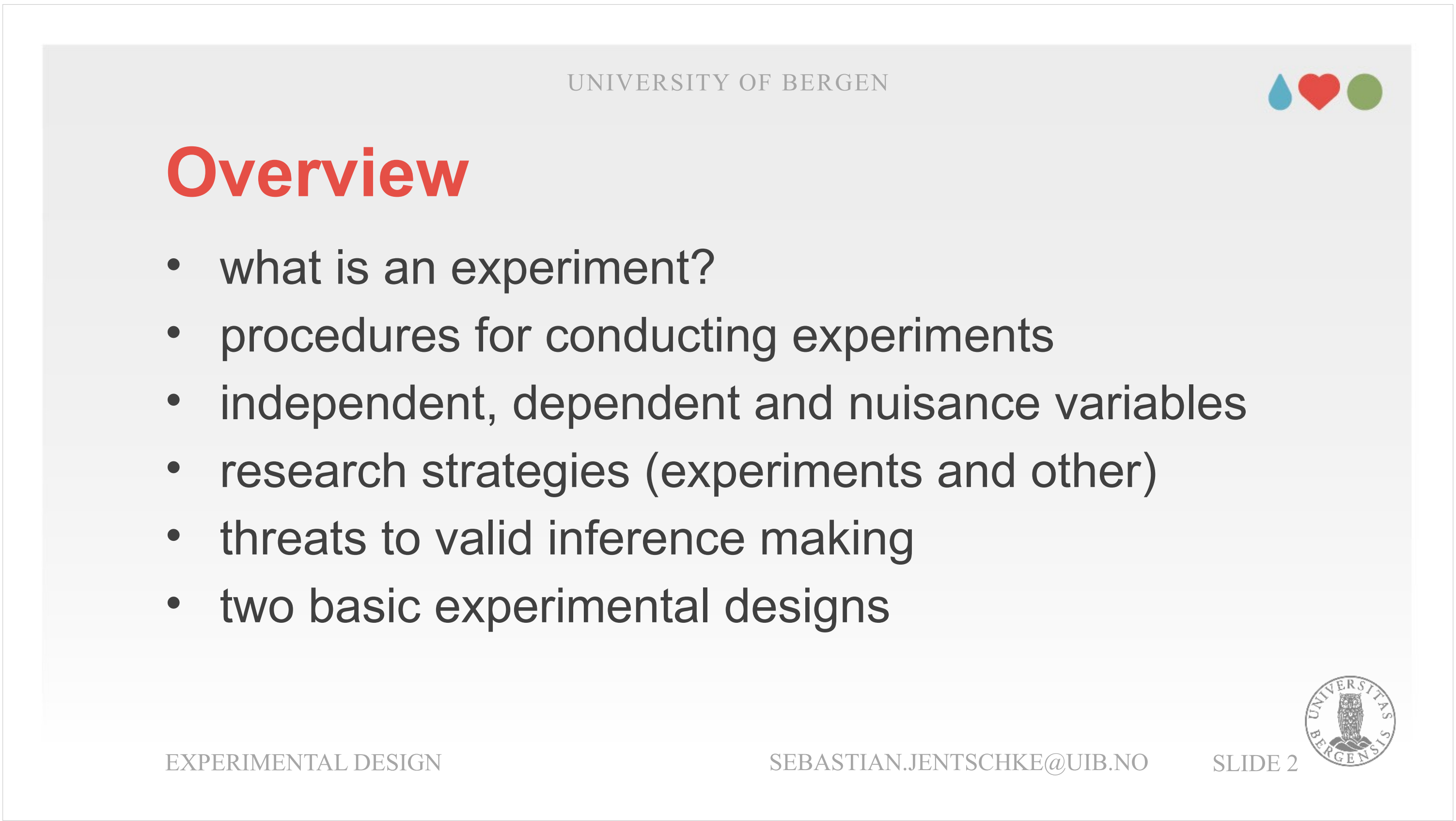
UNIVERSITY OF BERGEN

Experimental design

Sebastian Jentschke

UNIVERSITY OF BERGEN



Overview

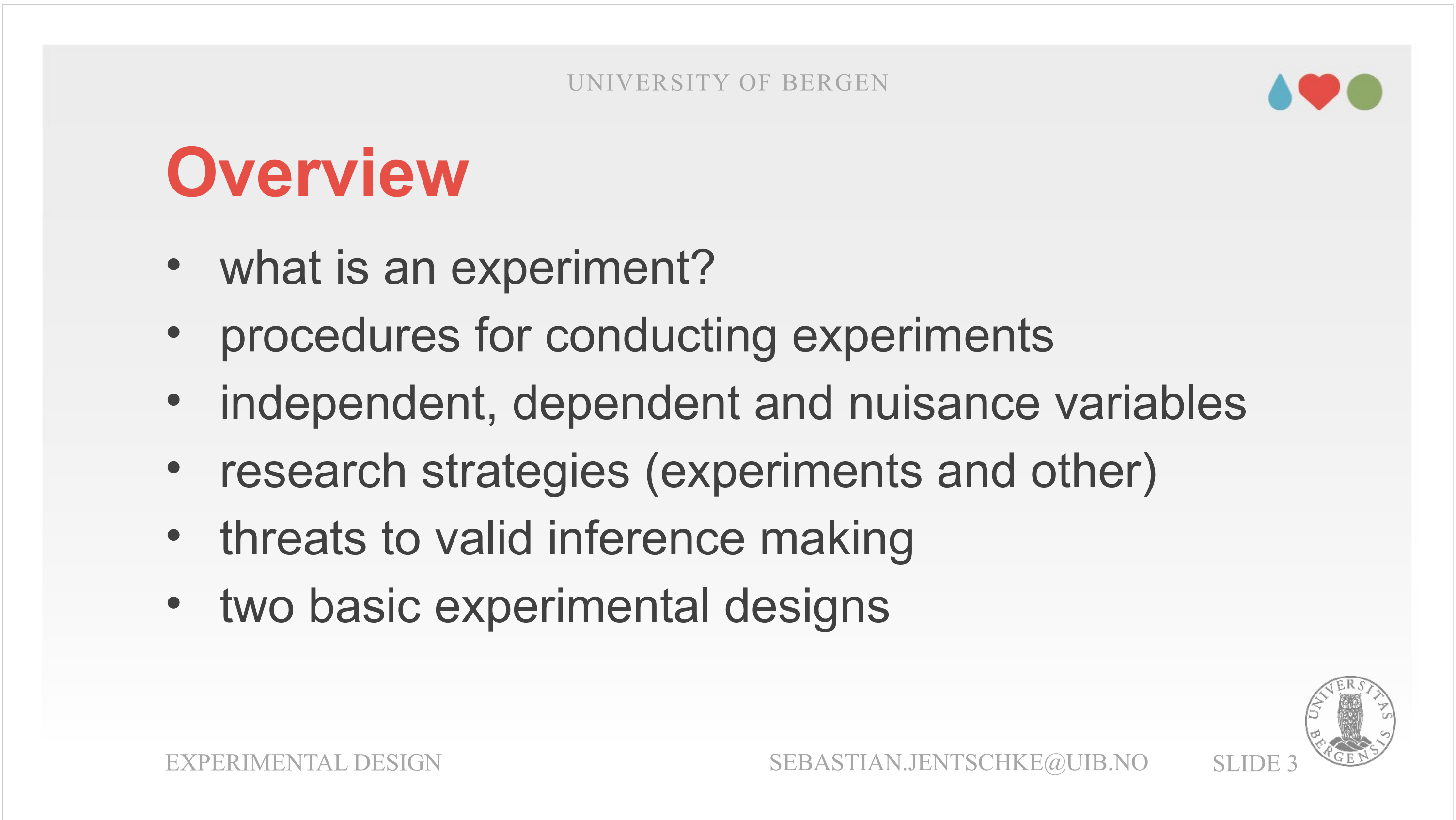
- what is an experiment?
- procedures for conducting experiments
- independent, dependent and nuisance variables
- research strategies (experiments and other)
- threats to valid inference making
- two basic experimental designs

In this lecture, we will speak about experiments and how to design them. As a first step, we will have a comparison between different accounts of what an experiment is.

This is followed by a description of which procedure is typically followed when an experiment is conducted.

Experiments aim to manipulate an independent variable to see an effect on a dependent variable while controlling for possible nuisance variables. It is explained what these variable classes are and how they interact. Furthermore, it is introduced how nuisance variables can be controlled for.

There are different research strategies. Experiments may be the most suitable strategy for assessing causal relationships. However, there are alternatives. Those are briefly introduced.



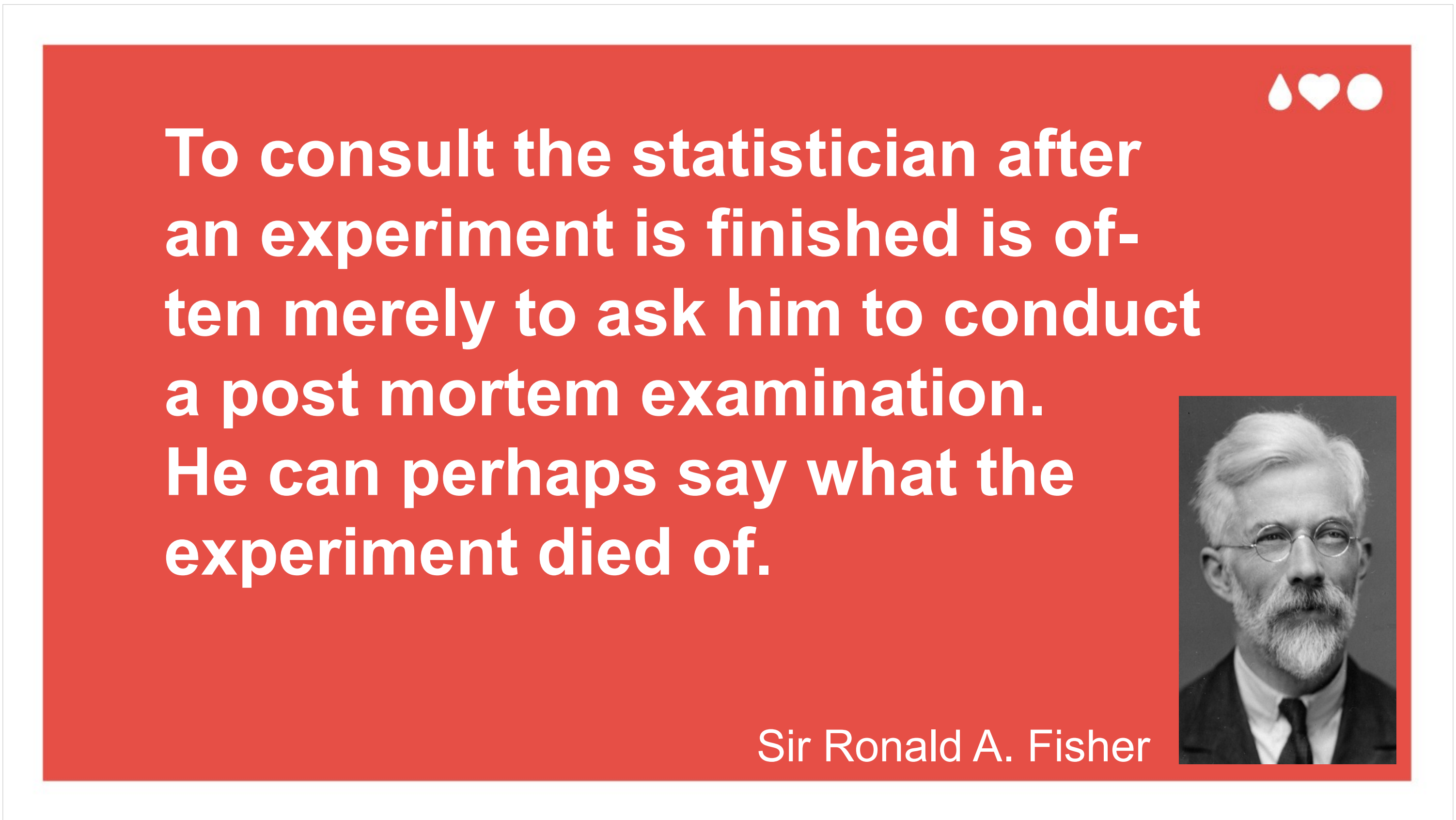
Overview

- what is an experiment?
- procedures for conducting experiments
- independent, dependent and nuisance variables
- research strategies (experiments and other)
- threats to valid inference making
- two basic experimental designs

Threats to valid inference may fall into four categories (statistical conclusion validity, internal validity, construct validity, and external validity). Threats within each of these categories are discussed. An additional section discusses possible approaches for minimizing those threats.

Finally, two experimental designs, the t-tests for Independent and Paired Samples are introduced.

The lecture is intended to clarify the “mechanics” behind experiments, and how they can be designed. Main focus is the rationale of designing them that way (e.g., to avoid particular threats to validity).



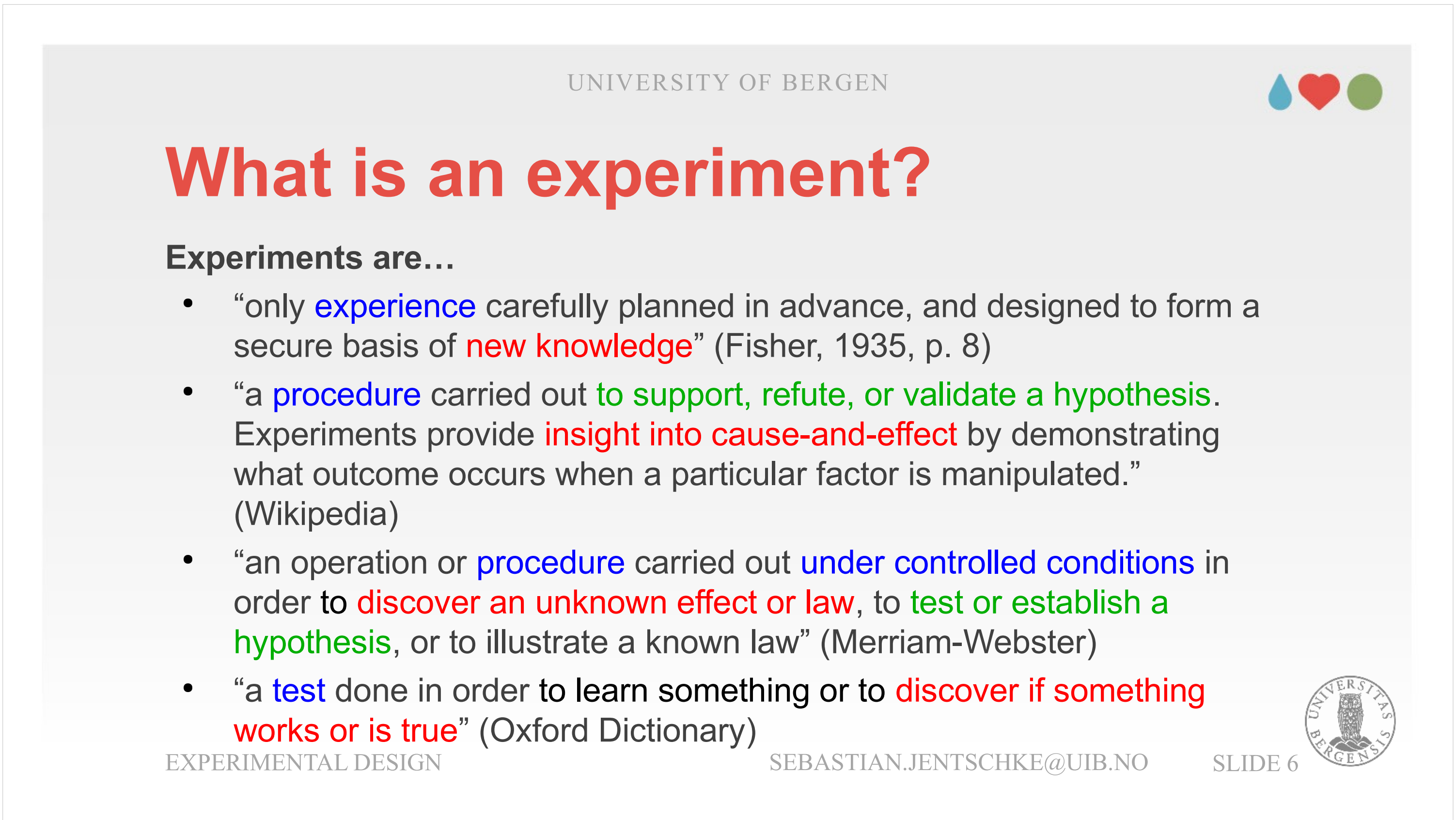
To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of.

Sir Ronald A. Fisher

One possibly shouldn’t explain jokes, but Fisher’s point denotes how crucial good experimental design is for advancing science and that all claims that we make based upon inference statistical analyses can only be as good as the experimental design under which the data were collected.

Another point is likely the quality of the manipulated or measured independent and dependent variables as well as to what degree you were able to control possible nuisance variables in your experiment. For the quality of independent and dependent variables it is, e.g., important that they are good operationalizations of your theoretical constructs or research questions, that they show enough variability (think of the problem with the decimals in the Anderson’s Iris examples or of floor and ceiling effects because a test may not differentiate well at the extremer ends of the spectrum of possible values).

What is an experiment?



What is an experiment?

Experiments are…

- “only experience carefully planned in advance, and designed to form a secure basis of new knowledge” (Fisher, 1935, p. 8)
- “a procedure carried out to support, refute, or validate a hypothesis. Experiments provide insight into cause-and-effect by demonstrating what outcome occurs when a particular factor is manipulated.” (Wikipedia)
- “an operation or procedure carried out under controlled conditions in order to discover an unknown effect or law, to test or establish a hypothesis, or to illustrate a known law” (Merriam-Webster)
- “a test done in order to learn something or to discover if something works or is true” (Oxford Dictionary)

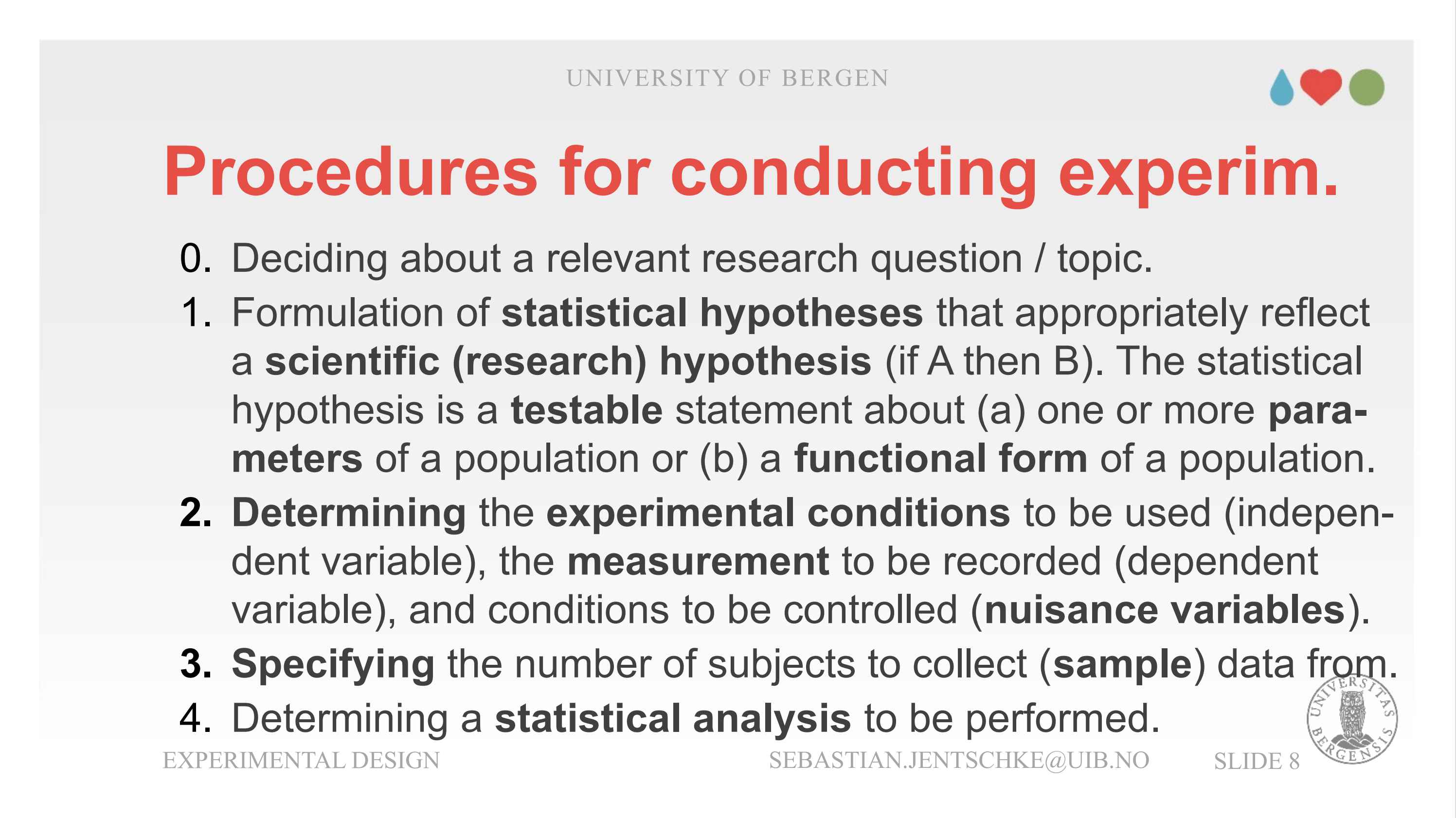
Before we speak about experimental design, we first should define what an experiment is. There is a couple of definitions collected on the page. You will see that they have some themes in common, but generally emphasize different aspects.

On common topic is that they generate new knowledge (red), often describing cause-effect-relationships.

A second common topic (blue) is that in order to gain that knowledge, experiments involve carrying out certain “tests” or “procedures” (or making a certain “experience”; this choice of word is making very clear that it is based upon a philosophical theory called empiricism).

Some definition further include that the test explores a hypothesis or that the test has to take place under controlled conditions.

Procedures for conducting experiments

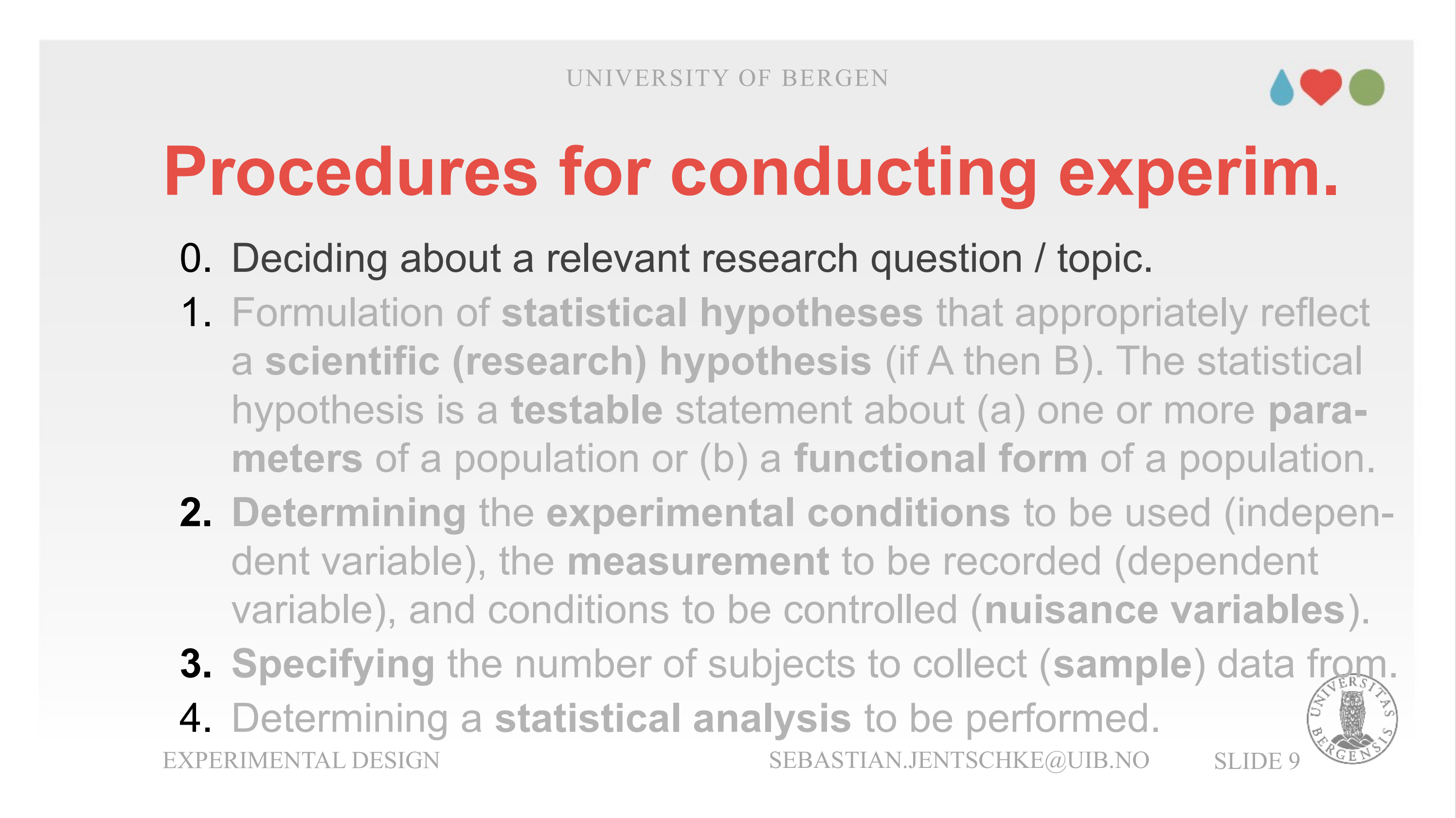


Procedures for conducting experim.

0. Deciding about a relevant research question / topic.
1. Formulation of **statistical hypotheses** that appropriately reflect a **scientific (research) hypothesis** (if A then B). The statistical hypothesis is a **testable** statement about (a) one or more **parameters** of a population or (b) a **functional form** of a population.
2. **Determining** the **experimental conditions** to be used (independent variable), the **measurement** to be recorded (dependent variable), and conditions to be controlled (**nuisance variables**).
3. **Specifying** the number of subjects to collect (**sample**) data from.
4. Determining a **statistical analysis** to be performed.

The second common point in the definitions from the previous slide indicated that an experiment involves some form of test or procedure. A typical succession of steps when conducting an experiment is shown on this slide.

Within that succession, point 2 is maybe the most central aspect when devising an experimental design. This part is typically called operationalization. However, when thinking about an experimental design, all the points above should be considered.



UNIVERSITY OF BERGEN

Procedures for conducting experim.

0. Deciding about a relevant research question / topic.
1. Formulation of **statistical hypotheses** that appropriately reflect a **scientific (research) hypothesis** (if A then B). The statistical hypothesis is a **testable** statement about (a) one or more **parameters** of a population or (b) a **functional form** of a population.
2. **Determining** the **experimental conditions** to be used (independent variable), the **measurement** to be recorded (dependent variable), and conditions to be controlled (**nuisance variables**).
3. **Specifying** the number of subjects to collect (**sample**) data from.
4. Determining a **statistical analysis** to be performed.

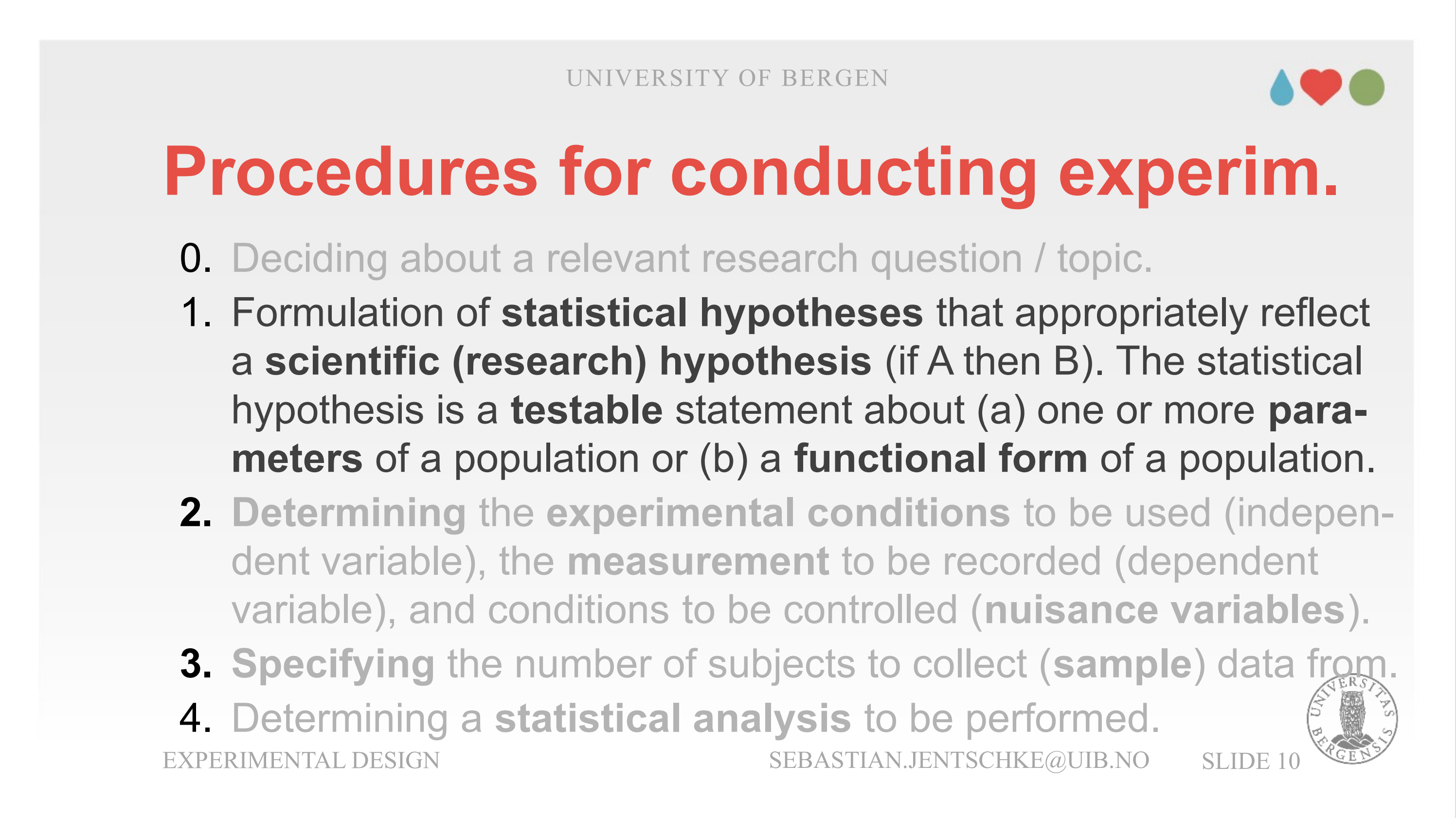
EXPERIMENTAL DESIGN SEBASTIAN.JENTSCHKE@UIB.NO SLIDE 9

Having been assigned 0. (because it is strictly speaking not an aspect of experimental design), is the most important facet: identifying relevant research questions.

This facet is related to the first common point within the definitions what an experiment is: The aim of an experiment is to generate knowledge. Such knowledge can only be thought as answer to a research question (or hypothesis) that we explore (using a scientific procedure).

Therefore, identifying questions that are relevant is decisive because if they weren’t relevant, we would just spend our time (without much benefit for us and others).

Considering whether a research questions is relevant and why always comes before questions regarding the practical implementation that are dealt with when deciding about an experimental design.



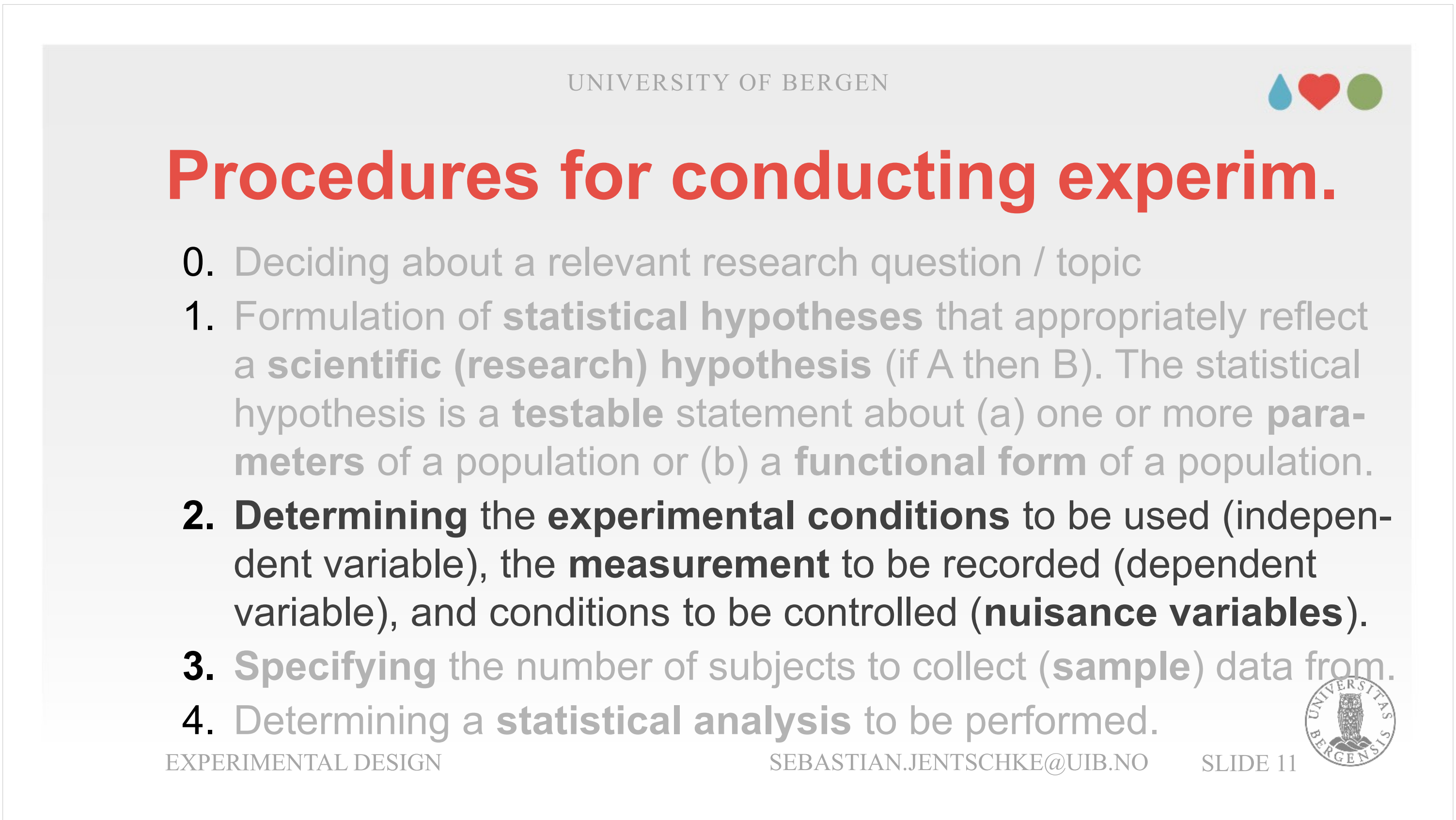
UNIVERSITY OF BERGEN

Procedures for conducting experim.

0. Deciding about a relevant research question / topic.
1. Formulation of **statistical hypotheses** that appropriately reflect a **scientific (research) hypothesis** (if A then B). The statistical hypothesis is a **testable** statement about (a) one or more **parameters** of a population or (b) a **functional form** of a population.
2. **Determining** the **experimental conditions** to be used (independent variable), the **measurement** to be recorded (dependent variable), and conditions to be controlled (**nuisance variables**).
3. **Specifying** the number of subjects to collect (**sample**) data from.
4. Determining a **statistical analysis** to be performed.

The first step after deciding about a research question is to formulate the research hypothesis precedes devising the experiment. Typically these hypotheses follow the form: If **A** , then **B**. This form describes a cause-effect-relationship.

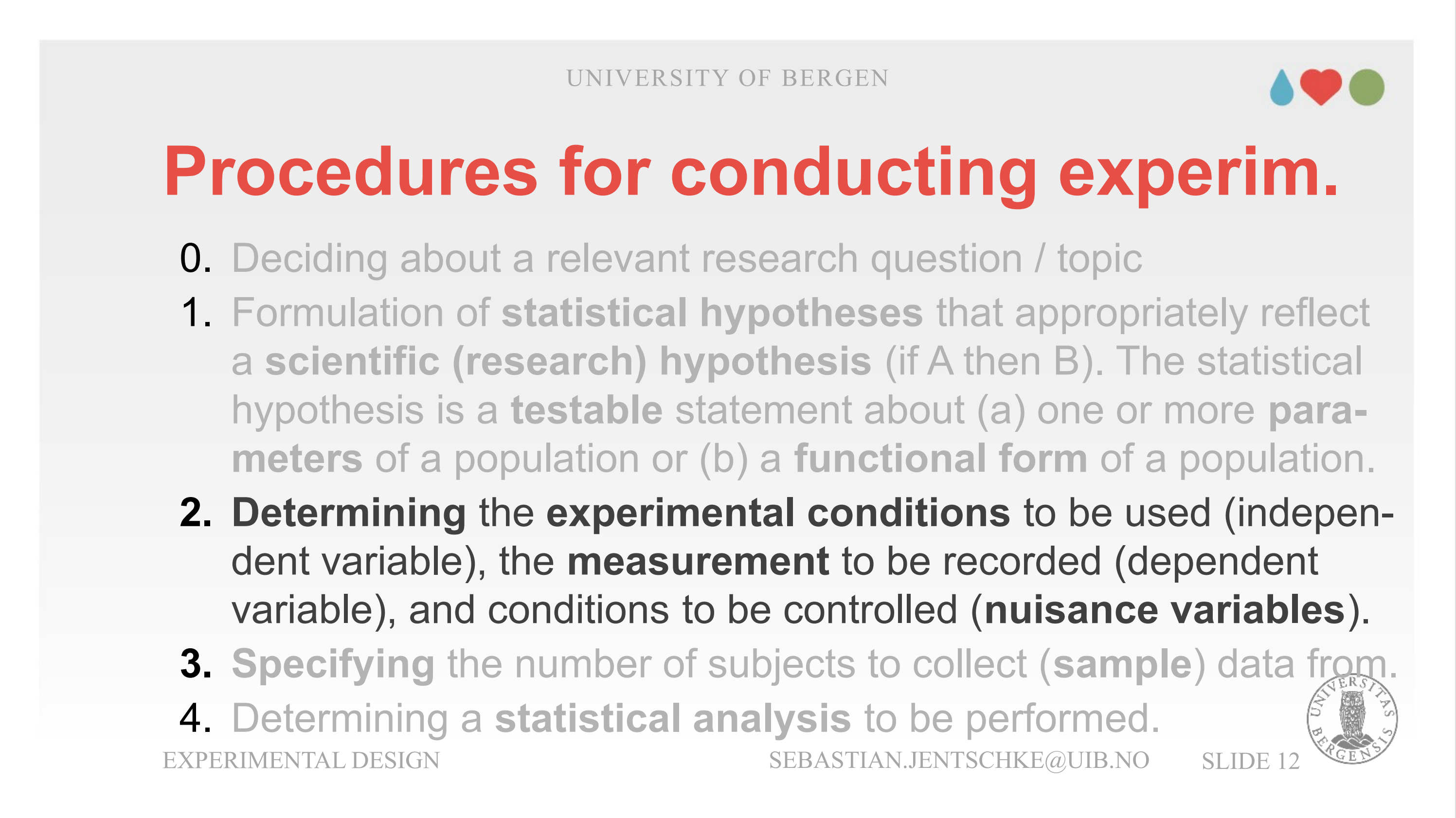
The experiment involves the manipulation of one or more variables by a researcher (**A**) to determine the effect of this manipulation on another variable (**B**). What logically follows from that is that in order to evaluate a research hypothesis, it is necessary to have procedures available to manipulate **A** and to measure **B** (or at least being able to develop procedures to do so; e.g., devising a new experiment or a new questionnaire).



Procedures for conducting experim.

0. Deciding about a relevant research question / topic
1. Formulation of **statistical hypotheses** that appropriately reflect a **scientific (research) hypothesis** (if A then B). The statistical hypothesis is a **testable** statement about (a) one or more **parameters** of a population or (b) a **functional form** of a population.
2. **Determining** the **experimental conditions** to be used (independent variable), the **measurement** to be recorded (dependent variable), and conditions to be controlled (**nuisance variables**).
3. **Specifying** the number of subjects to collect (**sample**) data from.
4. Determining a **statistical analysis** to be performed.

Core of experimental designs are to determine which experimental conditions should be manipulated (independent variable) and how this can be in the most efficient fashion (i.e., while minimizing potentially confounding effects). We also have to decide which dependent variables are going to be measured and to ensure at these are valid representations of the underlying theoretical concepts. This includes that we have to ensure to use instruments of good quality (e.g., enough variation, no floor or ceiling effects). Finally, we should carefully think about which nuisance variables could affect the causal relation we are interested in and how to control for them.

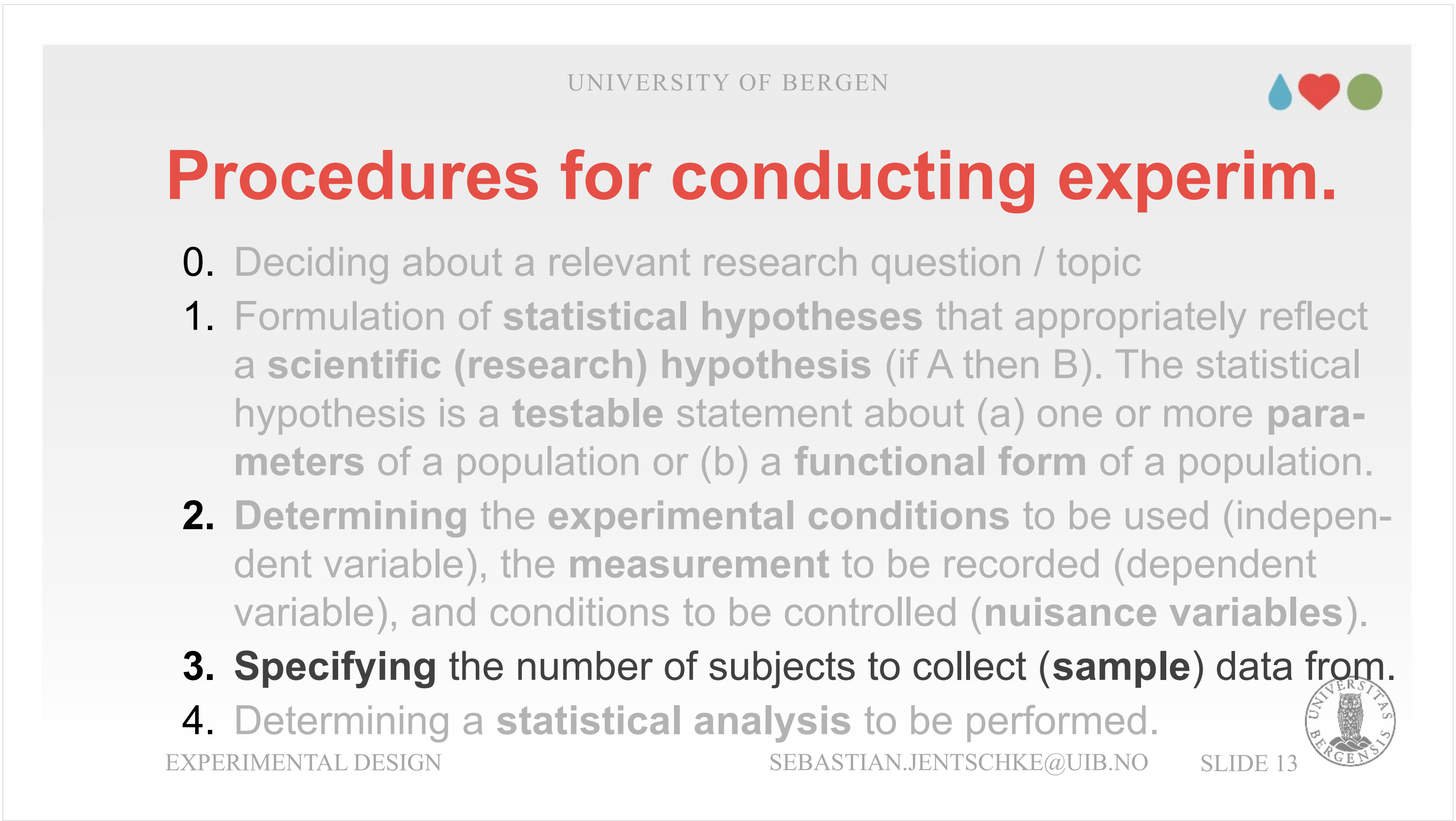


Procedures for conducting experim.

0. Deciding about a relevant research question / topic
1. Formulation of **statistical hypotheses** that appropriately reflect a **scientific (research) hypothesis** (if A then B). The statistical hypothesis is a **testable** statement about (a) one or more **parameters** of a population or (b) a **functional form** of a population.
2. **Determining** the **experimental conditions** to be used (independent variable), the **measurement** to be recorded (dependent variable), and conditions to be controlled (**nuisance variables**).
3. **Specifying** the number of subjects to collect (**sample**) data from.
4. Determining a **statistical analysis** to be performed.

Sometimes, we have to rely on nature doing the manipulations (instead of the researcher) because applying such manipulations would be unethical. Examples are studying the effects of certain conditions in the environment (e.g., prenatal malnutrition), or assessing the consequences of a certain medical condition (e.g., lack of oxygen supply) on certain behaviour (e.g., cognitive skills such as IQ or memory performance).

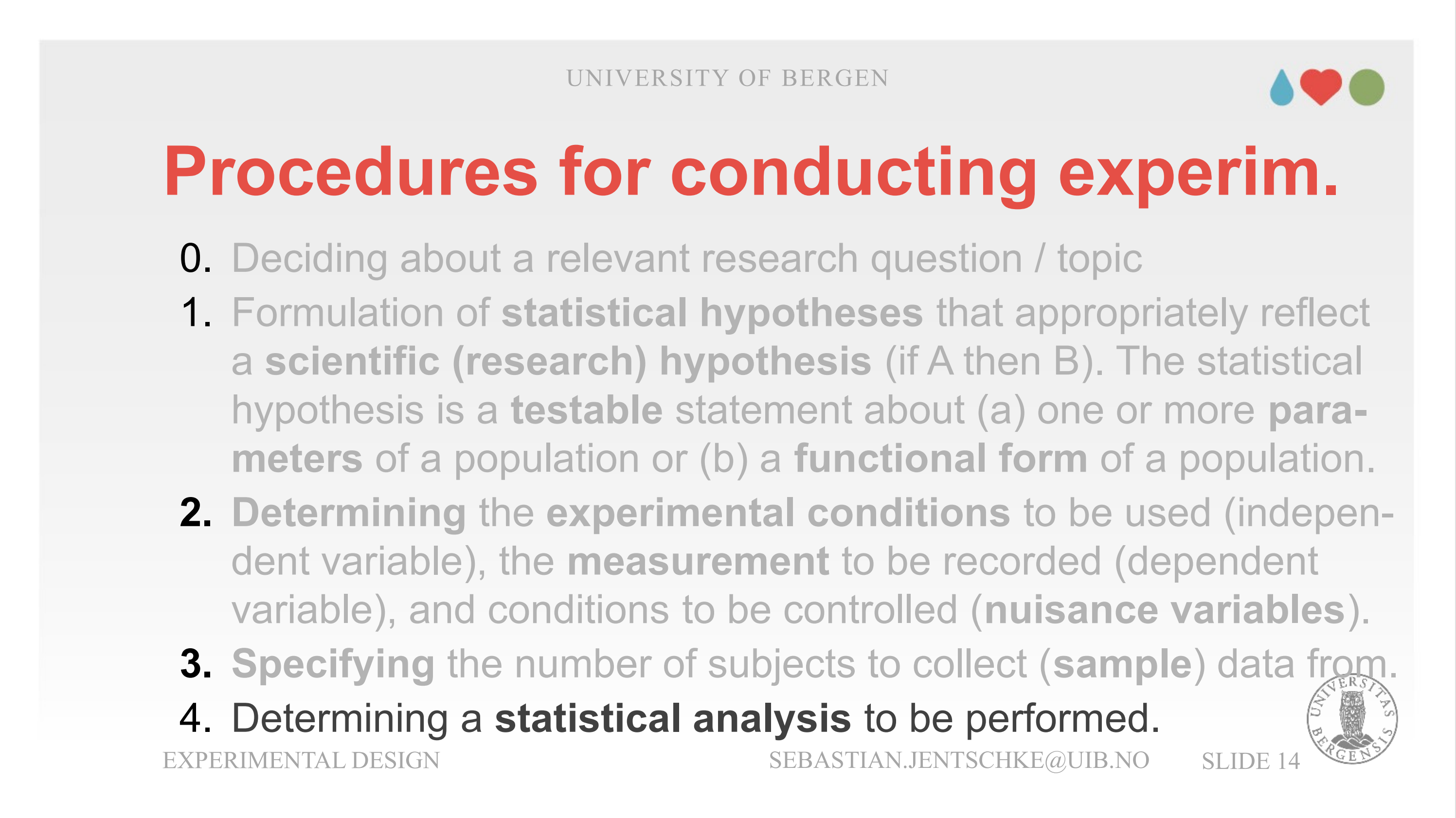
Ex-post-facto studies, surveys, case studies, and naturalistic observation (which are introduced later) fall within that category.



Procedures for conducting experim.

0. Deciding about a relevant research question / topic
1. Formulation of **statistical hypotheses** that appropriately reflect a **scientific (research) hypothesis** (if A then B). The statistical hypothesis is a **testable** statement about (a) one or more **parameters** of a population or (b) a **functional form** of a population.
2. **Determining** the **experimental conditions** to be used (independent variable), the **measurement** to be recorded (dependent variable), and conditions to be controlled (**nuisance variables**).
3. **Specifying** the number of subjects to collect (**sample**) data from.
4. Determining a **statistical analysis** to be performed.

After designing the experiment, we turn to specifiying the sample we wish to collect data from. Connected with that is the question of how to assign subjects to experimental conditions.



UNIVERSITY OF BERGEN

Procedures for conducting experim.

0. Deciding about a relevant research question / topic
1. Formulation of **statistical hypotheses** that appropriately reflect a **scientific (research) hypothesis** (if A then B). The statistical hypothesis is a **testable** statement about (a) one or more **parameters** of a population or (b) a **functional form** of a population.
2. **Determining** the **experimental conditions** to be used (independent variable), the **measurement** to be recorded (dependent variable), and conditions to be controlled (**nuisance variables**).
3. **Specifying** the number of subjects to collect (**sample**) data from.
4. Determining a **statistical analysis** to be performed.

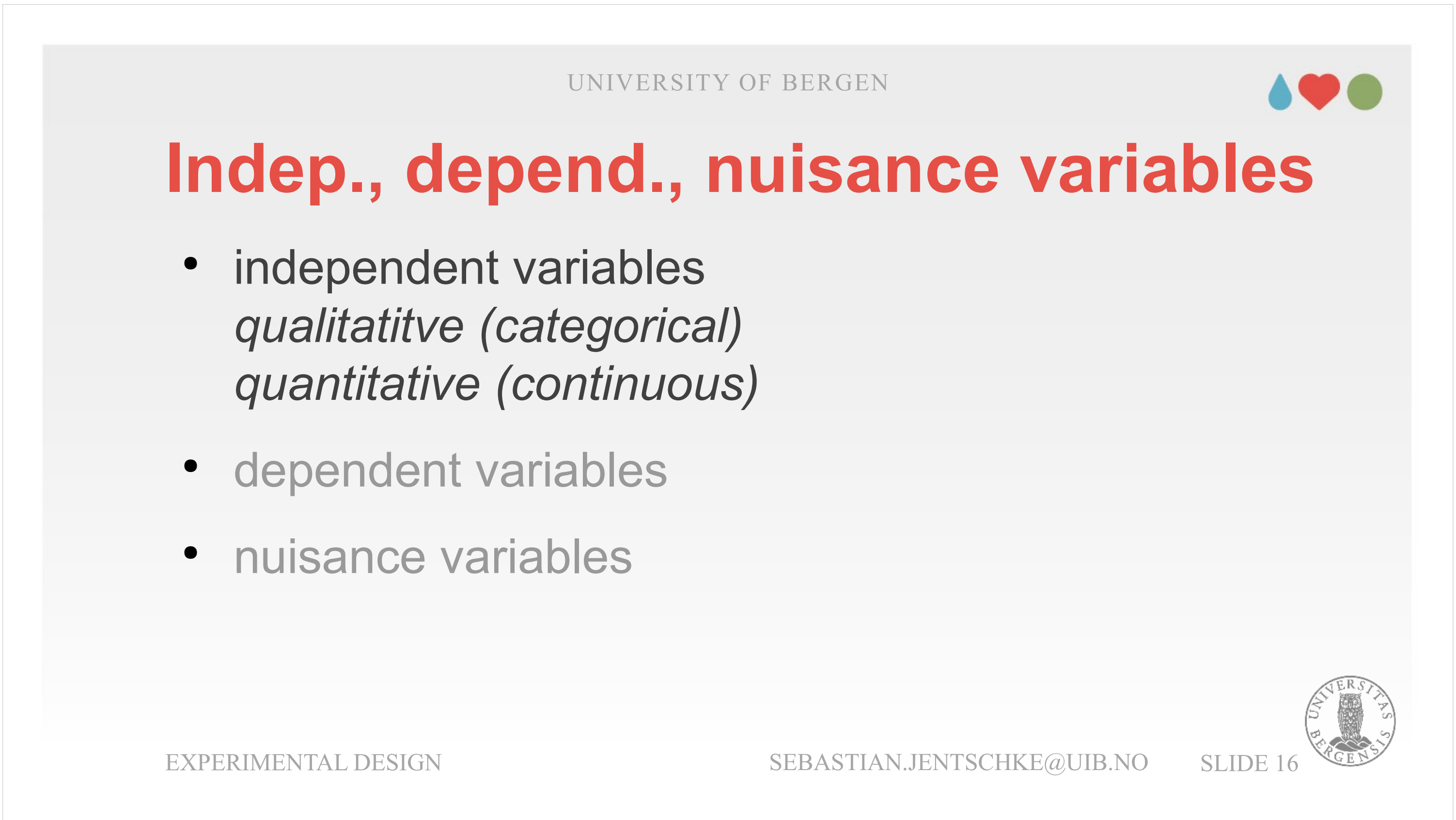
EXPERIMENTAL DESIGN SEBASTIAN.JENTSCHKE@UIB.NO SLIDE 14

The final step is to determine the statistical analyses associated with that plan. How you wish to evaluate your data is a crucial part of designing an experiment since the analyses may constrain your experimental conditions (e.g., when it comes to what type of variables you can collect).

All these steps prepare us for data collection which is the most costly and time-consuming aspect of an experiment. Careful conducting these steps helps to ensure that the data can be used to maximum advantage. If an experiment is badly designed, there is little likelihood of extracting useful information from it, even with advanced statistical analyses.

That several things can go wrong is also a strong argument for pilot testing. If we can detect mistakes at this stage, this prevents us from rendering the outcome of the experiment unanalyzable or invalid.

Independent, dependendent, and nuisance variables

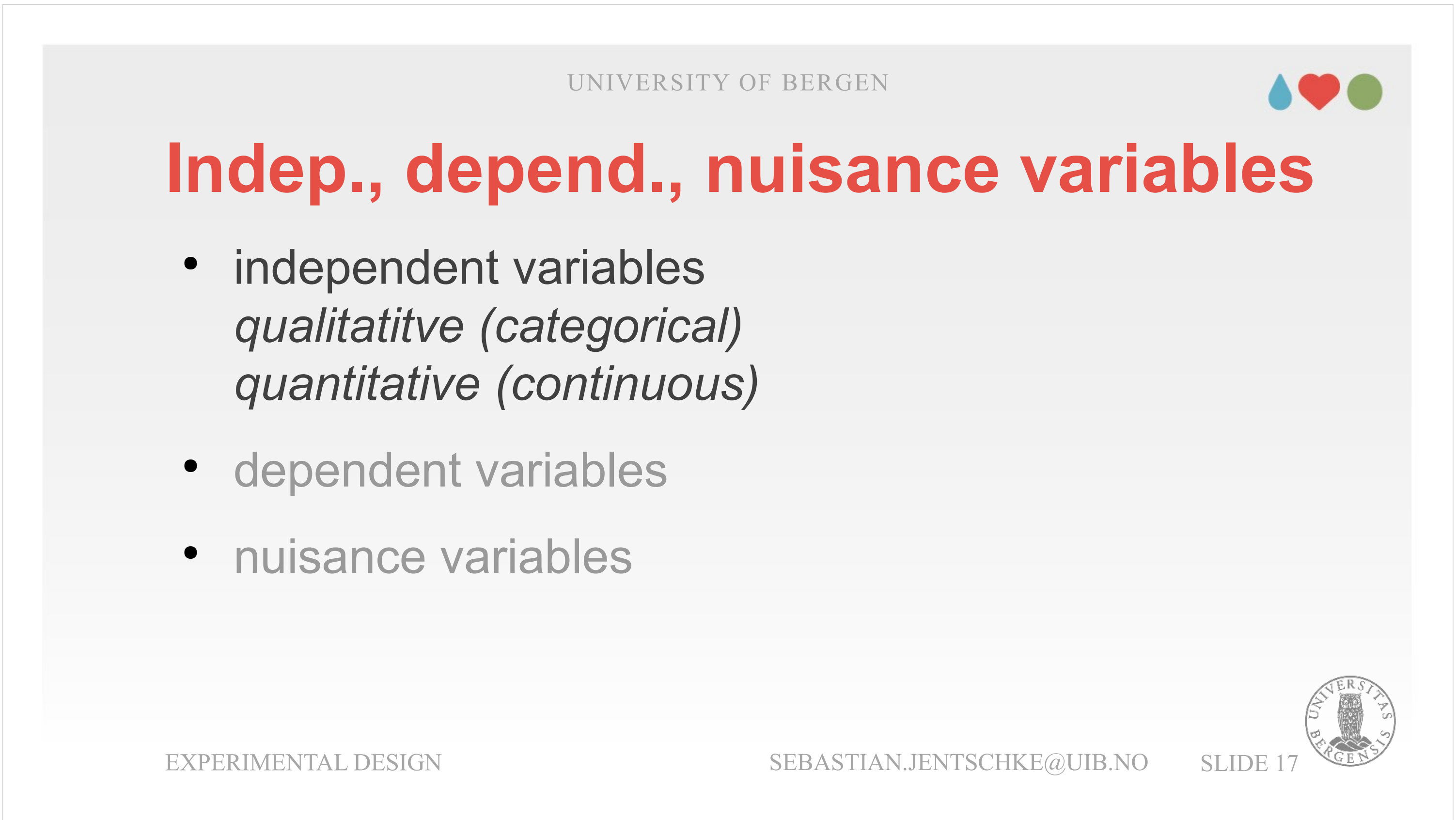


Indep., depend., nuisance variables

- independent variables
  qualitatitve (categorical)
  quantitative (continuous)
- dependent variables
- nuisance variables

Central to experimental design are the selection of one or more independent variables, and typically one dependent variable.

In addition, consideration is required which so-called nuisance variables need to be considered and controlled for.



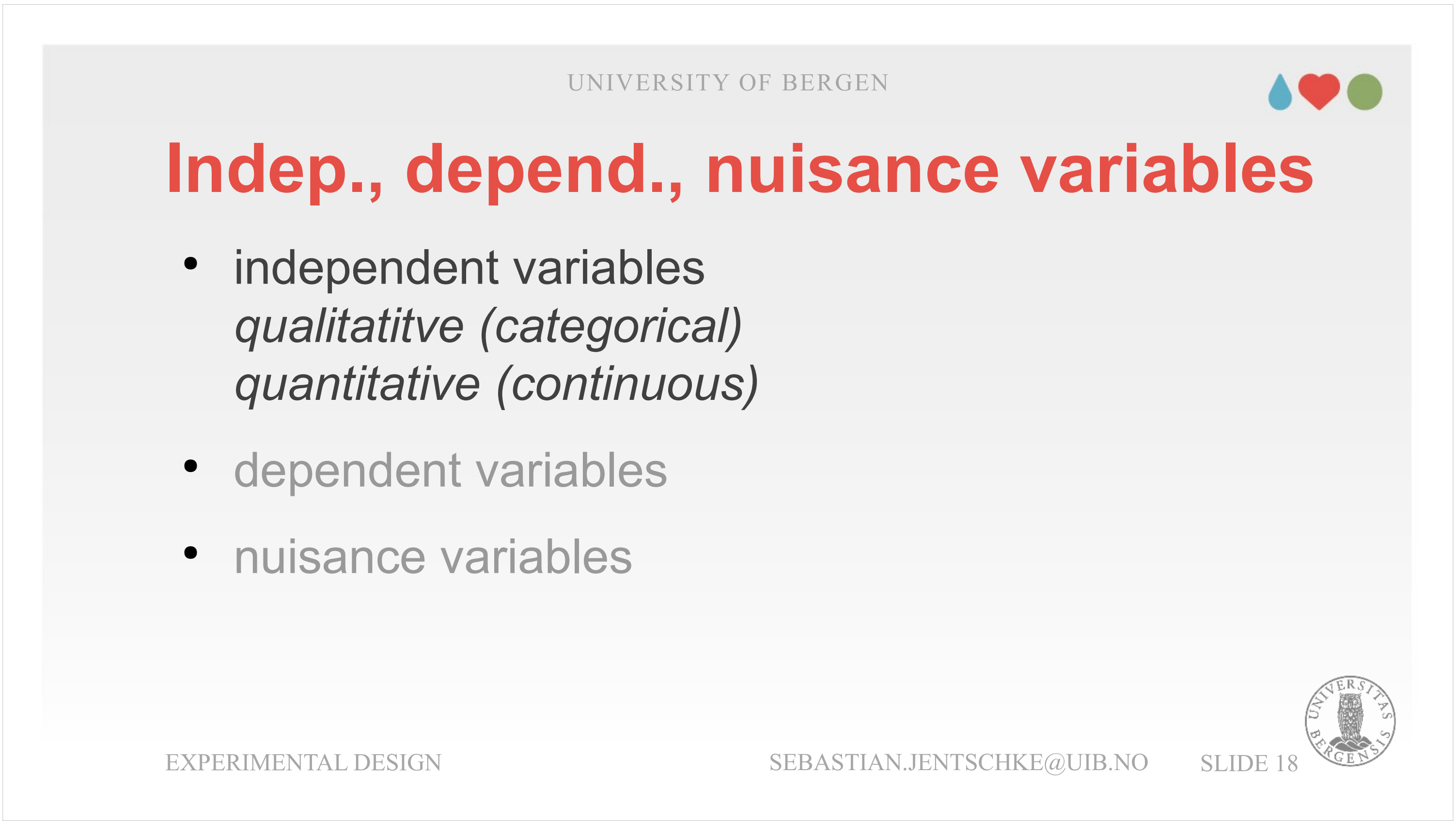
Indep., depend., nuisance variables

- independent variables
  qualitatitve (categorical)
  quantitative (continuous)
- dependent variables
- nuisance variables

The Independent Variable is what was denoted “A” (if “A”) on a previous slide.

It can be either qualitative or quantitative. If it is ***qualitative*** we have a ***categorical distinction***, i.e., the treatment levels represent different kinds of the independent variable (e.g., “treatment” vs. “control” / “placebo” or different types of interventions, e.g., medication vs. cognitive-behavioural therapy vs. hypnosis).

If it is ***quantitative***, it is best described as ***different amounts*** (e.g., different dosages of a medication, how often or how long an intervention is conducted, etc.).



Indep., depend., nuisance variables

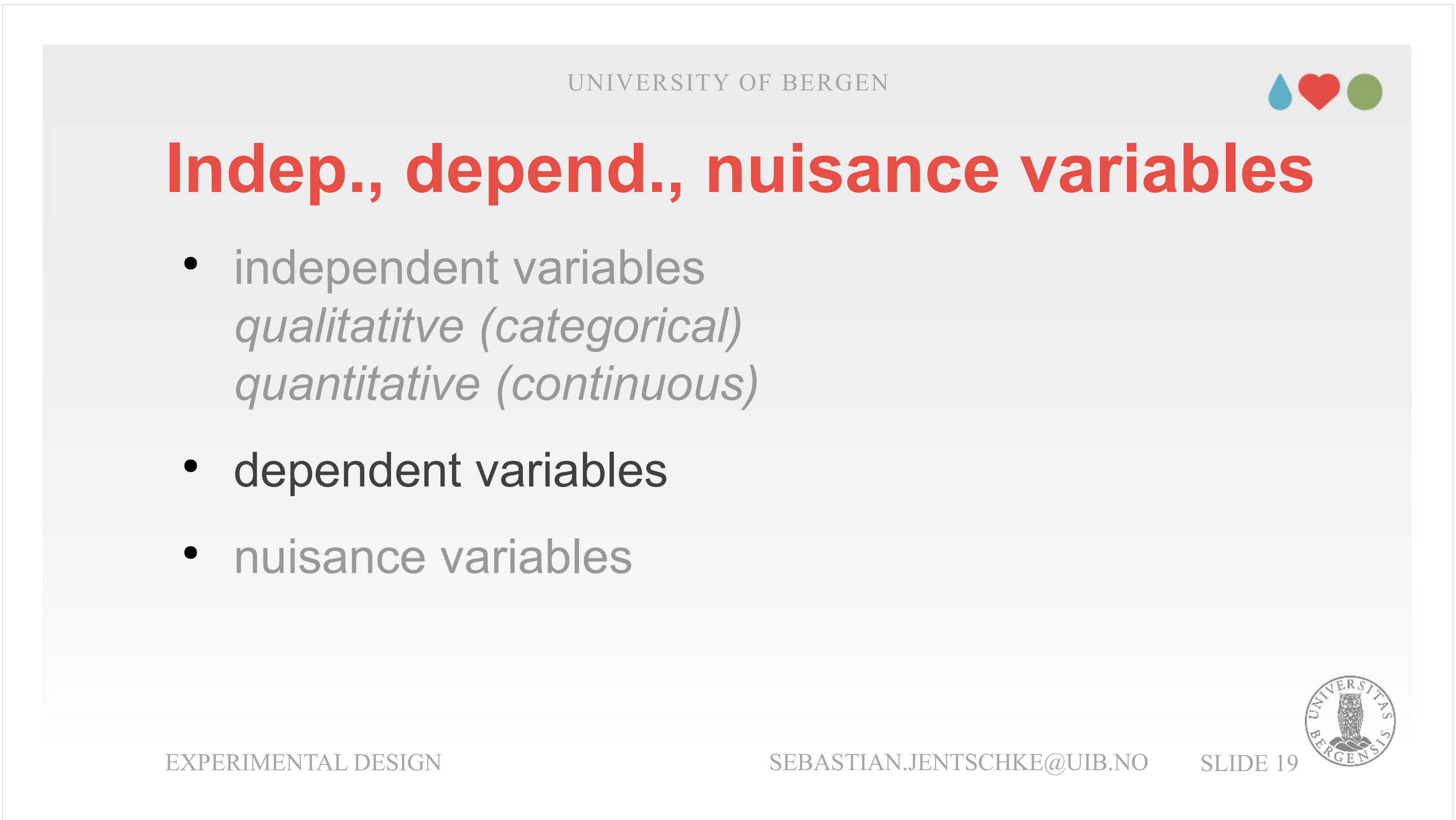
- independent variables
  qualitatitve (categorical)
  quantitative (continuous)
- dependent variables
- nuisance variables

It was mentioned before that a statistical hypothesis takes the form of either “(a) one or more ***parameters of a population*** or (b) a ***functional form of a population***.”

(a) typically assesses ***differences in the mean*** between conditions or groups and is related to qualitative independent variables.

(b) assesses ***relations*** (e.g., whether, if the dosage of medication A is increased, the number of hallucinations is reduced) and is related to quantitative independent variables.

This distinction is quite similar to the one I made in the introduction lecture even though I spoke of categorical and continuous variables: categorical denotes qualitative (independent) variables, continuous denotes quantitative (independent) variables.



UNIVERSITY OF BERGEN

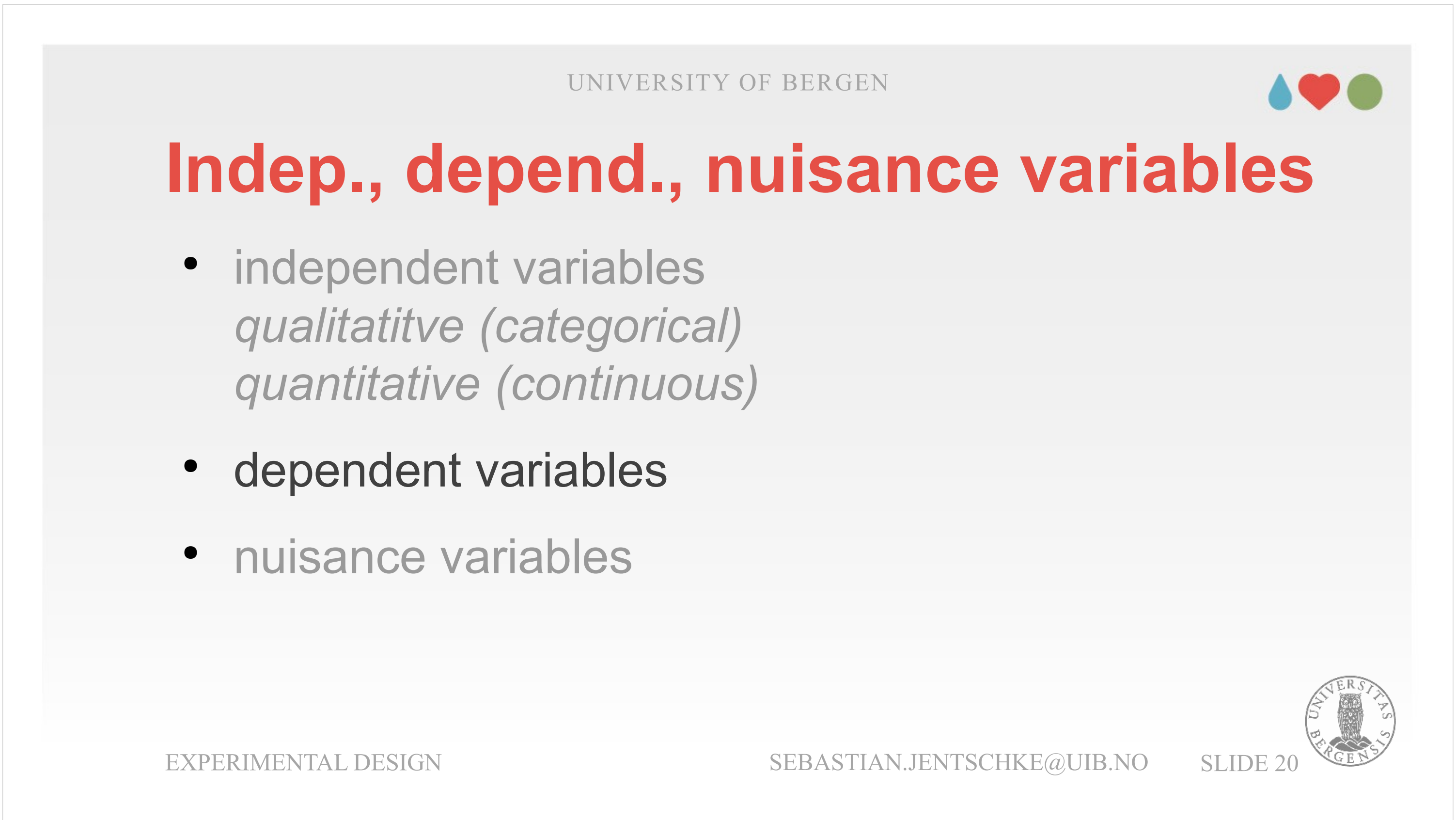
Indep., depend., nuisance variables

- independent variables
  qualitatitve (categorical)
  quantitative (continuous)
- dependent variables
- nuisance variables

EXPERIMENTAL DESIGN SEBASTIAN.JENTSCHKE@UIB.NO SLIDE 19

The **Dependent Variable** is what was denoted “B” (then “B”) on a previous slide. What an appropriate dependent variable may be is based on theoretical considerations as well as practical consideration ones.

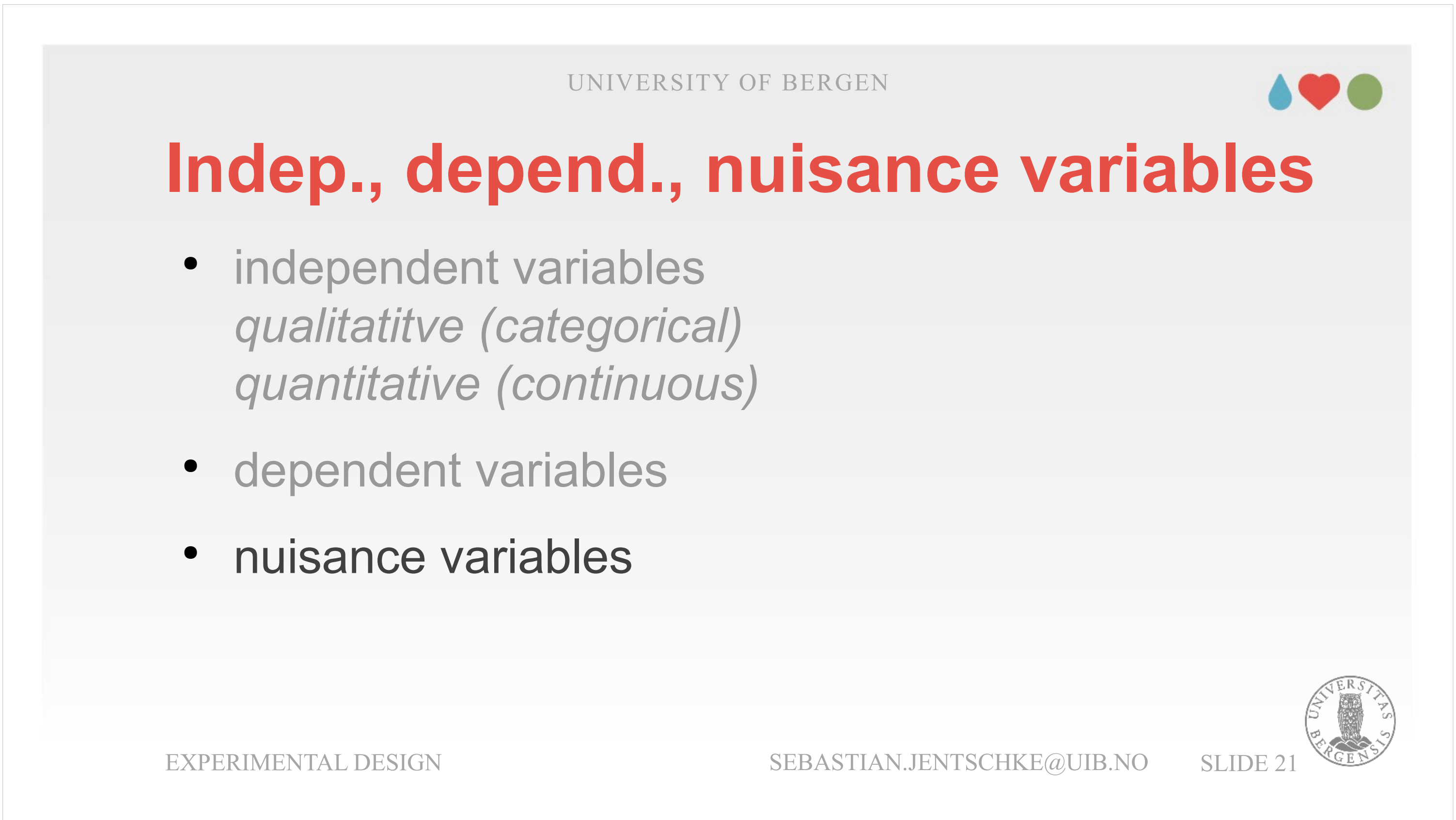
Theoretical considerations include two facets: one being a theoretical description of the property that is measured (e.g., intelligence) and how this property is measured (e.g., if somebody is intelligent, this is indicated by high scores in certain tests). Often such a theoretical consideration is called a psychological construct. Such construct is a label for a cluster or domain of covarying behaviours (if somebody reaches high test scores in specific tests or if somebody shows certain behaviour in a defined situation or under specific circumstances).



Indep., depend., nuisance variables

- independent variables
  qualitatitve (categorical)
  quantitative (continuous)
- dependent variables
- nuisance variables

Practical considerations may include whether there is already a validated instrument available (e.g., a standardized test) or whether we have to develop one; how “representative” a certain behaviour is for measuring a certain construct; how valid and reliable that measure is (more on validity and reliability later); and whether we can expect that this measure follows a normal distribution (which is required for most statistical tests).



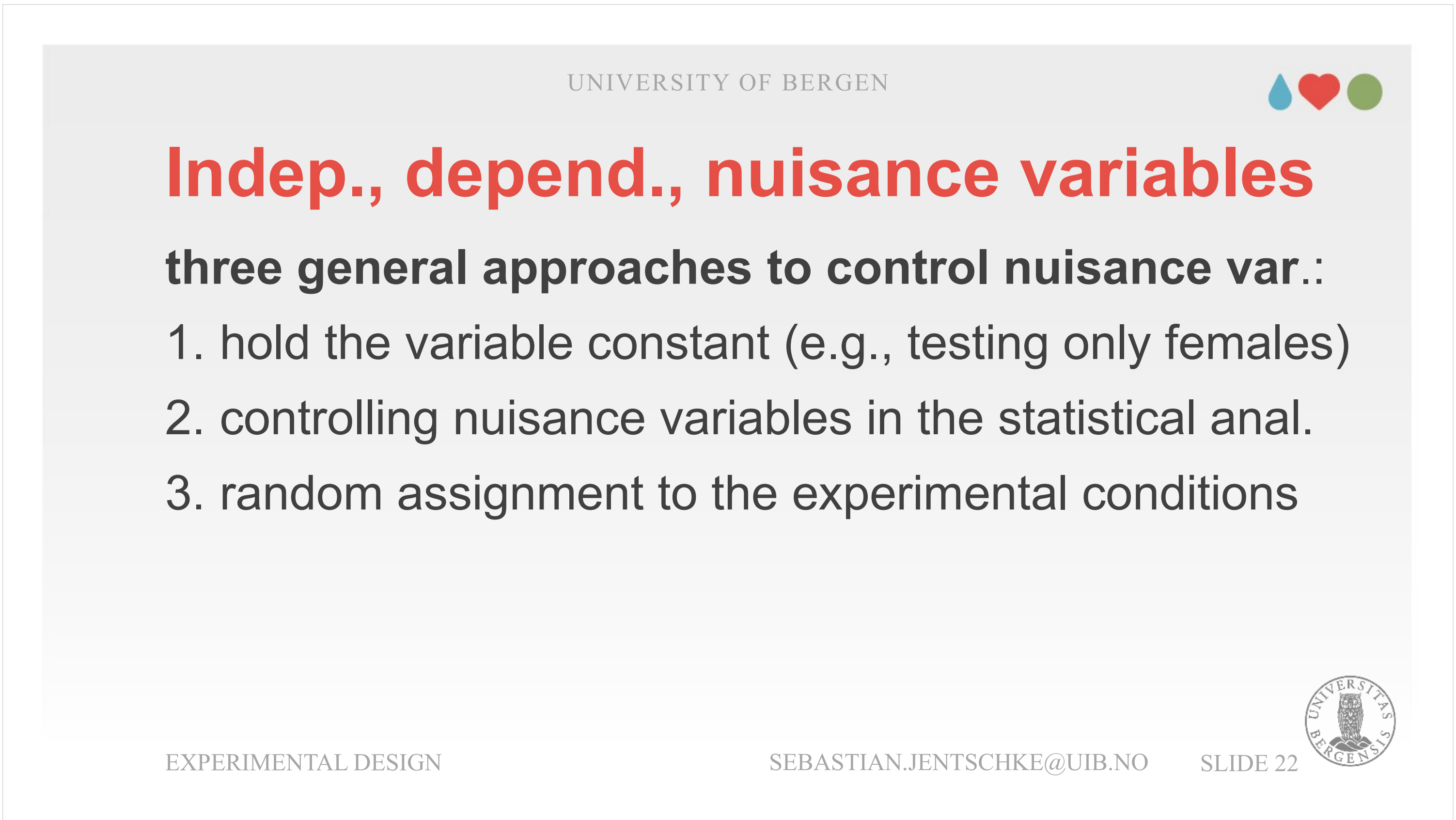
UNIVERSITY OF BERGEN

Indep., depend., nuisance variables

- independent variables
  qualitatitve (categorical)
  quantitative (continuous)
- dependent variables
- nuisance variables

EXPERIMENTAL DESIGN SEBASTIAN.JENTSCHKE@UIB.NO SLIDE 21

Nuisance Variables denote undesired sources of variation in an experiment that affect the dependent variable. As the name implies, nuisance variables are of no interest per se. These nuisance variables can either modulate the independent variable or affect the measurement of the dependent variable and include, e.g., the presentation of instructions, environmental factors such as room illumination, noise level, and room temperature, the calibration of the measurement instrument, the state participants are in (excited, a little frightened, bored), etc.



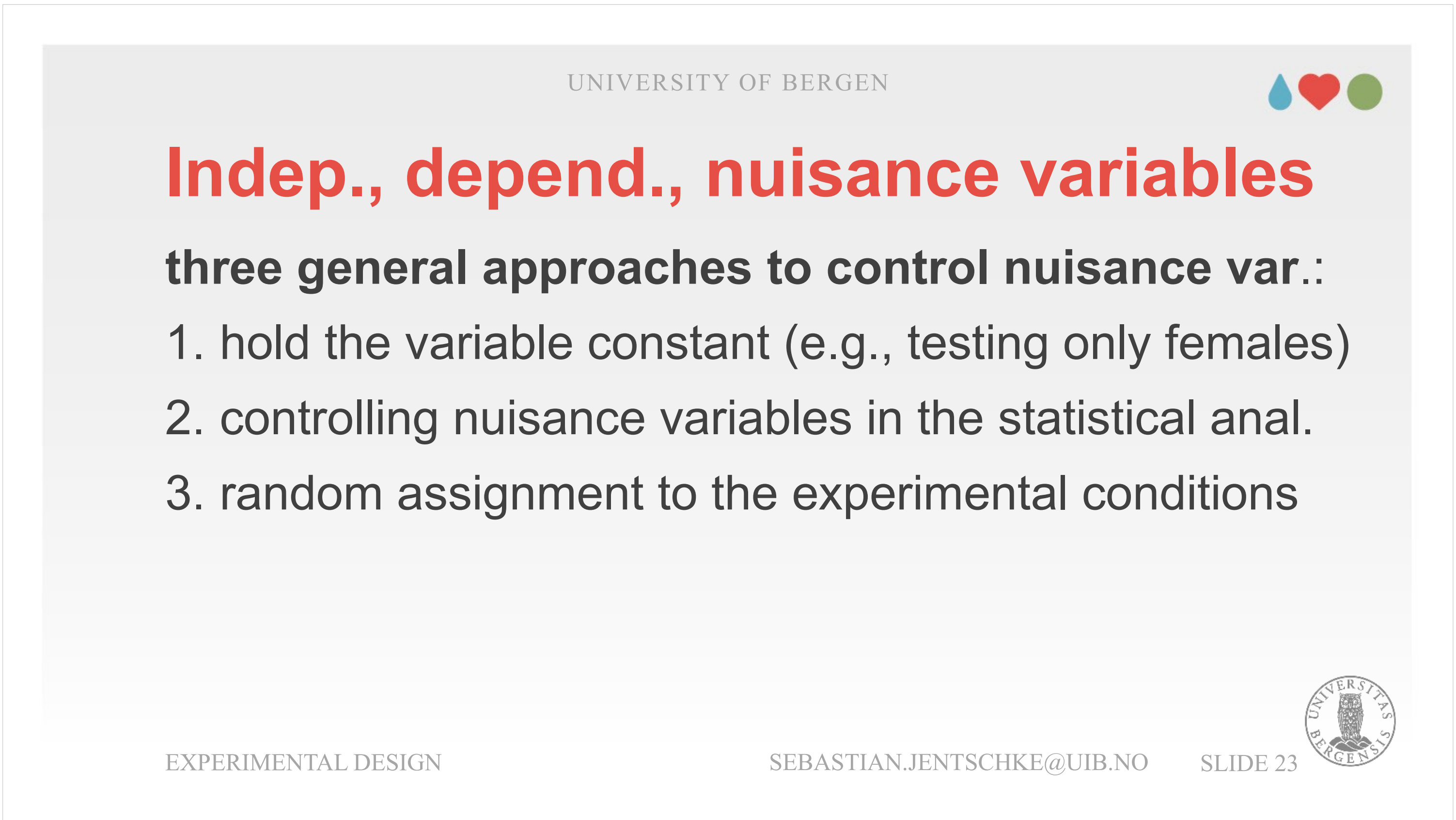
Indep., depend., nuisance variables

three general approaches to control nuisance var.:

1. hold the variable constant (e.g., testing only females)
2. controlling nuisance variables in the statistical anal.
3. random assignment to the experimental conditions

There are three general approaches to control Nuisance variables. Those include to hold the variable constant (e.g., testing only one sex, e.g., females), controlling nuisance variables by including them statistical analyses or by random assignment of participants to the experimental conditions.

The first two categories try to minimize or control the influence of nuisance variables, thereby reducing error variance. Minimizing would, e.g., include holding the experimental settings (e.g., room, noise level, instructions, etc.) constant. In addition, there may be nuisance variables we can not control but possibly measure, e.g. “natural” sources of variation such as gender differences. If a nuisance variable distorts results in a particular direction, the effect is called “bias”. Even when the effect of the nuisance variable is randomly distributed, it typically increases the error variance.

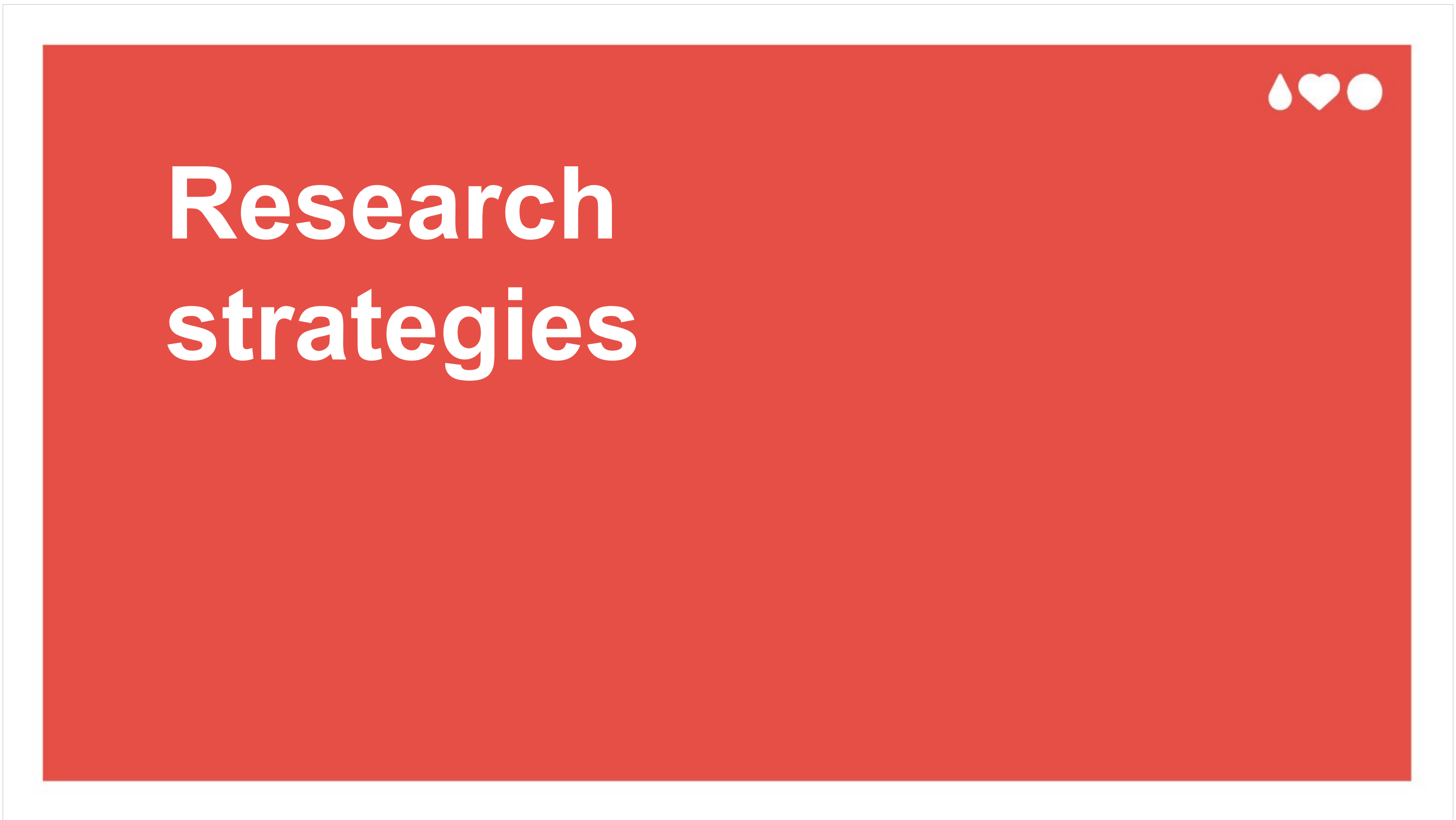


Indep., depend., nuisance variables

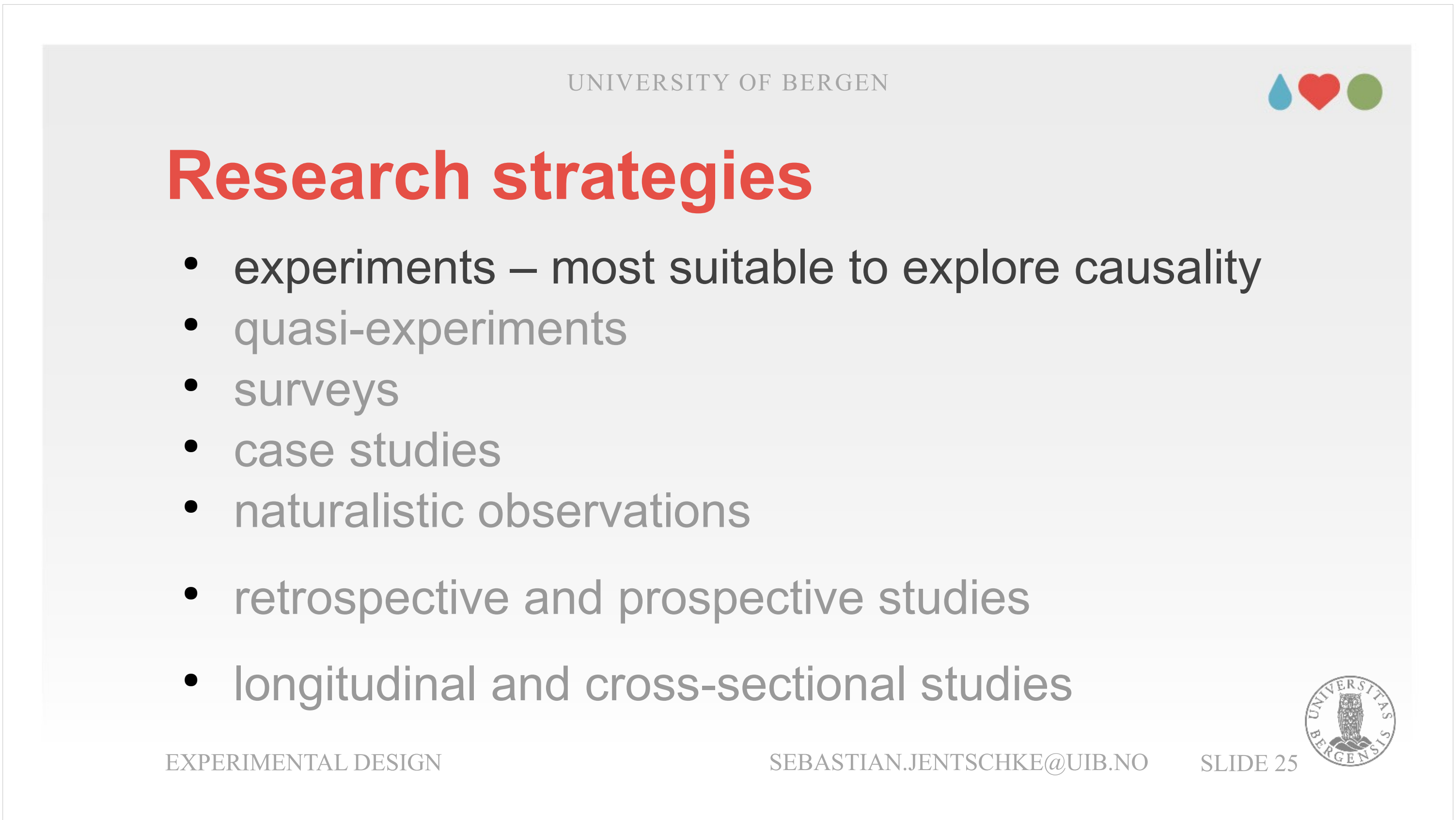
three general approaches to control nuisance var.:

1. hold the variable constant (e.g., testing only females)
2. controlling nuisance variables in the statistical anal.
3. random assignment to the experimental conditions

Error variance denotes the variability of the dependent variable that cannot be attributed to the effects of manipulating the independent variable. Randomization is another way of controlling for error variance. It aims to “spread” the error variance equally among the experimental conditions. Whereas the first two strategies only have an effect on variables that the experimenter is aware of (and that are controlled or minimized), is randomizing working for any kind of nuisance variable (i.e., even those unconsidered).



Research strategies

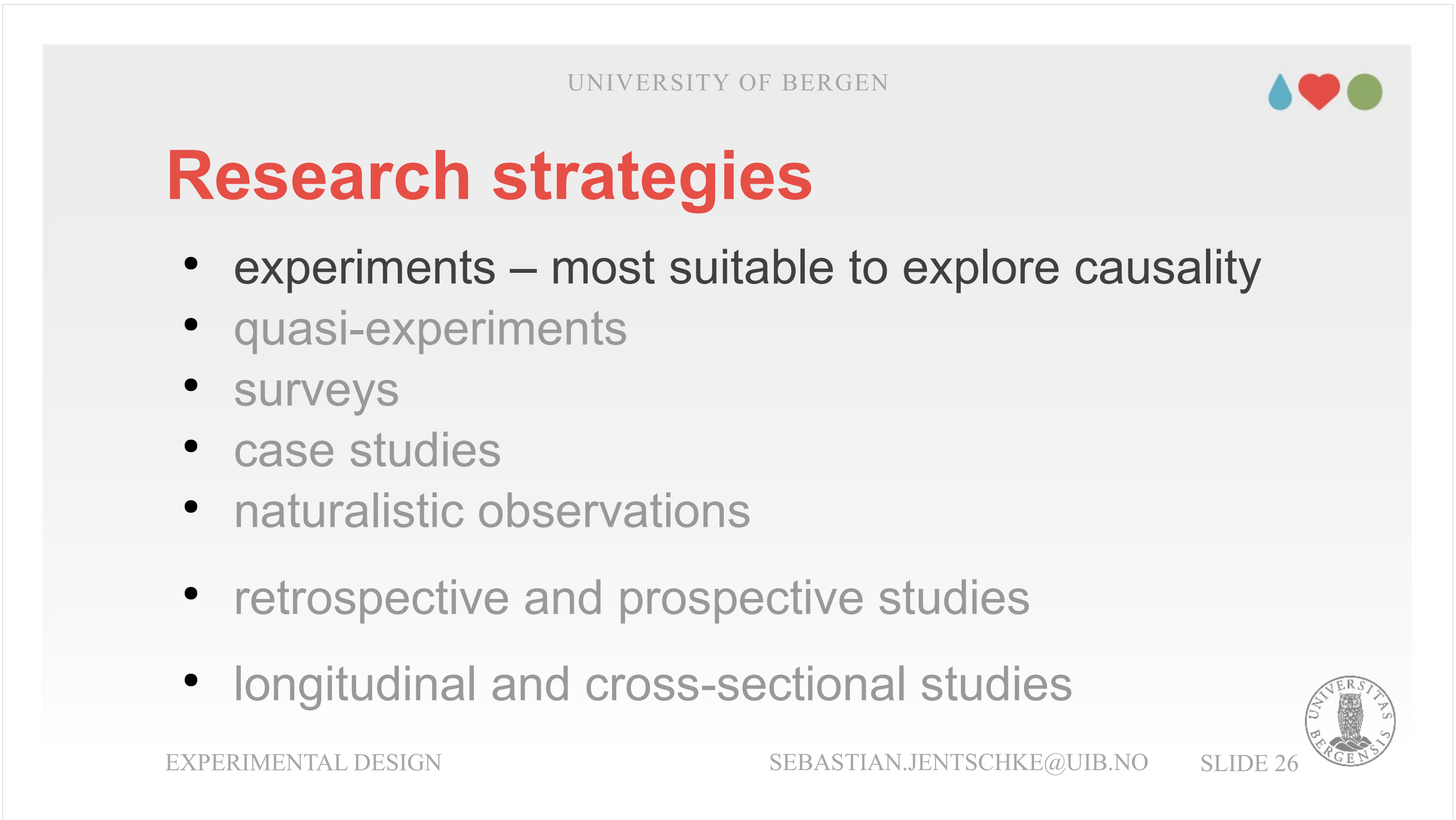


Research strategies

- experiments – most suitable to explore causality
- quasi-experiments
- surveys
- case studies
- naturalistic observations
- retrospective and prospective studies
- longitudinal and cross-sectional studies

There are different strategies for answering research questions.

Experiments represent the prime strategy. In accordance with the definitions at the very begin of this lecture, Kirk describes experiments as ***testing*** “a ***hypothesized relationship*** between an **independent** variable and a ***dependent variable*** by ***manipulating*** the ***independent*** variable […] performed in an ***environment*** that permits a ***high*** degree of ***control of nuisance variables***”.



UNIVERSITY OF BERGEN

Research strategies

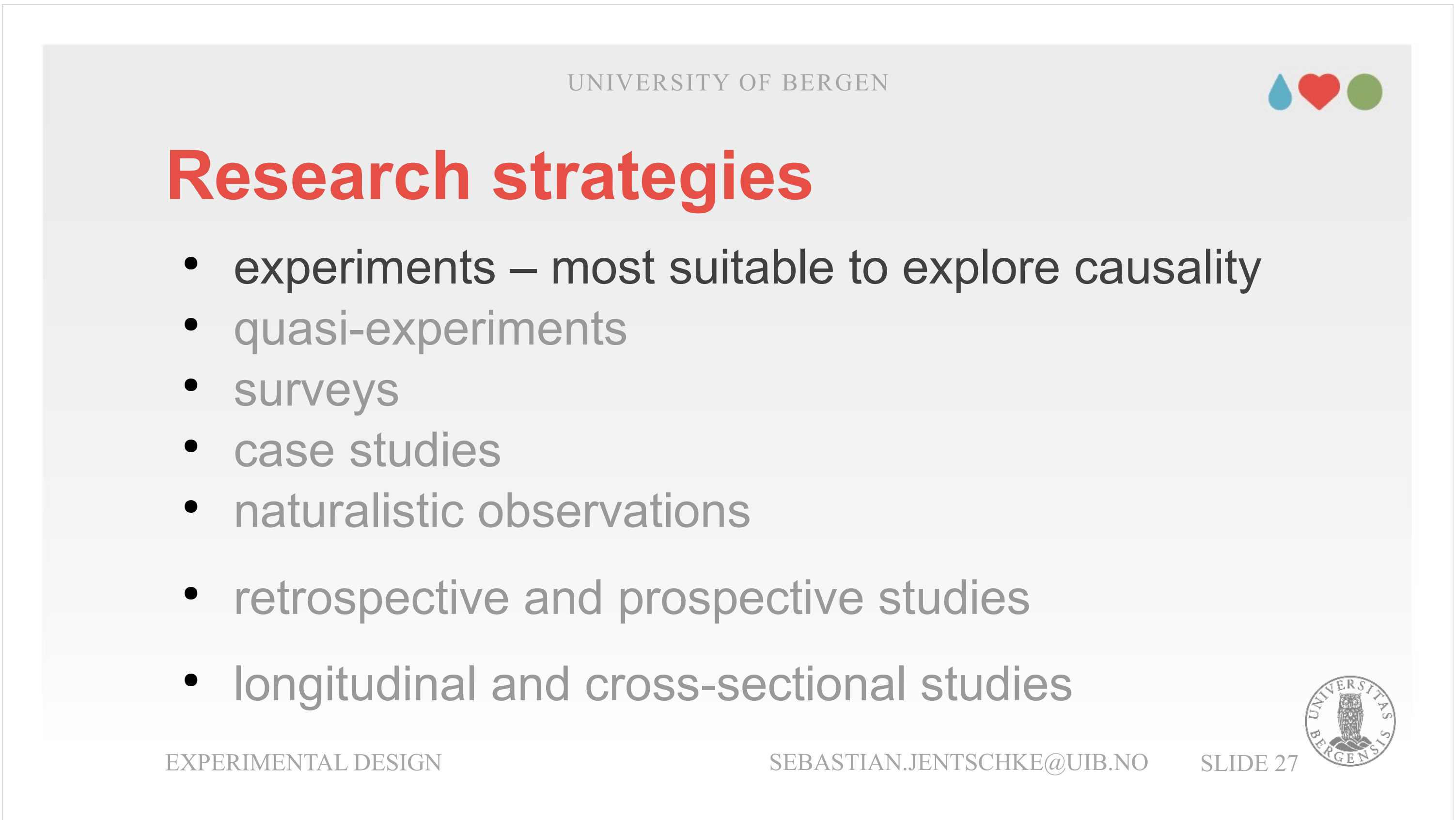
- experiments – most suitable to explore causality
- quasi-experiments
- surveys
- case studies
- naturalistic observations
- retrospective and prospective studies
- longitudinal and cross-sectional studies

EXPERIMENTAL DESIGN SEBASTIAN.JENTSCHKE@UIB.NO SLIDE 26

These claims can also be put in the form of a procedure: “(1) manipulation […] one or more independent variables, (2) use of controls such as randomly assigning subjects or experimental units to the experimental conditions, and (3) careful observation or measurement of one or more dependent variables”.

Among these steps, (2) relates to controlling nuisance variables, (1) and (3) to the cause and effect side of a causal relationship.

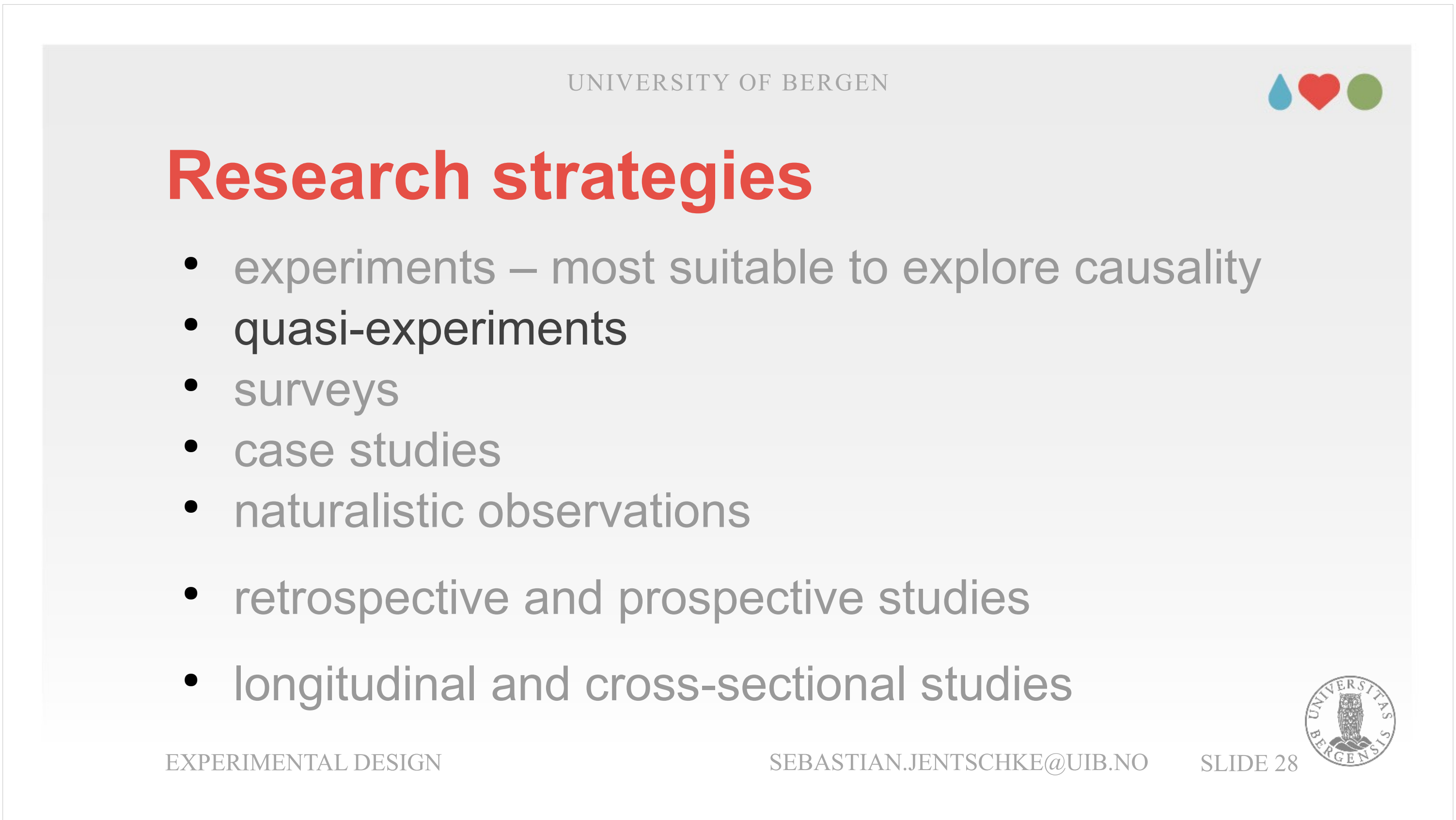
To infer causality, it is required that: (a) The ***cause***, the manipulation of the independent variable, ***precedes*** the ***effect*** on the dependent variable (called temporal precedence). (b) ***Whenever*** the ***cause*** is ***present***, a ***certain effect occurs*** (called sufficiency); and (c) the ***effect*** occurs ***only if*** the ***cause is present*** (called necessity).



Research strategies

- experiments – most suitable to explore causality
- quasi-experiments
- surveys
- case studies
- naturalistic observations
- retrospective and prospective studies
- longitudinal and cross-sectional studies

Using these procedures and rules, experiments have a considerable power to help gaining knowledge regarding cause-effect-relationships.



UNIVERSITY OF BERGEN

Research strategies

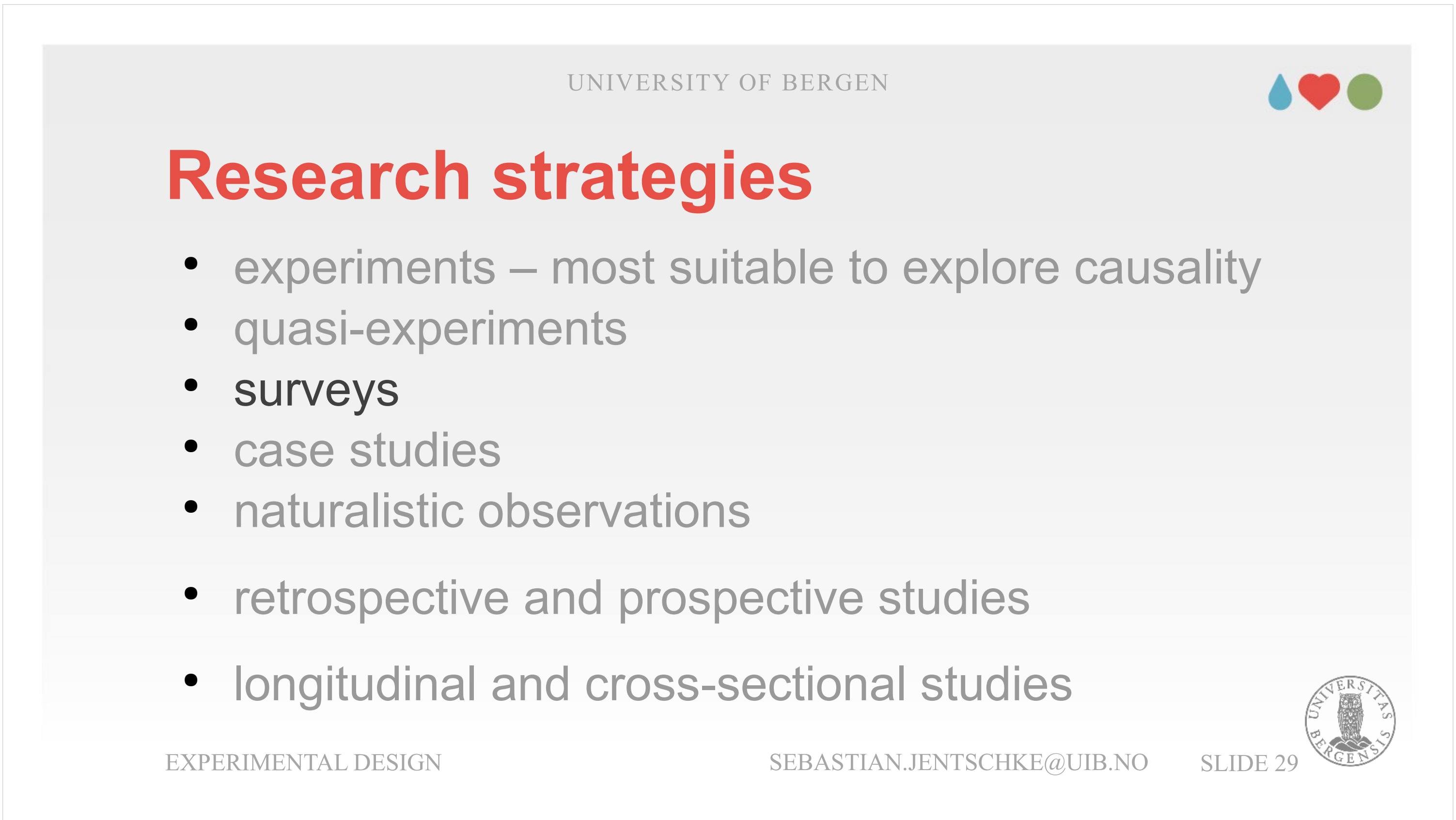
- experiments – most suitable to explore causality
- quasi-experiments
- surveys
- case studies
- naturalistic observations
- retrospective and prospective studies
- longitudinal and cross-sectional studies

EXPERIMENTAL DESIGN SEBASTIAN.JENTSCHKE@UIB.NO SLIDE 28

Quasi-experiments are much alike experiments. The crucial difference is, that they ***don't assign participants randomly*** to conditions and which is a disadvantage with respect to controlling for possible nuisance variables (randomization assumes that effects of nuisance variables distributes equally over experimental conditions).

Their interpretation is thus often less straightforward: It is difficult to rule out other influences than the manipulation of the independent variable as explanations for an observed difference.

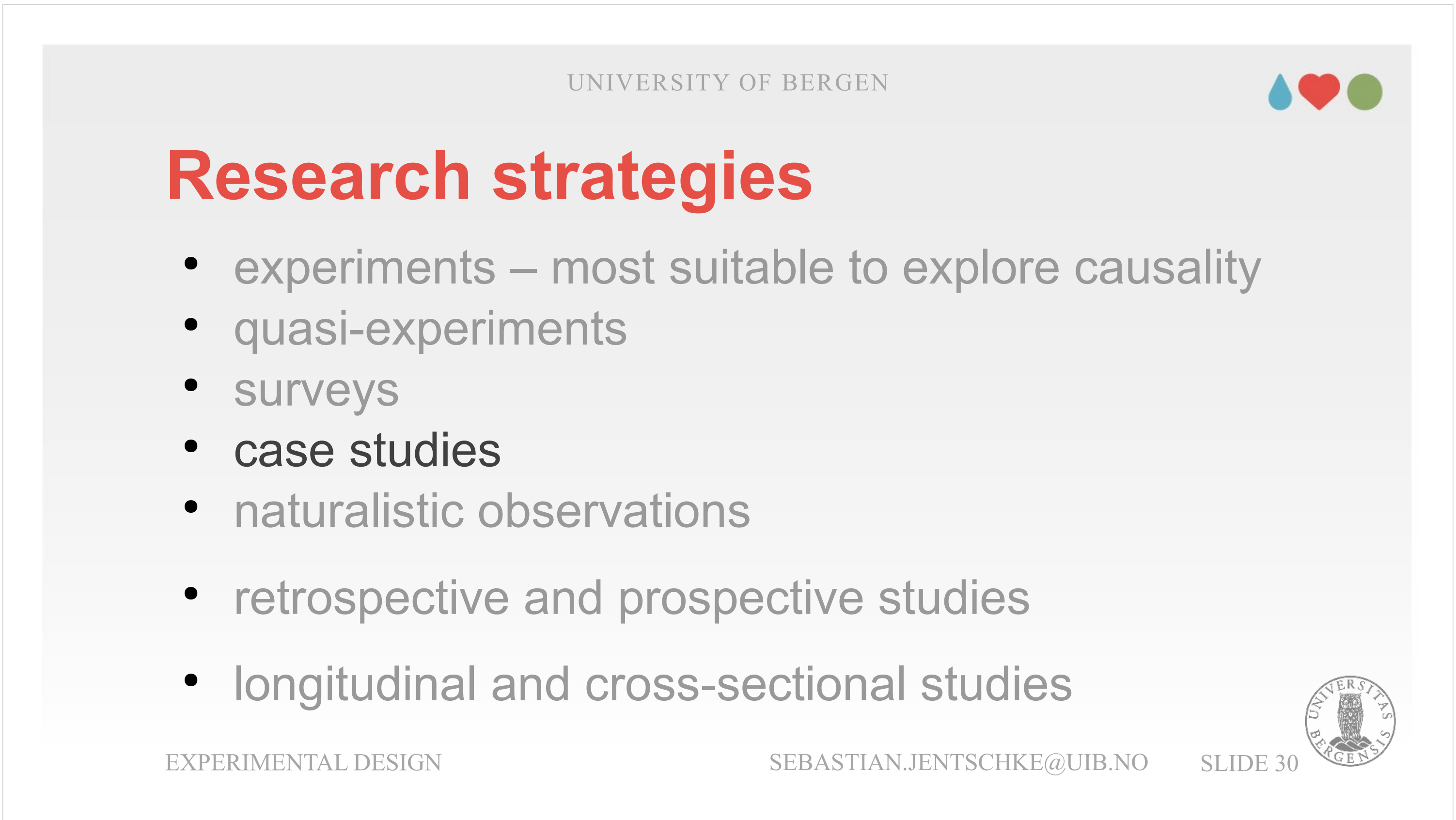
One possible strategy is to ***control for as many nuisance variables*** as possible. Generally, though, random assignment is the best safeguard against undetected nuisance variables. The better our randomization is (i.e., the more we can be certain that the participants are truly assigned randomly), the easier becomes interpreting the outcome of that research.



Research strategies

- experiments – most suitable to explore causality
- quasi-experiments
- surveys
- case studies
- naturalistic observations
- retrospective and prospective studies
- longitudinal and cross-sectional studies

Surveys rely on the technique of ***self-report*** to obtain information about variables such as people's ***attitudes, opinions, behaviors***, and ***demographic characteristics***. The data are collected by means of an ***interview*** or a ***questionnaire***.



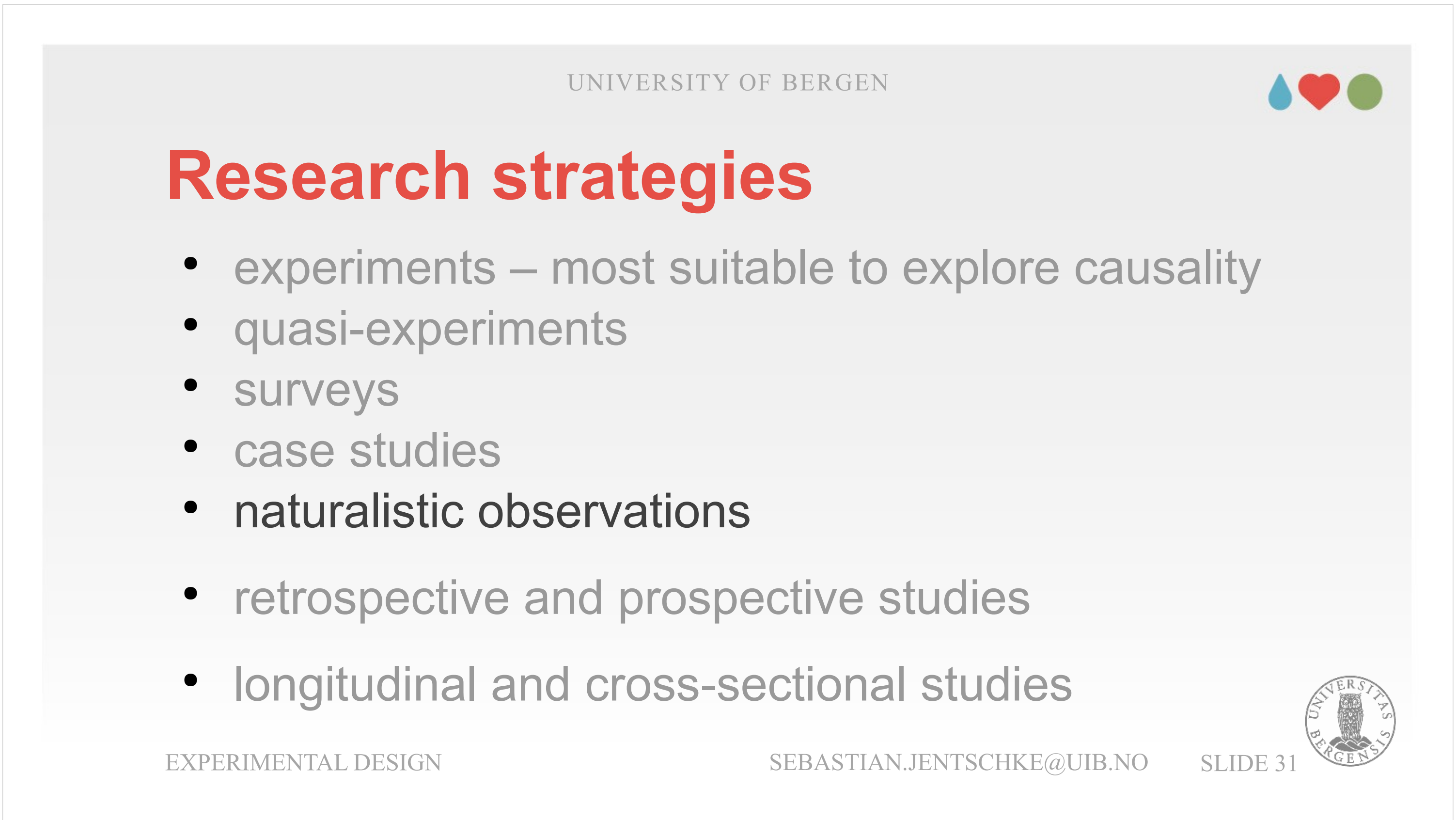
Research strategies

- experiments – most suitable to explore causality
- quasi-experiments
- surveys
- case studies
- naturalistic observations
- retrospective and prospective studies
- longitudinal and cross-sectional studies

Case studies explore selected aspects of a subject's ***behavior over*** a period of ***time***. The reason is that those ***cases often*** possess an unusual or ***noteworthy condition*** (have been exposed to a certain event or have a certain medical condition).

Both surveys and case studies ***can*** lead to ***insights*** that merit ***further investigation***. However, ***neither*** of the two methods can ***establish causality***. They rather explore, describe, classify, and establish relationships among variables.

Both are ***particularly susceptible*** to the effects of ***nuisance variables***, and (especially for case studies) questions may arise about the degree to which the findings can be generalized.



UNIVERSITY OF BERGEN

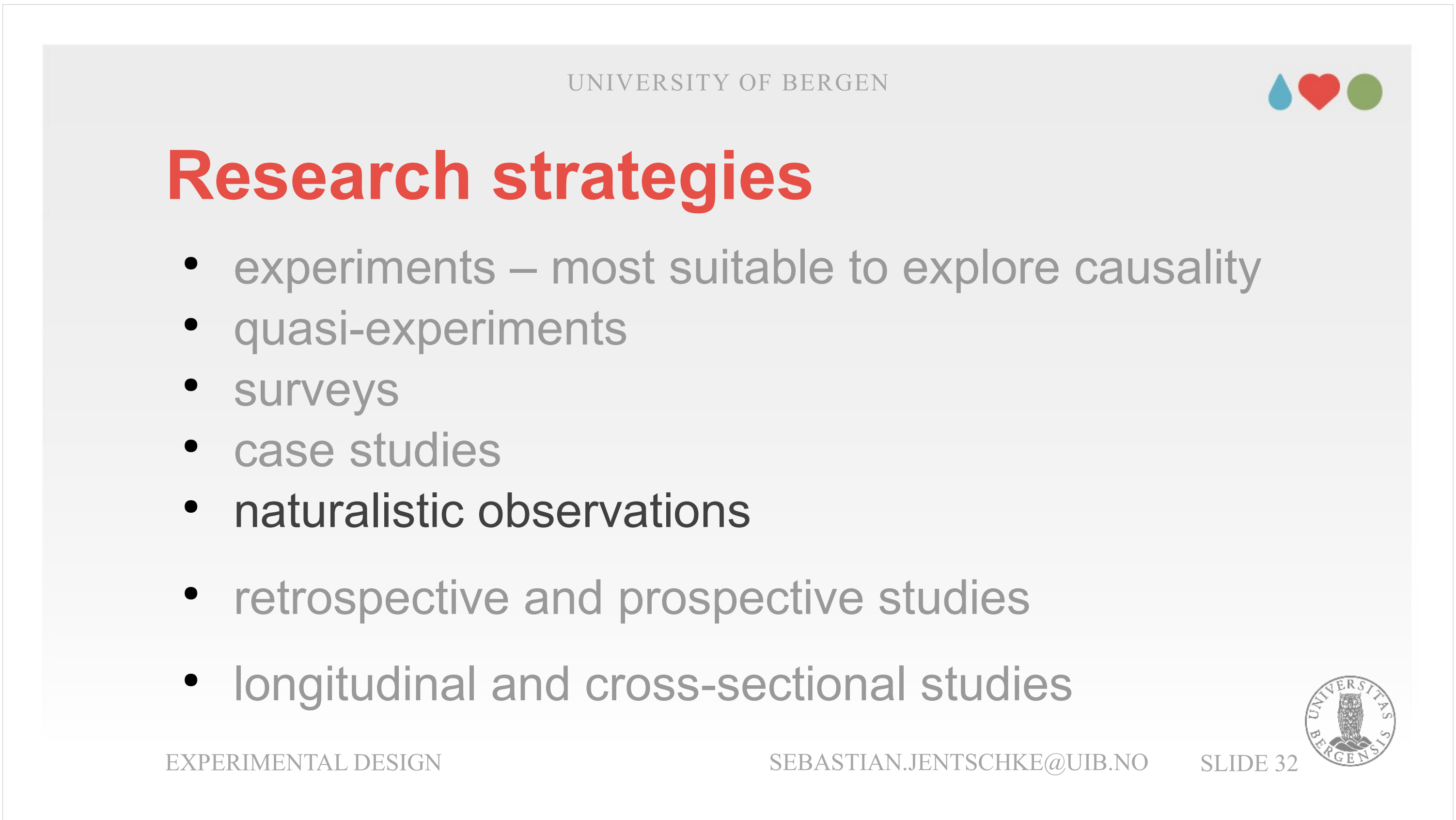
Research strategies

- experiments – most suitable to explore causality
- quasi-experiments
- surveys
- case studies
- naturalistic observations
- retrospective and prospective studies
- longitudinal and cross-sectional studies

EXPERIMENTAL DESIGN SEBASTIAN.JENTSCHKE@UIB.NO SLIDE 31

Naturalistic observations involves ***observing individuals*** or events in their ***natural setting***. Other than experiments, naturalistic observation ***neither*** involves the ***manipulation*** of an ***independent nor measuring*** the ***dependent variable in ways that intrude*** on the setting. ***Instead*** certain ***events*** are determined to be ***recorded***, and the researcher is an unobtrusive recorder of these events. Because a researcher can focus on only a finite number of events, decisions must be made concerning the events that will be observed. The data from naturalistic observations may therefore be difficult to analyze.

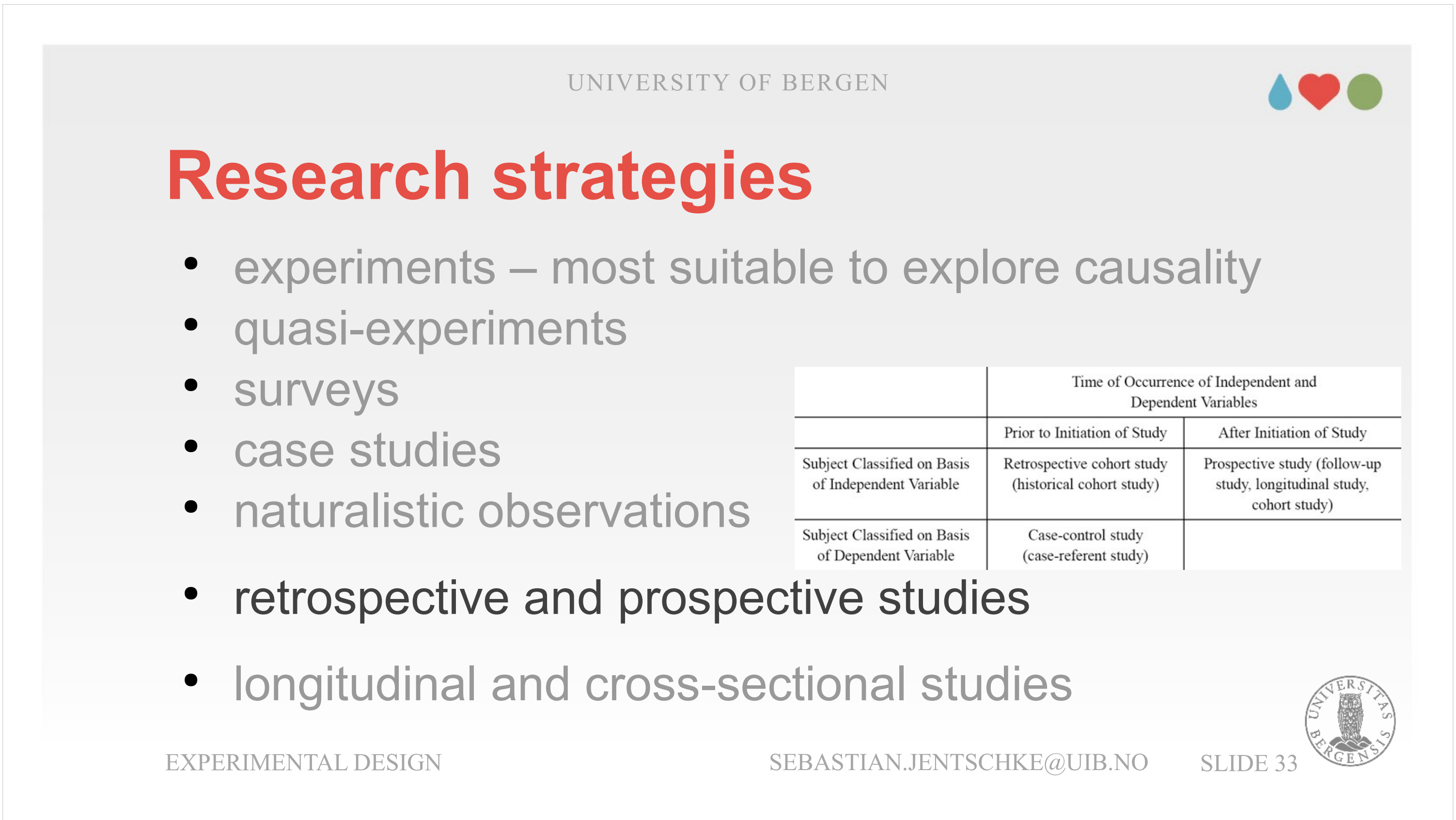
Naturalistic observation is one of the oldest methods for studying individuals and events. An example are Charles Darwin's voyages during which he compiled the data that led to the development of the theory of evolution.



Research strategies

- experiments – most suitable to explore causality
- quasi-experiments
- surveys
- case studies
- naturalistic observations
- retrospective and prospective studies
- longitudinal and cross-sectional studies

Naturalistic observations have two main advantages over more controlled strategies such as the experiment: (1) Findings generalize readily to other real-life situations. (2) The strategy avoids or strongly reduces participants being reactive to being measured.



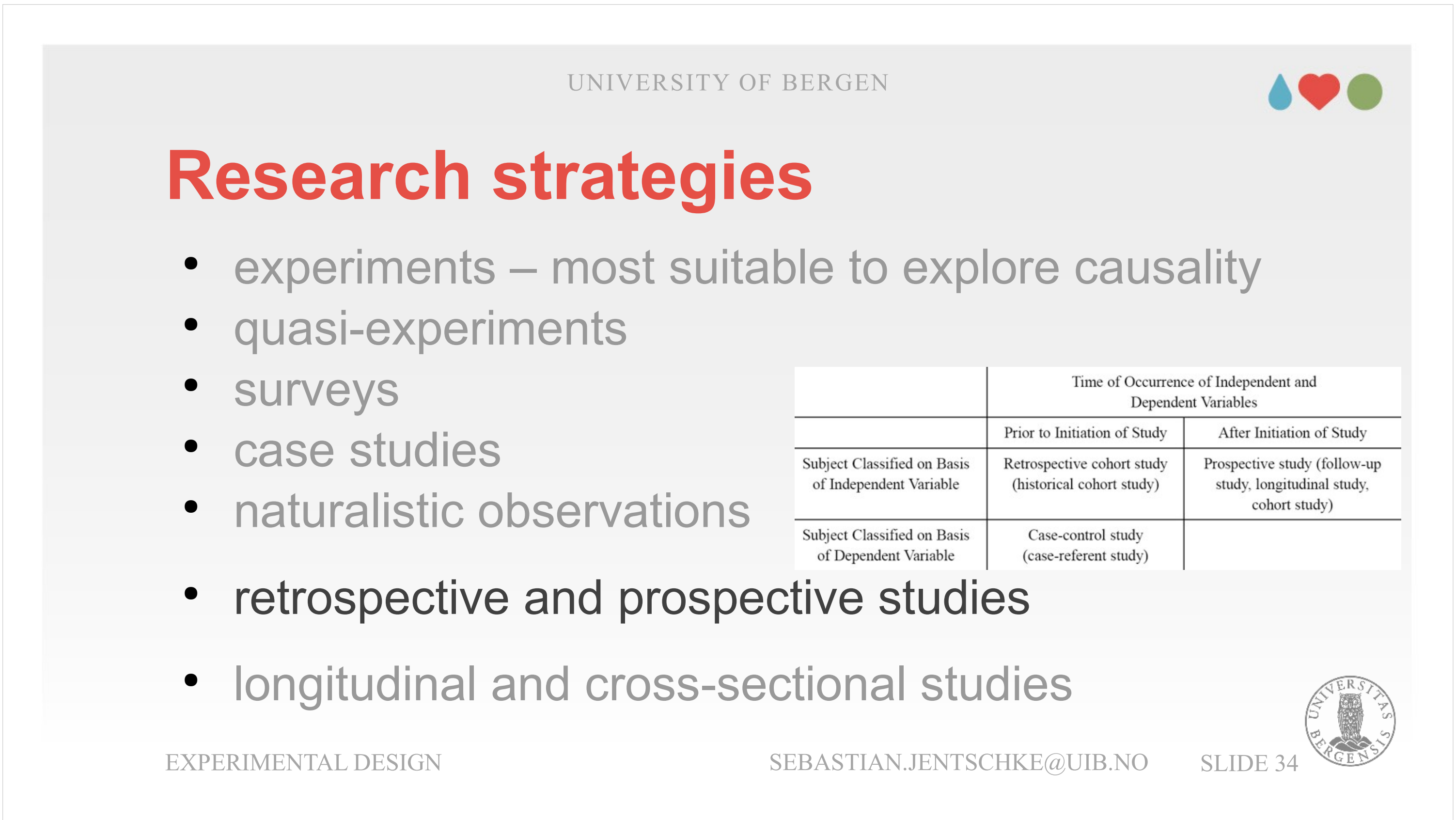
Research strategies

- experiments – most suitable to explore causality
- quasi-experiments
- surveys
- case studies
- naturalistic observations
- retrospective and prospective studies
- longitudinal and cross-sectional studies

	Time of Occurrence of Independent and Dependent Variables	
	Prior to Initiation of Study	After Initiation of Study
Subject Classified on Basis of Independent Variable	Retrospective cohort study (historical cohort study)	Prospective study (follow-up study, longitudinal study, cohort study)
Subject Classified on Basis of Dependent Variable	Case-control study (case-referent study)	

Retrospective and prospective studies are ***non-experimental*** research strategies in which the ***independent*** and ***dependent variables occur before or after***, respectively, the beginning of the ***study***. The table provides an overview.

(1) ***Retrospective*** studies ***look backward in time***. (2) ***prospective*** studies ***look forward in time***. ***Within retrospective*** studies there are two further categories according to: (a) whether ***people have*** or ***have not been exposed*** to the ***independent*** variable or (b) ***whether*** they ***exhibit*** a ***certain outcome*** in the dependent variable.



Research strategies

- experiments – most suitable to explore causality
- quasi-experiments
- surveys
- case studies
- naturalistic observations

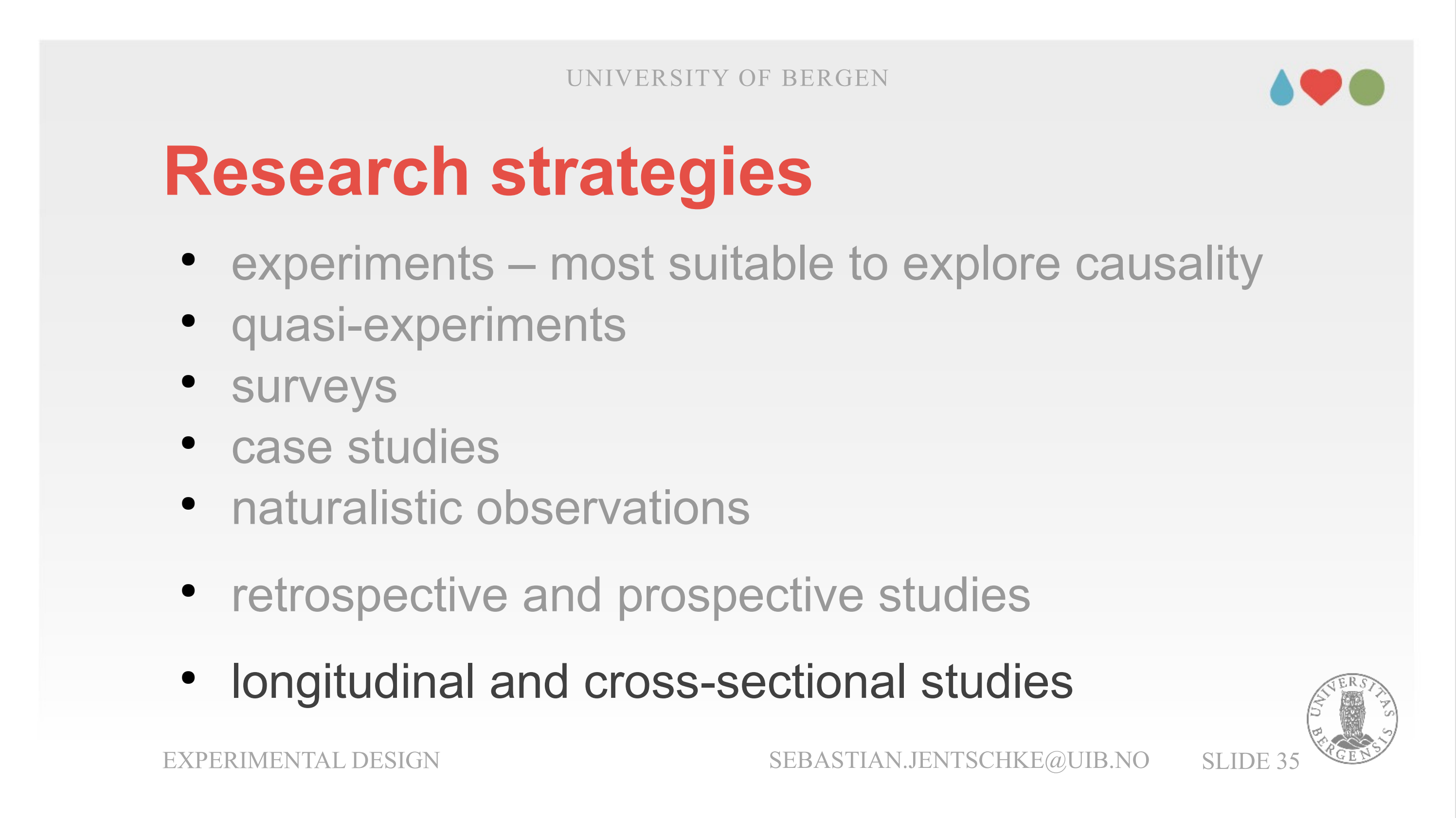
	Time of Occurrence of Independent and Dependent Variables	
	Prior to Initiation of Study	After Initiation of Study
Subject Classified on Basis of Independent Variable	Retrospective cohort study (historical cohort study)	Prospective study (follow-up study, longitudinal study, cohort study)
Subject Classified on Basis of Dependent Variable	Case-control study (case-referent study)	

- retrospective and prospective studies
- longitudinal and cross-sectional studies

(1a, top-left) ***Historical cohort studies*** explore ***participants*** who have had a ***certain*** medical ***condition*** (e.g. a heart infarct) ***or*** where exposed to a ***certain*** historical ***event*** (e.g., holocaust survivors). Consequences of that exposure are assessed.

(1b, bottom left) ***Case-control studies*** explore ***participants who developed*** a ***certain condition*** (e.g., cancer) and ***determine*** which ***independent variables*** (e.g., smoking, high blood pressure, etc.) could have accounted for that condition.

(2, top-right) ***Prospective*** / follow-up studies ***classify*** participants based on ***whether*** they have been ***exposed*** to a ***naturally occurring independent variable*** or not. These participant are ***followed up***. For example could participants not having a cardiovascular disease at outset be followed up to identify factors that contribute to them later developing such (cardiovascular) disease.



UNIVERSITY OF BERGEN

Research strategies

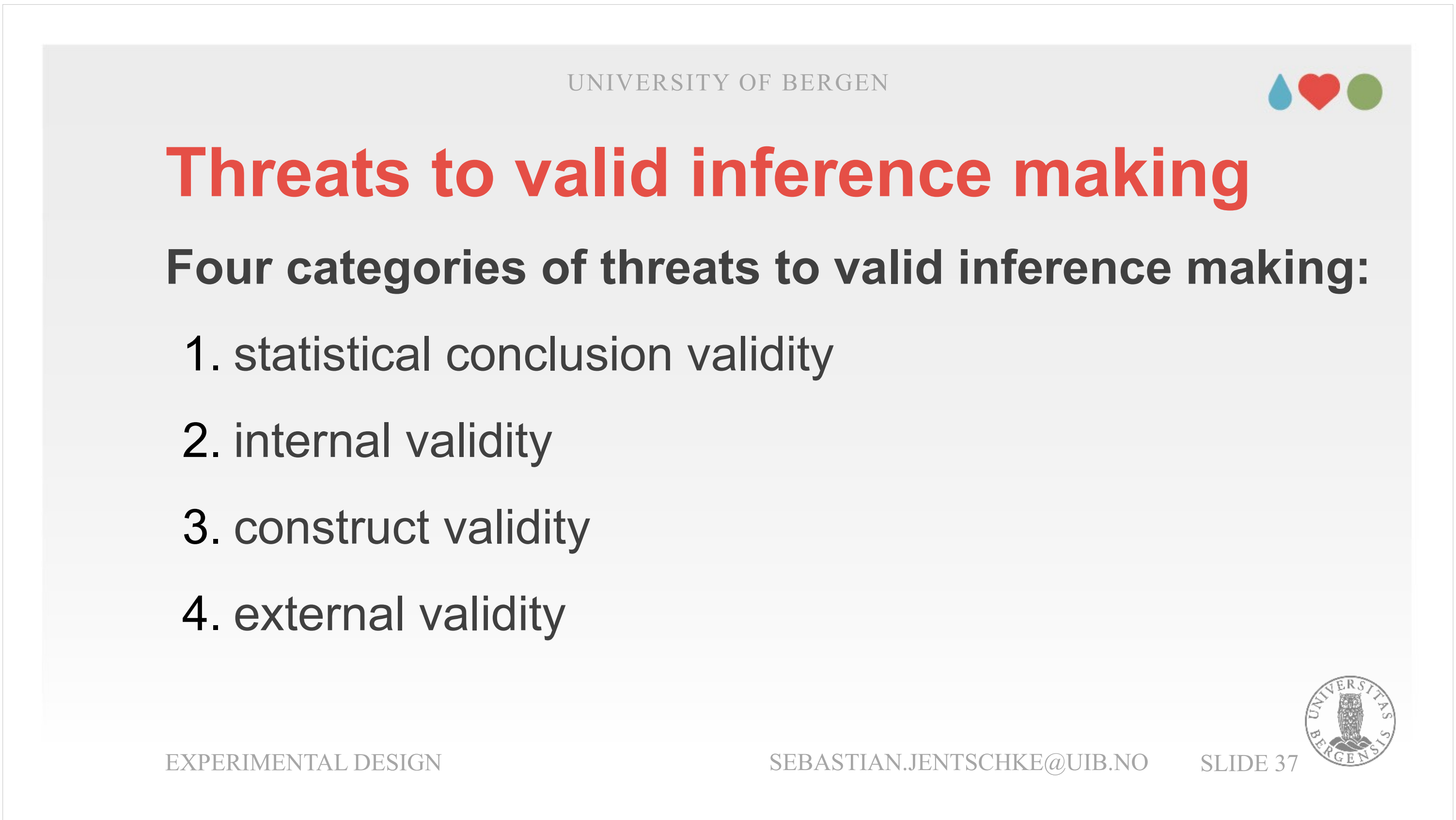
- experiments – most suitable to explore causality
- quasi-experiments
- surveys
- case studies
- naturalistic observations
- retrospective and prospective studies
- longitudinal and cross-sectional studies

EXPERIMENTAL DESIGN SEBASTIAN.JENTSCHKE@UIB.NO SLIDE 35

Prospective and retrospective studies are examples that may employ longitudinal studies where participants are followed over time. Developmental psychology also uses longitudinal studies to trace certain skills or behaviours in their development (e.g., of memory performance through adolescence) or their decay (e.g., cognitive decline in elderly people).

However, given that ***longitudinal studies require lots of time*** and ***resources*** and being subject to other ***problems*** such as drop-outs, ***cross-sectional*** design may be ***more economical***. Here, different age groups are evaluated at one measurement time point. Such cross-sectional design, however, have the disadvantage that they confound belonging to a certain age group with belonging to a certain cohort (e.g., being born 1980 vs. 2000; think what such a difference makes, e.g., with respect to exposure to the internet).

Threats to valid inference making



Threats to valid inference making

Four categories of threats to valid inference making:

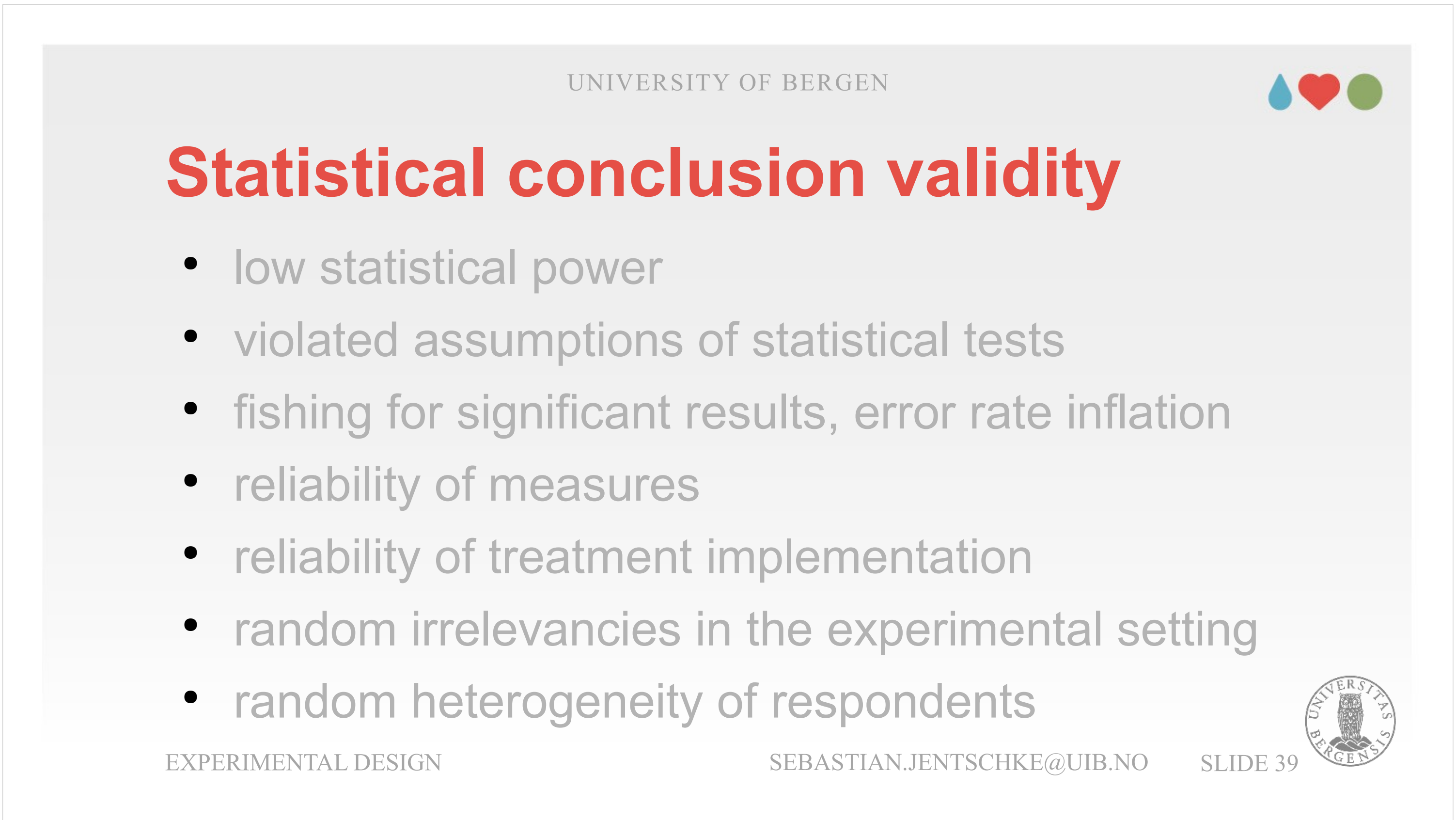
1. statistical conclusion validity
2. internal validity
3. construct validity
4. external validity

When devising a research design we are faced with certain threats to valid inference. These can be assigned to four different categories, asking the following questions:

1. How ***reliable and large*** is the ***relationship*** between presumed cause and effect?
2. Is the relationship between independent and dependent variable ***causal or*** could the same covariation have ***been obtained without*** or with another ***treatment***?
3. How well reflect the persons, treatments, observations and settings the ***underlying*** general ***theories or constructs***?
4. How ***generalizable*** is this ***causal relationship*** over varied persons, treatments, observations and settings?

With choosing specific designs or strategies, we aim to minimize the influence of such threats.

Threats to statistical conclusion validity



Statistical conclusion validity

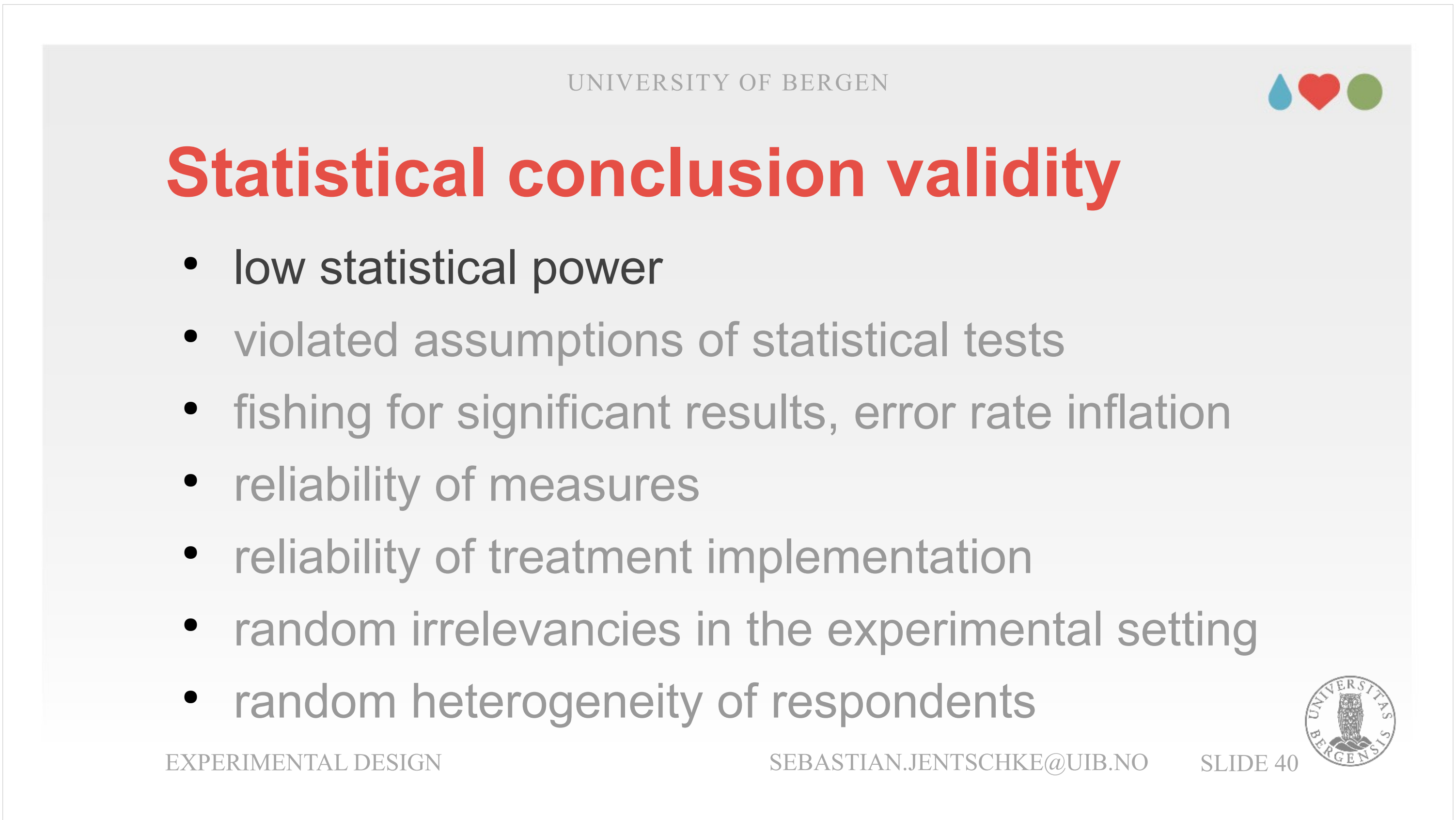
- low statistical power
- violated assumptions of statistical tests
- fishing for significant results, error rate inflation
- reliability of measures
- reliability of treatment implementation
- random irrelevancies in the experimental setting
- random heterogeneity of respondents

Statistical conclusion validity is the degree to which conclusions about the relationship among variables based on the data are correct or “reasonable”.

Fundamentally, two types of errors can occur within statistical tests: ***type I*** (finding a difference or correlation when none exists) and ***type II*** (finding no difference or correlation when one exists).

Statistical conclusion validity concerns qualities of the study that make these types of errors less likely. It ***involves*** ensuring the use of ***adequate sampling*** procedures, ***appropriate statistical tests***, and ***reliable measurement*** procedures.

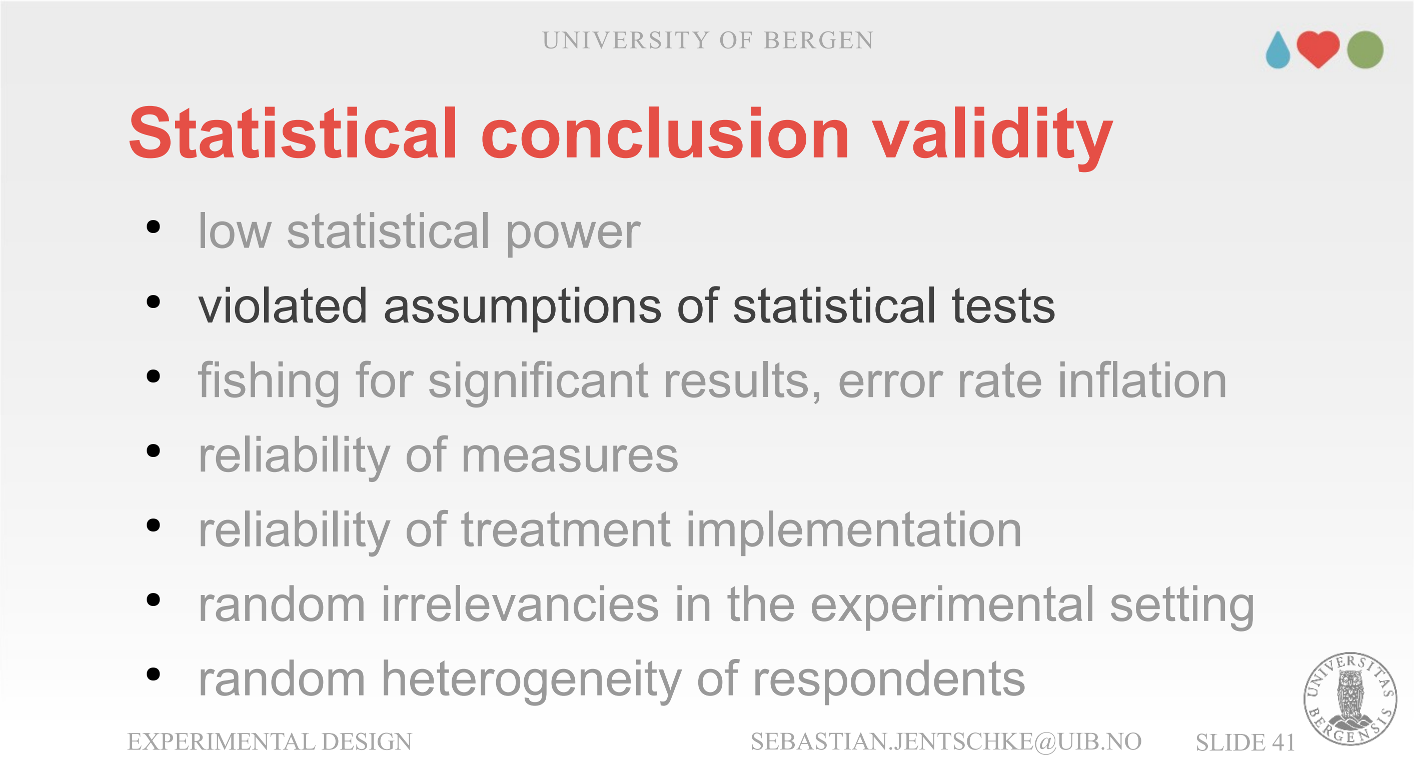
Threats to statistical conclusion validity fall into seven categories.



Statistical conclusion validity

- low statistical power
- violated assumptions of statistical tests
- fishing for significant results, error rate inflation
- reliability of measures
- reliability of treatment implementation
- random irrelevancies in the experimental setting
- random heterogeneity of respondents

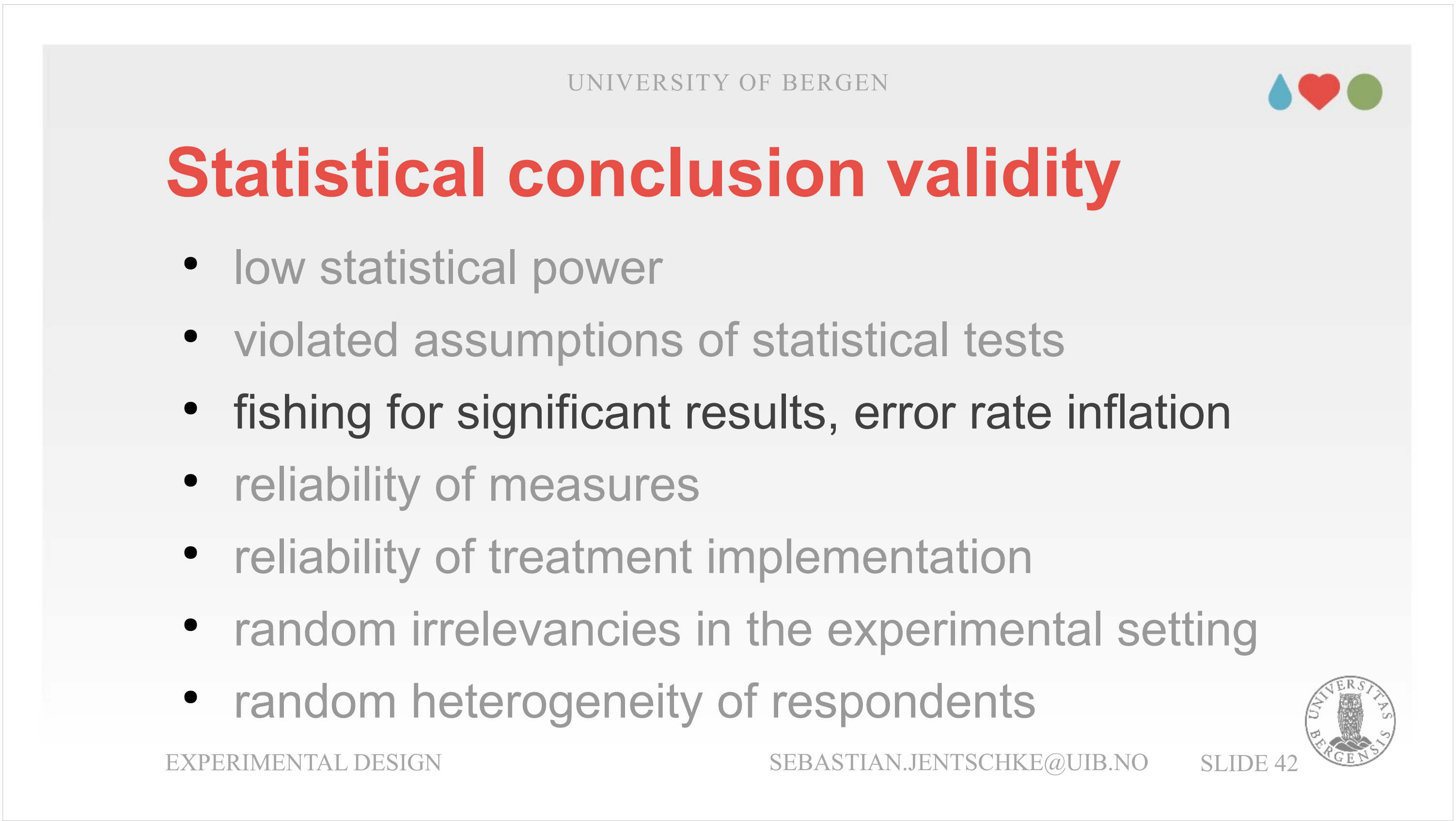
Low statistical power denotes the failure to reject a false null hypothesis because the sample size is inadequate, irrelevant sources of variation are not controlled or isolated, or inefficient test statistics are used.



Statistical conclusion validity

- low statistical power
- violated assumptions of statistical tests
- fishing for significant results, error rate inflation
- reliability of measures
- reliability of treatment implementation
- random irrelevancies in the experimental setting
- random heterogeneity of respondents

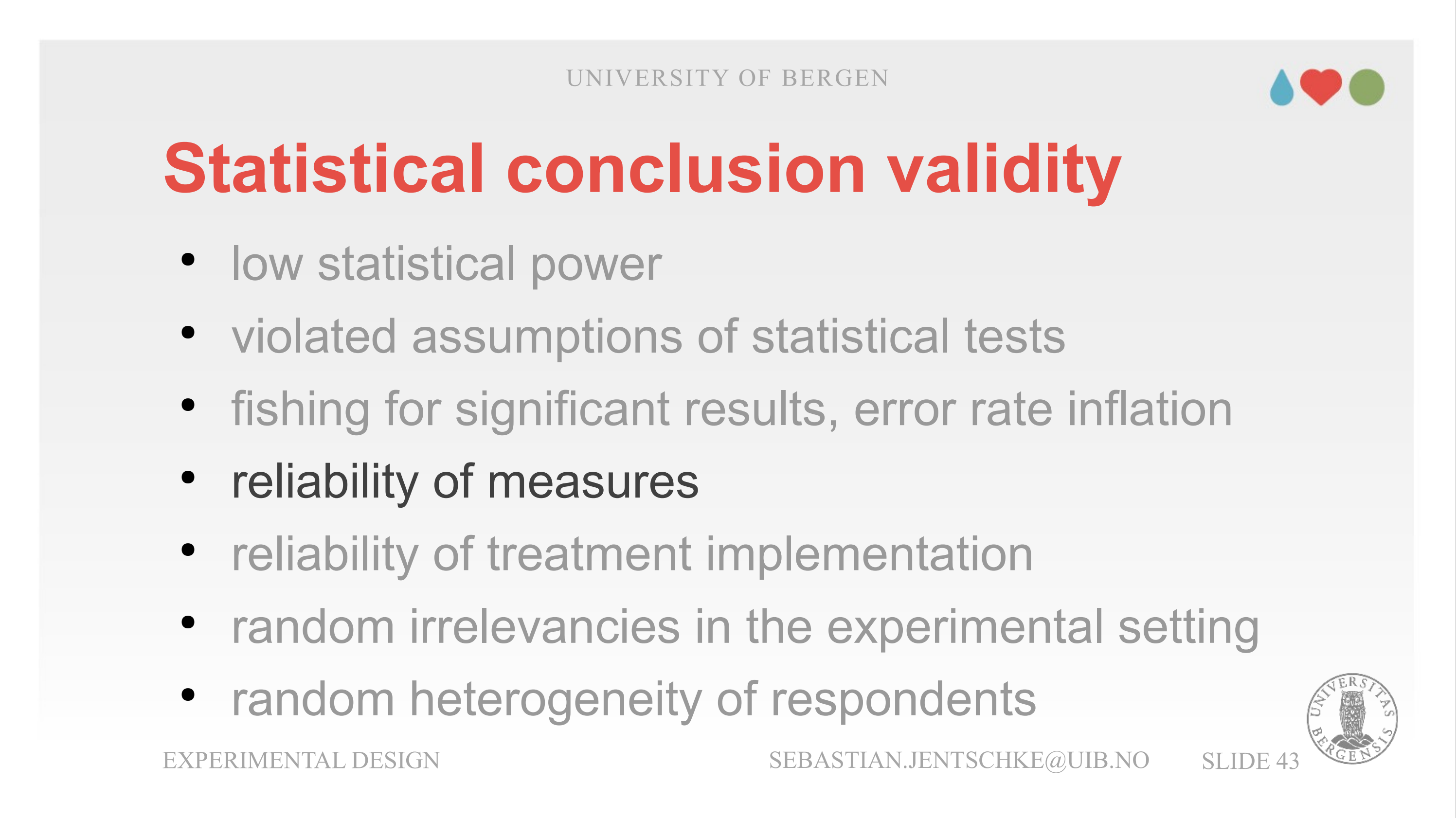
Violated assumptions of statistical tests (e.g., ***normality*** or ***equality of variances***) may cause incorrect inferences. Another aspect that is often not sufficiently considered is the demand that the ***measurements*** are ***independent***.



Statistical conclusion validity

- low statistical power
- violated assumptions of statistical tests
- fishing for significant results, error rate inflation
- reliability of measures
- reliability of treatment implementation
- random irrelevancies in the experimental setting
- random heterogeneity of respondents

A researcher may “fish” for significant results (i.e., conduct numerous tests with the data). As a consequence, the error probability is inflated: The probability of drawing erroneous conclusions increases as a function of the number of tests performed. There are opportunities to counteract this inflation (e.g., Bonferroni-correction for multiple comparis.).

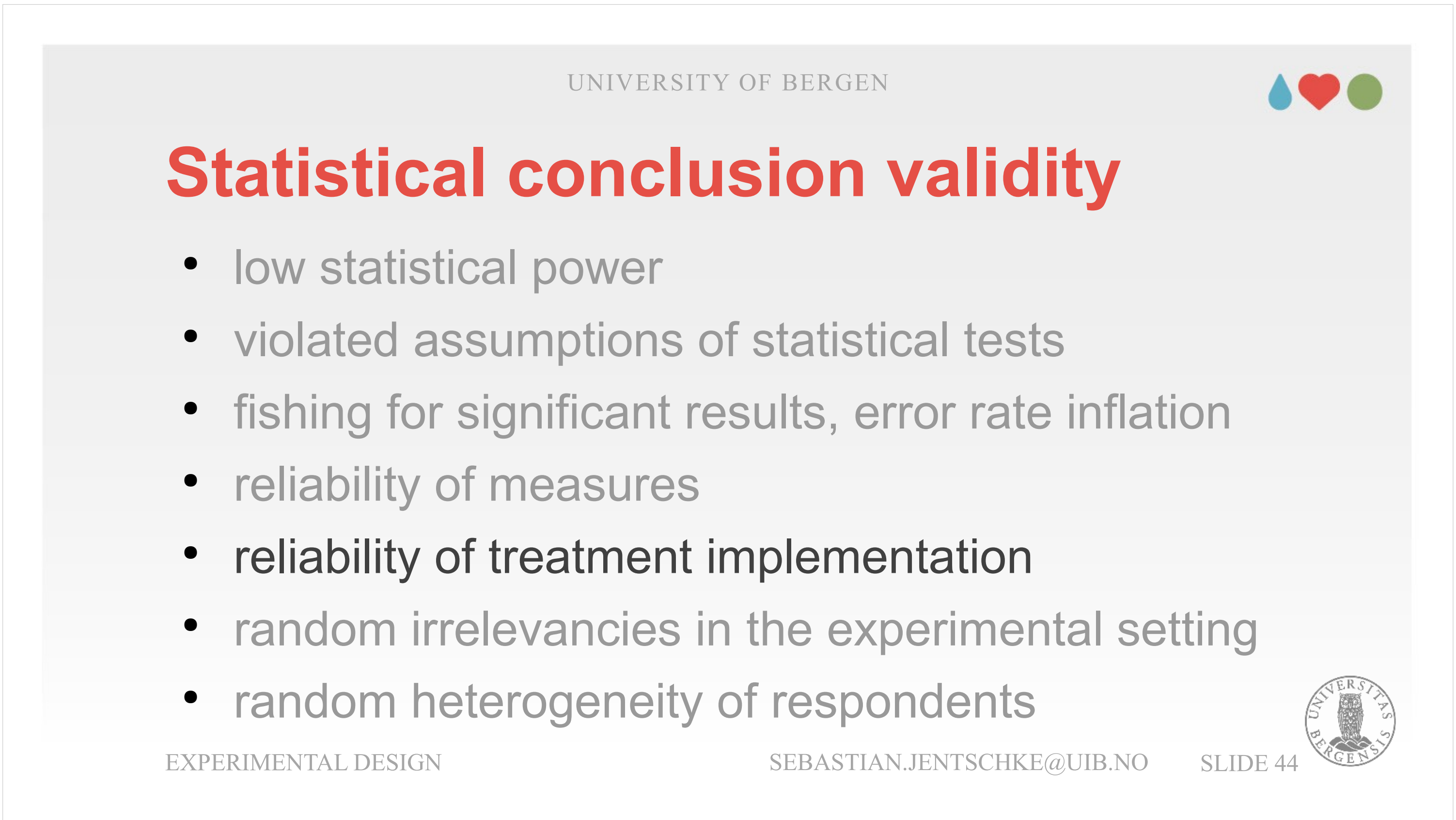


Statistical conclusion validity

- low statistical power
- violated assumptions of statistical tests
- fishing for significant results, error rate inflation
- reliability of measures
- reliability of treatment implementation
- random irrelevancies in the experimental setting
- random heterogeneity of respondents

The remaining four categories all have to do with an increase in error variance which may result in not rejecting a false null hypothesis.

Reasons may be that: The ***dependent variable can’t be measured reliably*** (e.g., because the used questionnaire has low reliability).



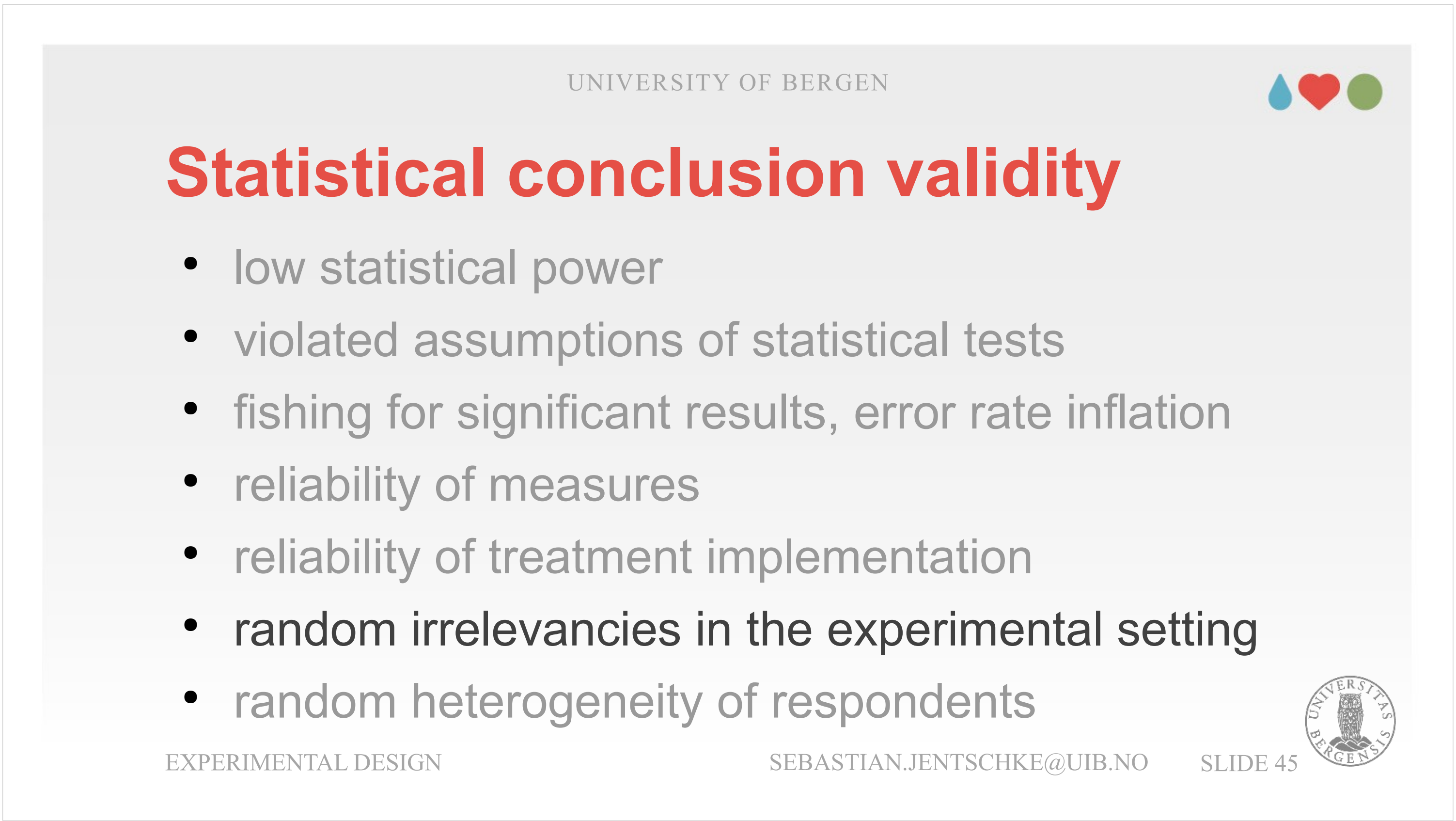
Statistical conclusion validity

- low statistical power
- violated assumptions of statistical tests
- fishing for significant results, error rate inflation
- reliability of measures
- reliability of treatment implementation
- random irrelevancies in the experimental setting
- random heterogeneity of respondents

There may be a ***failure to standardize the administration*** of the treatment levels.

Another point to consider here are ***floor and ceiling effects***: Sometimes, the manipulation of the quantitative independent variable has to exceed a certain value or threshold in order to observe an effect (e.g., a very low amount of a psychoactive substance – e.g. alcohol – might not affect attention to a degree that can be measured). This is called floor effect.

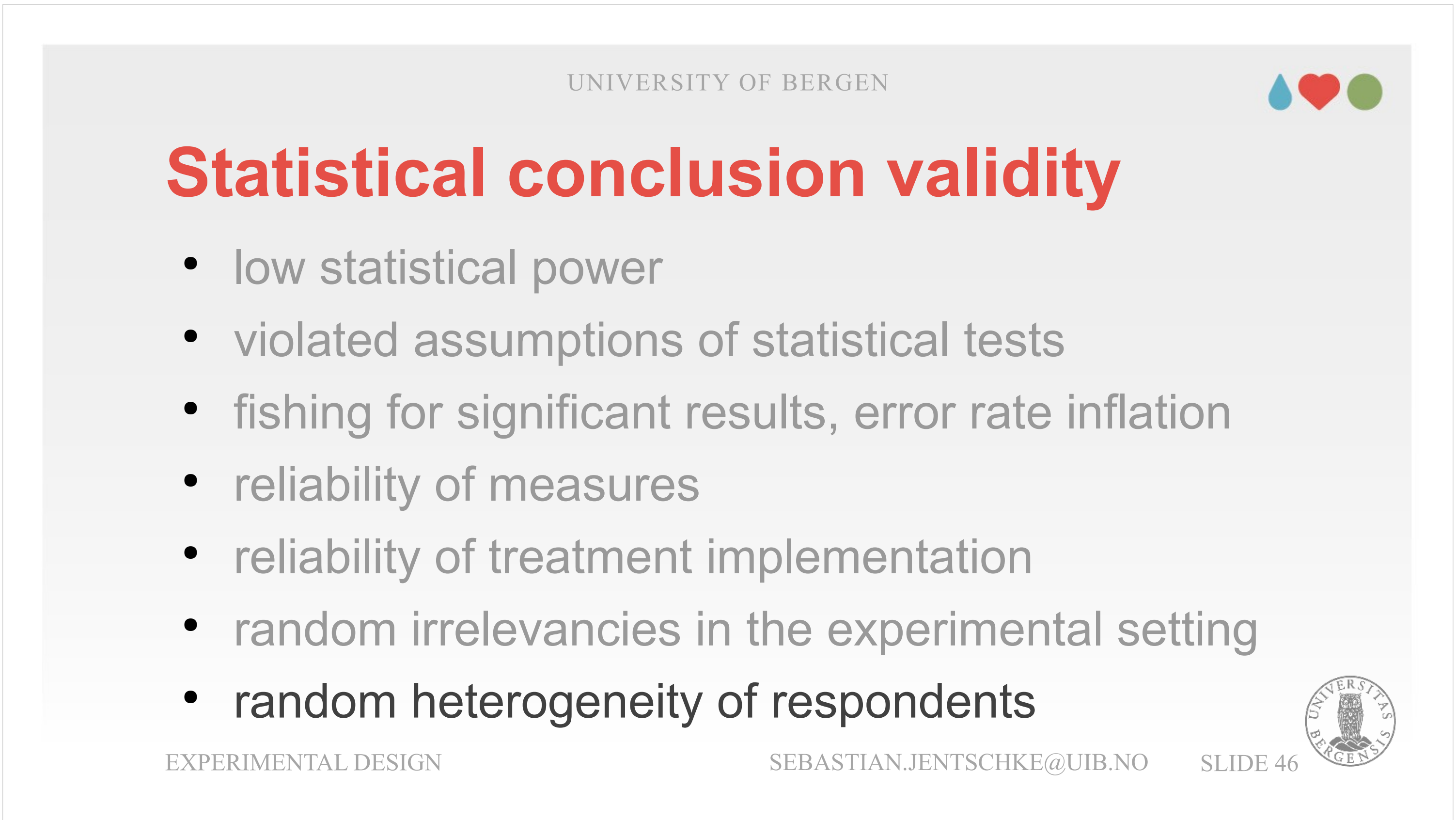
Vice versa, beyond a certain amount, a manipulation of the quantitative independent variable might not exert an effect anymore (e.g., a very high dosage of a headache medication might not serve to further reduce the pain). This is called ceiling effect.



Statistical conclusion validity

- low statistical power
- violated assumptions of statistical tests
- fishing for significant results, error rate inflation
- reliability of measures
- reliability of treatment implementation
- random irrelevancies in the experimental setting
- random heterogeneity of respondents

There may be a variation in the environment (physical, social, etc.) in which a treatment level is administered that affect the dependent variable.

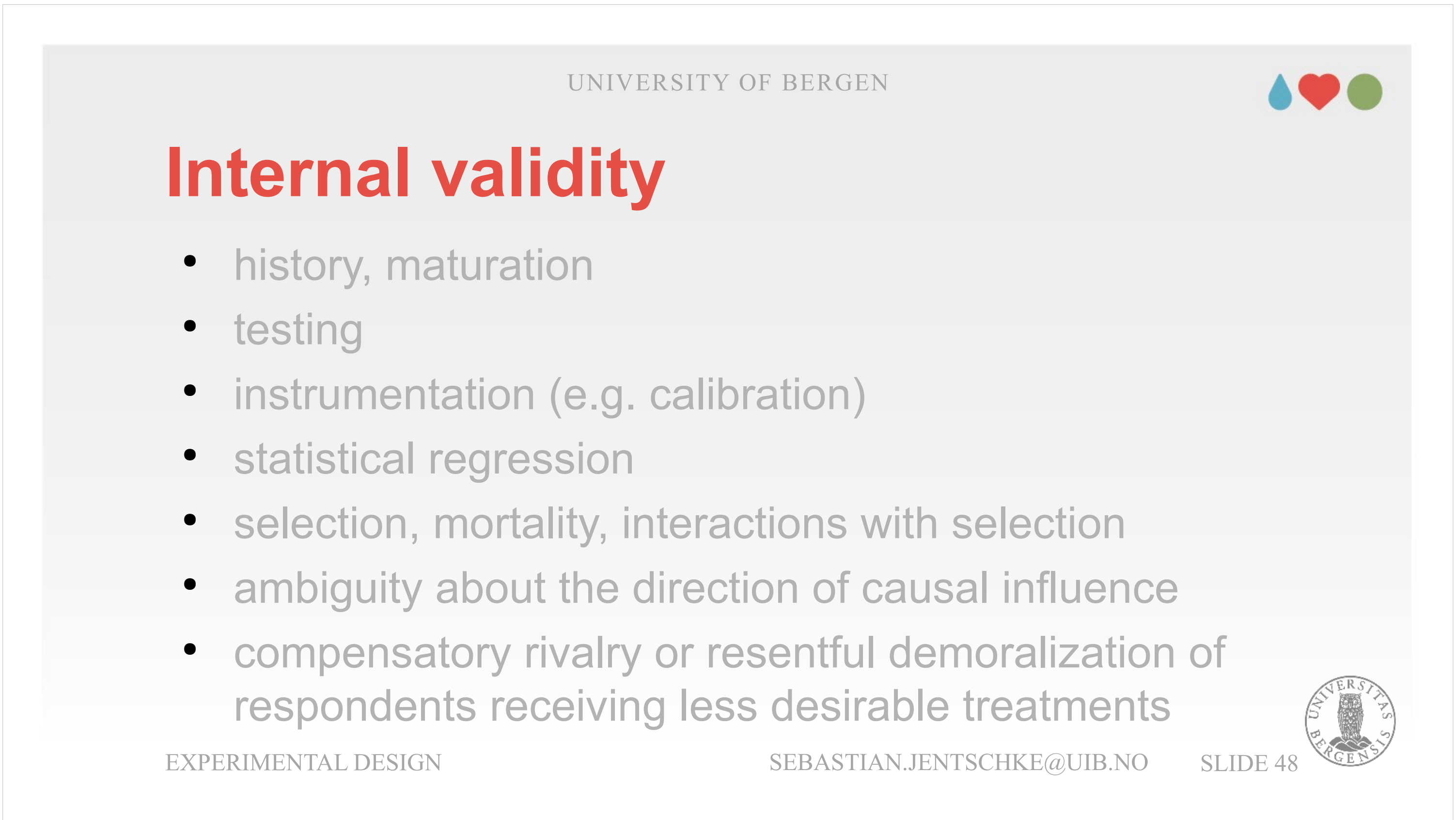


Statistical conclusion validity

- low statistical power
- violated assumptions of statistical tests
- fishing for significant results, error rate inflation
- reliability of measures
- reliability of treatment implementation
- random irrelevancies in the experimental setting
- random heterogeneity of respondents

Finally, there may be idiosyncratic characteristics of the subjects.

Threats to internal validity



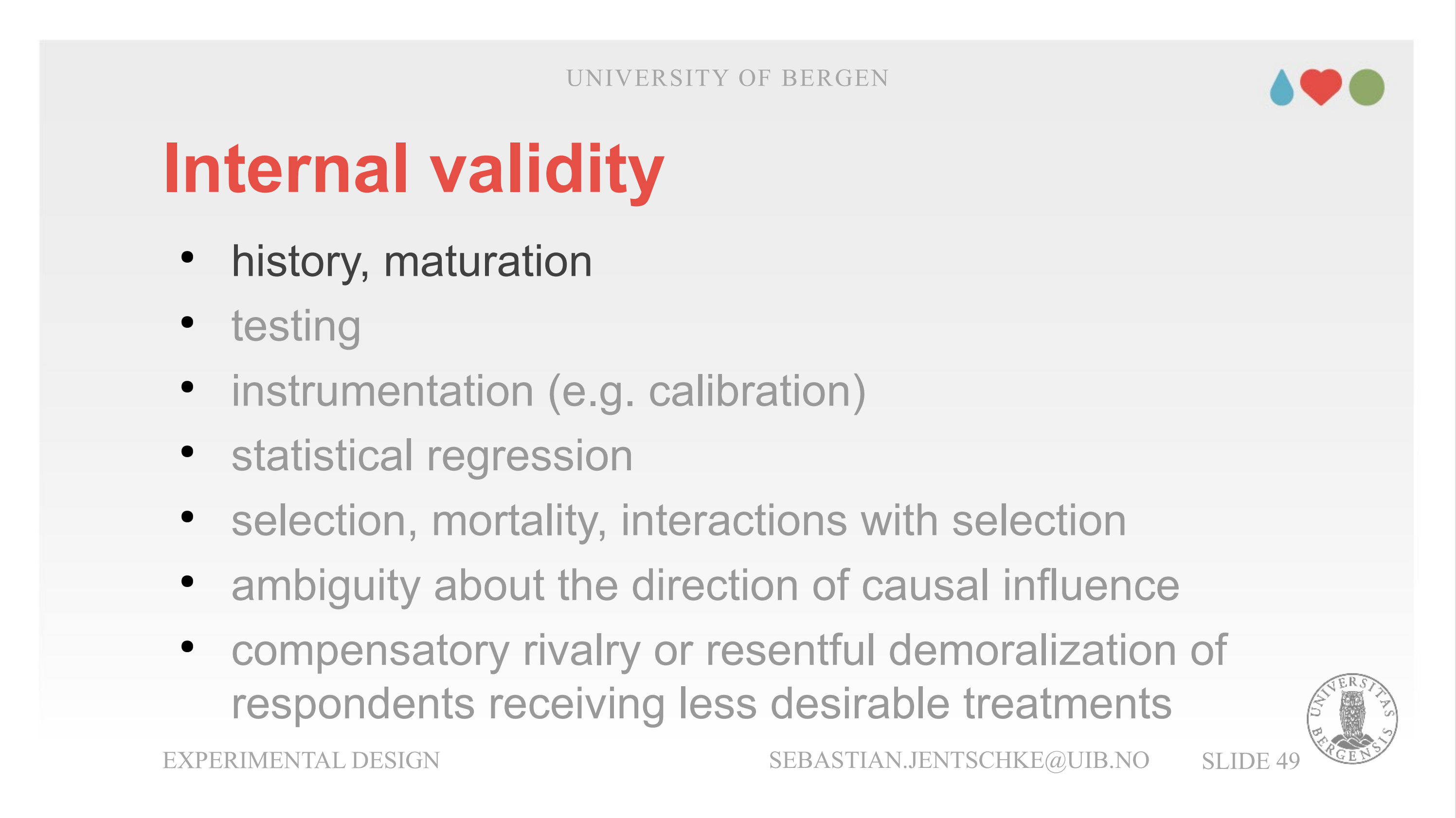
Internal validity

- history, maturation
- testing
- instrumentation (e.g. calibration)
- statistical regression
- selection, mortality, interactions with selection
- ambiguity about the direction of causal influence
- compensatory rivalry or resentful demoralization of respondents receiving less desirable treatments

Internal validity is the ***extent*** to which ***evidence supports a claim about a cause-effect-relationship*** based upon a particular study.

Internal validity is determined by how well a study can ***rule out alternative explanations*** for its findings (usually sources of systematic error or “bias”).

Common threats to internal validity fall within seven categories.



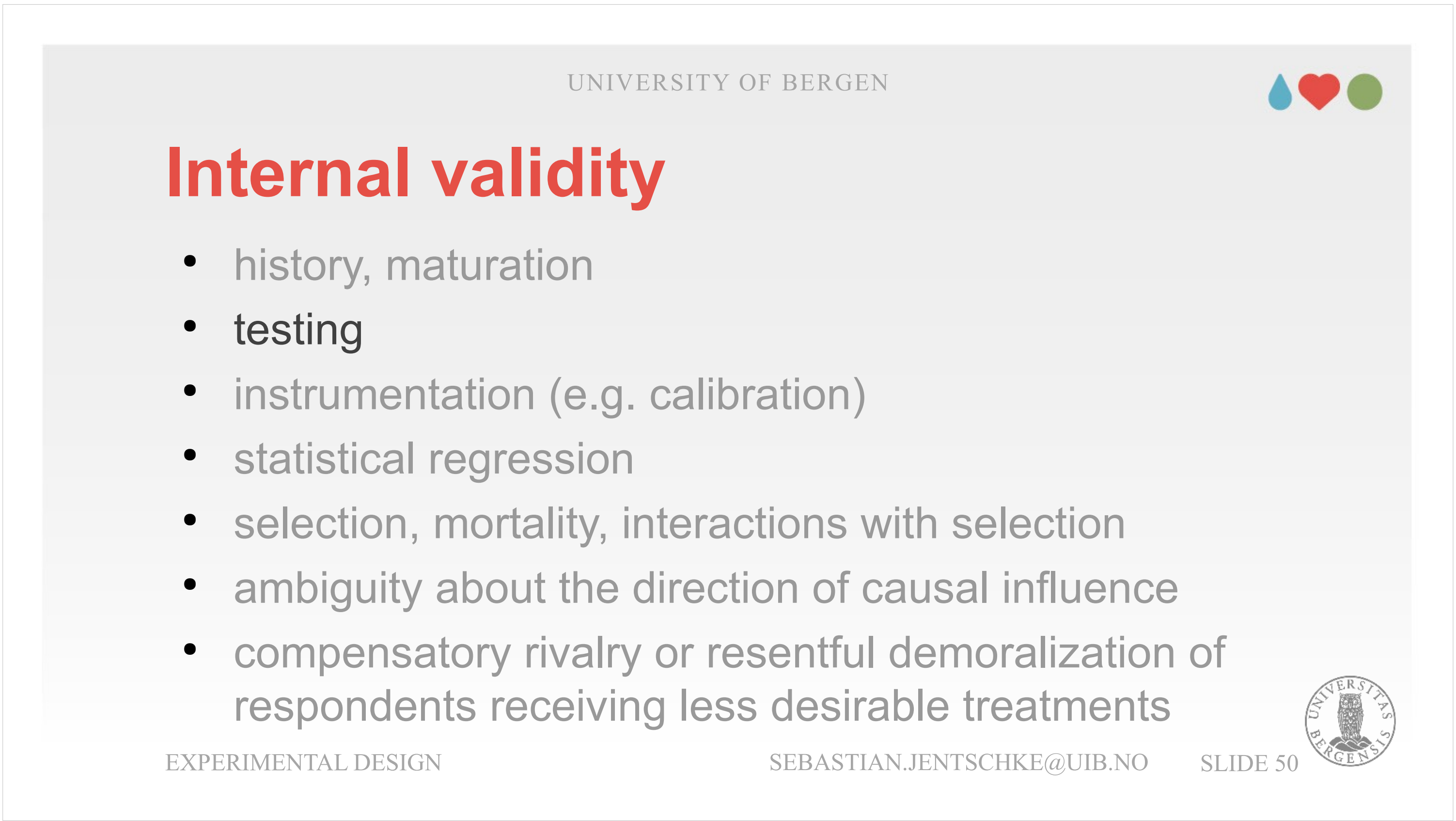
UNIVERSITY OF BERGEN

Internal validity

- history, maturation
- testing
- instrumentation (e.g. calibration)
- statistical regression
- selection, mortality, interactions with selection
- ambiguity about the direction of causal influence
- compensatory rivalry or resentful demoralization of respondents receiving less desirable treatments

EXPERIMENTAL DESIGN SEBASTIAN.JENTSCHKE@UIB.NO SLIDE 49

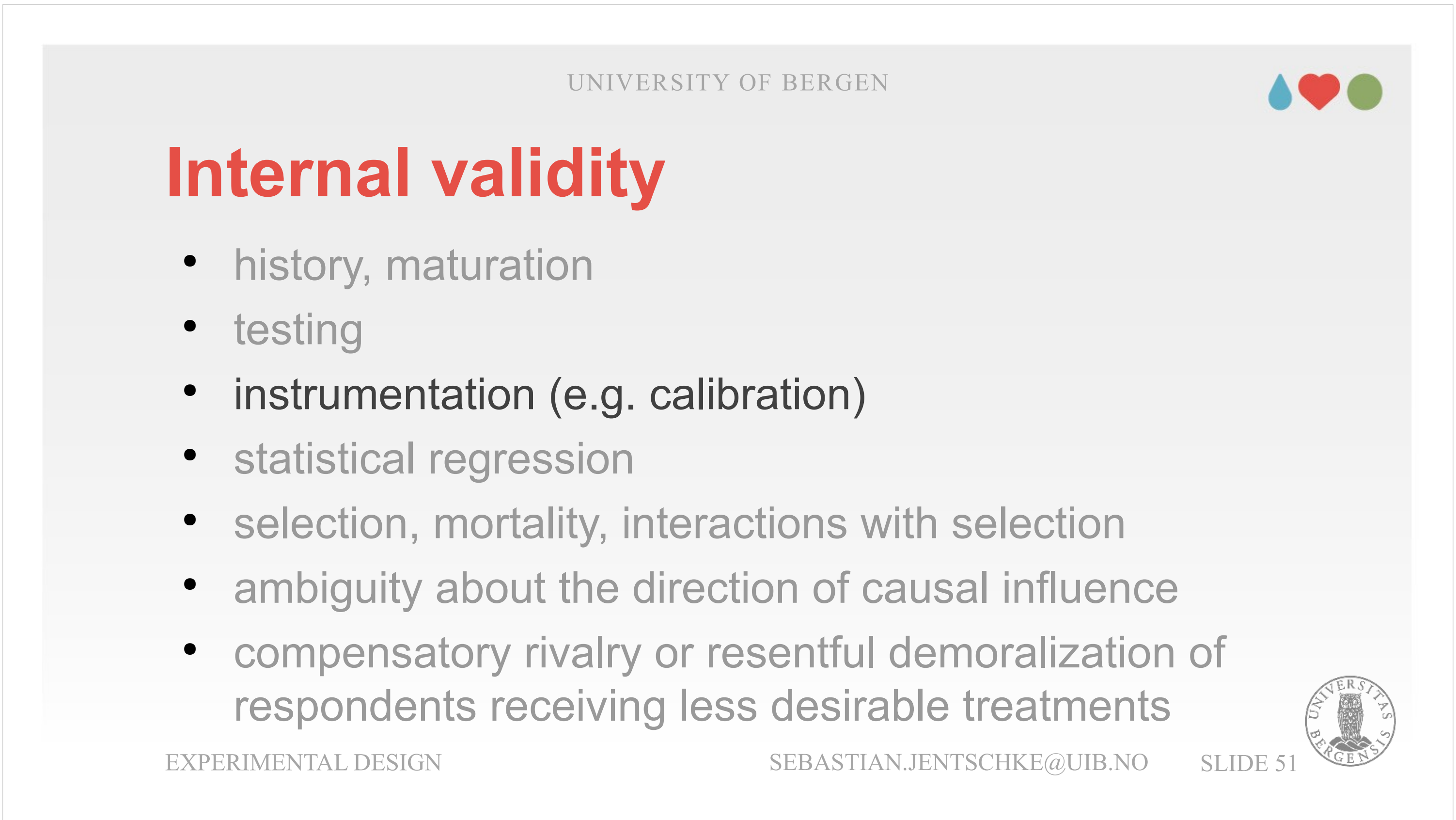
Changes that occur in the interval between the administration of a treatment level and the measurement of the dependent variable may affect the dependent variable. Such changes include history – events that occur in the outside world during that interval – as well as maturation or the passage of time – growing older, stronger, larger, more experienced, etc.



Internal validity

- history, maturation
- testing
- instrumentation (e.g. calibration)
- statistical regression
- selection, mortality, interactions with selection
- ambiguity about the direction of causal influence
- compensatory rivalry or resentful demoralization of respondents receiving less desirable treatments

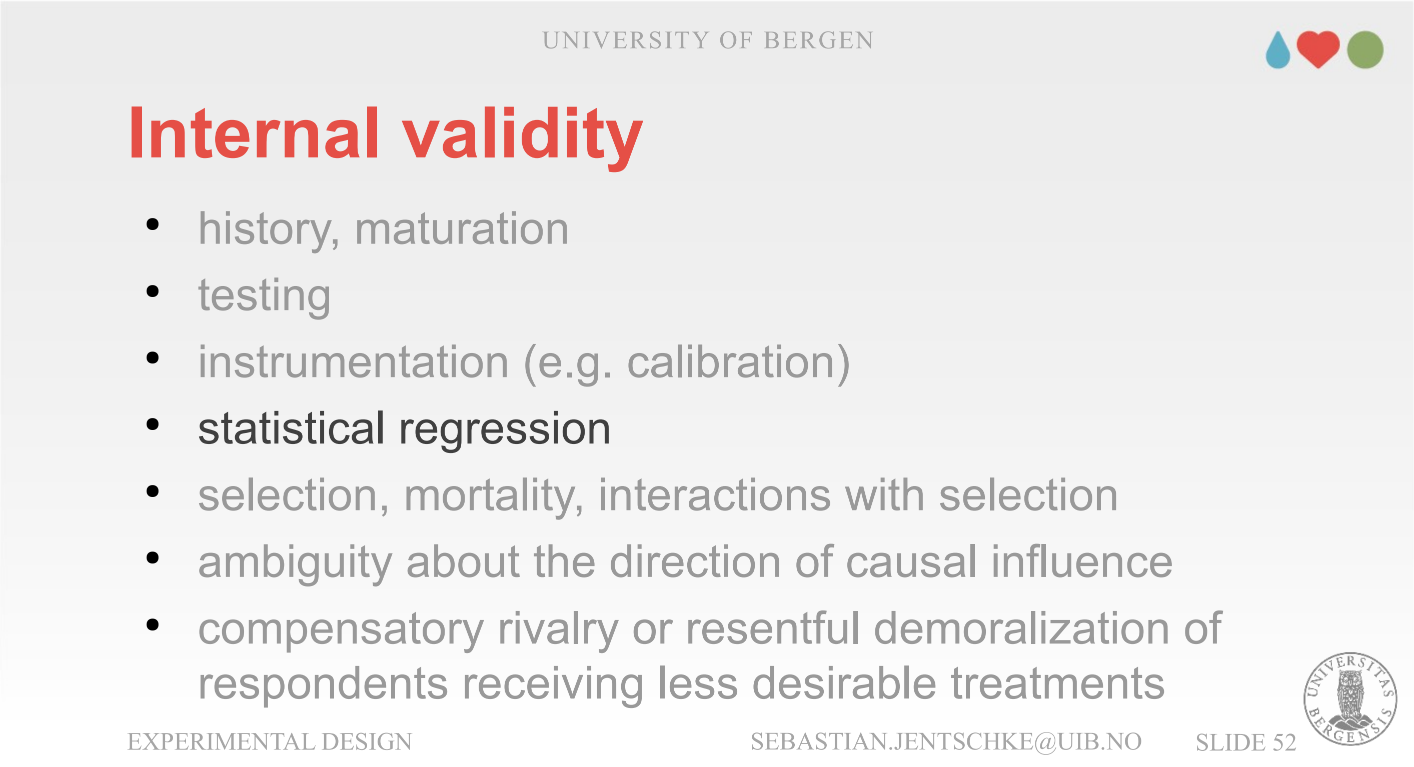
Repeated testing of subjects may result in familiarity with the testing situation or acquisition of information that can affect the dependent variable. This particularly affects tests of [cognitive] skills (intelligence, memory, etc.).



Internal validity

- history, maturation
- testing
- instrumentation (e.g. calibration)
- statistical regression
- selection, mortality, interactions with selection
- ambiguity about the direction of causal influence
- compensatory rivalry or resentful demoralization of respondents receiving less desirable treatments

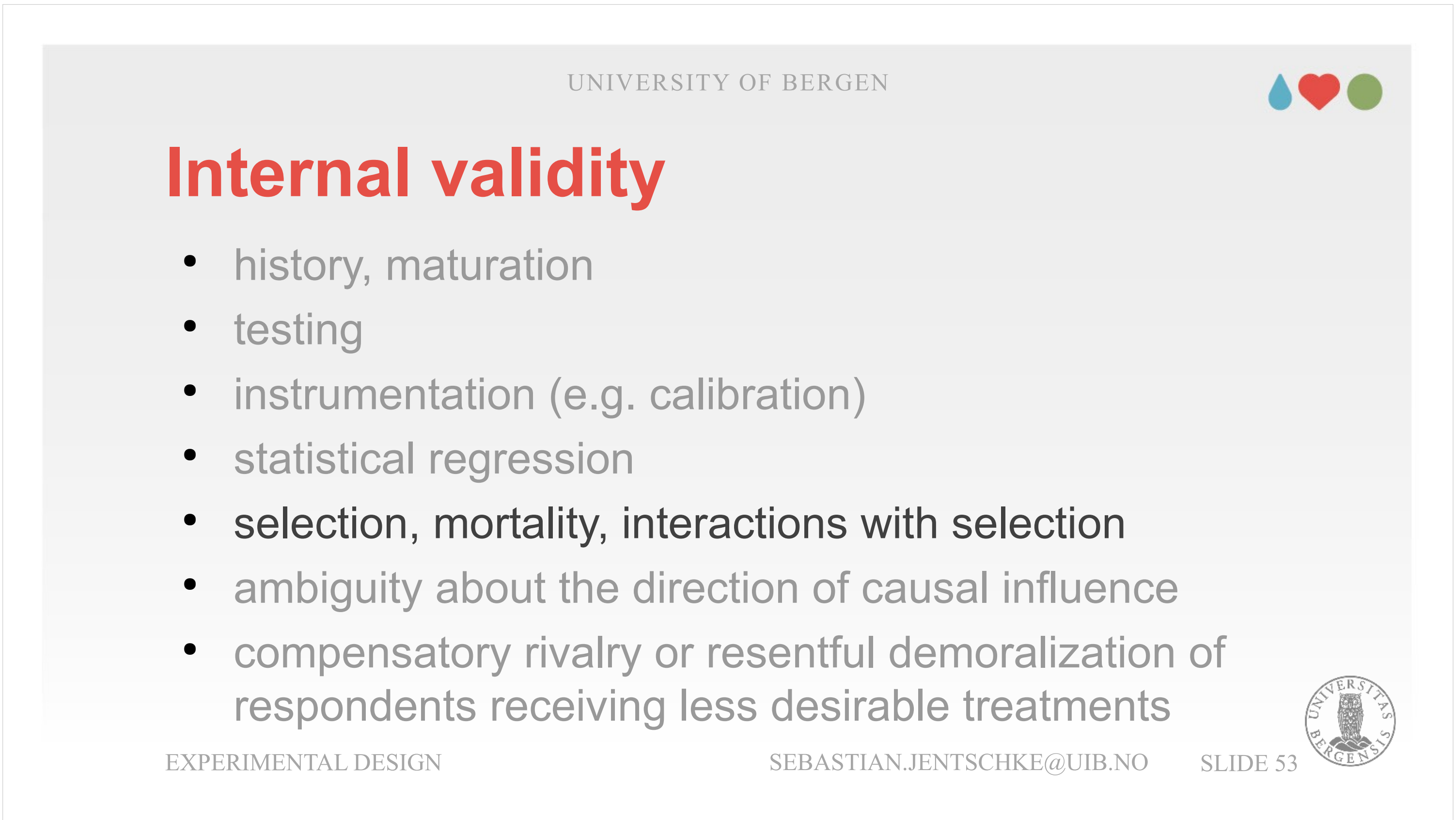
Changes in the calibration of a measuring instrument, shifts in the criteria used by observers and scorers, or unequal intervals in different ranges of a measuring instrument can affect the measurement of the dependent variable.



Internal validity

- history, maturation
- testing
- instrumentation (e.g. calibration)
- statistical regression
- selection, mortality, interactions with selection
- ambiguity about the direction of causal influence
- compensatory rivalry or resentful demoralization of respondents receiving less desirable treatments

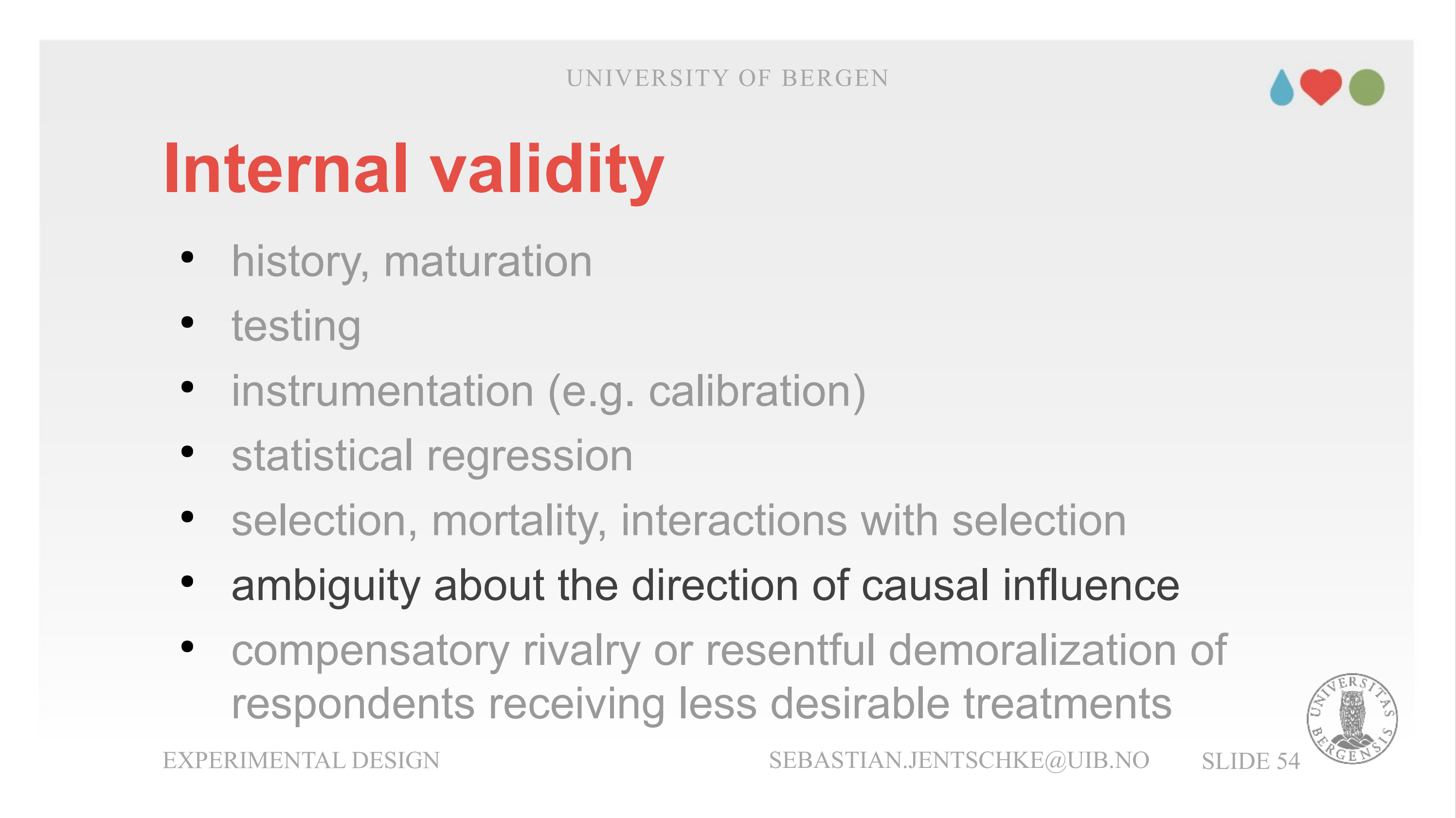
Statistical regression describes a tendency for extreme scores to regress or move toward the mean even when the measurement of the dependent variable is not perfectly reliable: Scores of subjects originally scoring very low on a test typically increase, those scoring very high on a test typically decrease, and those scoring around the mean of the test typically remains rather stable.



Internal validity

- history, maturation
- testing
- instrumentation (e.g. calibration)
- statistical regression
- selection, mortality, interactions with selection
- ambiguity about the direction of causal influence
- compensatory rivalry or resentful demoralization of respondents receiving less desirable treatments

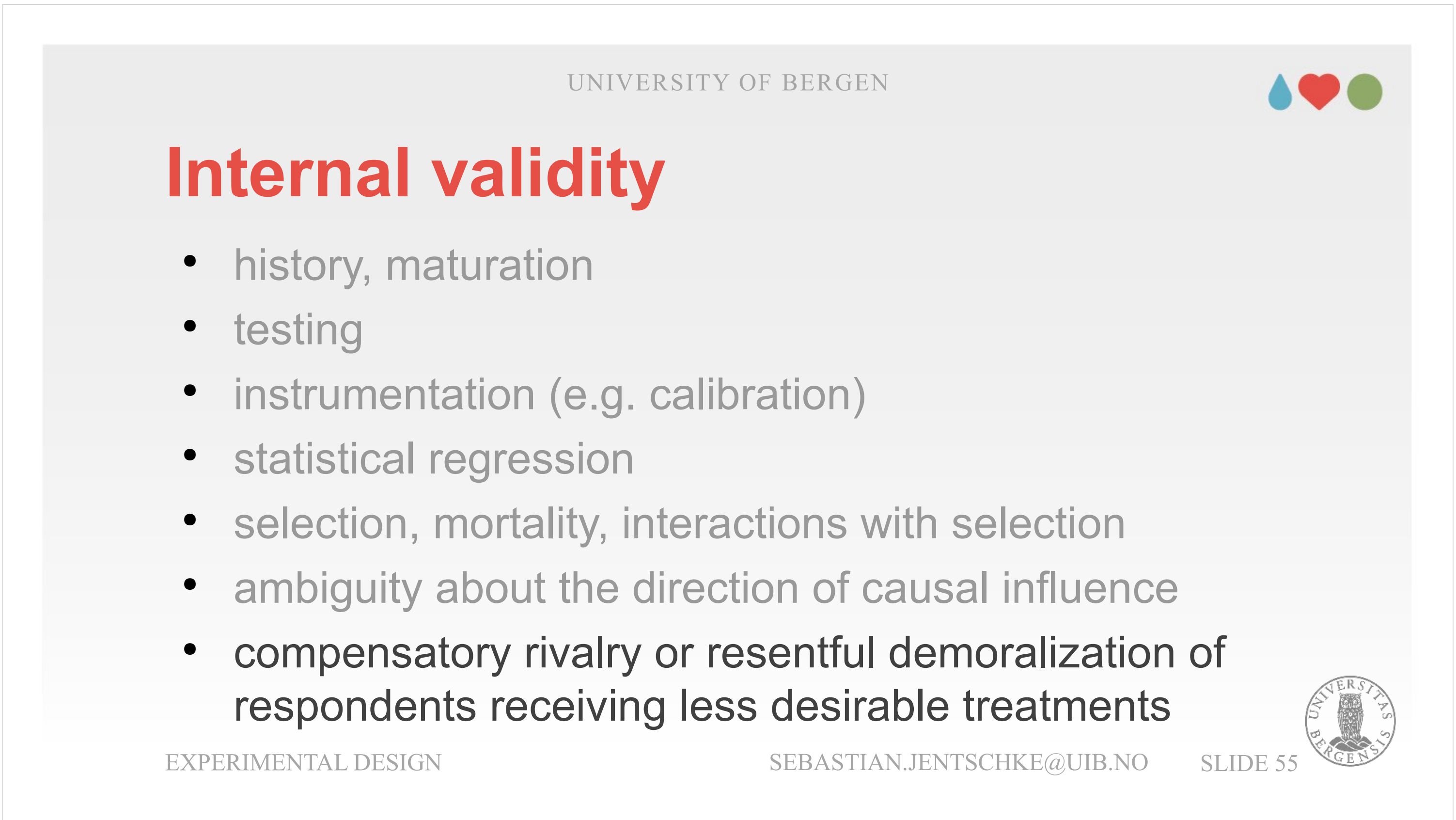
Differences among the group means of the dependent variable may reflect prior differences among the subjects assigned to the various levels of the independent variable. The loss of subjects in the various treatment conditions may also be selective and alter the distribution of subject characteristics across the treatment groups.



Internal validity

- history, maturation
- testing
- instrumentation (e.g. calibration)
- statistical regression
- selection, mortality, interactions with selection
- ambiguity about the direction of causal influence
- compensatory rivalry or resentful demoralization of respondents receiving less desirable treatments

In some types of research – for example, correlational studies – it may be difficult to determine the direction of causality. This ambiguity is not present when the manipulation of the independent variable is known to occur before measuring the dependent variable.



UNIVERSITY OF BERGEN

Internal validity

- history, maturation
- testing
- instrumentation (e.g. calibration)
- statistical regression
- selection, mortality, interactions with selection
- ambiguity about the direction of causal influence
- compensatory rivalry or resentful demoralization of respondents receiving less desirable treatments

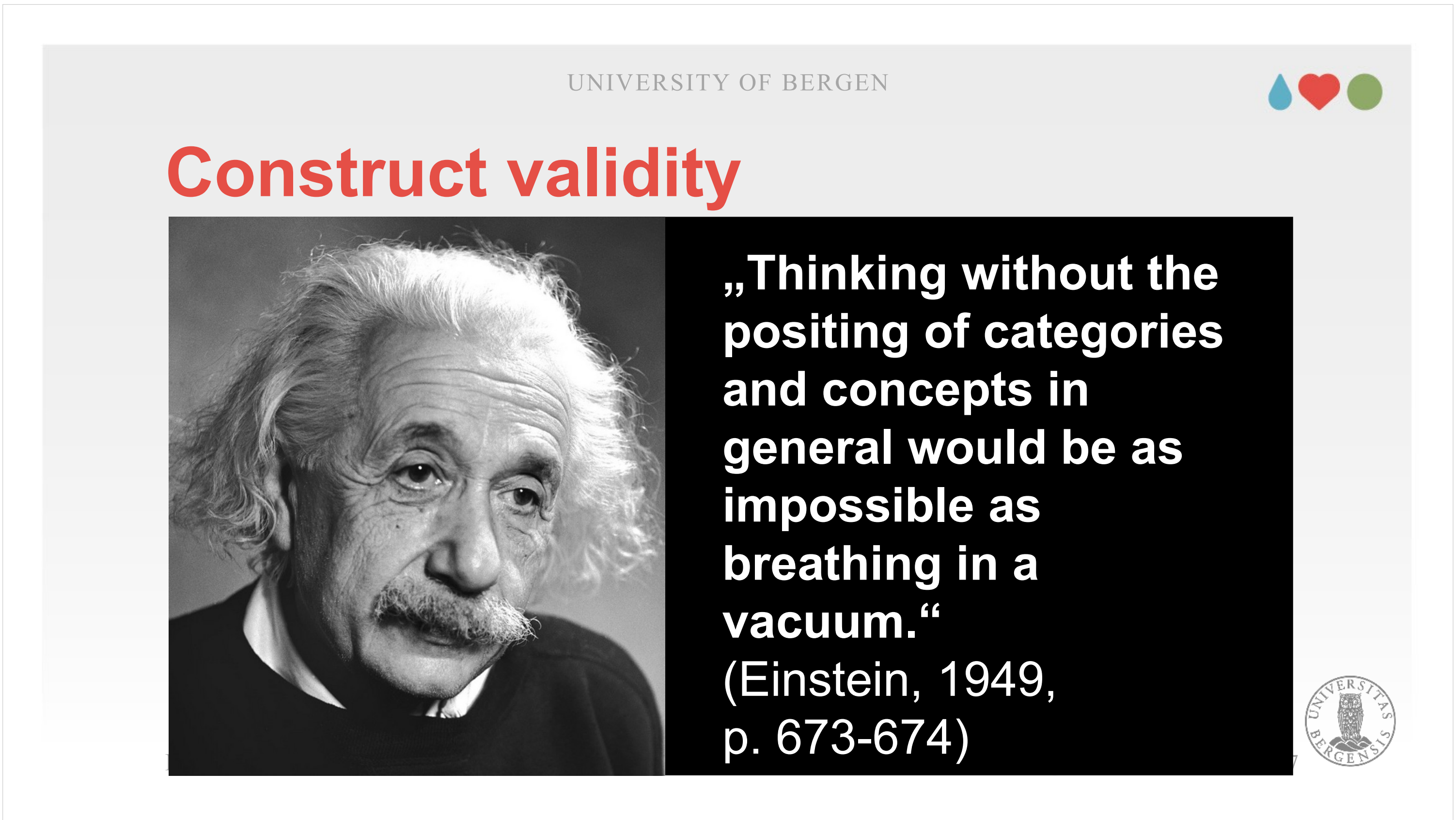
EXPERIMENTAL DESIGN SEBASTIAN.JENTSCHKE@UIB.NO SLIDE 55

Finally, social effects may interfere with applying the treatment or measuring of the outcome: How the independent variable is administered may affect the group who did not receive treatment. The treatment might be imitated if the participant from the different levels can communicate with one another.

There might be compensatory rivalry by respondents receiving less desirable treatments. Social competition may motivate these participants to attempt to overperform in order to reverse or reduce the anticipated effects of the desirable treatment levels.

There may also be effects in the opposite direction. Subjects receiving less desirable treatments may experience feelings of resentment and demoralization. That may result in them performing at an abnormally low level, thereby artificially increasing the magnitude of the difference between the experimental conditions.

Threats to construct validity



UNIVERSITY OF BERGEN

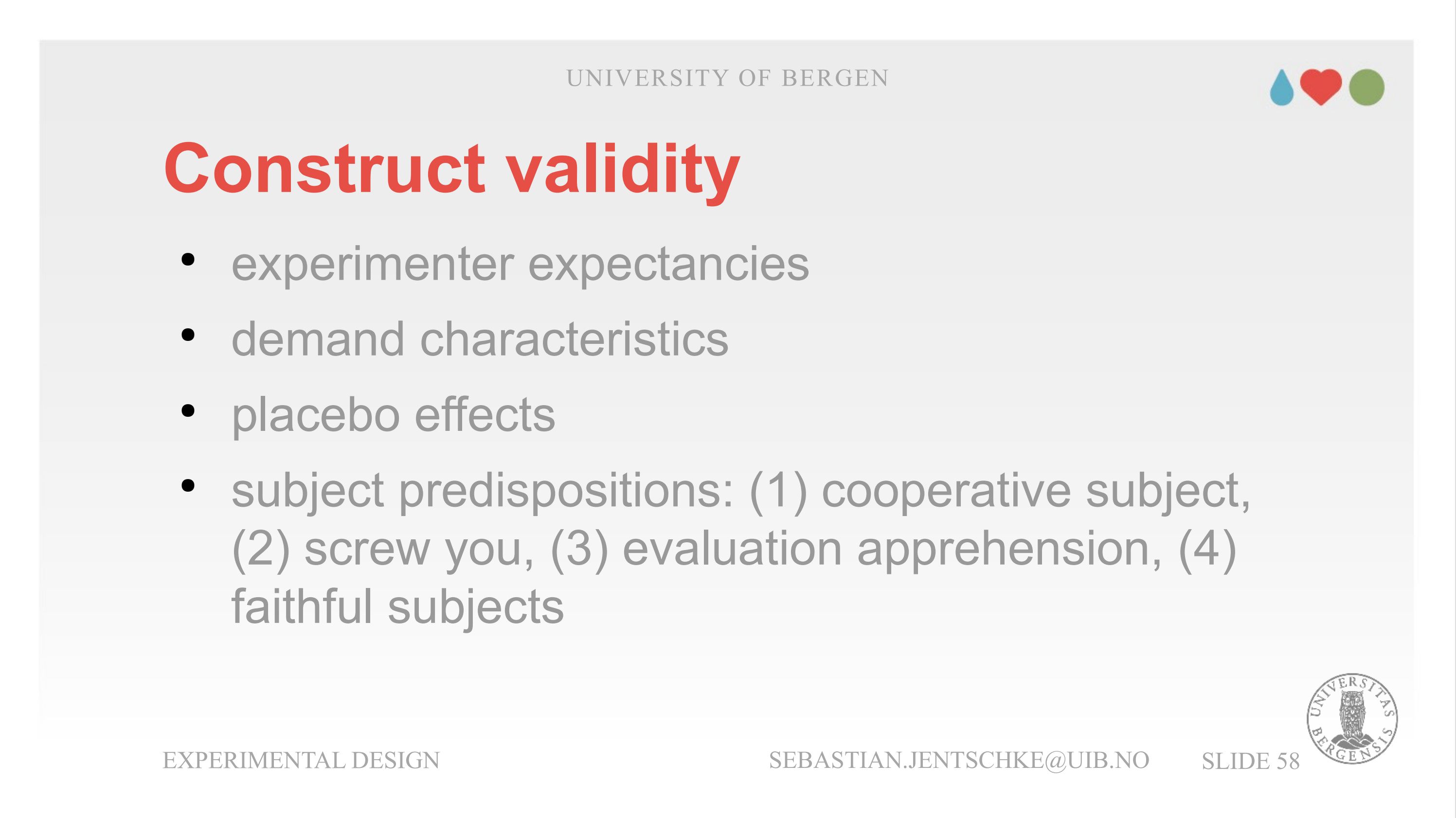
Construct validity

„Thinking without the positing of categories and concepts in general would be as impossible as breathing in a vacuum.“
(Einstein, 1949, p. 673-674)

What we do as scientists is to ***develop theories*** or ***hypotheses*** and ***put these to test***. By doing so, they accumulate knowledge and over time this knowledge is combined to form new theories or to integrate the knowledge with existing ones.

Such theories, also called constructs, are central means for connecting the operations used in an experiment to pertinent theory and language. In addition, constructs may have societal consequences and social, political and economic implications (shape perceptions, frame debates, and elicit support and criticism).

Threats to construct validity may include inadequate explication of constructs, i.e., constructs being not well enough defined, thereby leading to incorrect inferences or conclusions.



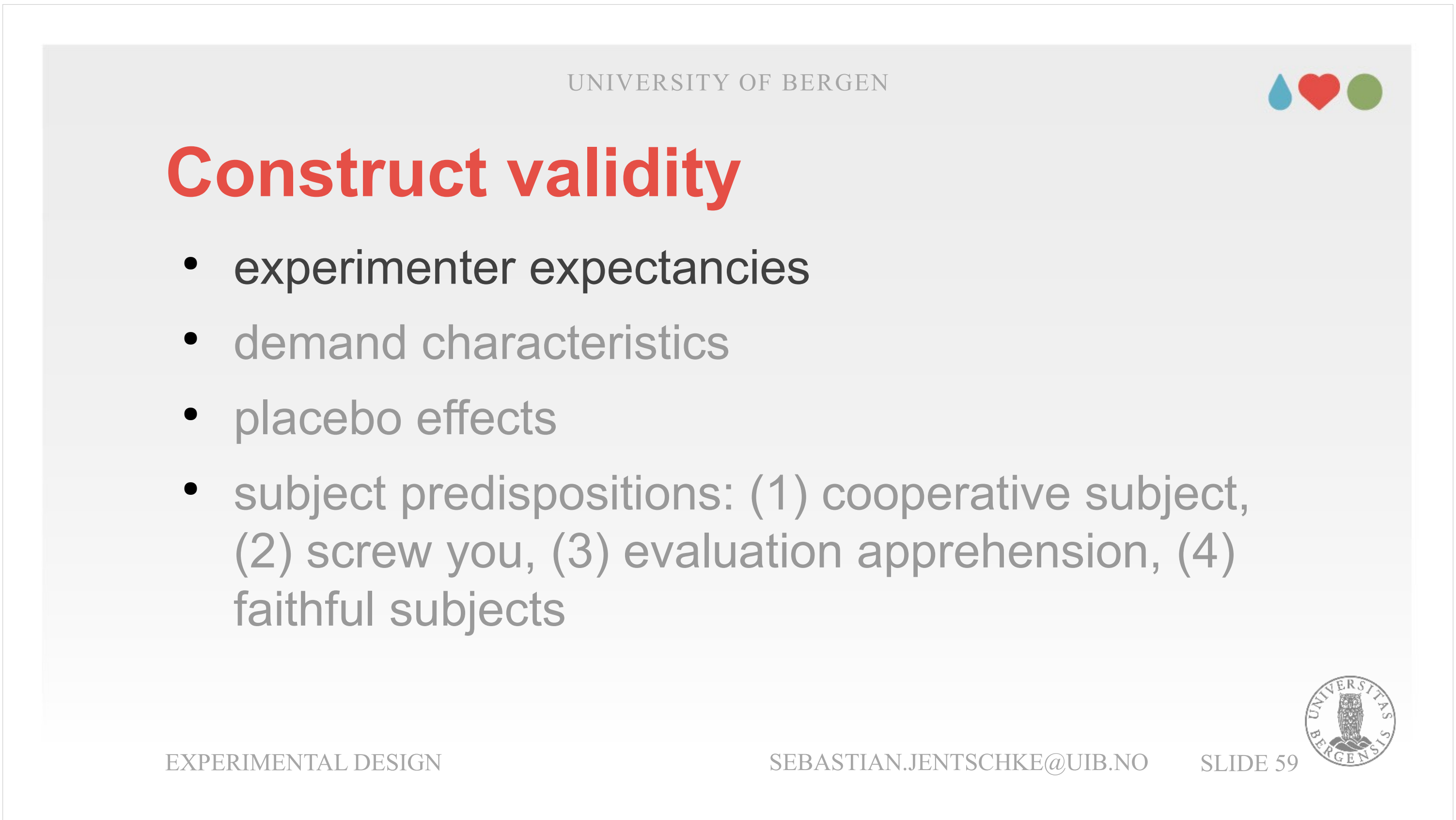
Construct validity

- experimenter expectancies
- demand characteristics
- placebo effects
- subject predispositions: (1) cooperative subject, (2) screw you, (3) evaluation apprehension, (4) faithful subjects

Unfortunately, theories about people, their behaviour and their inner workings are typically much more complex and much more subject to uncertainties than laws within physics.

As a consequence, a lot of unexpected effects may occur that are not covered by the theory laid out in the construct. Sources of such effects may be the participants themselves, the experimenter or the situation. The effects summarized on the slide therefore represent threats to construct validity.

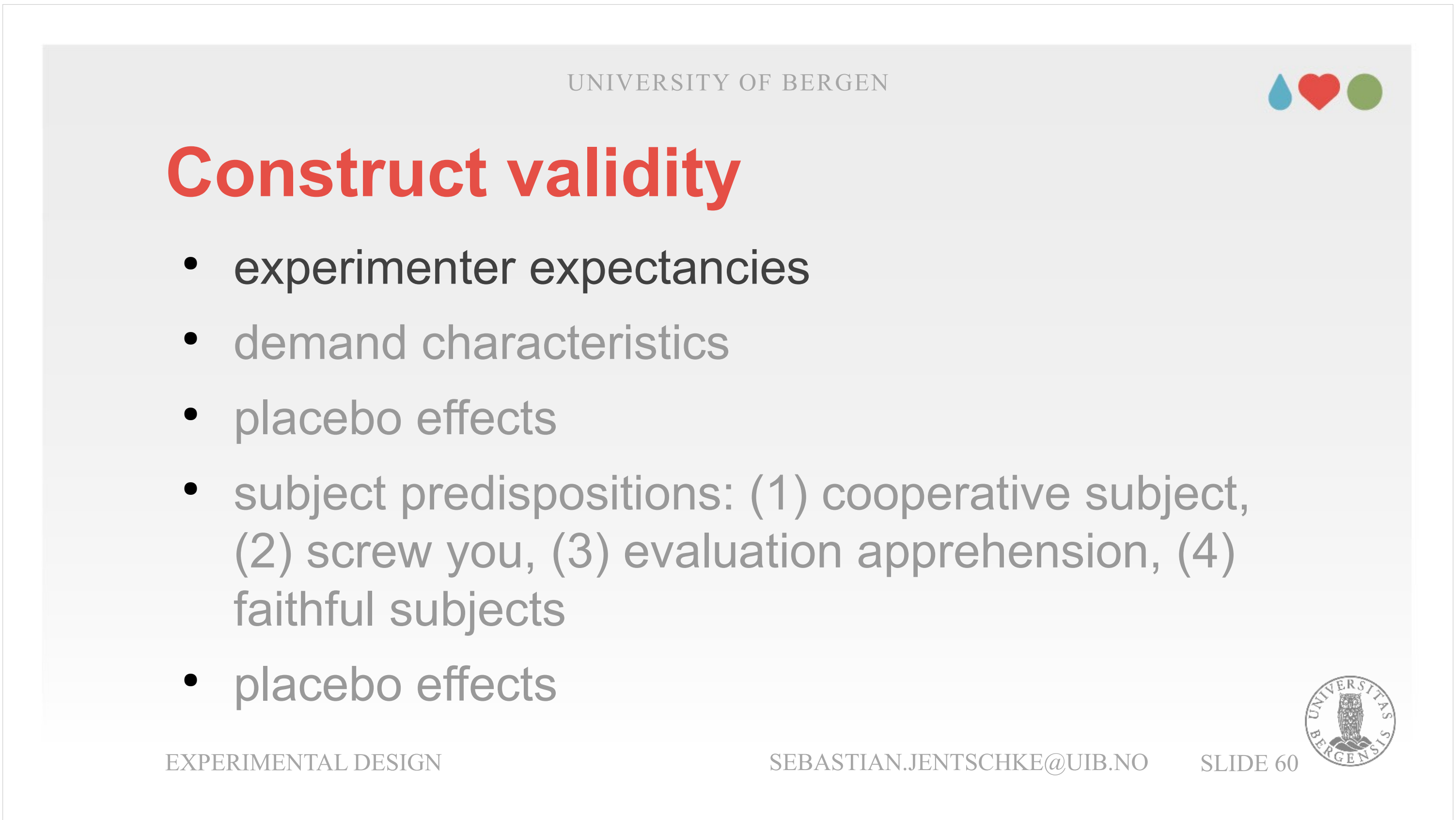
What furthermore makes them threats to construct validity is that they are all ***centered around*** what ***internal concepts*** or ***theories about*** what the ***experiment*** explores the ***experimenter*** or the ***participants*** may ***have***.



Construct validity

- experimenter expectancies
- demand characteristics
- placebo effects
- subject predispositions: (1) cooperative subject, (2) screw you, (3) evaluation apprehension, (4) faithful subjects

Experiments with human subjects are ***social situations*** in which one person behaves under the scrutiny of another. The two people have ***expectations*** about each other, ***communicate*** with each other, and form impressions about each other. The ***power*** in the situation is always unequal: The researcher requests a behaviour and the subject behaves. That might be accompanied by other more ***subtle requests and messages*** (using body language, tone of voice, and facial expressions). All can affect a subject's performance and contribute to a tendency to obtain data the researcher wants or expects to obtain (Rosenthal, 1963).



UNIVERSITY OF BERGEN

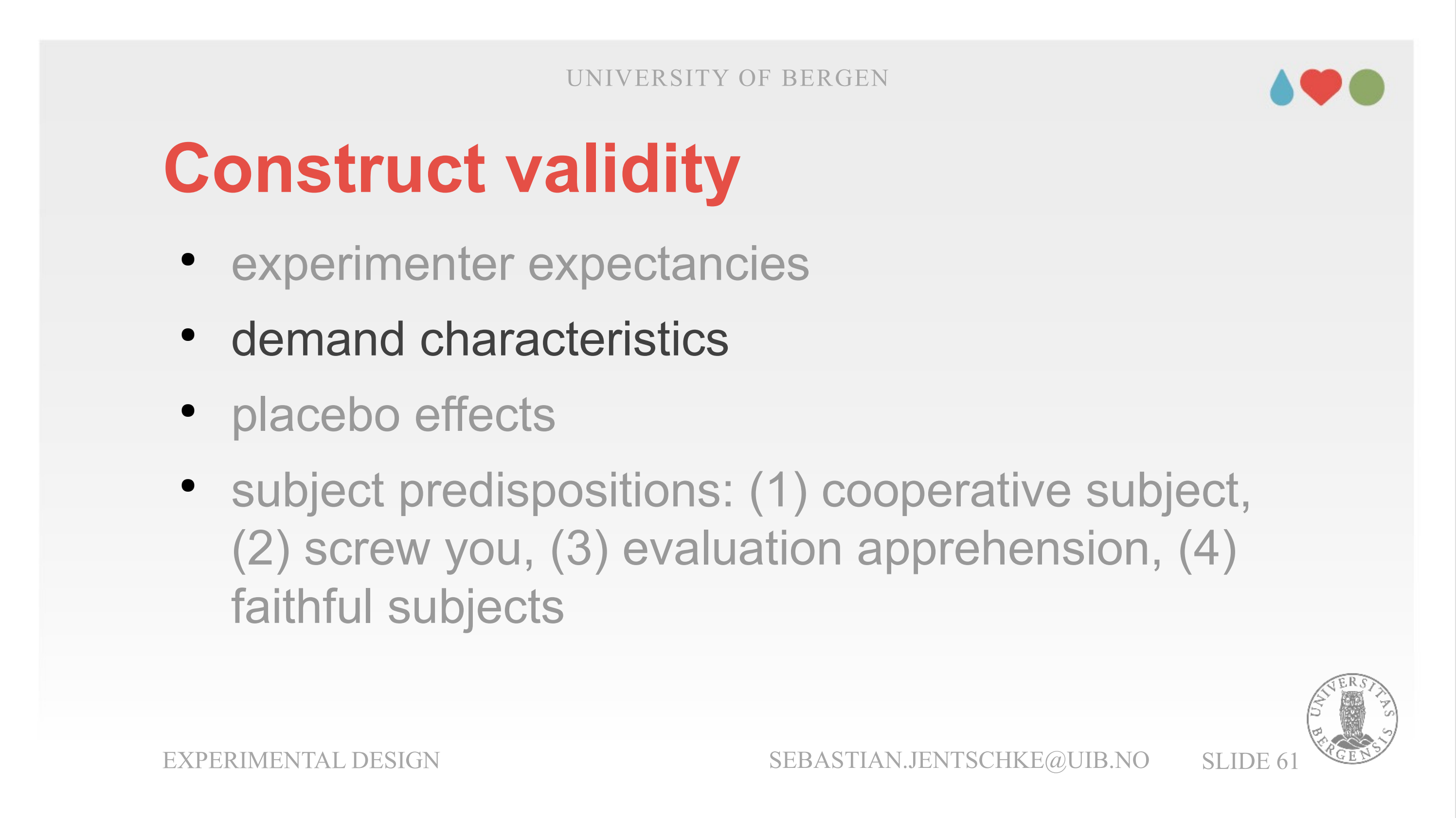
Construct validity

- experimenter expectancies
- demand characteristics
- placebo effects
- subject predispositions: (1) cooperative subject, (2) screw you, (3) evaluation apprehension, (4) faithful subjects
- placebo effects

EXPERIMENTAL DESIGN SEBASTIAN.JENTSCHKE@UIB.NO SLIDE 60

Beyond that, a researcher's expectations and desires also can influence the way the data are recorded, analysed, and interpreted.

For example, researchers are much more likely to recompute and double-check results that conflict with their hypotheses (Sheridan, 1976). Furthermore, observational or recording errors are typically in the direction of the hypothesis even though those errors are usually small and unintentional (Rosenthal, 1969, 1978).

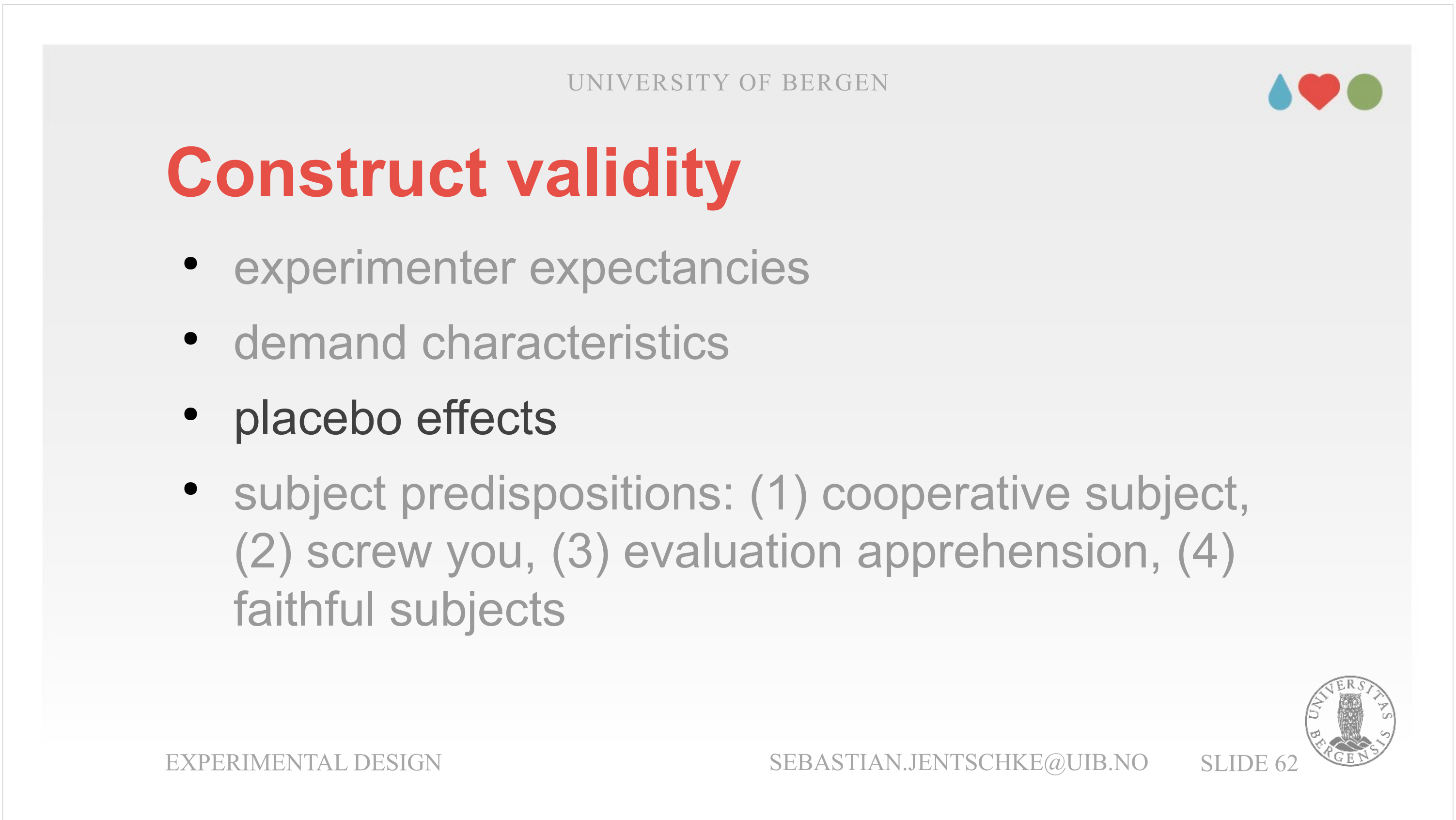


Construct validity

- experimenter expectancies
- demand characteristics
- placebo effects
- subject predispositions: (1) cooperative subject, (2) screw you, (3) evaluation apprehension, (4) faithful subjects

Demand characteristics are another source of bias in an experiment (Orne, 1962). They refer to any aspect of the ***experimental environment*** or ***procedure*** that ***leads a subject*** to make ***inferences*** about the ***purpose of an experiment*** and to respond in accordance with (or in some cases, contrary) to the perceived purpose.

Demand characteristics influence a subject's perceptions of what is appropriate or expected. They can result from rumours about an experiment, what subjects are told when they sign up for an experiment, the laboratory environment, or the communication that occurs during the course of an experiment.

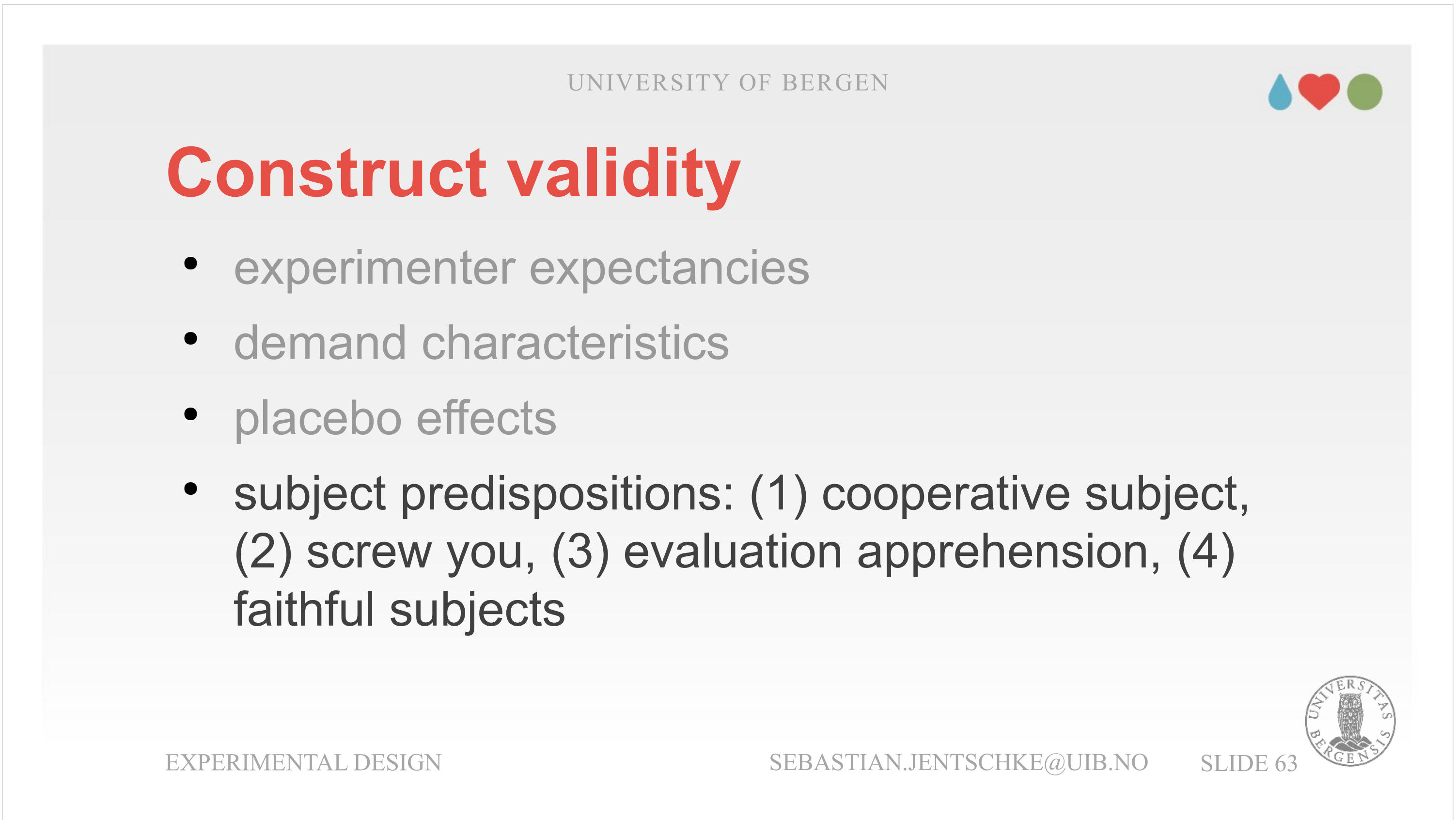


Construct validity

- experimenter expectancies
- demand characteristics
- placebo effects
- subject predispositions: (1) cooperative subject, (2) screw you, (3) evaluation apprehension, (4) faithful subjects

Placebo describes an inert substance or neutral stimulus that is administered as if it was the actual treatment condition. Any change in the dependent variable attributable to receiving a placebo is called the placebo effect.

Subjects in an experiment are not entirely naive. If they expect that an experimental condition will have a particular effect, they are likely to behave in a manner consistent with this expectation. If they believed a medication will relieve a particular symptom, they may report feeling better even though they have received a chemically inert substance.



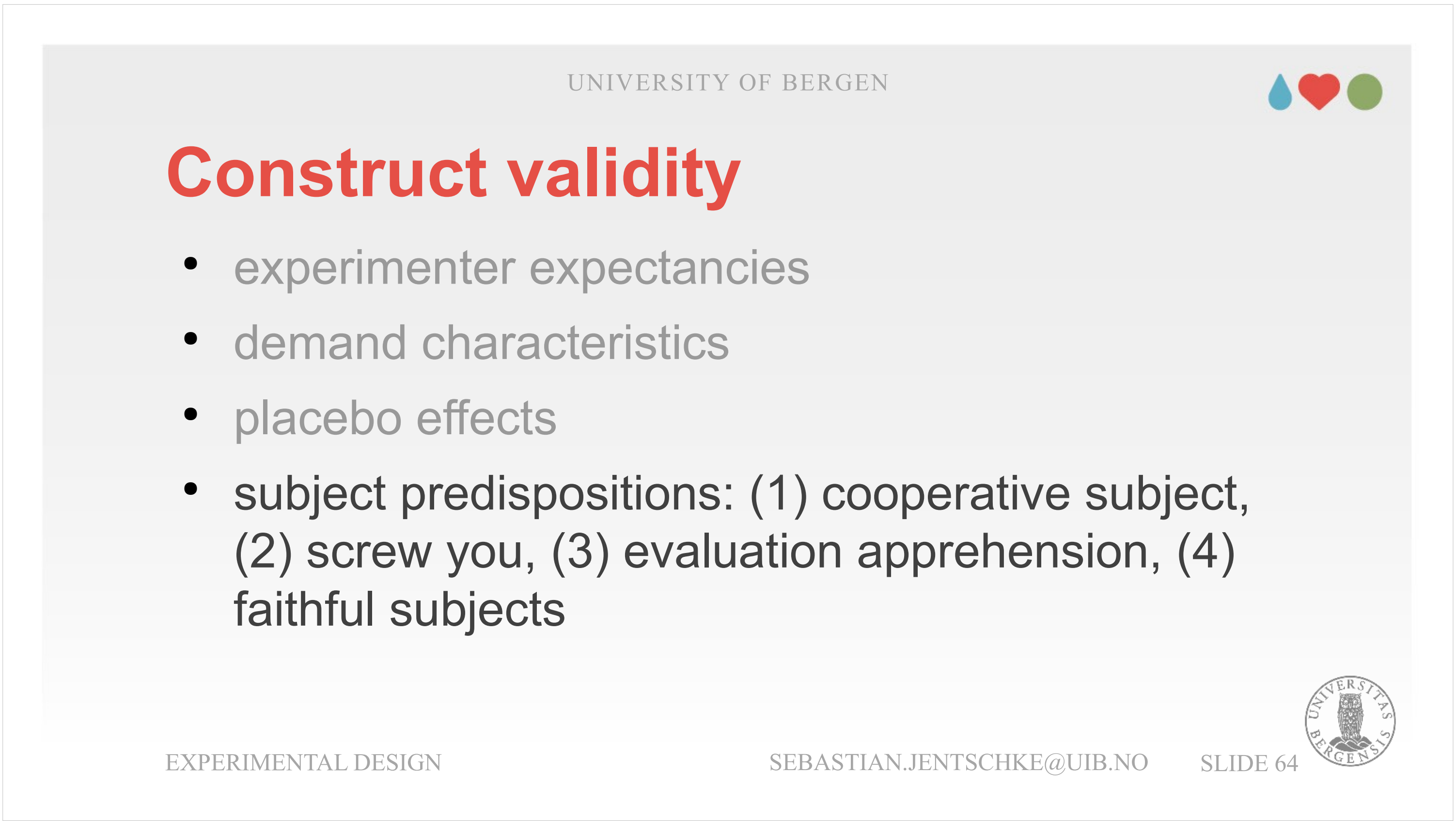
Construct validity

- experimenter expectancies
- demand characteristics
- placebo effects
- subject predispositions: (1) cooperative subject, (2) screw you, (3) evaluation apprehension, (4) faithful subjects

Subject may have predispositions to respond in a particular way. Those can be divided four categories:

(1) Cooperative subjects are concerned to ***please the researcher*** and be a “good subject.” and are therefore particularly susceptible to experimenter-expectancy effects. They try, consciously or unconsciously, to provide data that support the researcher's hypothesis.

(2) Subjects may be ***uncooperative*** or even try to sabotage the experiment. Reasons may include being ***required to participate***, ***bad experience*** in previous experiments (such as being deceived or made feel inadequate), or a ***dislike*** for ***persons associated*** with the ***experiment***. Uncooperative subjects may try, consciously or unconsciously, to provide data that do not support the researcher's hypothesis.



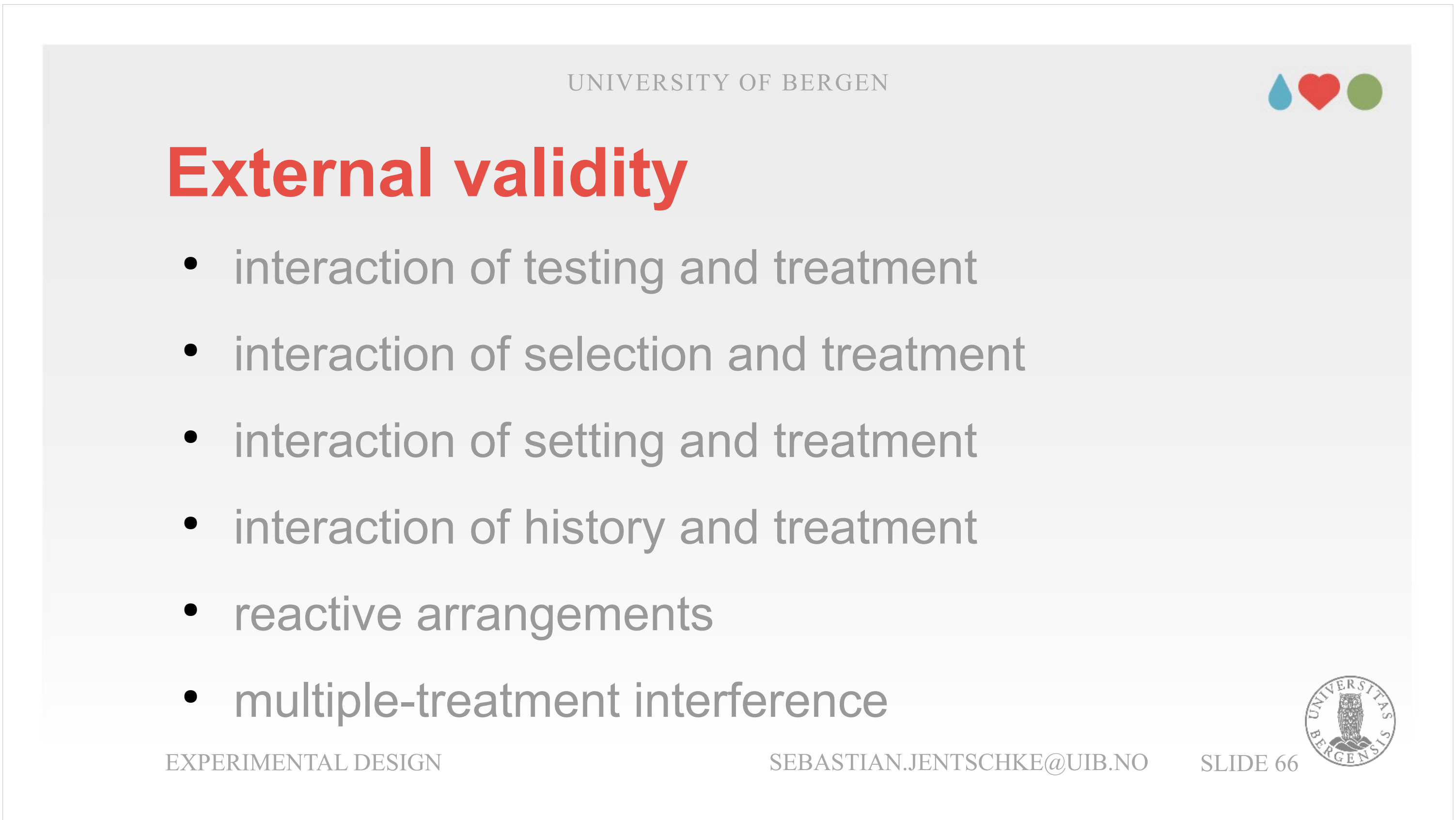
Construct validity

- experimenter expectancies
- demand characteristics
- placebo effects
- subject predispositions: (1) cooperative subject, (2) screw you, (3) evaluation apprehension, (4) faithful subjects

(3) Subjects may be apprehensive about being evaluated. Their primary concern is to gain a positive evaluation from the researcher, and therefore the aim to provide data that make them appear intelligent, well adjusted, etc. and to ***avoid revealing*** characteristics that they consider ***undesirable***.

(4) Faithful subjects try to put aside their own hypotheses about the purpose of an experiment and to ***follow*** the ***researcher's instructions*** to the letter. Often they are motivated by a desire to advance scientific knowledge.

Threats to external validity



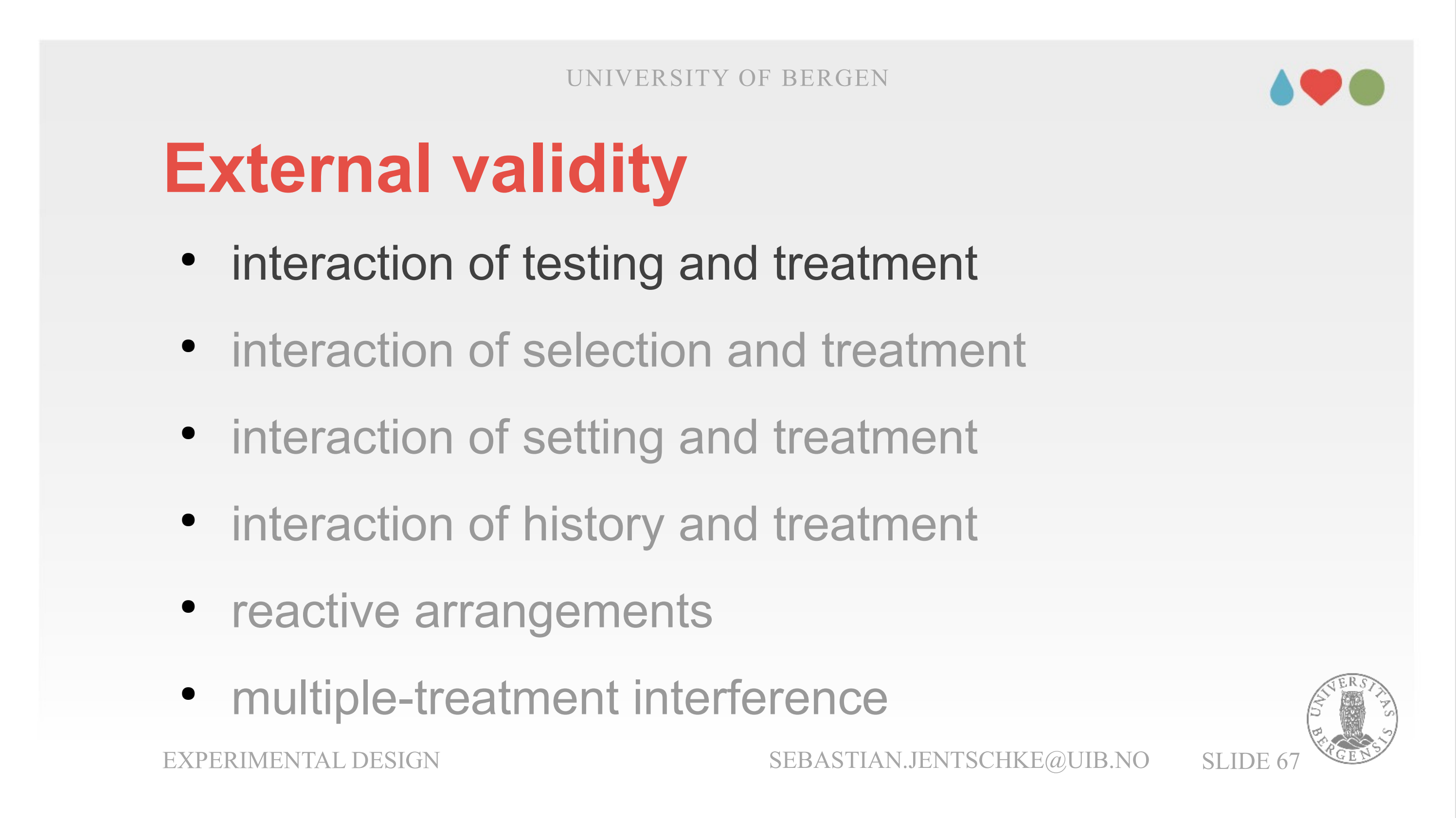
External validity

- interaction of testing and treatment
- interaction of selection and treatment
- interaction of setting and treatment
- interaction of history and treatment
- reactive arrangements
- multiple-treatment interference

External validity describes to what degree a causal relationship explored in an experiment also applies to the population, i.e., to what degree that relationship can be generalized.

It is the ***extent*** to which a ***causal relationship holds*** when it is ***taken from*** persons, settings, treatments and outcomes in the ***experiment*** and is ***applied*** to those who where not. Our goal must be to design experiments that are more valid externally (e.g., by testing whether treatment effects hold over different outcomes / measures or different kinds of persons).

However, the ***problem*** with such strategy is often an ***economical*** one: Given the ***typically heterogenous*** range of persons, treatments, outcomes and settings such strategy ***requires large samples*** to obtain adequate power (and collecting data is costly).



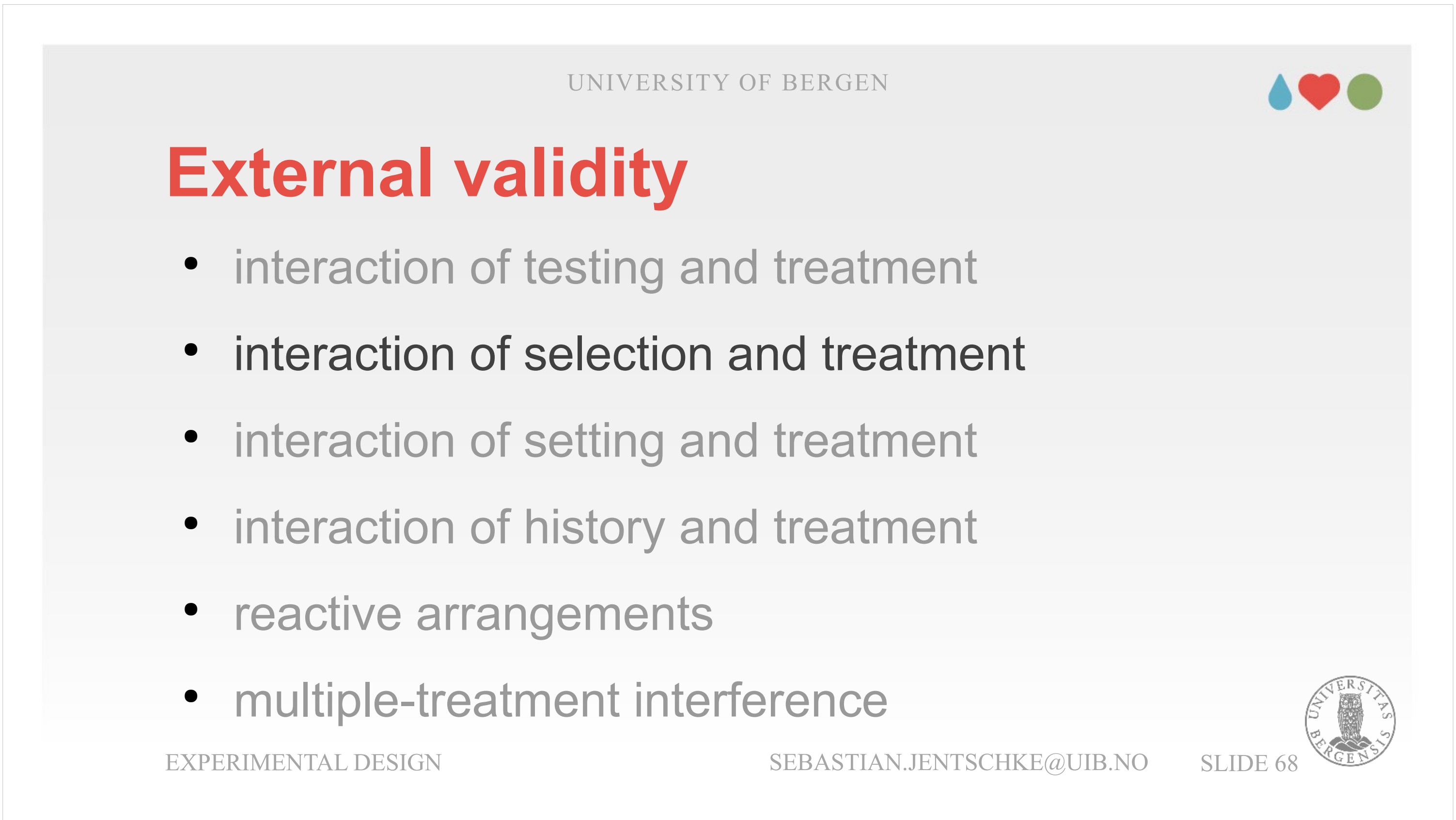
UNIVERSITY OF BERGEN

External validity

- interaction of testing and treatment
- interaction of selection and treatment
- interaction of setting and treatment
- interaction of history and treatment
- reactive arrangements
- multiple-treatment interference

When it comes to ***threats to external validity***, ***most*** of them can be described as ***interaction*** of some variable ***with the treatment***.

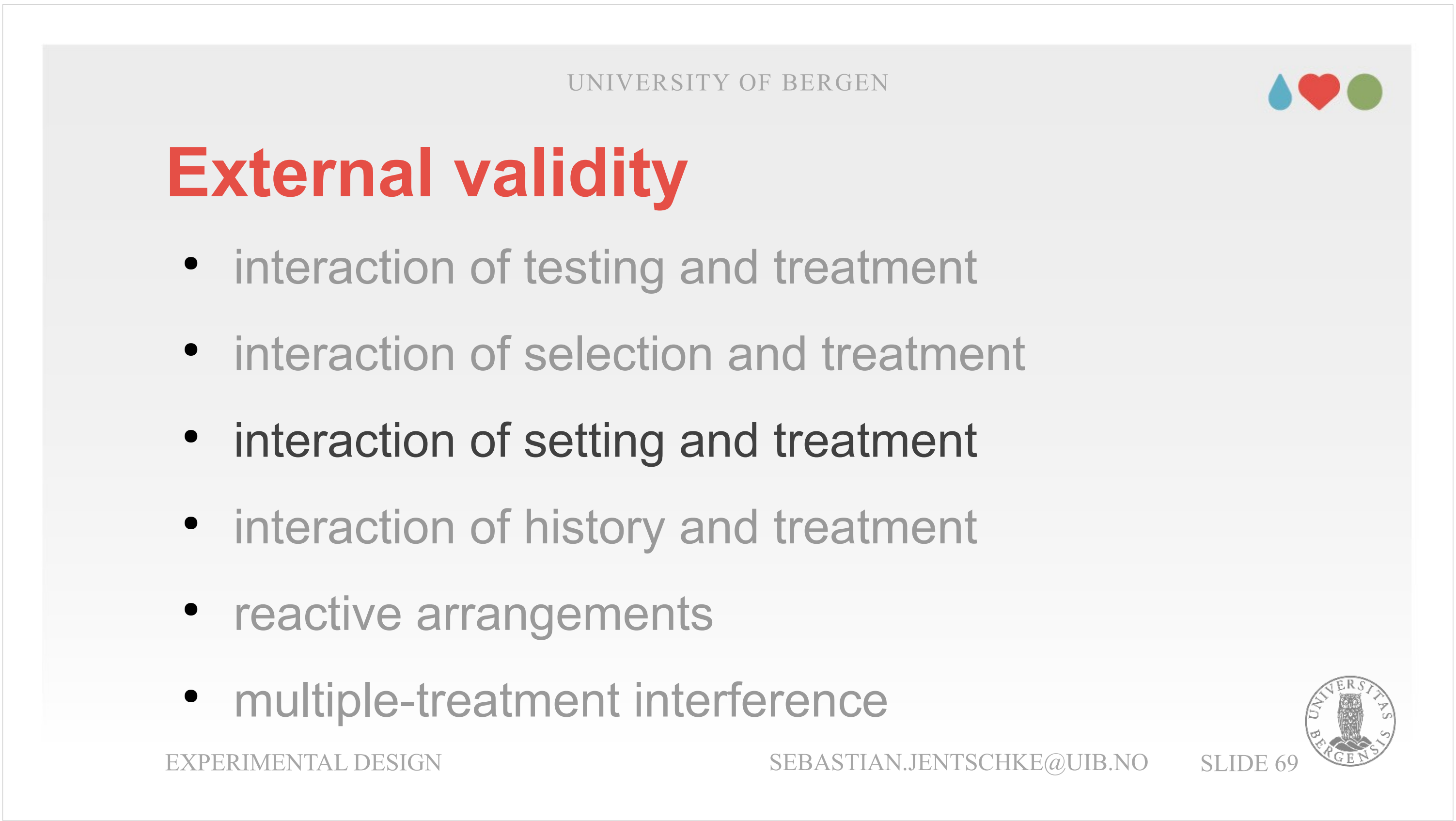
Repeated testing may affect the results and may not generalize to situations that do not involve repeated testing. The testing may ***sensitize*** subjects and, by focusing attention on the topic, ***enhance*** the effectiveness of a treatment. It can work either way and could ***also reduce the effectiveness*** of a treatment when the subjects' sensitivity to a topic is diminished (e.g., because of boredom as the test is applied several times or lack of interest).



External validity

- interaction of testing and treatment
- interaction of selection and treatment
- interaction of setting and treatment
- interaction of history and treatment
- reactive arrangements
- multiple-treatment interference

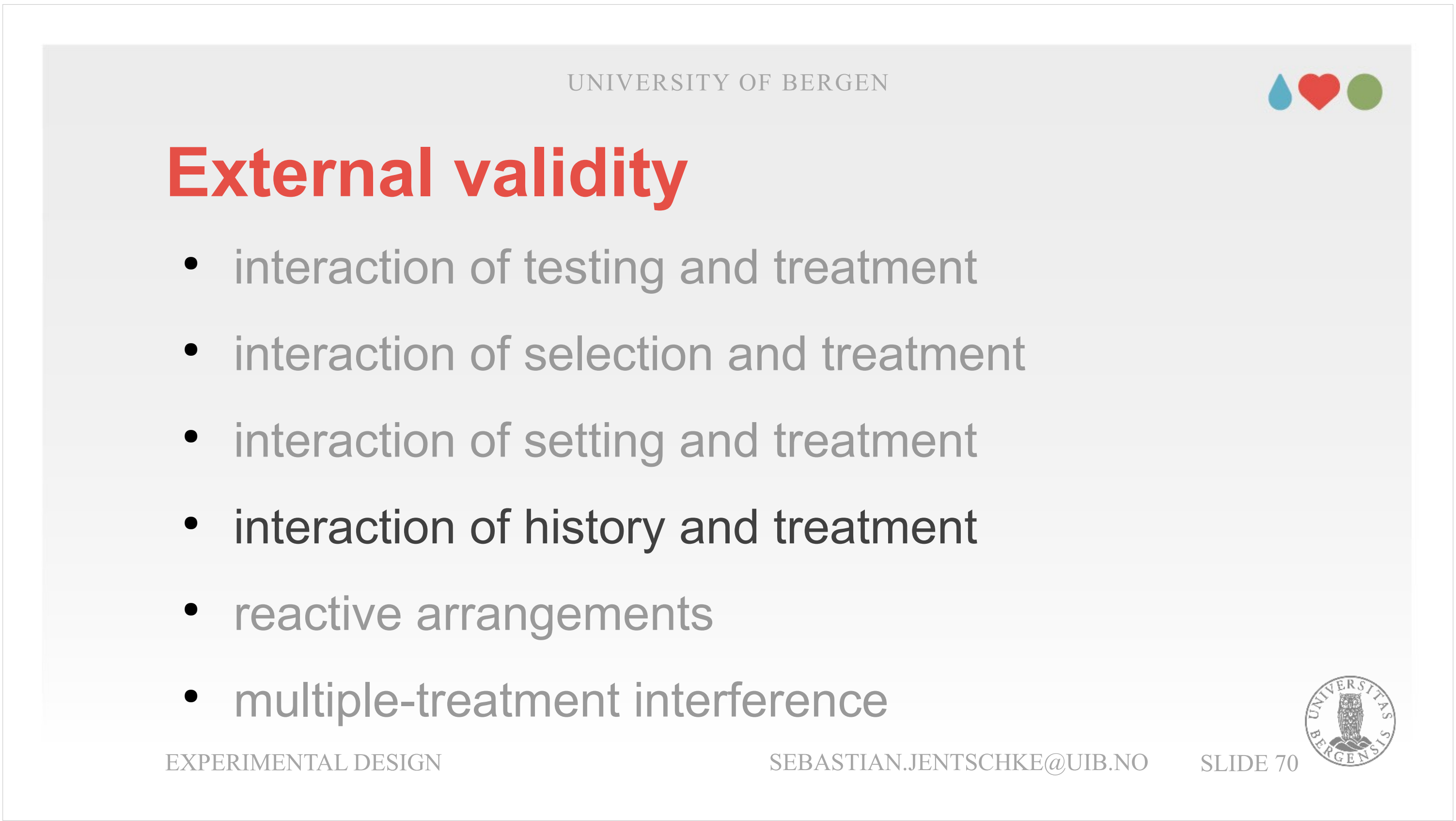
Factors affecting the ***availability of participants*** and their characteristics may also restrict the generalizability of results. Volunteers or students that participate for course credits may be examples where the results fail to generalize beyond these populations.



External validity

- interaction of testing and treatment
- interaction of selection and treatment
- interaction of setting and treatment
- interaction of history and treatment
- reactive arrangements
- multiple-treatment interference

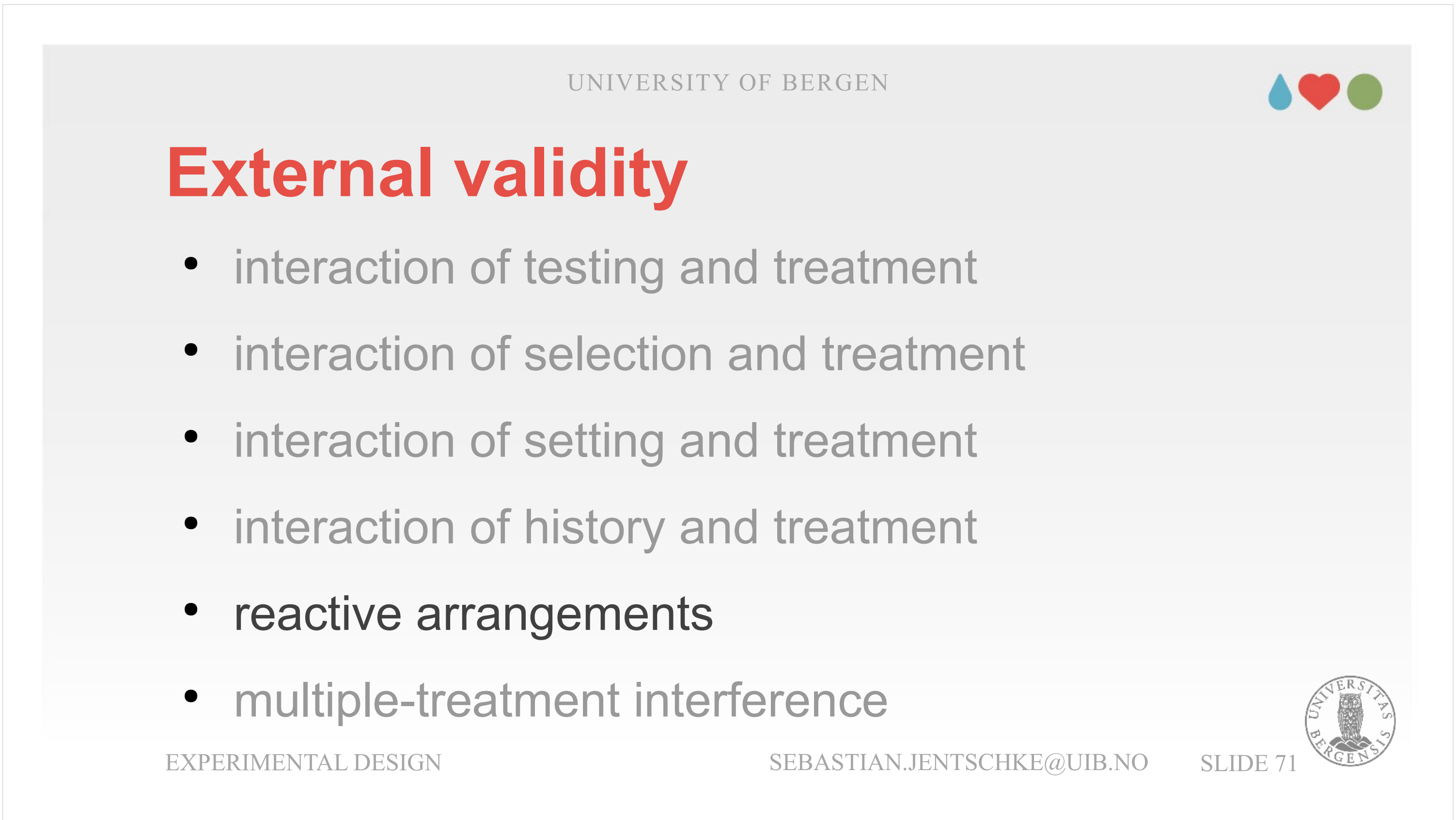
Unique ***characteristics of the experimental setting*** may restrict the generalizability of the results. Such cases are quite common since experiments are typically conducted in a laboratory under ***very controlled conditions*** and may fail to generalize to ***“real world”*** situations.



External validity

- interaction of testing and treatment
- interaction of selection and treatment
- interaction of setting and treatment
- interaction of history and treatment
- reactive arrangements
- multiple-treatment interference

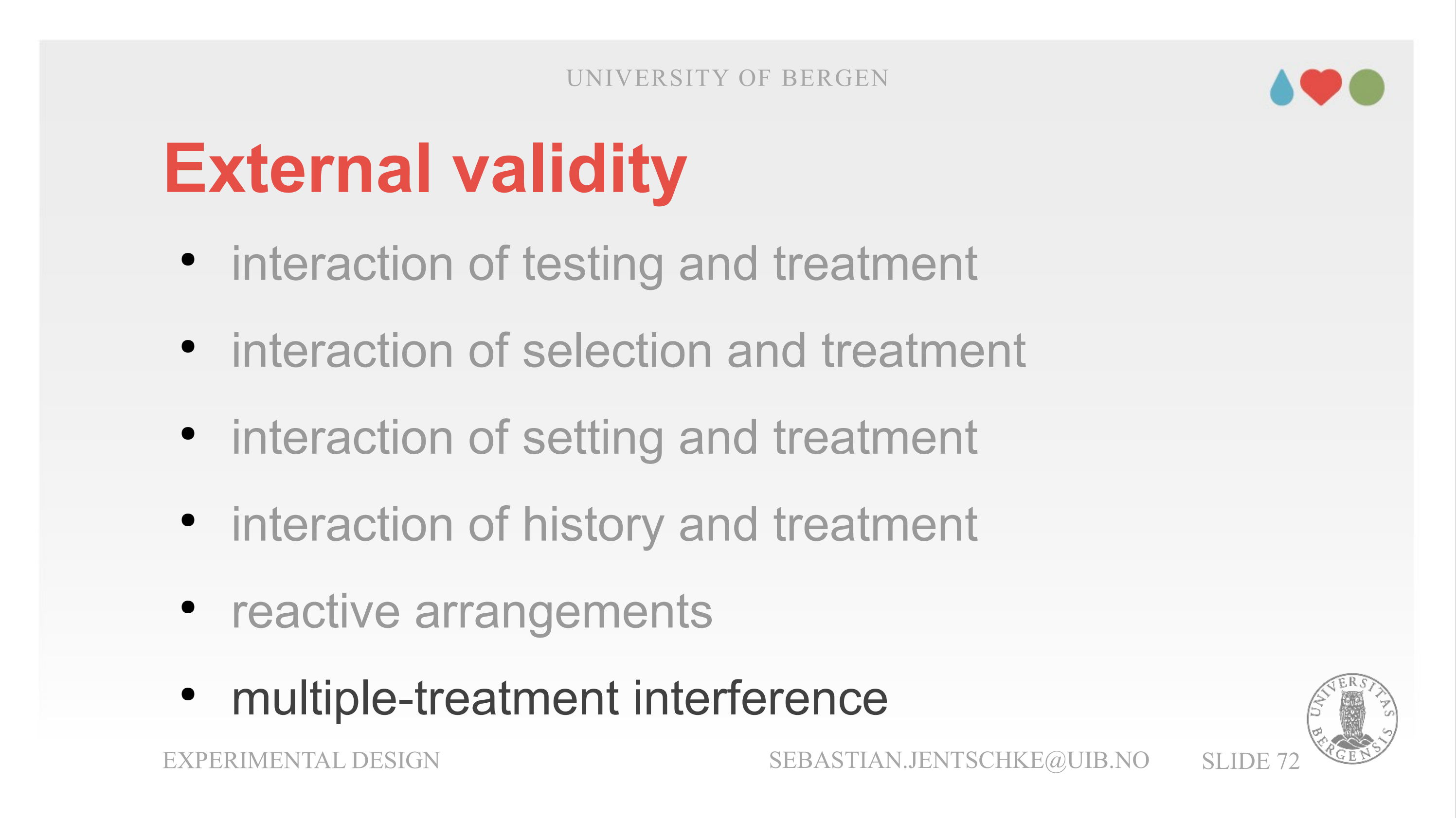
Occasionally results are obtained on the same day as an ***event*** that is particularly ***noteworthy*** to the participant. Such results may be different from results that would have been obtained in the absence of that noteworthy event. However, even smaller fluctuations (e.g., in mood or motivation) may leave consequences.



External validity

- interaction of testing and treatment
- interaction of selection and treatment
- interaction of setting and treatment
- interaction of history and treatment
- reactive arrangements
- multiple-treatment interference

Subjects who are ***aware*** that they are ***being observed*** may ***behave differently*** than subjects who are not aware that they are being observed.

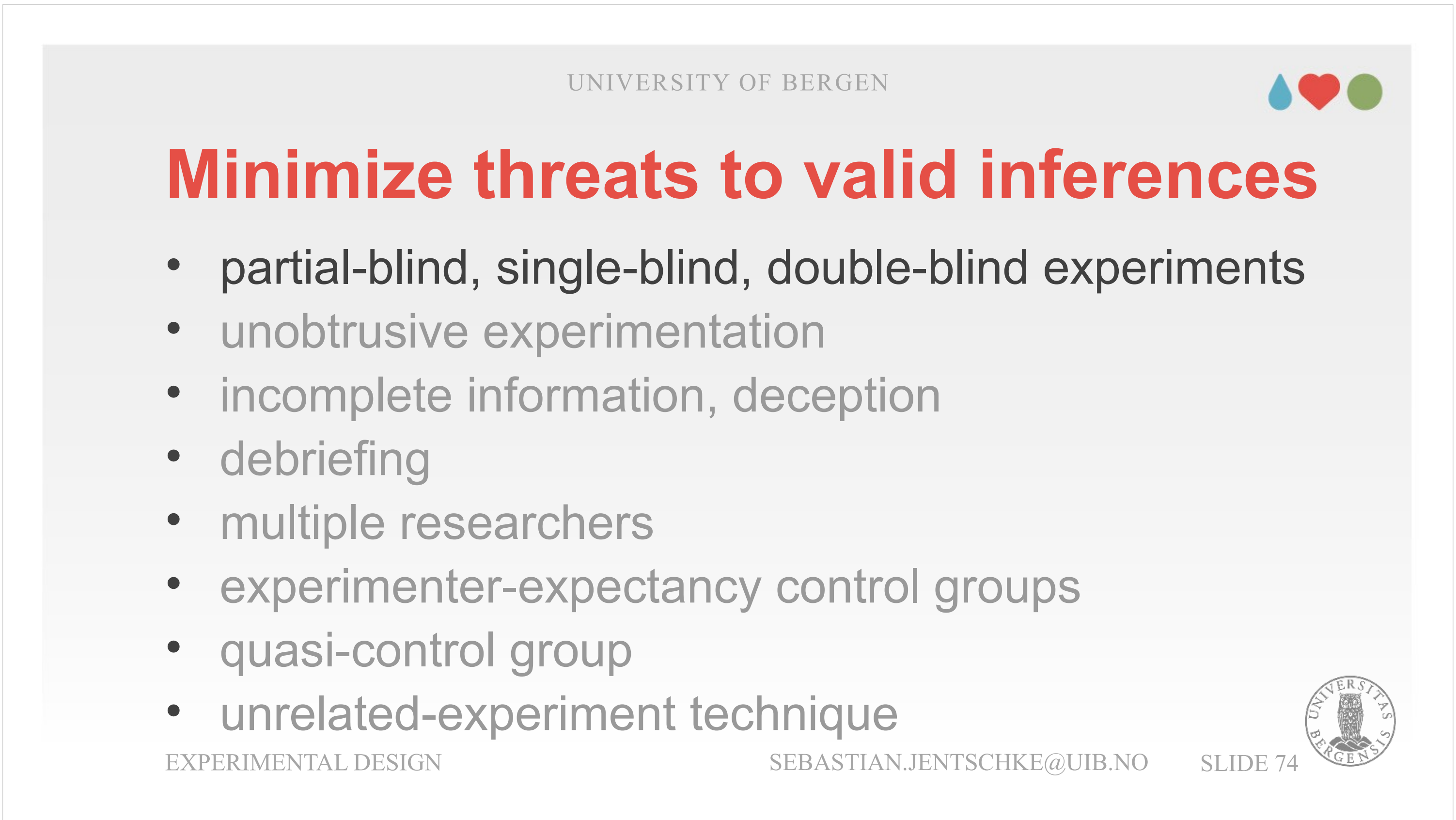


External validity

- interaction of testing and treatment
- interaction of selection and treatment
- interaction of setting and treatment
- interaction of history and treatment
- reactive arrangements
- multiple-treatment interference

When subjects are exposed to ***more than one treatment***, the results ***may not generalize*** to others ***not*** receiving the ***same combination*** of treatments.

How to minimize threats to valid inferences?



Minimize threats to valid inferences

- partial-blind, single-blind, double-blind experiments
- unobtrusive experimentation
- incomplete information, deception
- debriefing
- multiple researchers
- experimenter-expectancy control groups
- quasi-control group
- unrelated-experiment technique

Partial-blind, single-blind, and double-blind experimentation follow the same rationale: not revealing information about the nature of the treatment or the purpose of the experiment. All strategies aim to minimize the effects of experimenter expectancy and demand characteristics.

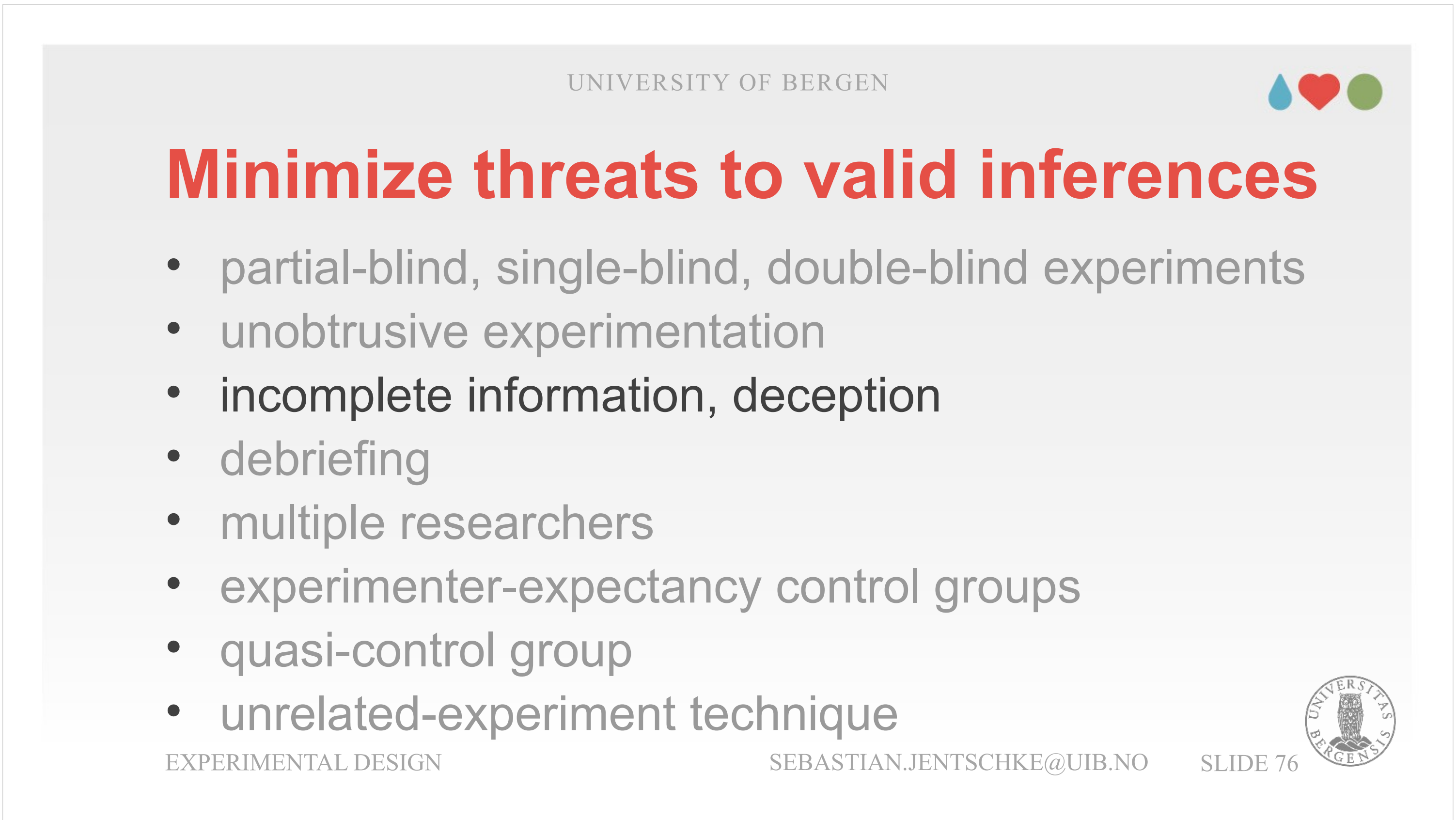
(a) Partial-blind experiments denote that the researcher does not know until just before administering the treatment level which level will be administered. (b) In single-blind procedures, the subjects are not informed. (c) In a double-blind experiment, neither the subjects nor the researcher are informed.

Even though double-blind is the most desirable of these options, practical considerations may limit what degree of blindness is possible: Many treatments are of such a nature that they are easily identified by a researcher (and possibly the participant). Informed consent requirements may also prevent withholding information.

Minimize threats to valid inferences

- partial-blind, single-blind, double-blind experiments
- unobtrusive experimentation
- incomplete information, deception
- debriefing
- multiple researchers
- experimenter-expectancy control groups
- quasi-control group
- unrelated-experiment technique

Unobtrusive experimentation denotes that ***subjects*** are ***not aware*** that they are ***participating*** in an experiment. It aims at minimizing the influence of reactive arrangements and demand characteristics.



UNIVERSITY OF BERGEN

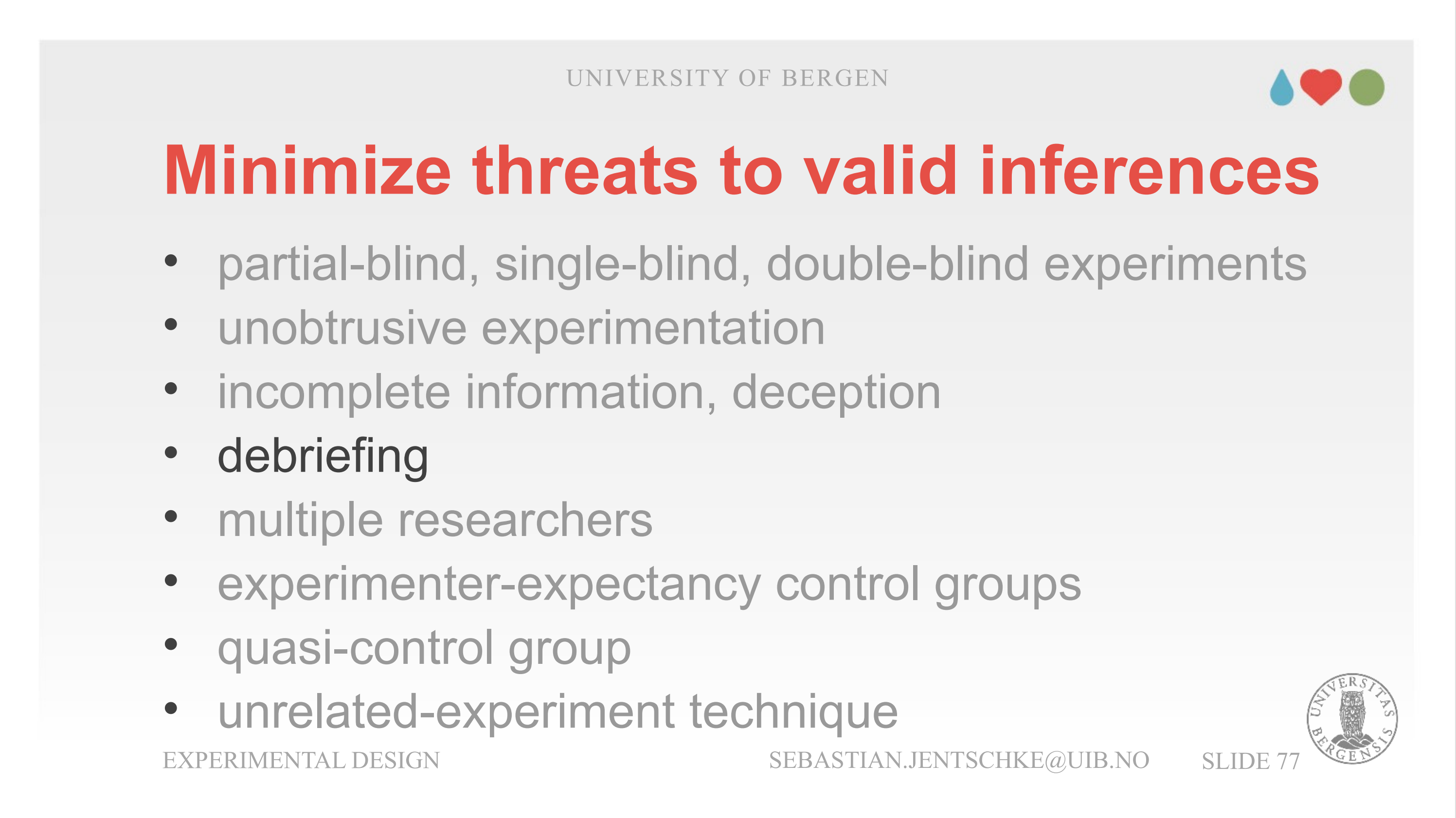
Minimize threats to valid inferences

- partial-blind, single-blind, double-blind experiments
- unobtrusive experimentation
- incomplete information, deception
- debriefing
- multiple researchers
- experimenter-expectancy control groups
- quasi-control group
- unrelated-experiment technique

EXPERIMENTAL DESIGN SEBASTIAN.JENTSCHKE@UIB.NO SLIDE 76

Incomplete information or even deception occurs when subjects are not told the relevant details of an experiment or when they are told that the experiment has one purpose when in fact the purpose is really something else.

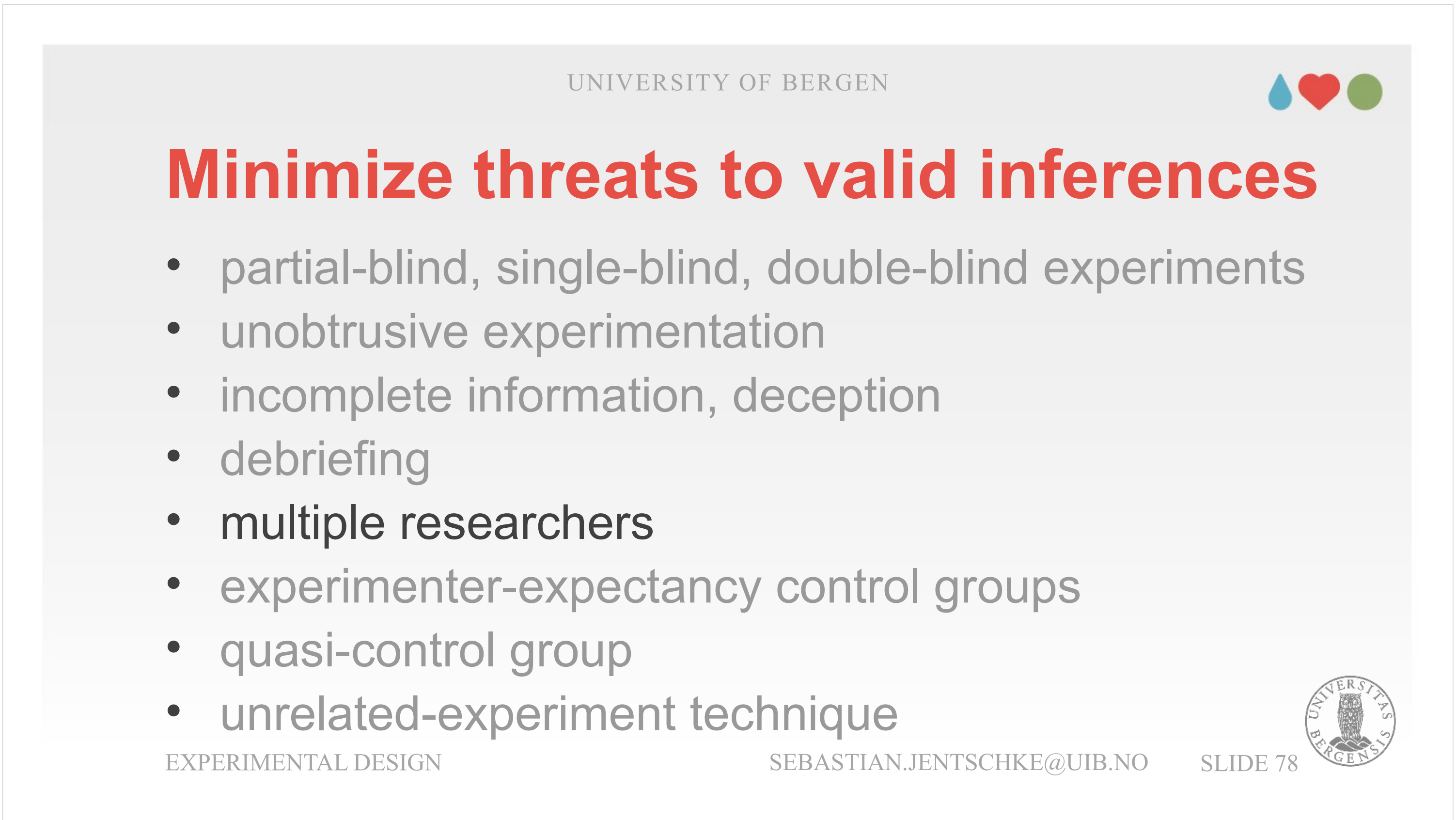
The aim is to minimize the effects of demand characteristics by ***directing a subject's attention away*** from the purpose of an experiment. However, deception should only be used after prior careful analysis of the ethical ramifications.



Minimize threats to valid inferences

- partial-blind, single-blind, double-blind experiments
- unobtrusive experimentation
- incomplete information, deception
- debriefing
- multiple researchers
- experimenter-expectancy control groups
- quasi-control group
- unrelated-experiment technique

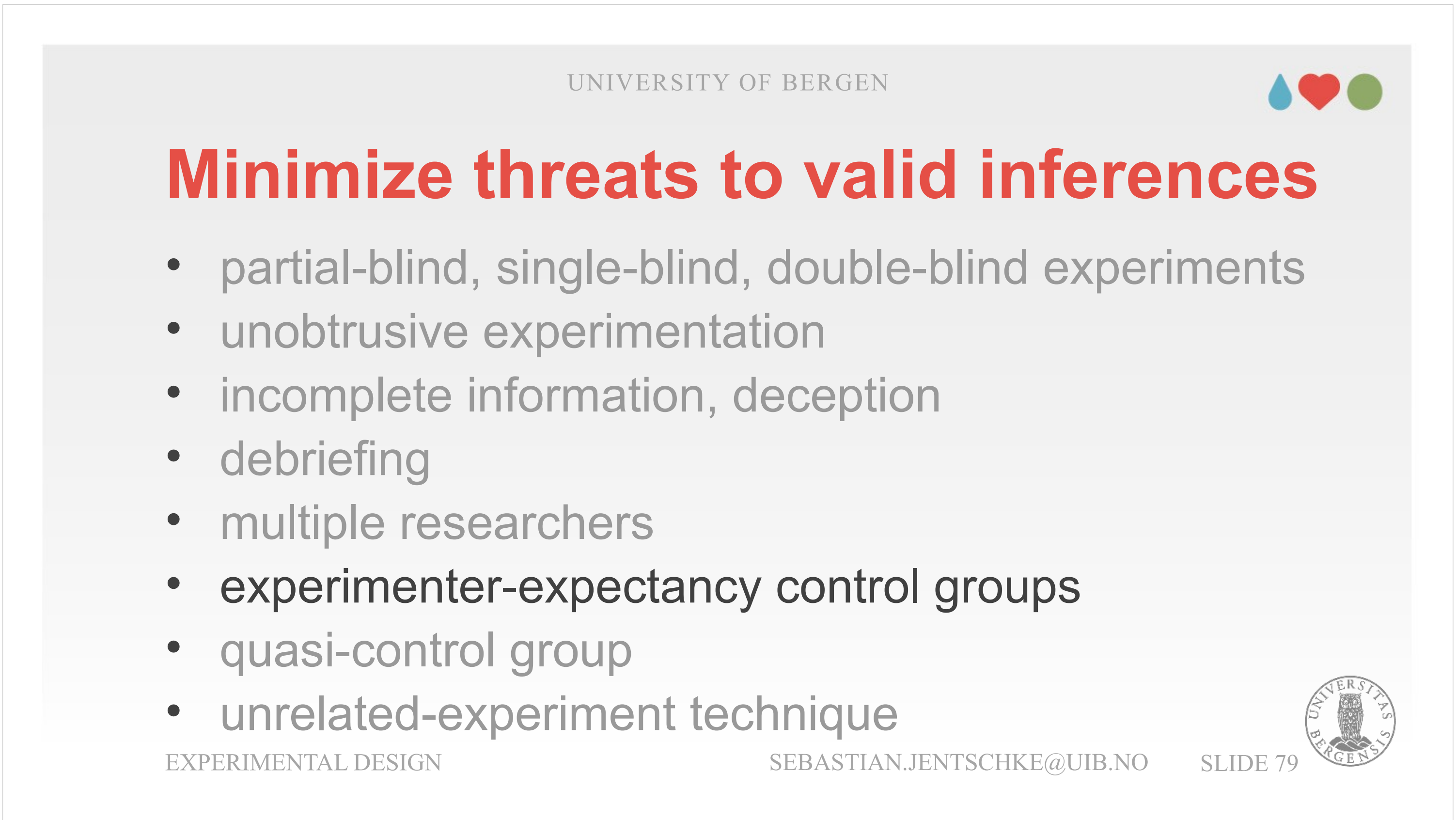
Debriefing is a common practice to ***share details about the experiment***. At the same time, it is possible to ***explore*** which ***beliefs*** and expectations ***subjects*** held ***about the experiment***. Information obtained at this time can be used to determine whether demand characteristics could have affected the results of the experiment.



Minimize threats to valid inferences

- partial-blind, single-blind, double-blind experiments
- unobtrusive experimentation
- incomplete information, deception
- debriefing
- multiple researchers
- experimenter-expectancy control groups
- quasi-control group
- unrelated-experiment technique

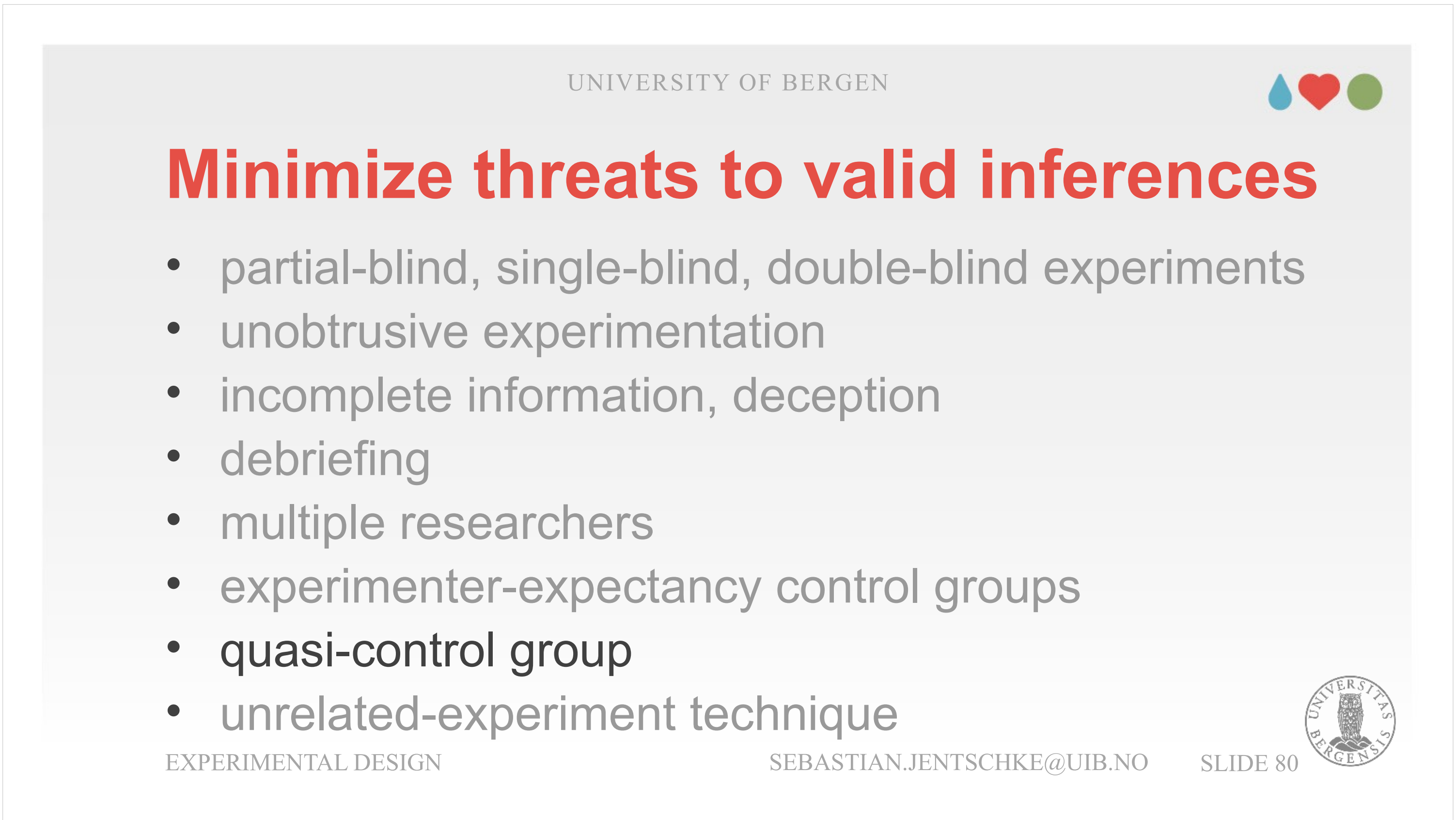
Characteristics of a researcher such as appearance, personality, (in-)experience, and so on can affect the results that are obtained and seriously limit the their generalizability. If ***several researchers*** are used, the researchers can be included as a ***nuisance variable*** and such influences can be statistically controlled for.



Minimize threats to valid inferences

- partial-blind, single-blind, double-blind experiments
- unobtrusive experimentation
- incomplete information, deception
- debriefing
- multiple researchers
- experimenter-expectancy control groups
- quasi-control group
- unrelated-experiment technique

If several groups of researchers are used it also is possible to estimate the magnitude of the experimenter-expectancy effects: ***One group*** of researchers is ***led to expect one experimental outcome***, a ***second group*** is led to ***expect the opposite outcome***, and a ***third group*** is ***led*** to believe that the ***treatment will have no effect*** on the dependent variable. Unfortunately, this procedure can be costly because it involves using numerous researchers and subjects.



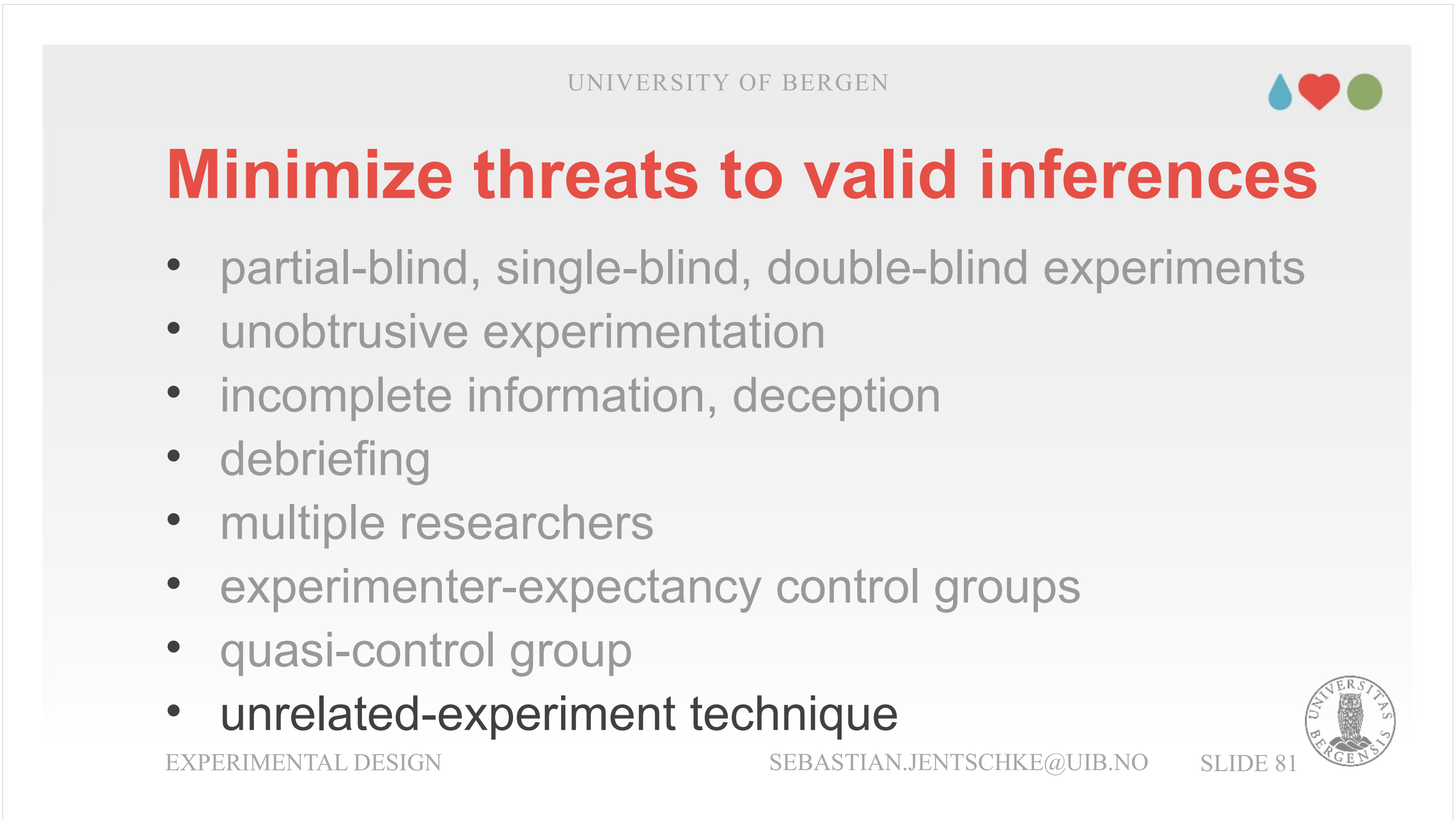
UNIVERSITY OF BERGEN

Minimize threats to valid inferences

- partial-blind, single-blind, double-blind experiments
- unobtrusive experimentation
- incomplete information, deception
- debriefing
- multiple researchers
- experimenter-expectancy control groups
- quasi-control group
- unrelated-experiment technique

A so-called quasi-control group may be used to assess the effects of experimenter-expectancy effects and demand characteristics.

The ***quasi-control group*** is ***exposed*** to all ***instructions*** and ***conditions*** given to the experimental group except that the ***treatment*** condition of interest is ***not administered***, neither does it receive a placebo. Instead, its members are ***asked to produce the data*** that they would have produced ***if*** they had actually ***received the treatment*** condition.

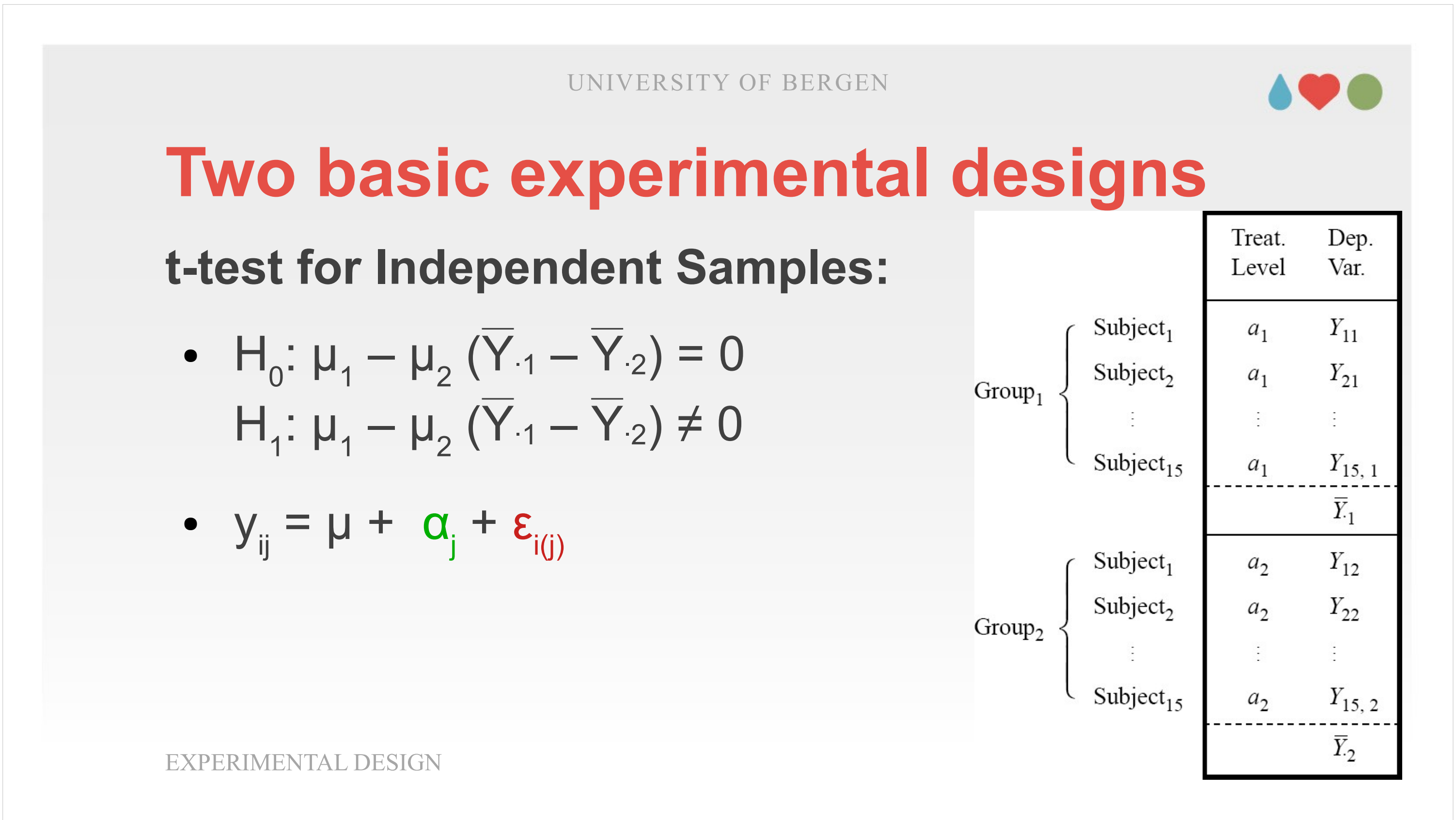


Minimize threats to valid inferences

- partial-blind, single-blind, double-blind experiments
- unobtrusive experimentation
- incomplete information, deception
- debriefing
- multiple researchers
- experimenter-expectancy control groups
- quasi-control group
- unrelated-experiment technique

By separating the presentation of the independent variable from the measurement of the dependent variable, the purpose of an experiment may be disguised and subject demand characteristics be minimized. Subjects ***receive*** the ***independent variable*** in a ***first*** experiment, and are ***later*** contacted and ***asked*** whether they are interested to ***participate in a second experiment*** during which the ***dependent variable*** is ***measured***. The participants should get the ***impression*** that the ***second experiment*** bears ***no relationship*** with the first experiment.

Two basic experimental designs



UNIVERSITY OF BERGEN

Two basic experimental designs

t-test for Independent Samples:

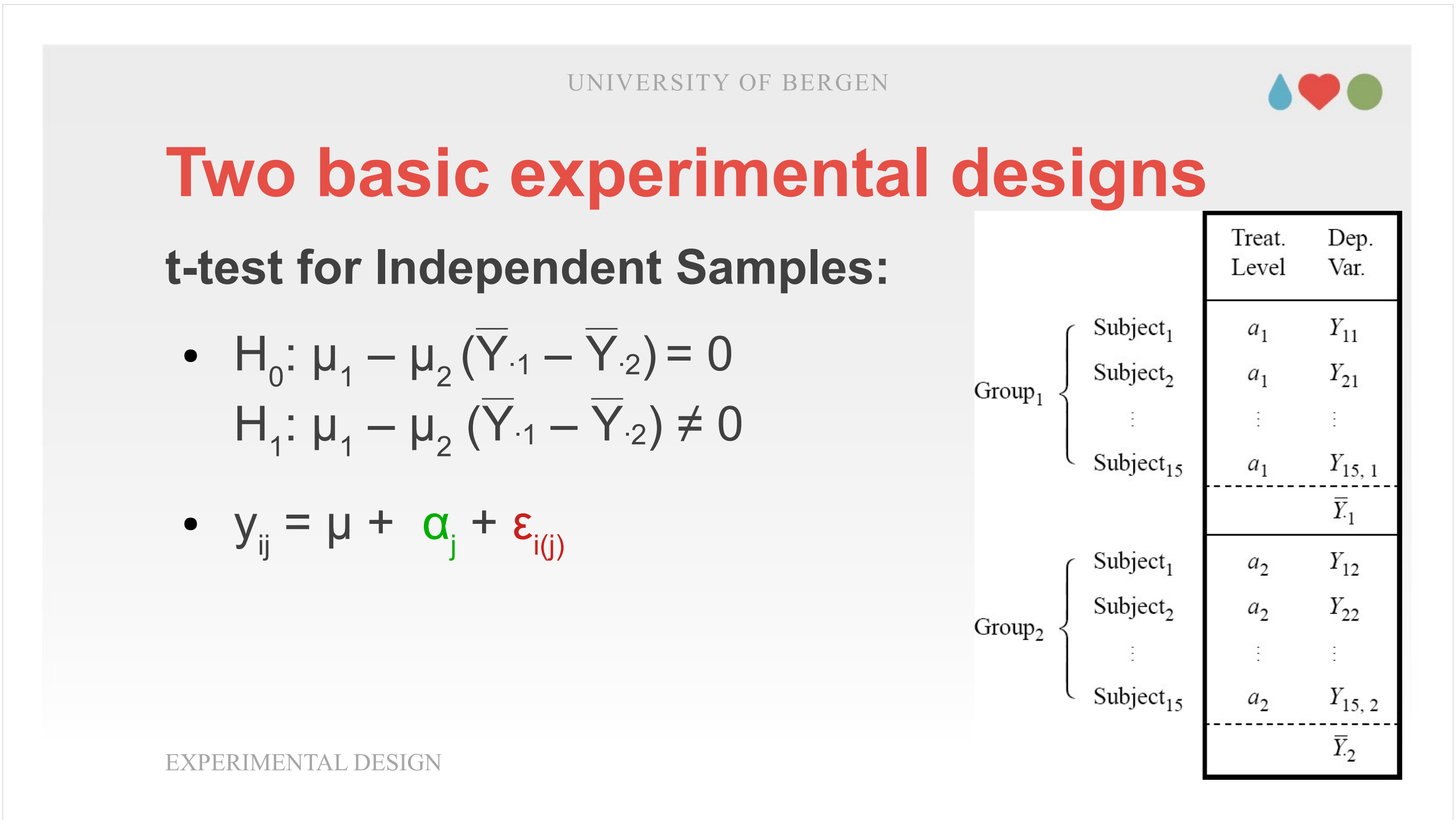
- H_0: $\mu_1 - \mu_2$ $(\overline{Y}_{\cdot 1} - \overline{Y}_{\cdot 2}) = 0$
  H_1: $\mu_1 - \mu_2$ $(\overline{Y}_{\cdot 1} - \overline{Y}_{\cdot 2}) \neq 0$
- $y_{ij} = \mu + \alpha_j + \varepsilon_{i(j)}$

		Treat. Level	Dep. Var.
Group_1	Subject_1	a_1	Y_{11}
	Subject_2	a_1	Y_{21}
	⋮	⋮	⋮
	Subject_{15}	a_1	$Y_{15,1}$
			$\overline{Y}_{\cdot 1}$
Group_2	Subject_1	a_2	Y_{12}
	Subject_2	a_2	Y_{22}
	⋮	⋮	⋮
	Subject_{15}	a_2	$Y_{15,2}$
			$\overline{Y}_{\cdot 2}$

EXPERIMENTAL DESIGN

When speaking about different experimental designs, those can be categorized according to several criteria.

One important distinction is whether the experimental manipulation of the Independent Variable occurs with randomly assigning participants to **different groups** or whether it occurs **within one person** / or matched pairs of persons with one measurement before and one measurement after the experimental manipulation.



UNIVERSITY OF BERGEN

Two basic experimental designs

t-test for Independent Samples:

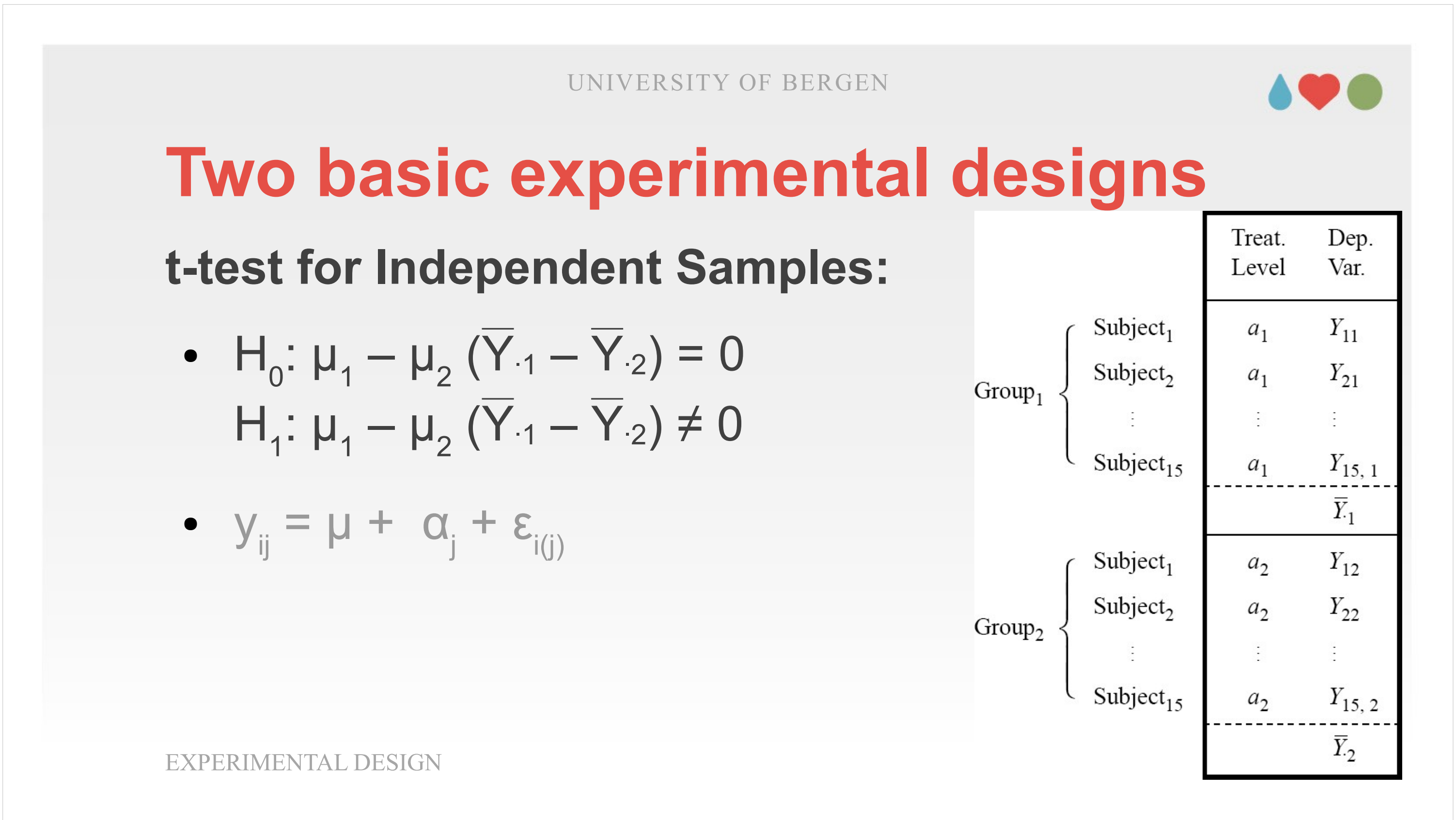
- H_0: $\mu_1 - \mu_2$ $(\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}) = 0$
  H_1: $\mu_1 - \mu_2$ $(\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}) \neq 0$
- $y_{ij} = \mu + \alpha_j + \varepsilon_{i(j)}$

		Treat. Level	Dep. Var.
Group_1	Subject_1	a_1	Y_{11}
	Subject_2	a_1	Y_{21}
	⋮	⋮	⋮
	Subject_{15}	a_1	$Y_{15,\,1}$
			$\bar{Y}_{\cdot 1}$
Group_2	Subject_1	a_2	Y_{12}
	Subject_2	a_2	Y_{22}
	⋮	⋮	⋮
	Subject_{15}	a_2	$Y_{15,\,2}$
			$\bar{Y}_{2}$

EXPERIMENTAL DESIGN

For the first category of experimental designs, we speak of a ***between-subjects factor*** or that the independent variable is manipulated between subjects. The simplest form of such an experiment is the t-test for Independent Samples. Each experimental condition is applied to one of two groups where participant get randomly assigned to. Afterwards it is tested whether the means within those groups differ from each other.

The design can easily be extended, e.g., to an analysis of variance (ANOVA) where the number of groups in increased beyond two. However, even though the same principle apply, the aim here is to keep this introduction very basic.



UNIVERSITY OF BERGEN

Two basic experimental designs

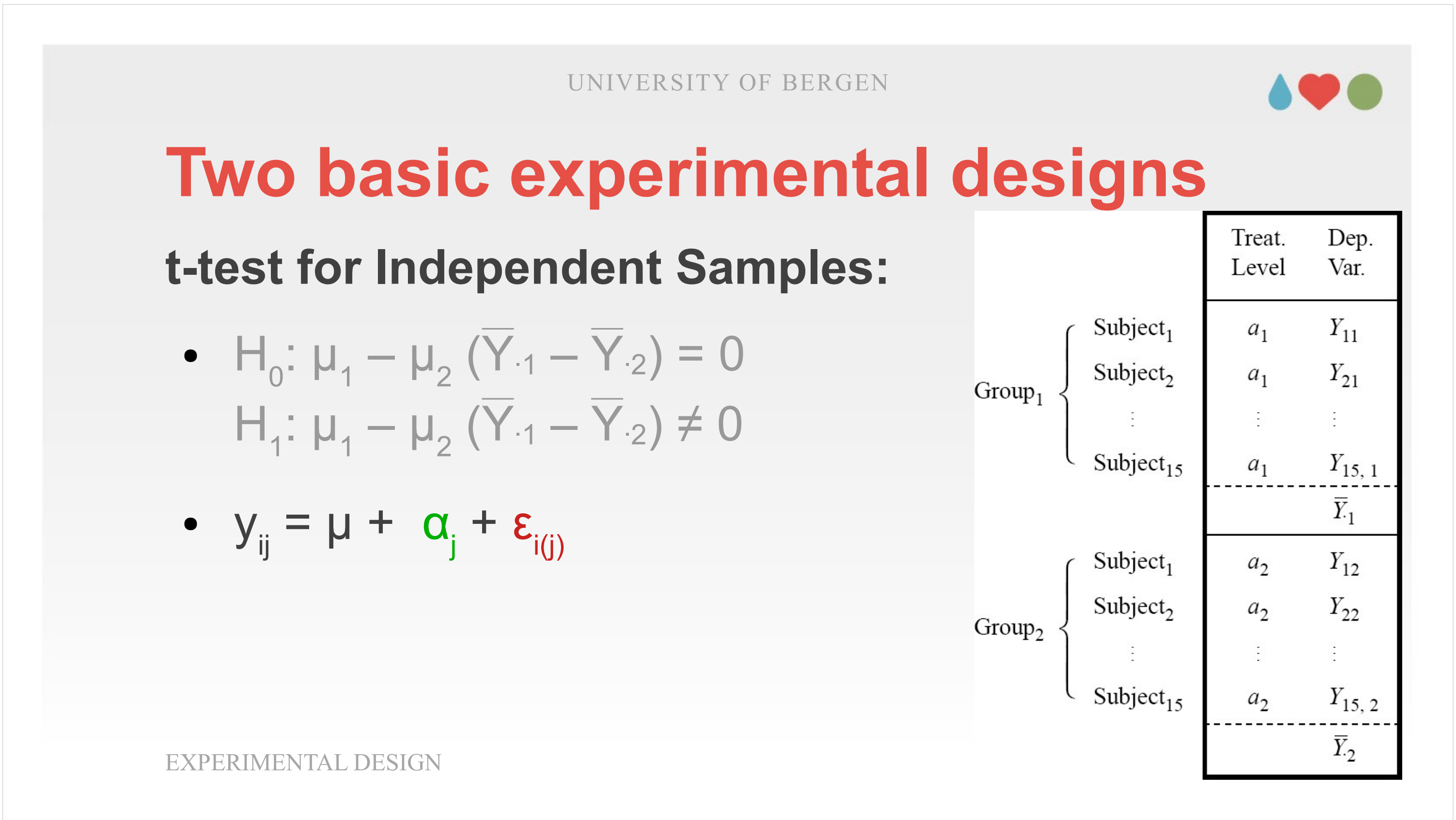
t-test for Independent Samples:

- H_0: $\mu_1 - \mu_2$ $(\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}) = 0$
  H_1: $\mu_1 - \mu_2$ $(\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}) \neq 0$
- $y_{ij} = \mu + \alpha_j + \varepsilon_{i(j)}$

		Treat. Level	Dep. Var.
Group_1	Subject_1	a_1	Y_{11}
	Subject_2	a_1	Y_{21}
	⋮	⋮	⋮
	Subject_{15}	a_1	$Y_{15,1}$
			$\bar{Y}_{\cdot 1}$
Group_2	Subject_1	a_2	Y_{12}
	Subject_2	a_2	Y_{22}
	⋮	⋮	⋮
	Subject_{15}	a_2	$Y_{15,2}$
			$\bar{Y}_{\cdot 2}$

EXPERIMENTAL DESIGN

The ***alternative hypothesis H_1*** assumes that our ***experimental manipulation*** had an ***effect*** and that, as a consequence, the ***two groups differ*** from each other. In some cases we possibly even define which direction we expect this effect to take (i.e., whether we expect μ_1 to be smaller than μ_2 or the other way round). In other cases, we don’t have (or make) assumptions about the direction of the effect, we just claim that the two groups er unequal in their means. The null hypotheses H_0 assumes that the two group either don’t differ from each other (i.e., the difference is 0) if we had no hypothesis about the direction of the effect. For the other case with a directed hypothesis, than the H_1 (i.e., if we expected μ_1 to be smaller than μ_2, our H_0 would claim μ_1 to be equal or larger than μ_2).



UNIVERSITY OF BERGEN

Two basic experimental designs

t-test for Independent Samples:

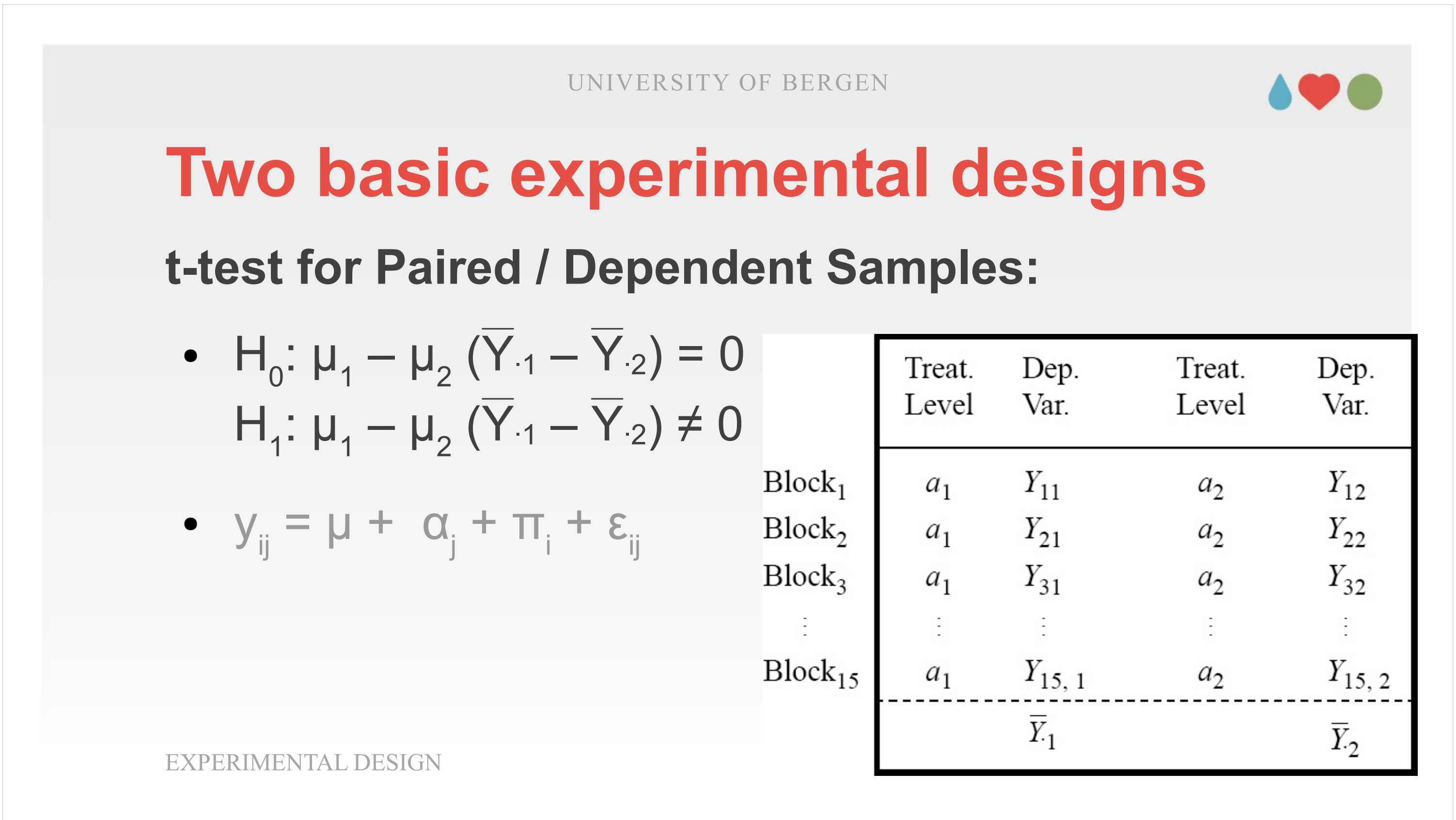
- $H_0: \mu_1 - \mu_2 \ (\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}) = 0$
  $H_1: \mu_1 - \mu_2 \ (\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}) \neq 0$
- $y_{ij} = \mu + \alpha_j + \varepsilon_{i(j)}$

		Treat. Level	Dep. Var.
Group_1	Subject_1	a_1	Y_{11}
	Subject_2	a_1	Y_{21}
	⋮	⋮	⋮
	Subject_{15}	a_1	$Y_{15,\,1}$
			$\bar{Y}_{\cdot 1}$
Group_2	Subject_1	a_2	Y_{12}
	Subject_2	a_2	Y_{22}
	⋮	⋮	⋮
	Subject_{15}	a_2	$Y_{15,\,2}$
			$\bar{Y}_{\cdot 2}$

EXPERIMENTAL DESIGN

Then we have the formula that “explains” the principles behind that. Each value of our dependent variable y_{ij} is composed of three components: (1) the mean of that variable in the population μ, (2) the effect of the experimental manipulation $\alpha_{j,}$ and (3) an error effect $\varepsilon_i({}_j)$. The notation $_{i(j)}$ indicates that the subject i appears only in one condition j (since it was assigned to one of the two groups.

α_j is the part we are most interested in because it is subject to our hypothesis: It is that proportion of the variation in the dependent variable that was caused by our experimental manipulation of the independent variable. Another way to describe α_j and $\varepsilon_i({}_j)$ is that $\boldsymbol{\alpha}_j$ is the ***part*** of the equation that ***we can explain*** whereas $\boldsymbol{\varepsilon_i({}_j)}$ is the ***part we can’t explain***. What we do in a hypothesis test is setting α_j and $\varepsilon_i({}_j)$ in relation. α_j “quantifies” the certainty with which we can make a decision in favour of our hypothesis, $\varepsilon_i({}_j)$ the uncertainty.



UNIVERSITY OF BERGEN

Two basic experimental designs

t-test for Paired / Dependent Samples:

- $H_0: \mu_1 - \mu_2 \ (\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}) = 0$
  $H_1: \mu_1 - \mu_2 \ (\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}) \neq 0$
- $y_{ij} = \mu + \alpha_j + \pi_i + \varepsilon_{ij}$

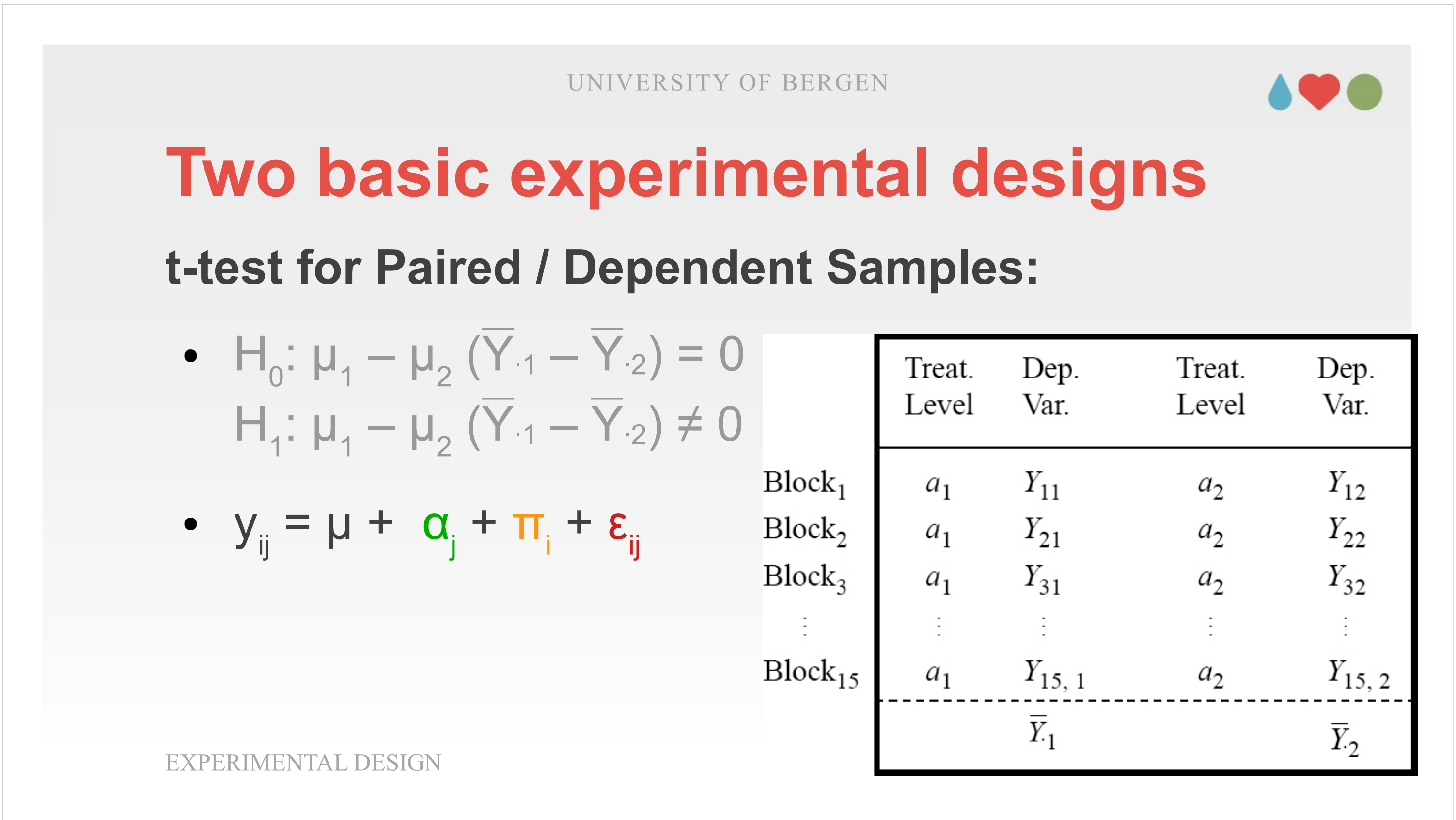
	Treat. Level	Dep. Var.	Treat. Level	Dep. Var.
Block_1	a_1	Y_{11}	a_2	Y_{12}
Block_2	a_1	Y_{21}	a_2	Y_{22}
Block_3	a_1	Y_{31}	a_2	Y_{32}
⋮	⋮	⋮	⋮	⋮
Block_{15}	a_1	$Y_{15,1}$	a_2	$Y_{15,2}$
		$\bar{Y}_{\cdot 1}$		$\bar{Y}_{\cdot 2}$

EXPERIMENTAL DESIGN

For the second category of experimental designs, we speak of a ***within-subject factor*** or that the independent variable is varied within a subject.

That means each participant is observed under each treatment level in the experiment by obtaining repeated measures on that participant. Alternatively, subject matching can be used where sets of subjects that are similar with respect to a nuisance variable that is correlated with the dependent variable. Of these subject pairs, one participant is randomly assigned to one condition, the second participant to the other condition.

The simplest for of such an experiment is the t-test for Paired / Dependent Samples. Null and alternative hypotheses are similar to what was described at the t-test for Independent Samples.



UNIVERSITY OF BERGEN

Two basic experimental designs

t-test for Paired / Dependent Samples:

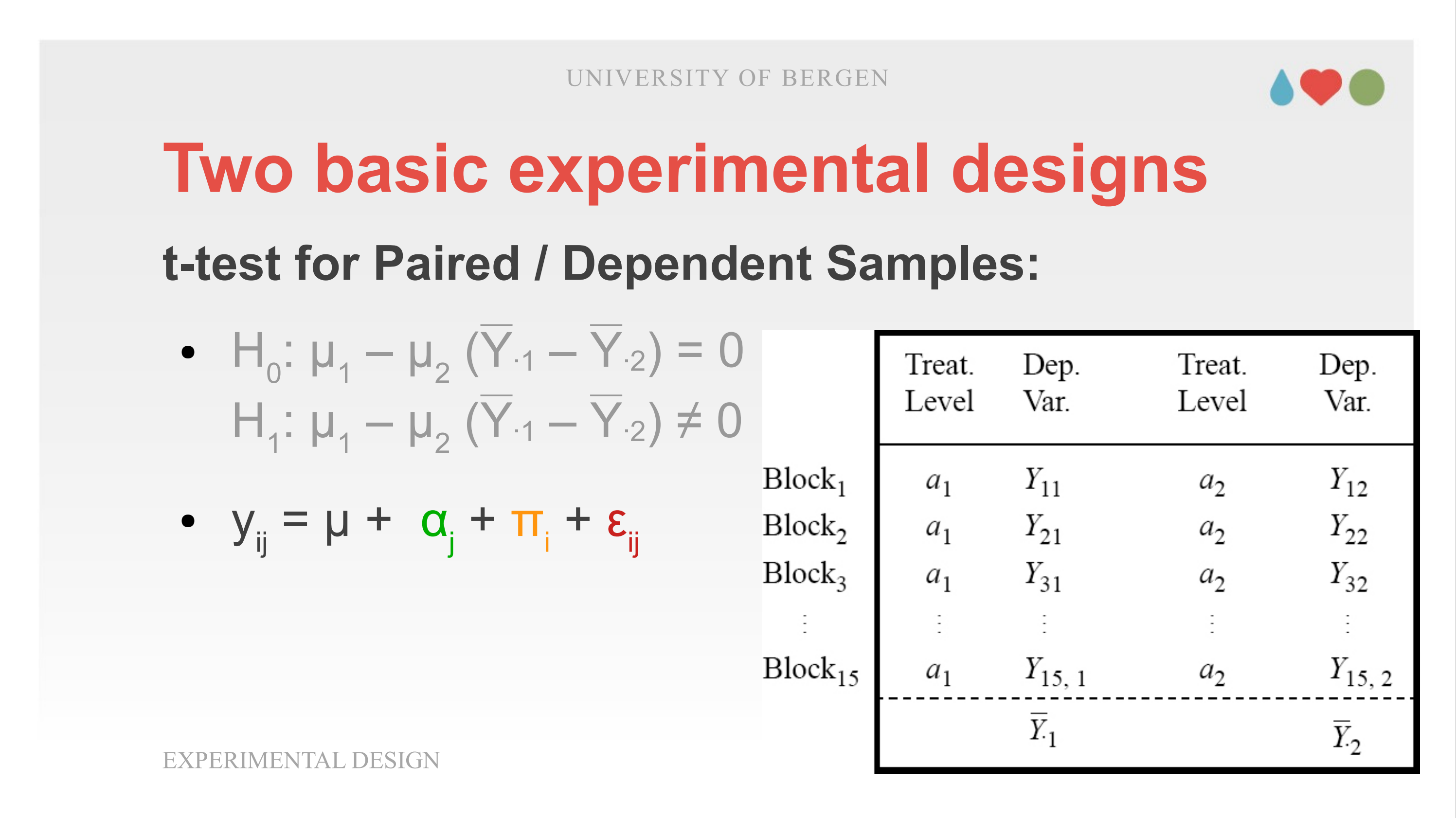
- $H_0: \mu_1 - \mu_2 \ (\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}) = 0$
  $H_1: \mu_1 - \mu_2 \ (\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}) \neq 0$
- $y_{ij} = \mu + \alpha_j + \pi_i + \varepsilon_{ij}$

	Treat. Level	Dep. Var.	Treat. Level	Dep. Var.
Block$_1$	a_1	Y_{11}	a_2	Y_{12}
Block$_2$	a_1	Y_{21}	a_2	Y_{22}
Block$_3$	a_1	Y_{31}	a_2	Y_{32}
⋮	⋮	⋮	⋮	⋮
Block$_{15}$	a_1	$Y_{15,\,1}$	a_2	$Y_{15,\,2}$
		$\bar{Y}_{\cdot 1}$		$\bar{Y}_{\cdot 2}$

EXPERIMENTAL DESIGN

Again, we have a formula that “explains” the principles behind the t-test for Paired / Dependent Samples.

This time, the value of our dependent variable y_{ij} is composed of four components: (1) the ***mean*** of that variable in the ***population*** μ, (2) the ***effect*** of the ***experimental manipulation*** $\boldsymbol{\alpha_j}$, (3) the ***contribution*** of the ***individual*** (its mean) $\boldsymbol{\pi_i}$, and (4) an error ***effect*** $\boldsymbol{\varepsilon_{ij}}$.



UNIVERSITY OF BERGEN

Two basic experimental designs

t-test for Paired / Dependent Samples:

- H_0: $\mu_1 - \mu_2$ $(\overline{Y}_{\cdot 1} - \overline{Y}_{\cdot 2}) = 0$
  H_1: $\mu_1 - \mu_2$ $(\overline{Y}_{\cdot 1} - \overline{Y}_{\cdot 2}) \neq 0$
- $y_{ij} = \mu + \alpha_j + \pi_i + \varepsilon_{ij}$

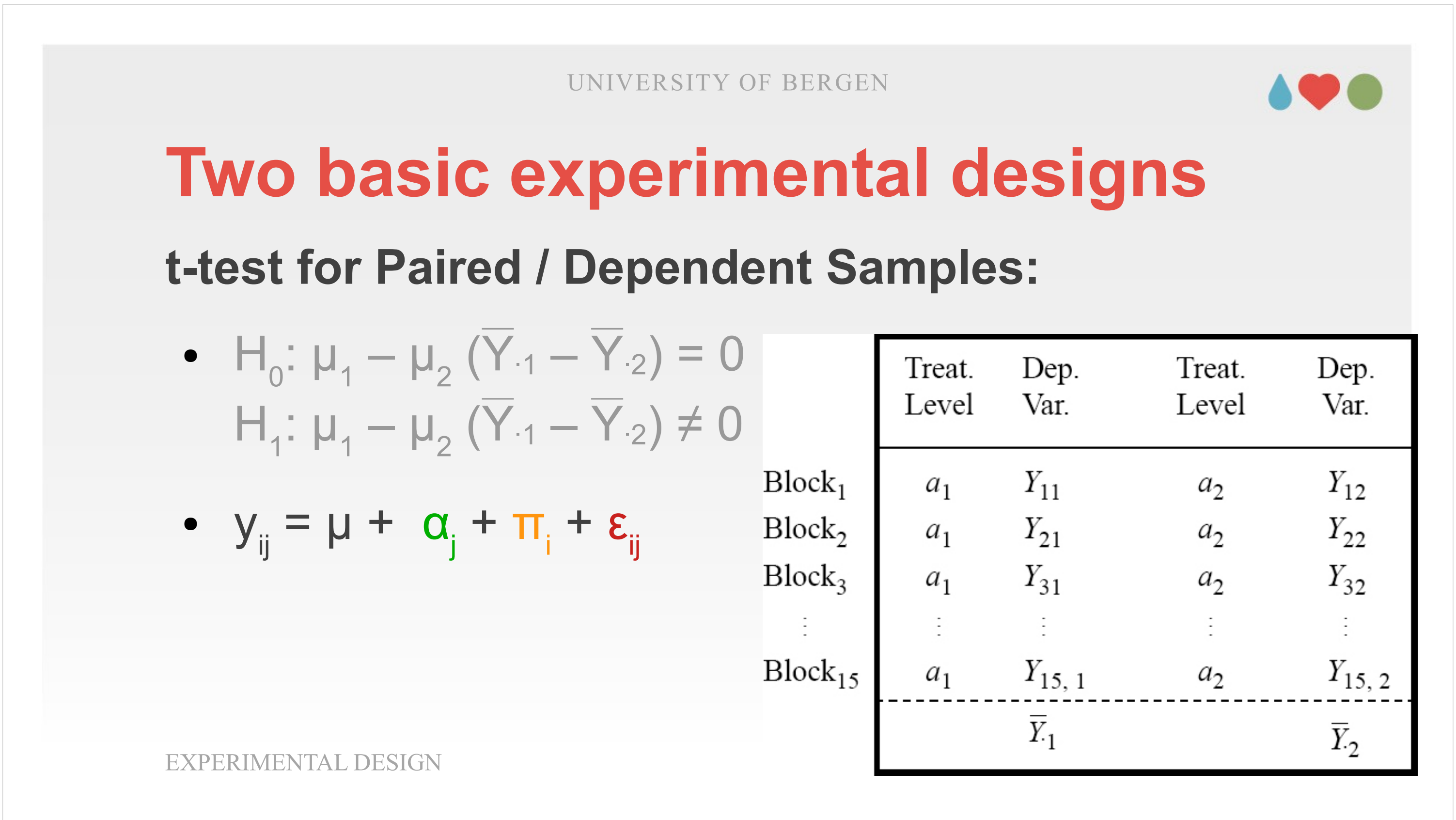
	Treat. Level	Dep. Var.	Treat. Level	Dep. Var.
Block_1	a_1	Y_{11}	a_2	Y_{12}
Block_2	a_1	Y_{21}	a_2	Y_{22}
Block_3	a_1	Y_{31}	a_2	Y_{32}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Block_{15}	a_1	$Y_{15,1}$	a_2	$Y_{15,2}$
		$\overline{Y}_{\cdot 1}$		$\overline{Y}_{\cdot 2}$

EXPERIMENTAL DESIGN

By measuring the same (or a matched) participant twice, we ***"control"*** for the influence of ***variation*** which is caused by that ***individual*** (π_i). By doing so, we ***reduce*** the ***error effect*** (ε_{ij}).

In the t-test for Independent samples, the error effect consists of both, the effect of natural variation between individuals plus other sources of variation unaccounted for.

In the t-test for Dependent samples, the effect of ***individual variation*** (π_i) ***is controlled for*** and thereby ***taken out*** of the ***error effect*** (ε_{ij}). That is, the error effect is smaller and with it the uncertainty with which we make a decision in favour of or against our hypothesis (where we set α_j and ε_{ij} in relation).



UNIVERSITY OF BERGEN

Two basic experimental designs

t-test for Paired / Dependent Samples:

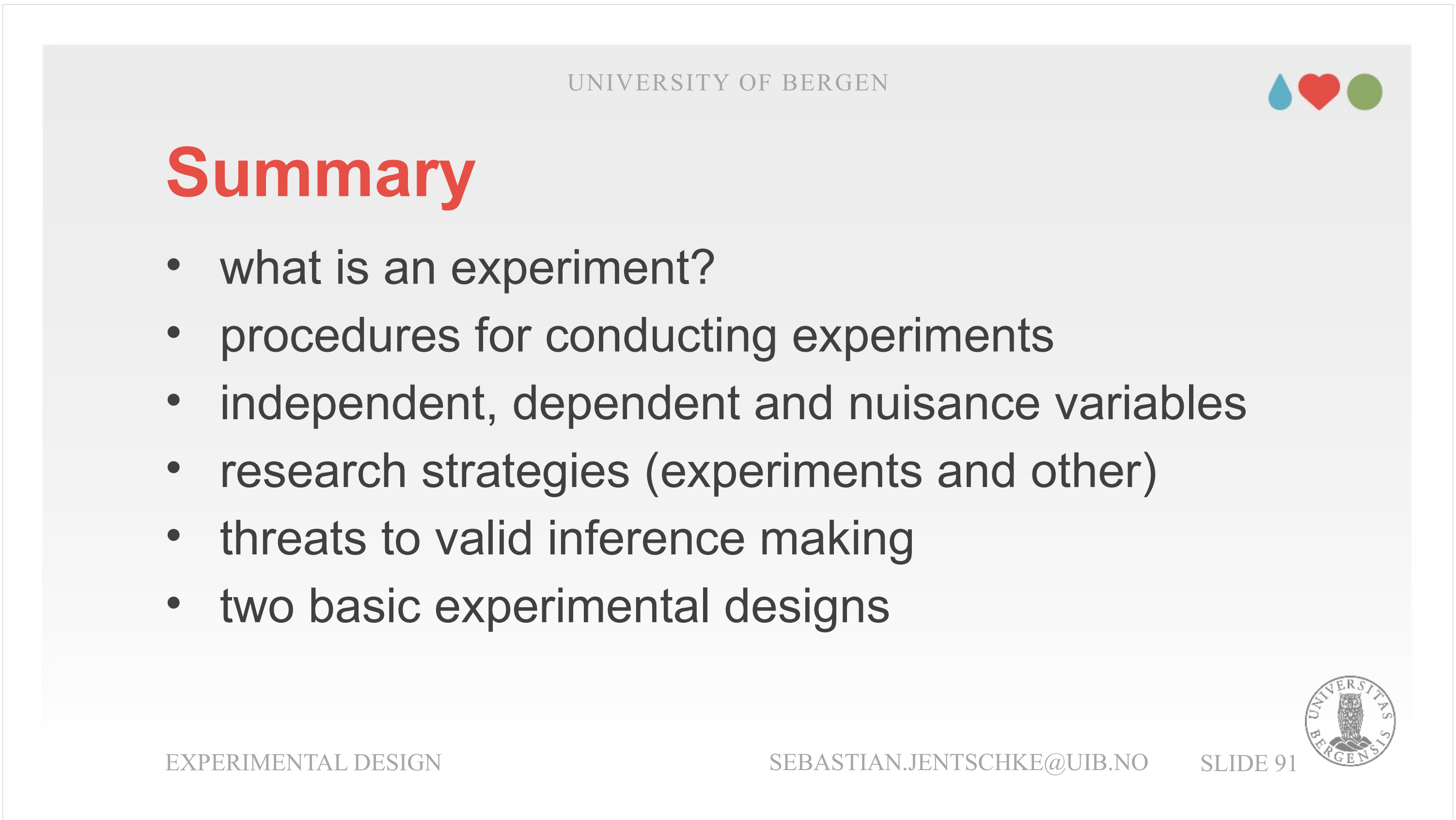
- $H_0: \mu_1 - \mu_2 \ (\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}) = 0$
  $H_1: \mu_1 - \mu_2 \ (\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}) \neq 0$
- $y_{ij} = \mu + \alpha_j + \pi_i + \varepsilon_{ij}$

	Treat. Level	Dep. Var.	Treat. Level	Dep. Var.
Block_1	a_1	Y_{11}	a_2	Y_{12}
Block_2	a_1	Y_{21}	a_2	Y_{22}
Block_3	a_1	Y_{31}	a_2	Y_{32}
⋮	⋮	⋮	⋮	⋮
Block_{15}	a_1	$Y_{15,1}$	a_2	$Y_{15,2}$
		$\bar{Y}_{\cdot 1}$		$\bar{Y}_{\cdot 2}$

EXPERIMENTAL DESIGN

This means, when using repeated-measurement-designs we are typically more likely to obtain a statistically significant result as a source of undesirable variation (π_i) is controlled for.

However, repeated-measurement-designs are ***not suitable for all situations***. If administering the same test twice when measuring the dependent variable is not suitable, such designs can’t be used. Certain tests will work only one time since participants understood the principle and are not “naive” at the second administration. In other situations, such learning may have less drastic but still recognizable differences, e.g., if participants “learn” from the first time the test is given and get higher scores in the second measurement. Finally, a condition for applying a test twice is that it is highly reliable: if there were a substantial variation from results obtained with one measurement to the next, we shouldn’t use it in a repeated-measures design.



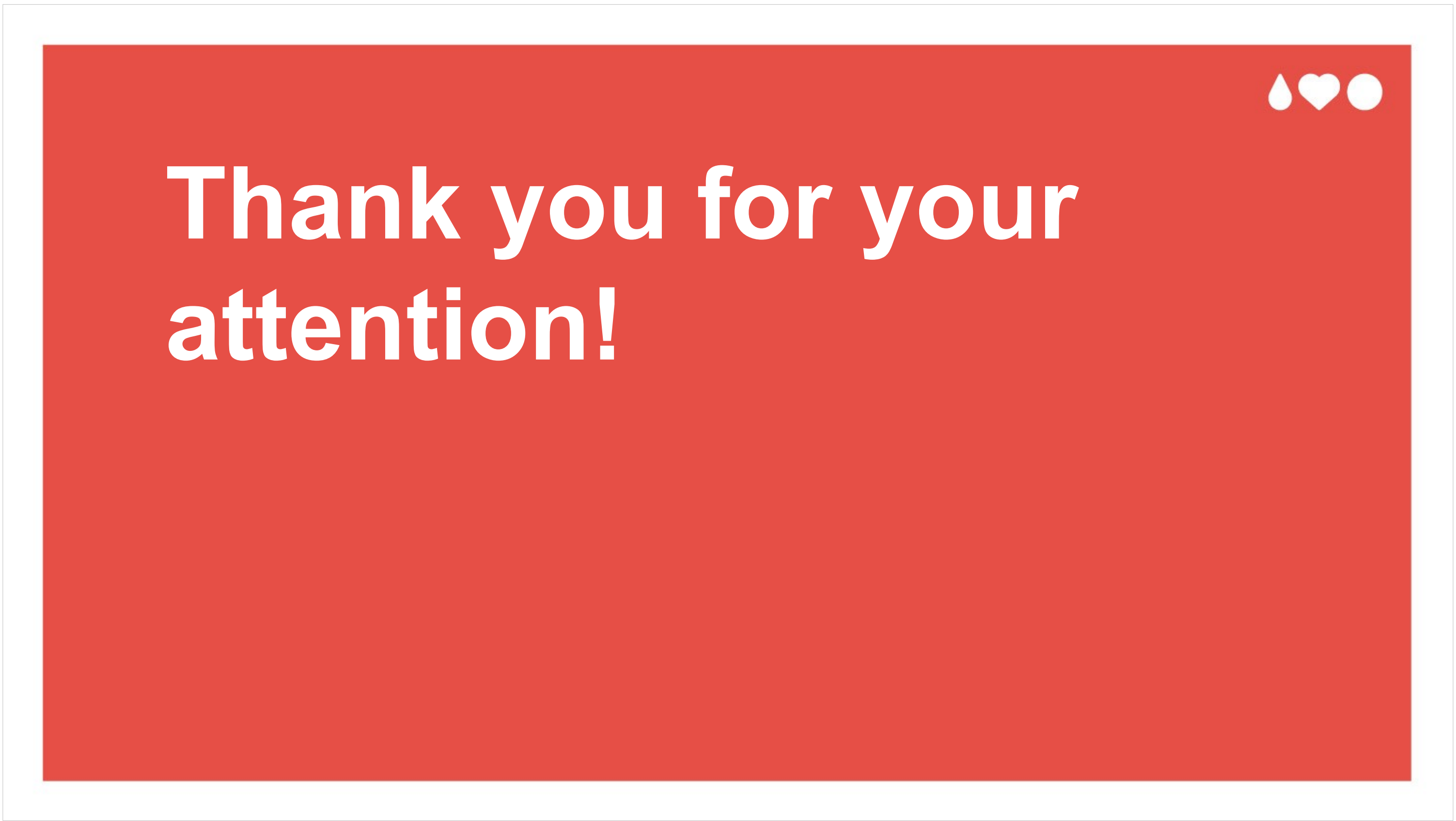
Summary

- what is an experiment?
- procedures for conducting experiments
- independent, dependent and nuisance variables
- research strategies (experiments and other)
- threats to valid inference making
- two basic experimental designs

I suppose, you felt, it was quite a comprehensive lecture, maybe even a bit too much. What the main aim was can be summarized quite simply: To raise awareness for what experiments and experimental design are and how they are used. I hope, you understood some key concepts such as what independent, dependent and nuisance variables are and how they interact in an experiment. And, most importantly, which possible threats you should consider when planning own experiments.

The slides with the threats are supposed to provide a kind of check list that can raise awareness and make you think what you could possibly do to avoid or minimize those threats.

A lot boils down to experience. So it is possibly wise to run pilot tests or small experiments where the main purpose is to learn how to experiment. The idea is to make mistakes on such occasions, i.e. before it counts (such as in your M.Sc. thesis).



Thank you for your attention!

Anyway, thanks for your attention. I hope, you got the feeling that you learned a thing or two…

UNIVERSITAS BERGENSIS

UNIVERSITY OF BERGEN

Eco-Lighthouse®