



TASKSHEET

ERRORS AND BIASES

CASE #1

Familiar with this sign?



Context: In 1920s imperfect car engines suffer from “knocking”.

Lead tetraethyl stops car engines from “knocking”.

In 1923 this new type of lead additive (lead tetraethyl) introduced into gasoline.

Engines stop “knocking”.

Levels of lead in Earth atmosphere increase up to 200 times normal levels.

Lead is a neurotoxin causing heavy and irreparable nerve damage in humans.

Lead additives producers fund studies that prove lead tetraethyl is not harmful to humans.

Study: Participants are asked to breathe in and swallow large doses of lead over a 5-year period. Feces and urine tested. No lead found in feces and urine. Participants do not suffer observable dramatic decline in health in the 5-year period of research. Researcher A concludes that lead is not harmful.

Fact: Lead is not excreted as a waste product. It accumulates in bone marrow and blood. The debilitating effects are felt many years after exposure.

What errors of biases can you see in Researcher A's study?

Researcher bias (this research got the result he wanted).

A little about the sad truth of commercially funded research: you are NOT paid to do research. You are paid to do research that fits the company's goals.

For more information on the subject: Google Thomas Midgley, Clair Cameron Patterson and lead tetraethyl.

CASE #2

Context: Paul Ekman publishes his papers on universal basic emotions around 1970s.

The idea of the universal nature of emotions propagates throughout the discipline of psychology.

Psycholinguists begin analyzing emotional speech patterns.

Emotion recognition in speech across languages becomes a popular topic of investigation in psychology.

Researchers want to investigate prosody alone, not the propositional content of speech.

Study: Participants speaking Hindi, Arabic, German and English who have been living in Canada for at least 3 years are asked to read out loud nonsense but plausible words and sentences in their native languages in an emotional



tone of voice. The recordings of these emotional expressions are played to other speakers of Hindi, Arabic, German, and English from the same population. Each language group only listens to recordings in their native language. They are asked to recognize the emotions by assigning them to one of seven categories (see table below).

	Sadness	anger	fear	happiness	neutral	disgust	surprise
English	91%	88%	87%	80%	76%	76%	72%
German	73%	88%	71%	60%	94%	77%	69%
Hindi	76%	74%	76%	67%	66%	64%	58%
Arabic	75%	63%	*	75%	63%	*	50%

*between 50 and 63%.

Researcher B concluded that basic emotions are universally recognized across cultures.

Fact: Emotion terms are not equivalent across cultures and languages.

What threats to validity and reliability can you define in this study?

Researcher bias.

(potentially) **Participant bias – obeying demand characteristics.**

Type 1 error due to lack of construct validity.

External validity in the study (construct validity to be precise) is compromised. The constructs of basic emotions are English-centric and were imposed on languages and cultures in which emotion constructs are different (Hindi, Arabic). Although the results in the study indicate recognition rate above chance level, the risk of participants' obeying demand characteristics is high.

CASE#3

Fact: Native speakers of Japanese have profound problems with distinguishing between green and blue, which is reflected in the lack of distinction between the colors in the language (we are aware of the increasing usage of 緑 midori for green; please disregard this trend for the purposes of this exercise).

Study: Researcher C conducts a color naming/discrimination task in which she finds native speakers of Japanese are in fact able to distinguish between the colors. Her sample size is: 4 participants (N=4).

What error might be responsible for Researcher C's result?

Type 1 error due to exceedingly small sample size.

(potentially) **Participant bias – obeying demand characteristics**, as the researcher invited her friends to the task.

EXPERIMENTAL TASKS

Study 1: using the Divided Visual Field (DVF) paradigm a researcher used the following word stimuli: *cactus*, *coctos*, *mirror*, *merror*, *rocket*, *rokcte*, *cocktail*, *cocktali*.



Study 2: a series of word pairs is displayed on screen, and the participants, who are all biology majors must decide whether the words have anything in common. Some of the stimuli are: *rose – violet, star – hammer, lilly – valley, moss – rock, lichen – tree, hawthorn – pavement, balloon – mountain.*

Study 3: a series of colored words are flashed on screen. Participants are asked to ignore the words and say the name of the color they are written in. The words are: *red, green, brown, fuck, yellow, crap, blue, shit, grey, cunt, black, faggot.*

Study 4: a series of colored words are flashed on screen. Participants are asked to ignore the words and say the name of the color they are written in. The words are: *blue, death, white, brown, sickness, red, yellow, wound, green, suicide, purple, maggot.*

Study	Task
Study 1	LDT
Study 4	Emotional Stroop
Study 2	SDT
Study 3	“Classic” Stroop

DATA TYPES

Task: Look at the hypotheses and decide what types of data will be collected for each of them. Some hypotheses will require more than one type of data.

Data: RT (ms), ACC/ER, qualitative, microSiemens (mS), BPM, ERPs, voxels.

Study hypothesis	Data
Processing picture stimuli activates larger volumes of brain cortex than processing word stimuli.	Voxels (fMRI technology)
Inducing positive mood increases the variety and fluency of vocabulary use in oral exams.	Qualitative
Novel metaphors are processed faster in L1 than in L2.	RTs
Positive mood makes people perform better on grammar tests.	ACC, (qualitative data may be analyzed as well depending on the established definition of “better”)
Verbal responses in emotional Stroop task are correlated with physiological measures of heart rate and skin conductance.	microSiemens (mS), BPM, qualitative (verbal responses), RT (may be analyzed), ACC (may be analyzed)
Brain reacts more strongly to concrete than to abstract words.	ERPs
Correct identification of irony is possible only if at least two channels of communication are available.	ACC, qualitative (may be analyzed depending on the manner of response selected)



for participants)

HYPOTHESES VS. RESEARCH QUESTIONS

Task: Identify research questions (R) and hypotheses (H).

NOTE: Normally, research questions are interrogatory sentences. Here, they were turned into declarative sentences on purpose. Because tasks aren't supposed to be too easy.

1. *Emotions are universal across languages.* **RQ**
2. *There is a relationship between the gender of the teacher of a speaking class and the performance on oral exams of the students.* **H**
3. *Classical music elevates retention of new information more than heavy metal music.* **H**
4. *There are differences in how L2 words are memorized depending on how familiar the denoted objects/phenomena.* **H**
5. *Negative gender stereotypes influence performance on mathematical aptitude tests.* **H**
6. *Drinking Coca Cola Light damages virility.* **H**
7. *People who eat French fries at least once a week are generally more happy.* **H**
8. *Eating French fries once a week increases levels of life satisfaction.* **H**

VARIABLES

Task: Identify the independent and dependent variables in the hypotheses given.

H1: N400 ERP will be greater in response to emotional environmental sounds than to non-emotional environmental sounds.

H2: Conventional metaphors are processed faster and more accurately than unconventional (novel) metaphors in schizophrenic patients.

H3: There are no significant differences in speed of processing of emotional prosody in L1 and L2 acquired late in life.

H4: Systematic formal musical training increases the phonological sensitivity scores in prepubescent children.

H5: Prolonged exposure to unpleasant audio stimuli of high frequency cause a temporary drop in retention skills.

H6: Physical warmth increases the levels of prosocial behavior.

H7: The ability to delay gratification in early childhood correlates strongly with educational and career success in later life.

	Independent variable(s)	Dependent variable(s)
H1	emotional/non-emotional environmental sounds	N400 ERP component
H2	conventional/unconventional metaphors	ACC, RT
H3	L1 and L2 (we manipulate this variable by choosing the kind of bilinguals our hypothesis pertains to)	RT (possibly also: emotional prosody – various types of



		<i>emotional prosody)</i>
H4	<i>systematic formal music training (we can select appropriate groups of children or compose a longitudinal study in which we train them)</i>	<i>ACC, qualitative (phonological sensitivity scores)</i>
H5	<i>prolonged unpleasant audio (we can manipulate the length of the unpleasant audio or the different types of unpleasant stimuli)</i>	<i>ACC (scores on retention tests)</i>
H6	<i>physical warmth (e.g. the temperature of the drink offered to participants who had been kept in a chilled room)</i>	<i>qualitative (analyzing the vocabulary used with respect to irritating people/places/events)</i>
H7	<i>This is a correlational study, so we cannot manipulate, only observe. The hypothesis is valid, but we cannot establish cause-effect relationship between the variables, we can only describe the outcome of the observed variable dynamics in a correlational relationships.</i>	

RELIABILITY AND VALIDITY

Task: What potential or real violations of validity and reliability can you see in the following study descriptions.

Study 1: A researcher wants to test H1 from the previous task. She performs her experiment on a sample of 30 people (acceptable in her field). She uses a population sample drawn from a local professional (vocational) university from a class of sound engineering. This sample is her only experimental group.

External validity violated. The participants are professionals working with sound. Because of that they can be one of two extremes: they will either react much more strongly or much more weakly to the unpleasant stimuli, because they are used to tuning in with the soundscape around them. In other words, *the sample is not representative of the general population and the results of this experiment are not generalizable.*

Study 2: A researcher wants to test H2 from the previous task. The hypothesis is confirmed in her experiment. It turns out her participants were participants of a Rosenhan experiment (they pretended to be hearing voices so that they would be committed to a mental institution).

Random error. The researcher did everything right, but she could not know about the Rosenhan group due to the nature of their presence in the mental institution where she had been conducting her research. Equally they could not tell her about the nature of their presence so as not to compromise their own experiment.

Study 3: A researcher wants to test H3 from the previous task. She performs an experiment and her data indicates that there are processing differences between L1 and L2 in late bilinguals. She used a group of Ukrainian-Russian bilingual speakers and conducted her study in early 2014 in Kiev.



External validity violated. The historical context forms a threat. Kiev in early 2014 was in the grip of a civil war between the pro-Russian government forces and anti-Russian, pro-European activists. Although Russian has long been one of the official languages in the Ukraine, the overwhelming anti-Russian sentiment in the experimental sample has influenced the way the Russian language is processed by the Ukrainians.

Study 4: A researcher wants to test H4 from the previous task. Her data support the hypothesis. To determine whether her participants had formal musical training she used a self-report questionnaires in which she asked whether the children had any formal training, how many years it lasted, and what the instrument was.

Internal validity violated. The concept of “systematic music training” is ill-defined and ill-controlled in this study. The researcher cannot rule out Type 1, Type 2 or Random Errors in her results. This hypothesis hinges on the construct of “systematic music training”, and the questions asked by the researcher in the self-report do not control the systematic aspect.

Study 5: A researcher wants to test H5 from the previous task. He had conducted his experiment in a controlled experimental environment, and the sounds which were generated on a computer were administered over headphones. He confirms the hypothesis. He concludes his paper on the study with a statement that in a noisy environment studying for exams is very ineffective.

Ecological validity threatened. The laboratory experiment using computer-generated sounds does not reflect real life to a satisfying degree.

Study 6: A researcher wants to test H6 from the previous task. It is a replication from another study, where the participants were kept in a cool room, subsequently they were offered a hot drink and engaged in a conversation about a series of irritating people and events. In the replication study the temperature in the waiting room wasn't controlled. The results did not reach significance levels.

Internal validity violated. The researcher did not control the critical variable of room temperature. Whatever results he got would not be valid because of that.

Study 7: A researcher wants to test H7 from the previous task. The researcher replicates an original study in which the correlation was observed. The replication is conducted in China among children who grow up to be civil servants. The procedure is replicated exactly, using the constructs of willpower and self-discipline. The constructs of civil obedience and subjecting to authority are not involved in the study. The researcher observes similar result, but the data is inconclusive.

Internal and external validity violated. Reliability compromised.

Internal > in a collectivist, strongly institutionalized and formalized society such as the Chinese the constructs of willpower and self-discipline are very hard to observe due to social desirability biases in the participants. Had the study been framed using the constructs of civil obedience and subjecting to authority, the results might have been different, but it would also be a different study.



External > the results in such a biased population sample cannot be generalized in any way to non-collectivist (individualistic) societies.

Reliability > reliability of the original study is compromised as it was not caveated to account for the differences b-n the individualist and collectivist societies.

CORRELATION VS. EXPERIMENTS

Task: Categorize the studies described below into correlational studies and experiments.

Study 1: An observation was made that when lights are switched off the number of pregnancies significantly increases. **CORRELATION**

Study 2: An observation was made that the existing methods of schizophrenia diagnosis were flawed. A group of volunteers went to psychiatric institutions and reported one vague symptom. They were all committed. **EXPERIMENT** (for details Google the Rosenhan study)

Study 3: An observation was made that when given orders by a person of authority individuals follow the orders even to the extent of seriously harming an innocent individual. **EXPERIMENT** (for details Google the Milgram experiment)

Study 4: An observation was made that individuals habitually using marijuana show lower expectations for academic achievement, lesser religiosity, less compatibility between friends and parents and greater actual involvement in other problem behaviors such as drunkenness. **CORRELATION**

Study 5: An observation was made that the bigger a person's brain the more intelligent they are. **CORRELATION**

Study 6: An observation was made that people readily conform to the social roles they are asked to play and that these roles can shape their behaviour and attitudes. **EXPERIMENT** (for details Google the Stanford prison experiment)