Tweet

Mars Buggy

Amazon Mechanical Turk | Carnegie Mellon University

saucers.cups.cs.cmu.edu/yacot/mnt/wtk/survey/index.php?t=1&i=A2NUXAJFPAX4Z2

This is a test version of the CMU Online Games Evaluation Study. You are currently using Microsoft Windows 7.

1. Were you able to play the game? *

- (•) Yes
- ( ) No **(you will be assigned another game to evaluate)**

Pleas

A

Please answer th

Have you ever

Do you think t

Did the game ha

- ( ) Yes (pleas
- ( ) No

Were you able to play the game?

- O Yes
- O No (you will be assigned another game to evaluate)

Please enter a one-sentence description of the game you played

Have you ever played this game before?

Do you think this game is fun?

Carnegie Mellon ... | sc11 - Paint | EN | 9:13 PM 10/9/2012

Have you ever played this game before?	○	◉
Do you think this game is fun?	◉	○

Did the game have any visual glitches, such as stalls in animations or overlapping windows, when running on your computer/browser? *

- ○ Yes (please explain briefly) [ ] *
- ◉ No

Did you see any

- ○ Yes (pleas
- ◉ No

Was there any other aspect of the game you thought could have been improved?

Was there any other aspect of the game that you thought could have been improved? *

- ○ Yes (please explain briefly) [ ] *
- ◉ No

Next

Amazon Mechanical Turk | Carnegie Mellon University

saucers.cups.cs.cmu.edu/yacot/mnt/wtk/survey/index.php?t=1&i=A2NUXAJFPAX4Z2

This is a test version of the CMU Online Games Evaluation Study. You are currently using Microsoft Windows 7.

Online games evaluation survey

Assigned game #2: Tom and Jerry Refrigerator Raid Game

Instructions to e

1. Click on the
2. Wait for the
3. Return to this survey to answer the questions below.

Assigned game #2: Tom and Jerry Refrigerator Raid Game
http://www.free-online-games-to-play.net/games/kidsgames/onlineflashgame/751/?i=A2NUXAJFPAX4Z2

Attention: The website whose URL appears above is external to this study. Our researchers **do not** control its content.

2. Were you able to play the game? *

- Yes
- No **(you will be assigned another game to evaluate)**

Next

Carnegie Mellon ... | sc13 - Paint | EN | 9:14 PM 10/9/2012

Add to Favorites

Tom and Jerry Refrigerator Raid Game

stars (3973)

2. Were you able to play the game? *

- (●) Yes
- ( ) No **(you will be assigned another game to evaluate)**

Please enter here a one-sentence description of the game you played (between 10 and 50 words): *

A boring Tom-and-Jerry game, may be fun for kids.

Please answer the following questions about the game you played: *

	Yes	No
Have you ever played this game before?	( )	(●)
Do you think this game is fun?	( )	(●)

Did the game have any visual glitches, such as stalls in animations or overlapping windows, when running on your computer/browser? *

- ( ) Yes (please explain briefly) [ ] *
- ( ) No

Online games evaluation survey

Assigned game #3: Colliderix Level Pack

Instructions to e

1. Click on the
2. Wait for the game to load. When it's fully loaded, play the game "Colliderix Level Pack" for about 2 to 3 minutes.
3. Return to this survey to answer the questions below.

Assigned game #3: Colliderix Level Pack
http://www.yourgamefactory.net/wtk/games/index.u1.php?i=A2NUXAJFPAX4Z2

Attention: The website whose URL appears above is external to this study. Our researchers **do not** control its content.

4. Were you able to play the game? *

- Yes
- No **(you will be assigned another game to evaluate)**

Next

YOUR game factory.net

ADD TO FAVORITES | SET AS HOMEPAGE

Username ••••••• Login

FORGOT PASSWORD? SIGN UP

ONLINE GAMES | DOWNLOAD GAMES (FREE) | GAME CLUB | MMORPG GAMES | MULTIPLAYER GAMES

SHOOTING | RACING | PUZZLE | ACTION | SPORT | DRESS UP | KIDS | CLASSIC | BOARD | MISC | NEW

Games / Puzzle Games / Colliderix Level Pack

Search...

This game requires the latest version of Microsoft Silverlight™ (v5.1.2). Silverlight is either missing or out of date.

Access being requested, please wait.

Description: Beloved Colliderix is back, equipped with levels that will break your mind!

Rate it:

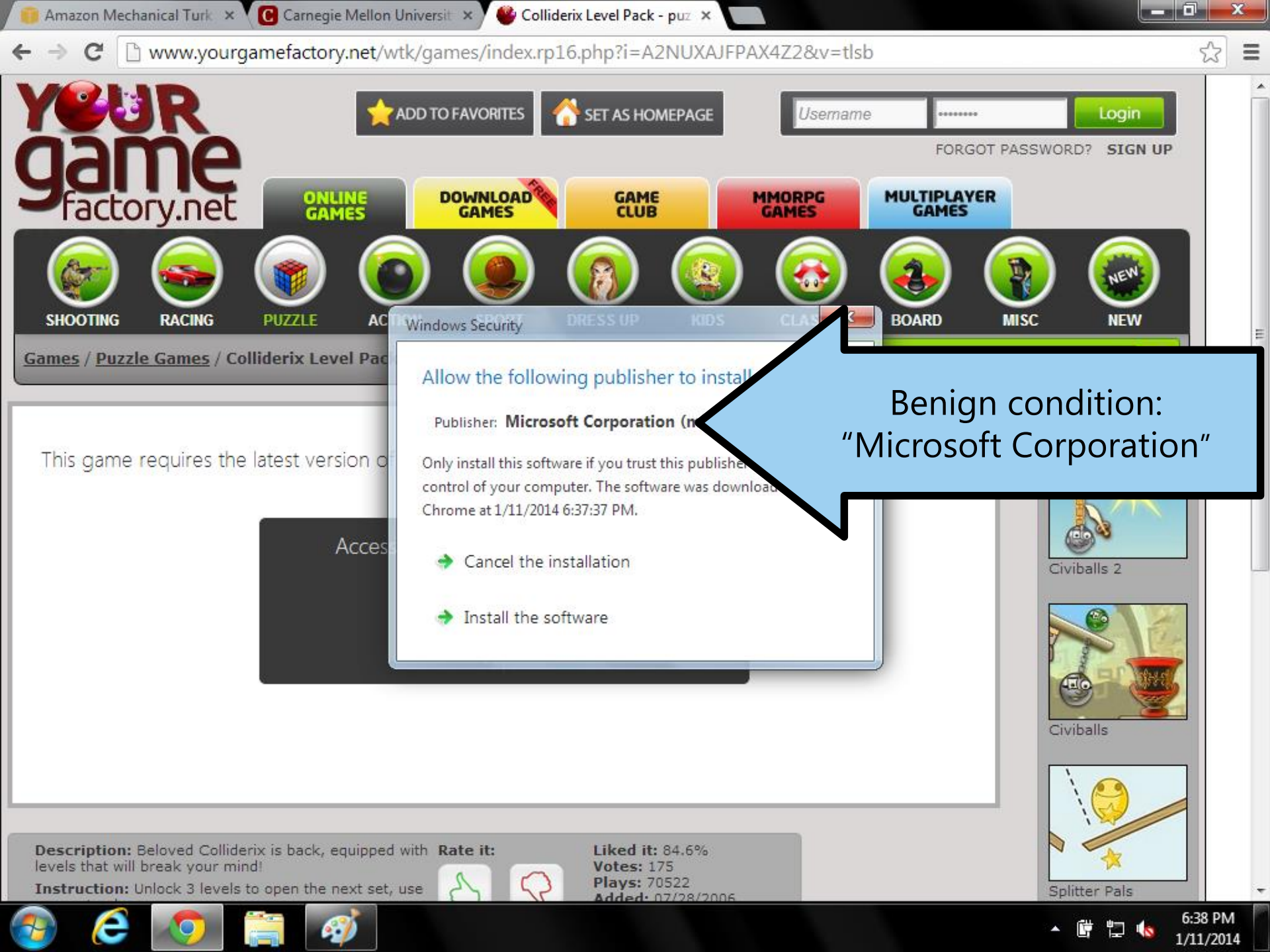
Liked it: 84.6%
Votes: 175
Plays: 70522
Added: 07/28/2006

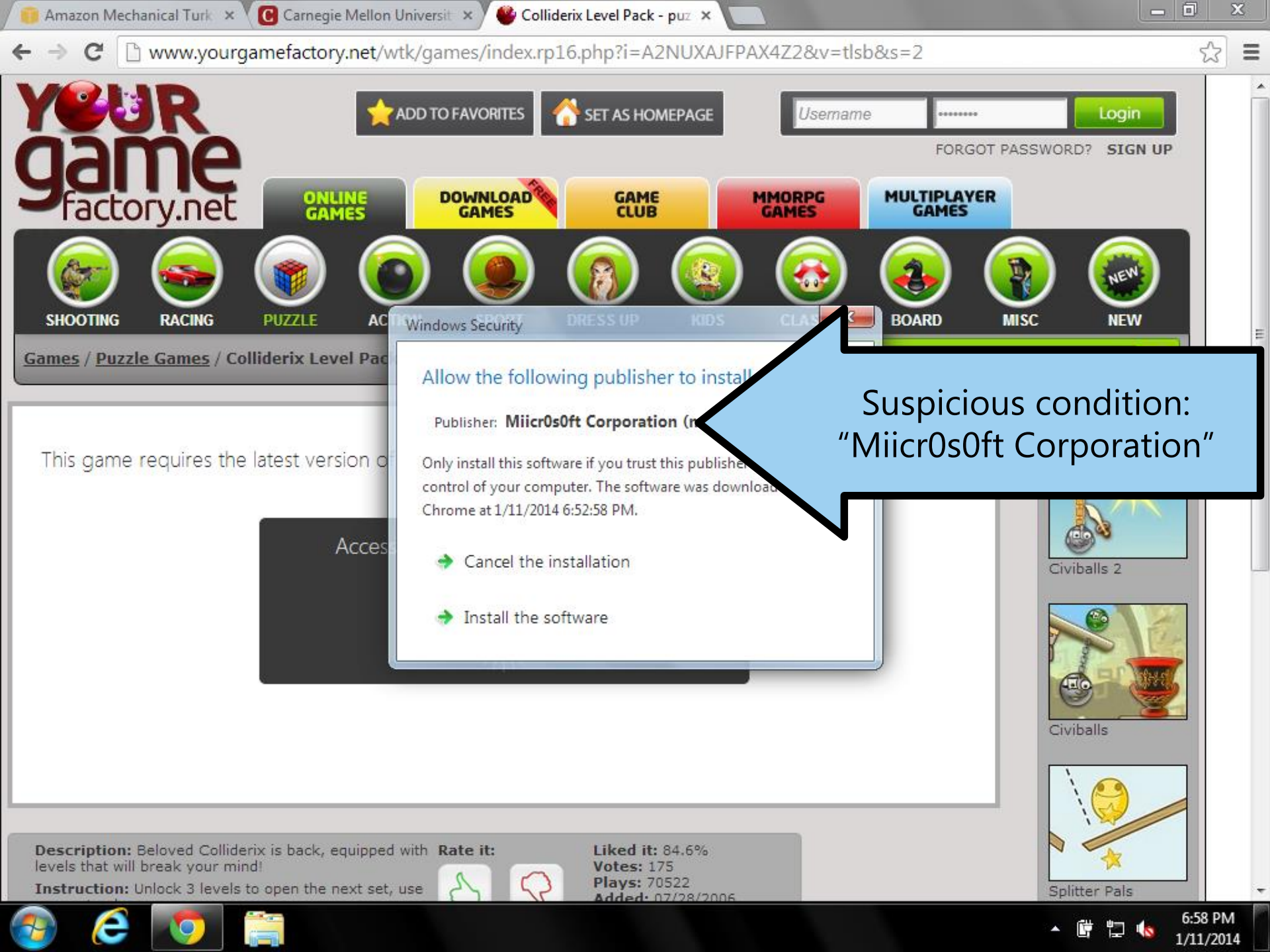
Related Games


Civiballs 2


Civiballs


Splitter Pals


Amazon Mechanical Turk
Carnegie Mellon Universit
Colliderix Level Pack - puz
www.yourgamefactory.net/wtk/games/index.rp16.php?i=A2NUXAJFPAX4Z2&v=tlsb
ADD TO FAVORITES
SET AS HOMEPAGE
Username
Login
FORGOT PASSWORD?
SIGN UP
ONLINE GAMES
DOWNLOAD GAMES
GAME CLUB
MMORPG GAMES
MULTIPLAYER GAMES
SHOOTING
RACING
PUZZLE
BOARD
MISC
NEW
Games / Puzzle Games / Colliderix Level Pac
Windows Security
Allow the following publisher to install
Publisher: Microsoft Corporation
Only install this software if you trust this publisher
control of your computer. The software was downloaded
Chrome at 1/11/2014 6:37:37 PM.
Cancel the installation
Install the software
Benign condition:
"Microsoft Corporation"
This game requires the latest version
Civiballs 2
Civiballs
Splitter Pals
Description: Beloved Colliderix is back, equipped with levels that will break your mind!
Instruction: Unlock 3 levels to open the next set, use
Rate it:
Liked it: 84.6%
Votes: 175
Plays: 70522
Added: 07/28/2006
6:38 PM
1/11/2014


Amazon Mechanical Turk
Carnegie Mellon Universit
Colliderix Level Pack - puz
www.yourgamefactory.net/wtk/games/index.rp16.php?i=A2NUXAJFPAX4Z2&v=tlsb&s=2
YOUR game factory.net
ADD TO FAVORITES
SET AS HOMEPAGE
Username
Login
FORGOT PASSWORD?
SIGN UP
ONLINE GAMES
DOWNLOAD GAMES
GAME CLUB
MMORPG GAMES
MULTIPLAYER GAMES
SHOOTING
RACING
PUZZLE
BOARD
MISC
NEW
Games / Puzzle Games / Colliderix Level Pac
Windows Security
Allow the following publisher to install
Publisher: Miicr0s0ft Corporation
Only install this software if you trust this publisher
control of your computer. The software was downloaded
Chrome at 1/11/2014 6:52:58 PM.
Cancel the installation
Install the software
Suspicious condition:
“Miicr0s0ft Corporation”
This game requires the latest version
Civiballs 2
Civiballs
Splitter Pals
Description: Beloved Colliderix is back, equipped with levels that will break your mind!
Instruction: Unlock 3 levels to open the next set, use
Rate it:
Liked it: 84.6%
Votes: 175
Plays: 70522
6:58 PM
1/11/2014

Participant decision design

- Workers in Amazon's Mechanical Turk aim to:
  - Complete the tasks they accept (otherwise, don't earn money)
  - Minimize the time and effort in each task (each accepted task has an opportunity cost)
- Our message to participants:
  - "You may skip a game. If you do, we will assign you another"
- The decision was designed to gamble time/money for security:
  - Install → Take small risk, play the game, finish sooner
  - Not install → Not take any risks, not play the game, waste time

Results are encouraging

- 2,227 participants encountered dialogs
- Benign scenario
  - Installation not prevented
  - But some approaches slowed people down
- Suspicious scenario
  - Our new dialogs reduced installations
  - Swipe, type, and delay were particularly effective

Debrief is crucial for ethics!

- Explain what the actual purpose of the study was

Statistics!

- The main idea and building blocks
- Major tests you'll see
- Non-independent data

Important Note

- In some cases in discussing stats, we will intentionally be imprecise (and sometimes not technically accurate) about certain concepts. We are trying to give you some intuition for these concepts without extensive formal background.

BUILDING BLOCKS

Statistics

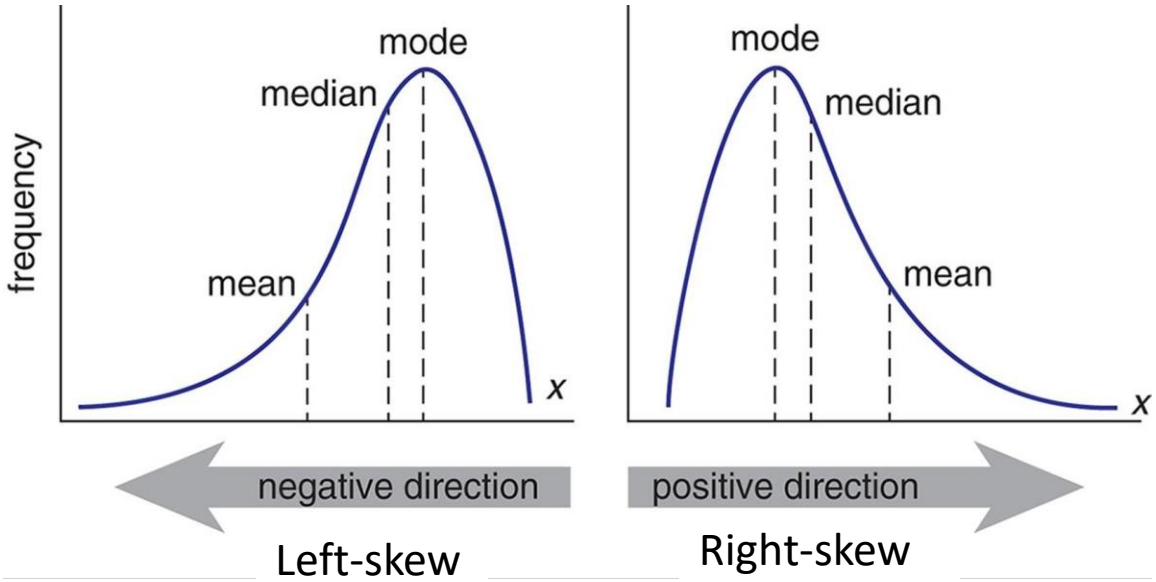
- In general: analyzing and interpreting data
- Statistical hypothesis testing: is it unlikely the data would look like this unless there is actually a difference in real life?
- Statistical correlations: are these things related?

What kind of data do you have?

- Quantitative
  - Discrete
  - Continuous
- Categorical
  - Nominal (no order)
  - Ordinal (ordered)

Exploratory Data Analysis (EDA)

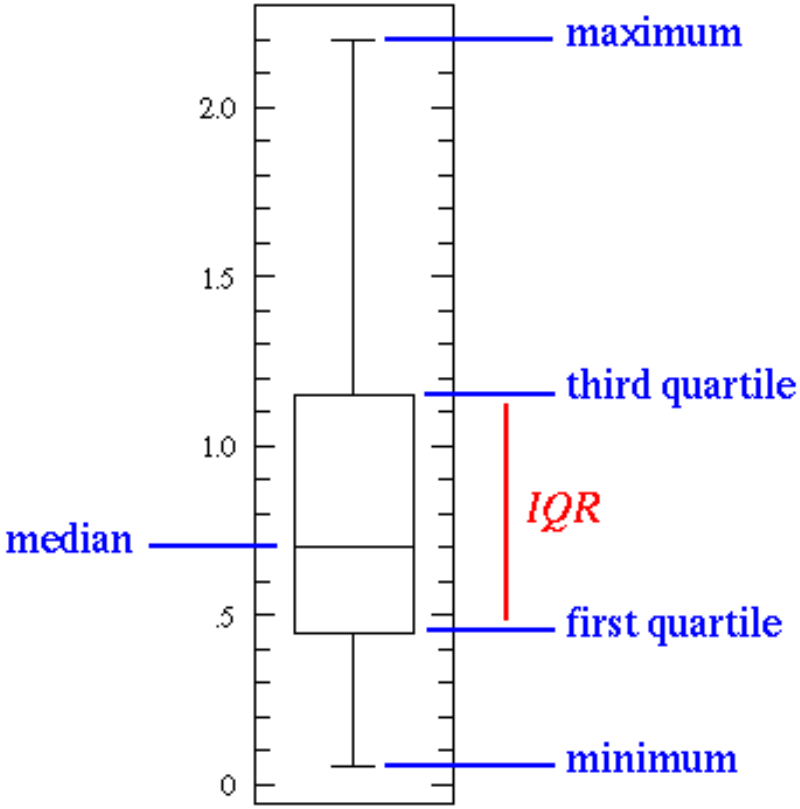
- Shape
- Center
  - Mean
  - Median
  - Mode



http://schoolbag.info/physics/physics_math/physics_math.files/image790.jpg

EDA Continued

- Spread
  - Standard Deviation
  - Variance
  - Interquartile range



http://www.physics.csbsju.edu/stats/simple.box.defs.gif

Hypothesis testing

- **Causation** (X causes Y)
  - vs. **correlation** (X is related to Y)
- Develop a hypothesis
  - Assign to conditions (include a **control**)
  - Terminology: “Condition” = “Treatment”
- H_0 (null hypothesis): there is no effect
- H_A or H_1 (alternative hypothesis): there is an effect

Hypothesis testing variables

- Independent variables: the thing(s) you assign / vary
- Dependent variables: the thing(s) you measure for evidence of an effect
- Co-variates: other aspects of a participant that might explain some of the effect (e.g., age, technical expertise, etc.)

P values and statistics

- Much of hypothesis testing involves calculating an appropriate statistic
- p value: probability of observing an effect at least as extreme as observed assuming the null hypothesis is true (i.e., no effect)
- α (alpha): cutoff for rejecting H_0
  - Treat this as a binary decision
  - Often $\alpha = .05$ in usable security

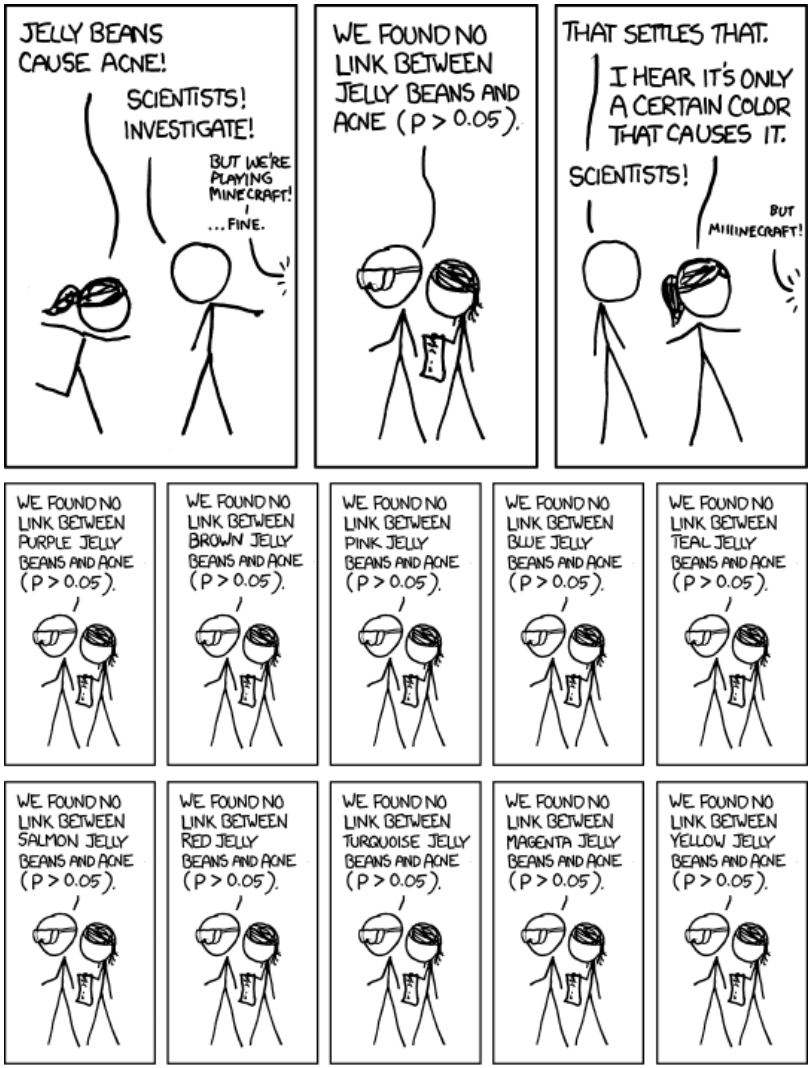
Is testing for significance enough?

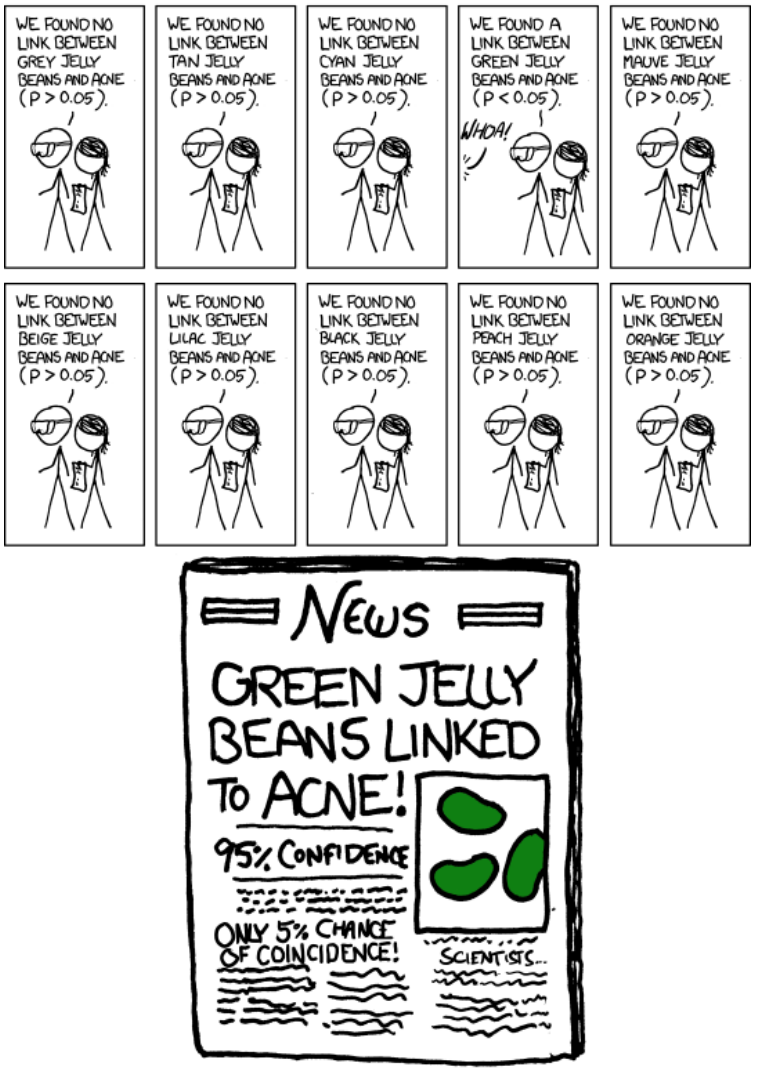
- No! Consider:
  - Effect size (magnitude of the effect of the manipulation)
  - Power (long-term probability of rejecting H_0 if there really is a difference)
- **Type 1 error: wrongly reject H_0 even if there is** no effect (**α**)
- Type 2 error: wrongly <u>fail</u> to reject **H_0** even if there is an effect (**β**)

Type I Errors

- Type I error (false positive)
  - You would expect this to happen 5% of the time if $\alpha = 0.05$
- What happens if you conduct a lot of statistical tests in one experiment?

Contrasts





https://xkcd.com/882/

Contrasts

- If we determine that the variables are dependent, we may compare conditions
- Planned vs. unplanned **contrasts**
  - You have a limited number of planned contrasts (depending on the DF) for which you don't need to correct p values.
- Bonferroni correction (multiply p values by the number of tests) is the easiest to calculate but most conservative

Type II Errors

- Type II error (false negative)
  - There is actually a difference, but you didn't see evidence of a difference
- Statistical power is the probability of rejecting the null hypothesis if you should
  - You could do a **power analysis**, but this requires that you estimate the effect size

PICKING THE RIGHT TEST

Not all tests are created equal

Different types of dependent and independent variables?

- Different tests!

Different data distributions?

- Different **assumptions**

→Different tests!!

Parametric vs non-parametric

Which tests are we learning about today?

Focusing on parametric tests!

		Independent Variable	
		Categorical	**Quantitative**
Dependent Variable	**Categorical**	Chi-Squared Test Fisher's Exact Test	Logistic Regression
	Quantitative	t-Test ANOVA	Correlation Linear Regression

Independence

- Why might your data not be independent?
  - Non-independent sample (bad!)
  - The inherent design of the experiment (ok!)
- If you have two data points of unicorns' race completion times (before and after some treatment), can you actually do a single test that assumes independence to compare conditions?

Picking a test

- http://webspace.ship.edu/pgmarr/Geo441/Statistical%20Test%20Flow%20Chart.pdf
- http://abacus.bates.edu/~ganderso/biology/resources/statistics.html
- http://med.cmb.ac.lk/SMJ/VOLUME%203%20DOWNLOADS/Page%2033-37%20-%20Choosing%20the%20correct%20statistical%20test%20made%20easy.pdf

What can we conclude statistically

- X varies in a way that's related to Y
  - As the age of a unicorn increases, its max speed decreases
  - Pearson's correlation / Spearman's correlation
- Assignment to X impacts Y (category)
  - Unicorns randomly assigned to eat vegan food (as opposed to non-vegan food) are more likely to be rated as successful (as opposed to unsuccessful)
  - χ^2, Fisher's exact test

What can we conclude statistically

- Assignment to X impacts Y (numerical)
  - Unicorns randomly assigned to eat vegan food (as opposed to non-vegan food) are more likely to take a shorter time to run a race
  - ANOVA, Kruskal-Wallis, etc.
- Lots of factors impact Y (category)
  - Logistic regression
- Lots of factors impact Y (numerical)
  - Regression