



University
of Glasgow

User Interaction

COMPSCI2031

Dr Ilyena Hirskyj-Douglas

ilyena.hirskyj-douglas@glasgow.ac.uk



Recap: What we did last yesterday

- Lecture on Human Factors on Perception and Capabilities
- Reading: None
- Went over assignment



User Interaction Topics

- ✓ HCI History and Introduction
- ✓ Usability and Heuristics
- ✓ Heuristic Evaluation and Human Cognition
- ✓ Human Perception and Capabilities
 - Experimental Design & Variables Research
 - Personas and Scenarios
 - Surveys in HCI
 - Ethnography
 - Statical Methods
 - Theories in HCI
 - Models of Interaction
 - Large Scale and Mobile HCI
 - User-Centered Design
 - Ethics in User Testing
 - Revision & Example Exams



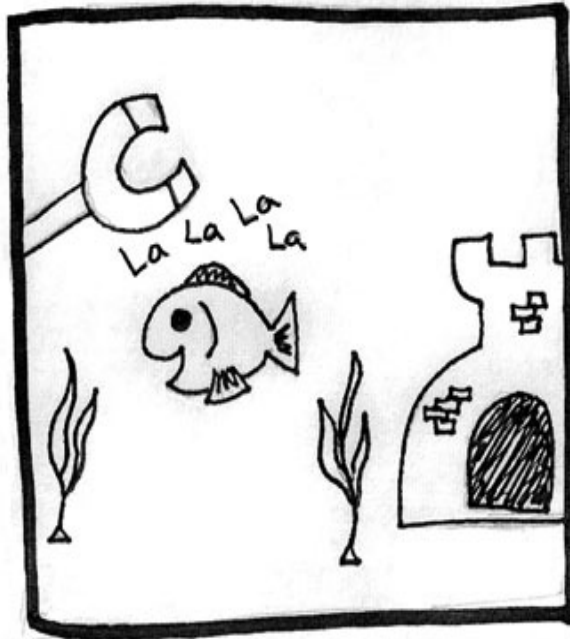
University
of Glasgow

Experimental Design & Variables Research

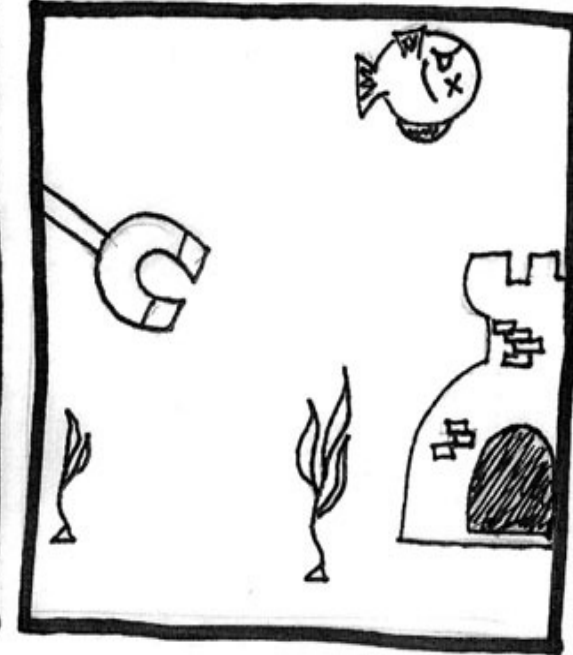
Lecture 5



The Importance of Experimental Design



Let's see if the subject responds to magnetic stimuli... ADMINISTER THE MAGNET!



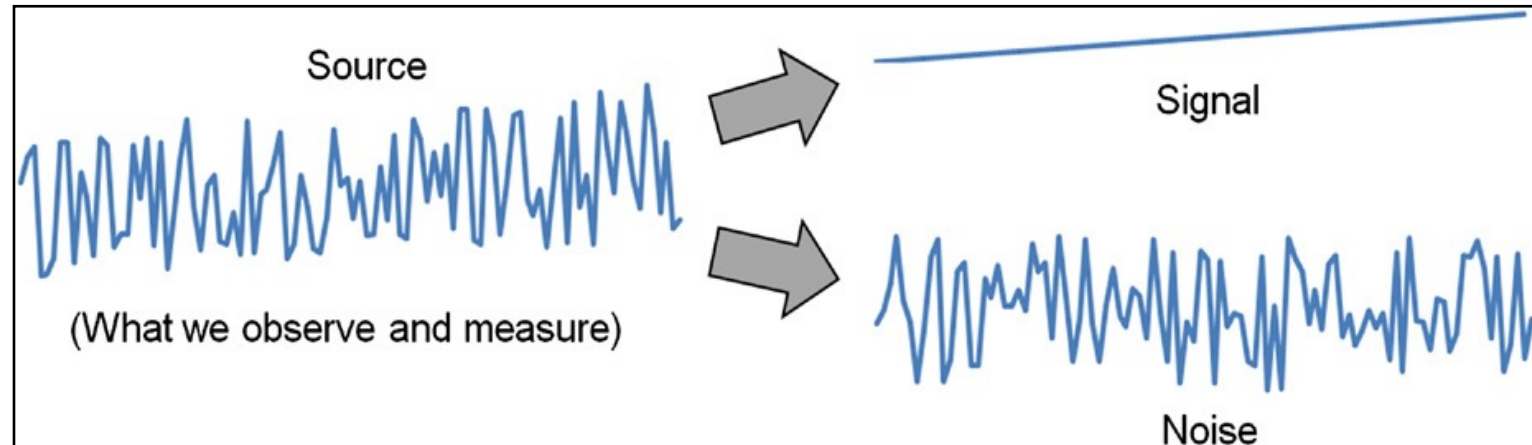
Interesting...there seems to be a significant decrease in heart rate. The fish must sense the magnetic field.



Method / Methodology

Methodology is the way an experiment is designed and carried out.

Sound methodology is critical to allow us to understand what is really going on (signal) in a noisy and messy world (noise):



“Science is method. Everything else is commentary.” Allen Newell



Why does this matter?

- Will help you run good studies in your workplace
- If you want to publish a scientific paper testing hypotheses you need to understand experimental design
- Even if you don't end up running studies (e.g. you go into data science field) you need to know how the data was collected in order to understand how to handle the data
- It's a good life skill – it gives you extra tools to think critically when you read the latest science story such as 'chocolate can cure cancer.' Sound methodology often separates good science from the bad.



Validity

- Internal validity
 - Is effect observed due to varied condition(s)?
- External validity
 - Are experimental results generalisable to other people/situations?
 - Sampling
 - Realistic conditions
- Often tension between internal vs external: one at expense of other?



Independent Variables

Also called *factors*

Experiments with independent variables are often called *factorial experiments*.

These can be naturally occurring or directly manipulated by the experimenter

Characteristics:

e.g. of computer interface – input device, feedback modality, display size

e.g. of participants – gender, handedness, expertise

OR Circumstances:

e.g. background noise, room lighting

Levels: each test condition. E.g. 'mouse' and 'trackpad' are *levels* of independent variable 'input device'



Dependent Variables

In HCI the DV is often a measured human behaviour. The measurement *depends* on what the participant does.

Any observable, measurable behaviour can be a DV:

e.g. typing speed, eye movements, 'negative facial expressions', 'read text events', how respond to questionnaire

You can be creative here to get the measurement that tells you most about the impact of the independent variable.



Effect numbers

You can have (and will often want) more than one IV and DV in a study. However it is important to limit these as the more IVs you have the more comparisons there are.

These increase rapidly:

Independent variables	Effects					Total
	Main	2-way	3-way	4-way	5-way	
1	1	-	-	-	-	1
2	2	1	-	-	-	3
3	3	3	1	-	-	7
4	4	6	3	1	-	14
5	5	10	6	3	1	25

FIGURE 5.2

The number of effects (main and interaction) increases as the number of independent variables increases.

Increasing IVs quickly accelerate your number of observed effects, so **limit to 1 – 3.**



Control Variables

These are not under investigation (i.e. are not IVs) but they might influence participant behaviour (DVs)

e.g. keyboard angle, chair height, display size

Experimenters control these variables to prevent their influence by setting up their study in a controlled environment and recruiting with strict inclusion / exclusion criteria.

e.g. Right handed only, Experienced users only

Increases internal validity but reduces external



Random Variables

Instead of attempting to control for everything, it is often better to allow some variables to vary randomly in order to generalise the results (and because controlling for everything is probably impossible).

e.g. height, hand/finger size, social disposition

Each study will require judgement about the trade off between control and allowing random variation.

e.g. using a questionnaire of motion sickness and recruiting only those under a threshold:

- 1) In a study investigating the acceptability of two in-car interaction techniques for the general population
- 2) In a study comparing the use of a VR headset in a moving vehicle with different VR conditions to mediate sickness level



Confounding Variables

In some cases a circumstance or condition will change ***systematically*** with the independent variable.

e.g. practice, different types of measurement for levels of the IV, prior experience with an interface (e.g. when comparing Google to anything)

Such variables are confounding because they prevent the possibility of a cause and effect relationship being inferred from the results. So they need to be controlled for. **This is a central skill in experimental design.**

Prior experience will almost definitely confound our attempts to find the best keyboard...

QWERTY KEYBOARD

~	!	@	#	\$	%	^	&	*	()	-	+	Delete
	1	2	3	4	5	6	7	8	9	0		=	
Tab	Q	W	E	R	T	Y	U	I	O	P	{	}	
Caps	A	S	D	F	G	H	J	K	L	:	"		Enter
Shift		Z	X	C	V	B	N	M	<	>	?	/	Shift
Ctrl			Alt								Alt		Ctrl

DVORAK Keyboard Layout

	!	@	#	\$	%	^	&	*	()	-	+	Delete
	1	2	3	4	5	6	7	8	9	0		=	
Tab	?	<	>	P	Y	F	G	C	R	L	{	}	
	A	O	E	U	I	D	H	T	N	S	"		Enter
Shift	:		Q	J	K	X	B	M	W	V	Z		Shift
Ctrl			Alt								Alt		Ctrl

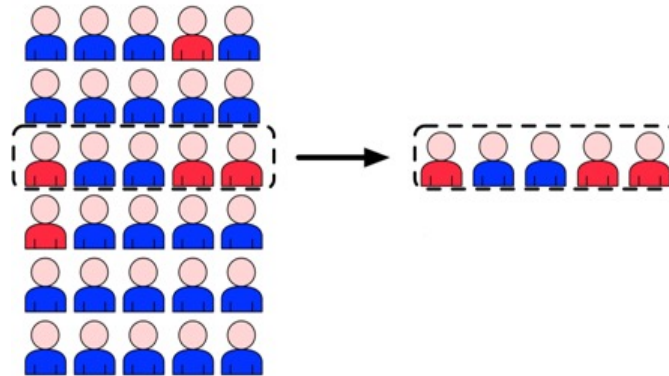


Participants

In order to correctly assume that research results apply to people other than those recruited you must:

Recruit people from the population you want to investigate

Recruit a sufficient number of participants

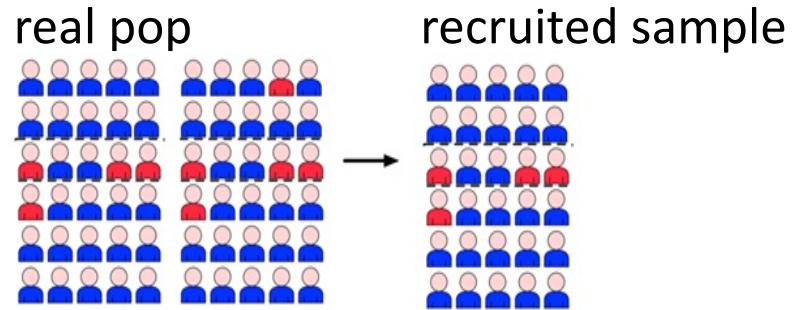




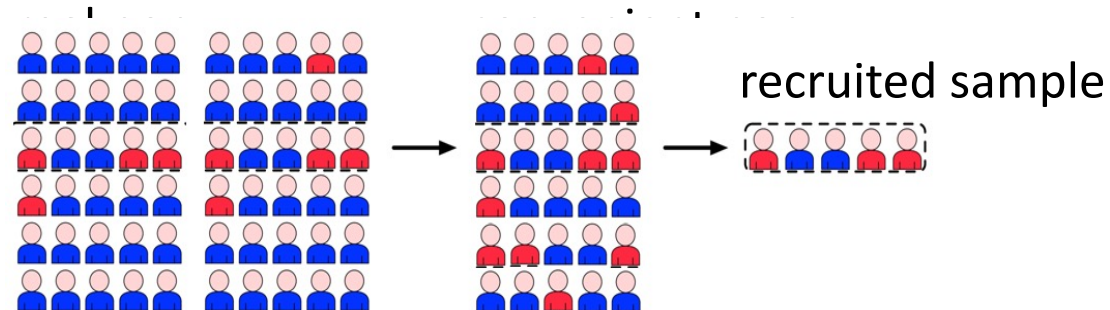
Participants

Recruitment methods

Ideal = participants drawn at random from a population



In practice = convenience sampling



How might these people differ from the population you are trying to generalise your findings to?



Participants

How many people should you recruit?

More is better

However there is a balance between representing the population and practical, and sometimes ethical considerations

Practical: Not much time to recruit (e.g. in student project), population difficult to access, more testing delays product going to market

Ethical: Study puts participants under burden; continuing study beyond necessary delays useful intervention or technique that can improve access.



Participants

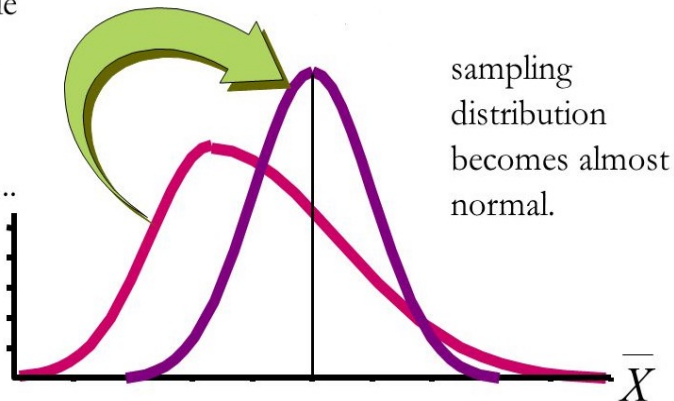
Central Limit Theorem:

As sample size increases to ≥ 30 becomes approximately normally distributed.

This means that, if you are measuring abilities that are normally distributed in general population (e.g. typing accuracy on a task) then after around 30 people are recruited this will be reflected in the data.

Therefore there has been a rule of thumb in psychology research that 30 people is a good number

As sample
size gets
large
enough
($n \geq 30$) ...



Within or Between subjects

Take the example comparing the two keyboard types:

QWERTY KEYBOARD

~	!	@	#	\$	%	^	&	*	()	-	+	Delete
Tab	Q	W	E	R	T	Y	U	I	O	P	{	}	
Caps	A	S	D	F	G	H	J	K	L	:	"	'	Enter
Shift	Z	X	C	V	B	N	M	<	>	?	/	Shift	
Ctrl		Alt									Alt		Ctrl

DVORAK Keyboard Layout

	!	@	#	\$	%	^	&	*	()	-	+	Delete
Tab	?	<	>	P	Y	F	G	C	R	L	{	}	
	A	O	E	U	I	D	H	T	N	S	"	'	Enter
Shift	:	Q	J	K	X	B	M	W	V	Z	Shift		
Ctrl		Alt									Alt		Ctrl

ComputerHope.com

You have 20 willing participants for a study, a typing task and way of calculating typing speed and accuracy. You could approach the comparison in two ways:

Within-subjects (repeated measures)

All 20 participants use both keyboards and you compare the average difference in performance between the two.

Between-subjects

You split the participants into two groups of 10 and compare performance of QWERTY group to Dvorak group

Order Effects

In general within-subject designs are favoured in HCI due to the nature of the research – this means we need to do something about the interference between test conditions or ***order effects***:

Learning effects

Order effects are usually a problem because of learning – people's performance on a task improves as the study goes on and this makes it hard to know if observed differences are due to the IV under investigation (e.g. keyboard type) or this confounding variable (learning leading to performance improvement).

Fatigue effects

These are more rare but it is possible that people get worse at a task the longer it goes on due to tiredness or because they lose interest. This is also a type of order effect.



Counterbalancing

Balance the order in which participants do each level

So counterbalancing is really easy if you have 2 levels of an IV, but what if you have more?

A **Latin square** is an $n \times n$ table that allows conceptualisation (and generating) of counterbalanced conditions



Counterbalancing

Latin Squares

Assign users to groups. Each group gets different order of conditions

Ensure equal number of people in each group

Below we can use Latin squares to ensure that each condition in a text entry experiment is presented first for each group:

	swype	standard	swype	standard	prediction	swype	standard	prediction	voice
Group 1	1	2	1	2	3	1	2	3	4
Group 2	2	1	2	3	1	2	3	4	1
Group 3			3	1	2	3	4	1	2
Group 4						4	1	2	3

Counterbalancing versus Randomisation

- When Latin Squares become needlessly complicated or impractical, randomisation is another technique to mitigate order effects
 - Consider a survey with 30 questions
 - Question order effects?
 - Full 30×30 Latin Square would get complicated
 - Might need a lot of users, to place equal numbers in each group
 - Could just completely randomise



Key point when counterbalancing:

If you are only interested in removing the order effects then you need to make sure that each possible combination is represented an equal number of times.

Don't need to worry about the size of groups that do each combination of condition orders, as long as all group sizes are same

i.e. when Order effect = a confounding variable

If you want to understand order effects, not just cancel them out, then you need a large enough sample of your participants to do each order combination

Group size has to be large enough to compare condition orders

i.e. when Order effect = an Independent variable in the study.

Usually you only want to control for order effects. If that's all you do then you can't make any conclusions about the impact of condition order.



Longitudinal studies

In the keyboard example there is a clear confounding variable – everyone is more likely to be familiar with QWERTY than other keyboard types.

However that is not to say it is the best layout overall, experience level being equal. To test this we need to use a longitudinal study to allow people to become equally experienced with both QWERTY and Dvorak methods.

QWERTY KEYBOARD

~	!	@	#	\$	%	^	&	*	()	-	+	Delete
1	2	3	4	5	6	7	8	9	0				
Tab	Q	W	E	R	T	Y	U	I	O	P	{	}	\
Caps	A	S	D	F	G	H	J	K	L	:	"	'	Enter
Shift	Z	X	C	V	B	N	M	<	>	?	/		Shift
Ctrl		Alt										Alt	Ctrl

DVORAK Keyboard Layout

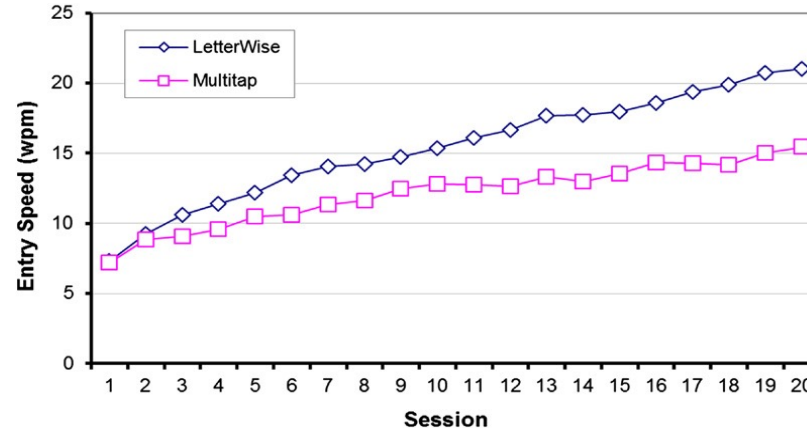
	!	@	#	\$	%	^	&	*	()	-	+	Delete
	1	2	3	4	5	6	7	8	9	0			
Tab	?	<	>	P	Y	F	G	C	R	L	{	}	\
	/	.	.										
	A	O	E	U	I	D	H	T	N	S	"	'	Enter
Shift	:	:	Q	J	K	X	B	M	W	V	Z		Shift
Ctrl		Alt										Alt	Ctrl

ComputerHope.com

Longitudinal studies are often used in HCI to investigate learning effects over time. This is very important considering the ubiquity of technology use in everyday life.

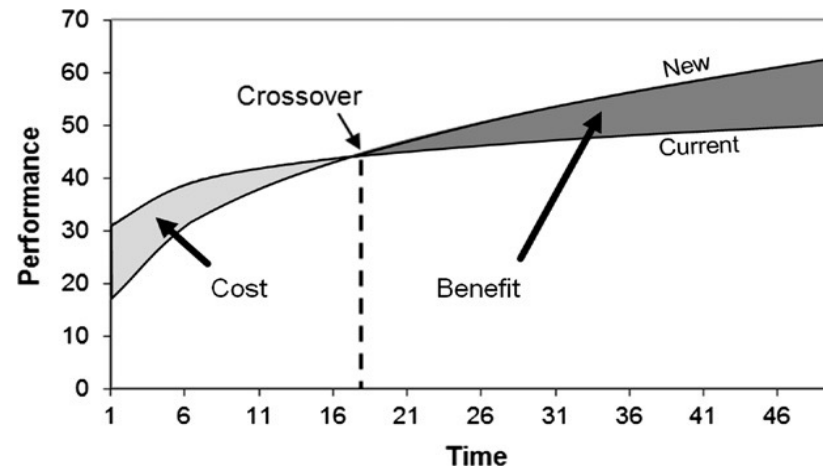
Longitudinal studies

Effects only seen after long term use:



Crossover:

In the case of a new product, it may be that performance on the traditional system will start off better, but that this crosses over after long term use:





Running the Experiment

Always pilot your study before running it!

You won't believe the issues that can arise – technical issues, no-one can understand your task instructions, the study takes 3 hours to finish...

Use consent forms and make sure the participant understands the study

Be consistent

Have a neutral manner and use a script if needed to make sure you give the same instructions each time. Running a research study is an acquired skill.

Also be aware of bias – is your desire for the hypothesis to be met making you inadvertently encourage participants to behave in the desired way?



University
of Glasgow



Questions?
Comments?
Concerns?



User Interaction Topics

- ✓ HCI History and Introduction
- ✓ Usability and Heuristics
- ✓ Heuristic Evaluation and Human Cognition
- ✓ Human Perception and Capabilities
 - Experimental Design & Variables Research
 - Personas and Scenarios
 - Surveys in HCI
 - Ethnography
 - Statical Methods
 - Theories in HCI
 - Models of Interaction
 - Large Scale and Mobile HCI
 - User-Centered Design
 - Ethics in User Testing
 - Revision & Example Exams



Task: Variable & Study Design

I have an experiment idea of wanting to discover how people feel about other people playing with smartwatches and using mobile phones while in a conversation. For this, I give ten people smartwatches, ten people mobile phones, and ten people who have nothing and invite them to a party to eat some snacks and drinks and use their mobile devices. After the party, I gave people questionnaires on their experiences while at the party on how well they felt someone listened to them and engaged in the conversation. The key research question is:

- RQ1: Does using mobile phones and smart watches in conversations affect how people feel while in conversation?
- RQ2: Do people prefer other people to use smartwatches or mobile phones during conversations?
- From this experiment, in groups *identify*:
 - problems with internal validation and external validity
 - Identify the independent variables and dependent variables of the study
 - Controlled, random and confounding variables
 - If this is a with or between-subjects design

Task Answer: Variable & Study Design

- Identify problems with internal validity and external validity
 - **Internal:** Effect could be due to other conditions (e.g., how much they have prior used devices)
 - **External:** Giving it to users so under more real conditions (longitudinal study), might be problems with sampling
- Identify the independent variables and dependent variables of the study
 - **Independent:** phone or smartwatch (device feedback and screen size if different (probably not in this case), age and gender of participants, prior relationships)
 - **Dependent:** usage of device (time spent and in conversation usage),
- Identify if you have taken into account the effect number
 - **Effect:** Didn't use both devices at once, could do A/B sampling (latin square)
- Controlled and confounding variables
 - **Controlled:** Type of device. **Confounding:** prior usage of devices
- With or between subjects
 - **Between:** Only single use smart phone or smart watch



University
of Glasgow



Questions?
Comments?
Concerns?



User Interaction Topics

- ✓ HCI History and Introduction
- ✓ Usability and Heuristics
- ✓ Heuristic Evaluation and Human Cognition
- ✓ Human Perception and Capabilities
- ✓ Experimental Design & Variables Research
 - Personas and Scenarios
 - Surveys in HCI
 - Ethnography
 - Statical Methods
 - Theories in HCI
 - Models of Interaction
 - Large Scale and Mobile HCI
 - User-Centered Design
 - Ethics in User Testing
 - Revision & Class Test



University
of Glasgow

Reading: MacKenzie, Chapter 5: Designing HCI Experiments

