Research Design

No Readings

Choosing a research design

1. Know your goal …
2. Divide your goal into measurable questions …
3. Identify behaviors to investigate …
4. Identify potentially important variables …
5. Identify commonly used methods for measuring variables (if they exist) …

Background: Maturity in scientific fields

- New scientific fields/areas: **Exploratory data collection and analysis**
  - Classifying behaviors, identifying potentially important variables, exploring
  - Example: User Experience in Games

- Established fields: **Hypothesis testing**: evaluating potential explanations for observed relationships
  - Example: monetization in F2P

Eye Tracking experiment

- A Dynamic Lighting System can be used to dynamically channel visual attention
- Non-gamers' do not spot enemies fast enough to respond
- Dynamic Lighting system will enable non-gamers to respond faster – better gameplay

Was this exploratory or hypothesis based?

Experiment

- Participants:
  - 26 students 300-level course
  - Only 16 were usable
    - 3 Novice gamers
    - 10 casual gamers
    - 3 core gamers

Experiment – Method

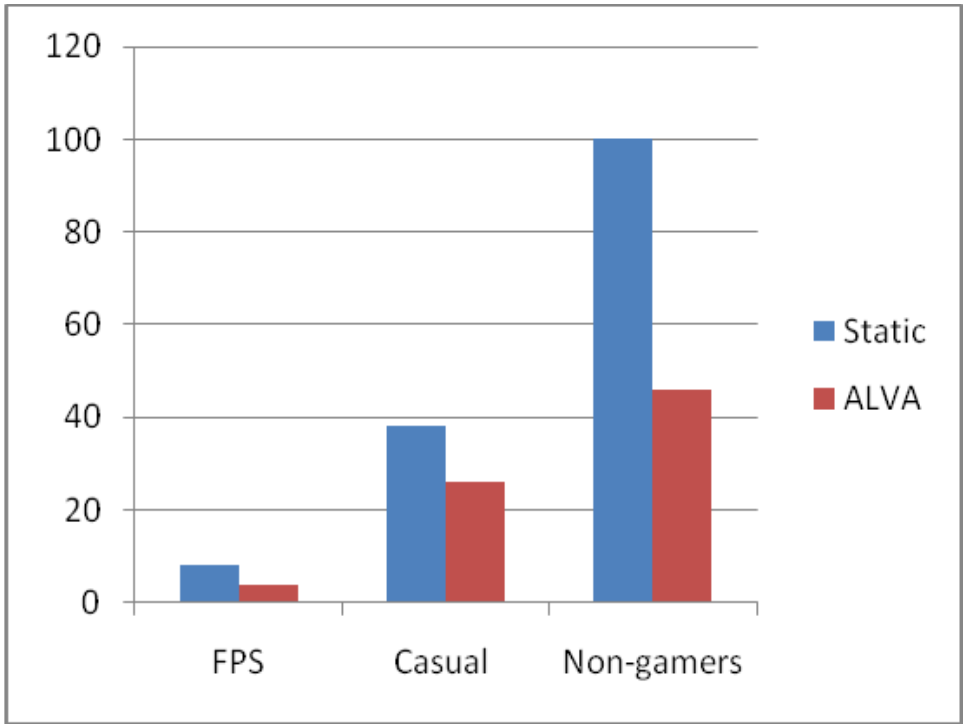
- Using ISCN ETL-500 head-mounted eye tracker
- Recorded video with superimposed cursor
- Analyzed video manually
- Variables measured:
  - time to spot
  - time of player death

Experiment – Method

- Step 1: Participants Play *Soul Caliber II*
- Step 2: Configure eye tracker
- Step 3: 2 Play sessions (10 min. each)
  - Using UT2003
  - Using UT2003 with ALVA
- Step 4: DVD signal => AVI file
- Step 5: analyze AVI file

Results

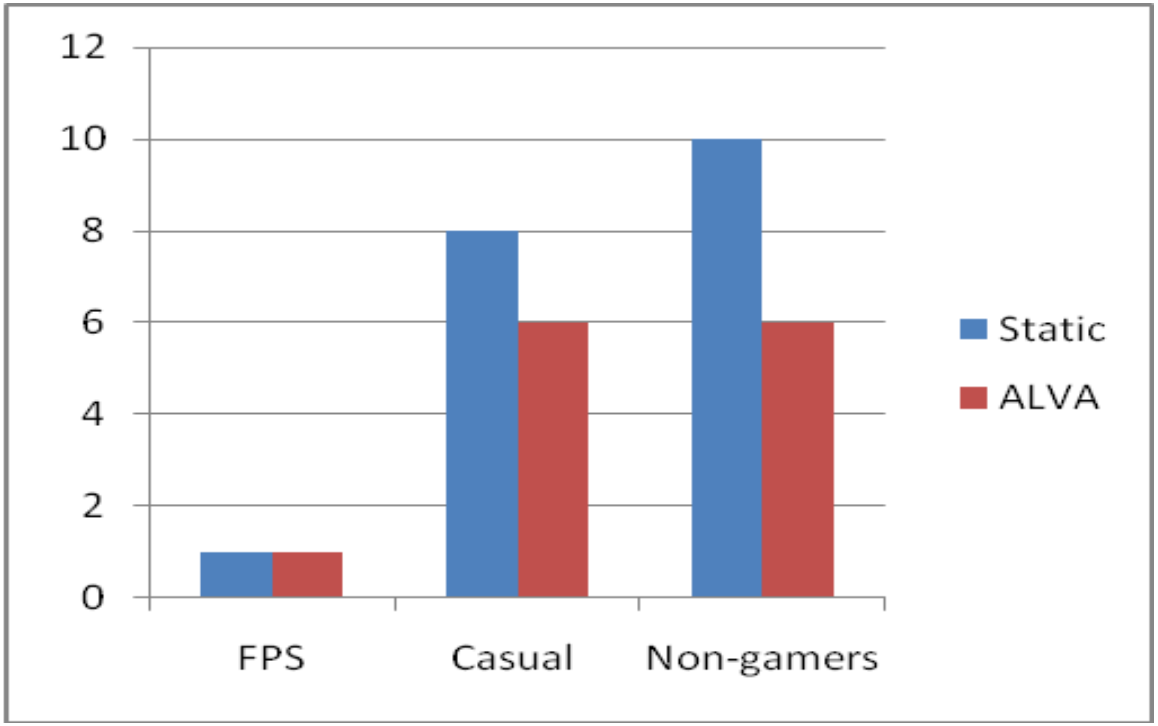
- Difference in spotting times between FPS gamers, Casual gamers, and Non-gamers.



95% statistically significant for casual and non-gamers

Results

- Difference in deaths times between FPS gamers, Casual gamers, and Non-gamers.



95% statistically significant for casual and non-gamers

Conclusions

- Evidence of bottom-up process
- For casual & novice gamers:
  - significantly faster spotting time with ALVA
  - Players didn’t die as quickly with ALVA
  - Less reported frustration
- For gamers: no significant difference
  - In fact they said “the game was too easy”

Answering questions

- Two general ways in which to answer questions:
  - Observe what happens naturally in the world
    - **Correlational** and **observational** methods
  - Manipulate an aspect of the environment and observe what happens
    - **Experimental** methods

Eye Tracking experiment

- A Dynamic Lighting System can be used to dynamically channel visual attention
- Non-gamers' do not spot enemies fast enough to respond
- Dynamic Lighting system will enable non-gamers to respond faster – better gameplay

Was this observational or experimental?

Variables

- Scientists are interested in how **variables** **change**, and what **causes** the change
- Anything that we can measure and which changes, is called a **variable**
- Variables can take many **forms**, i.e. numbers, abstract values, etc.
- Variables can be continous or discrete

Eye Tracking experiment

- A Dynamic Lighting System can be used to dynamically channel visual attention
- Non-gamers' do not spot enemies fast enough to respond
- Dynamic Lighting system will enable non-gamers to respond faster – better gameplay

What are the variables here?

Correlational vs. Causal research

- Main difference:

- Correlational and Causal research (experimental research) **treats variables differently**

- **Experimental research: manipulates variables**
- **Correlational/Observational research: observes variables**

Correlational vs. Causal Relationships

- **Causal:** Changes in one variable directly influence another
  - E.g. meteor hits earth, size of meteor determines numbers of dead dinos
  - If no meteor, no sudden death of dinos -> causality

- Causal relationships can be **unidirectional**: A -> B
  - Meteor kills dinosaurs
- Or **bidirectional**: A <-> B
  - Exercise influence body weight (negatively), and body weight influence exercise (negatively)

Correlational vs. Causal Relationships

- **Correlational:** Variables covary, but not necessarily causal relationship
  - E.g. sales of ice cream increase during period of warm weather
  - E.g. number of mosquitoes and number of baseball games (both occur during Spring)

- **Correlational research is non-experimental**
  - We do not manipulate variables ->
  - **Cannot infer causality**

  - Many types of non-experimental but empirical methods
  - E.g. calculating average value of SAT scores

- Karl Popper
  - Distinguished between **scientific-** and **non scientific statements**
  - Scientific statements can be verified with reference to empirical **testing**
    - ”Beating children is morally wrong” – non-scientific
    - ”On Earth, gravitational forces pulls objects with mass towards the center of the planet” – Scientific

Testing theories

To test a theory (answer a question) you must:

1. Rule out other explanations of the supposed cause
   - **Control the conditions** of experiment
   - **Minimize risk** of random/unknown factors influencing result
   - **Randomize** the procedure

2. Gain confidence that one theory is correct and another is not

Testing theories

1) Ruling out other explanations:

We need to control the conditions of the experiment

- To verify if eating **candy** makes you **fat**, we need one experimental setup where candy is present, one where it is not
- The condition where the cause is absent is known as a **control condition** or **baseline**

Testing theories

- In the simplest situation, the cause is either **present** or **absent (either we eat candy or not)**

- There can also be multiple **levels** – (0 pieces of candy per day, 2, 4, 7, 10, 10000 etc.)

- The variable being manipulated is the **independent variable** – it *depends only on the experimenter* (**candy**)

- The variable not manipulated is the **dependent variable** – it´s value depend on the value of *other variables* in the experiment (how **fat** you get)

Testing theories

B) Minimize risk of **random factors**

- We should compare situations that are identical in all respects apart from the proposed cause (causal/dependent variable)

- All random factors should be held **constant**
  - Everyone should eat the same candy
  - Eat it at the same time of day
  - Be the same gender
  - Etc.

Testing theories

C) Randomizing procedure

- We can rule out many random influences by **randomizing** parts of the experimental study
  - randomly alllocating participants to experimental and control groups – this spreads attributes randomly
  - Do not permit any **systematic bias** to enter the experiment

Testing theories

2) Gain confidence that one theory is correct and another is not

- The difference between the experimental group (cause is present) and the control group (cause is absent) must be ”large” so as not to have occured by **chance**

- This is where the **statistics** come in

The 4 aims of research

1. **Reliability:** Results can be replicated by others

2. **Validity (internal):** Results show what we intend them to show
   1. Ability of a research design to test the hypothesis it was designed to test
   2. Measure what we want to measure

3. **Generalizability (external validity):** Results have a wider application than merely the participants and the circumstances of the test

4. **Importance:** Results should be important (subjective).

   Results are never important if not reliable, valid and generalizable

Validity

- **Dependent variable** should be measured as **precisely** as possible

- Need to **define what it is that is measured**
  - Easy for some things: ”Number of dinosaurs killed”
  - Hard for others: ”Love”

- Popular solution: **Operational definition –** A well-argued definition for the particular study
  - Others may not agree, but at least you are clear

Validity

- Good experimental designs **maximise validity**

- **Internal validity:**
  - Extent to which we can be sure that changes in the DV are due to changes in the IV [meteor kills dinosaurs].
  - Requires confounding variables are eliminated

- **External validity (generalizability):**
  - Extent to which we can generalise from our participants to other groups (e.g. to real-life situations).

Validity

- **Ecological validity**
  - Extent to which research results can be applied to real life situations outside research settings
  - But focused on the degree to which findings can be **observed in the real world**
  - To have ecological validity, a research design must closely **mimic** the **real life situation under investigation**
  - (Ecology = science of interaction between organism and its environment)

Summary

Experiments are the best tool for establishing **cause** and **effect**

A **good experimental design** ensures that:

1) The *only* variable that varies is the independent variable/-s chosen by the experimenter

2) The effects of alternative *confounding variables* are eliminated (or at least rendered unsystematic by randomisation)

Choosing specific variables

- Choosing variables and defining these depend on multiple factors, e.g.:
  - **Research tradition of your field:** allows comparison across experiments
  - **Theory:** measures may have been defined previously
  - **Availability of new techniques:** new technology may allow new types of measures (e.g. EEG scans)
  - **Availability of equipment:** What do you have?

Choosing measures

- So …. we now know the variables and have defined them

- How do we **measure** them?

Choosing measures

- **Reliability** of measures
  - Ability to produce similar results when repeated measurements are made under identical conditions -> how much variance?
  - Bathroom weight +/- 1 point
  - Population estimates =/- 3%

- **Accuracy** of measures
  - E.g. 0.1 gram precision vs. 0.0001 gram precision.
  - **Bias:** if a measure on average measures more or less than the real value

Choosing measures

- **Validity** of measures
  - *The extent to which a measure measures what it is intended to measure*
  - E.g. head size as measure of intelligence?

  - Validity of measures especially problematic in psychology -> we cannot **directly** measure what goes on in people's heads

Choosing measures

- General concerns:

- Choosing **existing validated**, reliable measures …
  - Frees up time -> need not test reliability/validity
  - Binds you to existing theory/dogma

- If existing methods are not available, developing **new measures** …
  - Frees you from existing theory/dogma
  - New measures must be evaluated for reliability and validity

Measurement scales

- **Variables are measureable in different scales**

- Measuring size of variables is important for **comparing results** between studies/projects
- Different measures provide different quality of data:

- **Nominal** (categorical) data
- **Ordinal** data
- **Interval** data
- **Ratio** data

Non-parametric

Parametric

Measurement scales

- **Nominal data (categorical, frequency data)**
- When numbers are used as **names**
- No relationship between the size of the number and what is being measured
- Two things with same number are equivalent
- Two things with different numbers are different

Measurement scales

- E.g. Numbers on the shirts of soccer players
- Nominal data are only used for **frequencies**
  - How many times ”3” occurs in a sample
  - How often player 3 scores compared to player 1

Measurement scales

- **Ordinal data**
- Provides information about the **ordering** of the data
- Does not tell us about the relative **differences** between values

Measurement scales

- For example: The **order** of people who complete a race – from the winner to the last to cross the finish line.
- Typical scale for **questionnaire** data

Measurement scales

- **Interval** data
- When measurements are made on a **scale** with **equal intervals** between points on the scale, but the scale has **no true zero point.**

Measurement scales

- **Ratio** data

- When measurements are made on a scale with equal intervals between points on the scale, and the scale has a **true zero point**.
- e.g. height, weight, time, distance.
- Measurements of relevance include: Reaction times, numbers correct answered, error scores in usability tests.

Common user-oriented measures (HCI, Psychology …)

- **Behavioral measures**
  - Recording activity of users, e.g. frequency, error rates, latency
  - Example: game analytics

- **Physiological measures**
  - Recording heart rate, respiration rate, brain activity etc.
  - Example: game user research

Common user-oriented measures (HCI, Psychology …)

- **Self-Report Measures**
  - Participants report themselves, e.g. via survey
  - **Likert Scales** (disagree-agree scale to statements, e.g. 1 (strongly disagree) to 5 (strongly agrees)
  - Versatile and easy to use, but suffer from validity and reliability problems
    - Are participants being truthful?
    - Are participants happy/sad today?
    - People tend to respond in a *socially desirably way*

Common user-oriented measures (HCI, Psychology …)

- **Implicit measures**
  - Measuring responses not under conscious control of the participant
  - E.g. *if playing a video game users are more likely to shoot men than women?* (in situation where decision needs to be instantaneous)
  - Implicit measures require that participant must act fast to decouple conscious decision making

Manipulation checks

- **Manipulation checks**
  - Checking whether manipulation of IV had intended effect on participants
  - If participants perceived the experiment in the intended manner
  - E.g. understood survey questions correctly

Pilot studies

- **Pilot studies**
  - Small-scale version of a study used to establish procedures, materials and parameters
  - Always a good idea – can save resources as it points to problems/flaws in research designs early on

Class Assignment

- Candy Crush Experiment Design

Goal: I want to know if the game is addictive beyond first impression.

  - What are your research questions?
  - What is your approach (Hypothesis or Exploratory)?
  - Experiment Design
    - Who are the participants
    - How many
    - What are the conditions
    - What are the dependent and independent variables or observable variables?