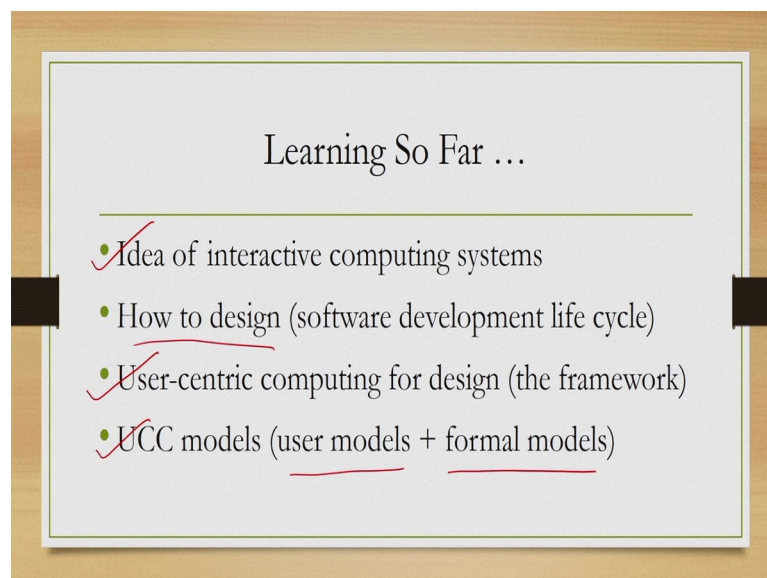


User-Centric Computing for Human-Computer Interaction
Prof. Samit Bhattacharya
Department of Computer Science and Engineering
Indian Institute of Technology, Guwahati

Lecture - 26
Introduction and research question formulation

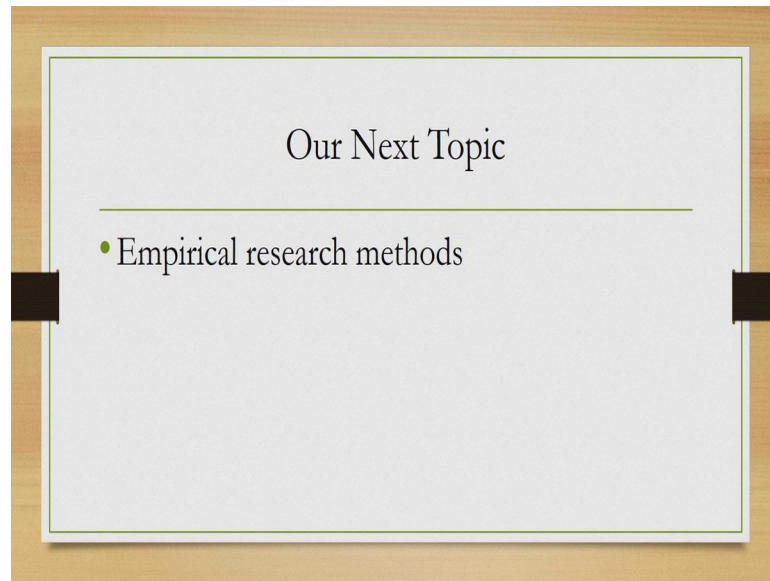
Hello and welcome to lecture number 26, in the course User-Centric Computing for Human Computer Interaction . First, let us quickly recap what we have learned so far this will be followed by our actual discussion for today's lecture.

(Refer Slide Time: 00:57)



So, we have so far covered, these components of the course namely the idea of interactive computing system, it was followed by discussion on how to design an interactive system, namely the software development lifecycle stages. And, then we introduced one user centric computing framework and had a detailed discussion on different components of the framework including, the major component namely the user centric computing models. So, there we discussed two types of models; the computational user models and formal models.

(Refer Slide Time: 01:48)



Today, we are going to start our discussion on another important topic of this course, namely empirical research methods. So, first let us try to understand why we required empirical research methods. Remember that we have discussed many user models. For example, the Fitts law or say the steering law or say the Hick Hyman law. Now, how these models were developed?

During the discussion on these models we briefly mentioned that these models were based on empirical data. Data were collected from human users and those were used to analyze and build the relationship between say for example, the movement time and the distance and width of the target in the context of Fitts law or the choice reaction time and the number of elements in the context of Hick Hyman law and so on.

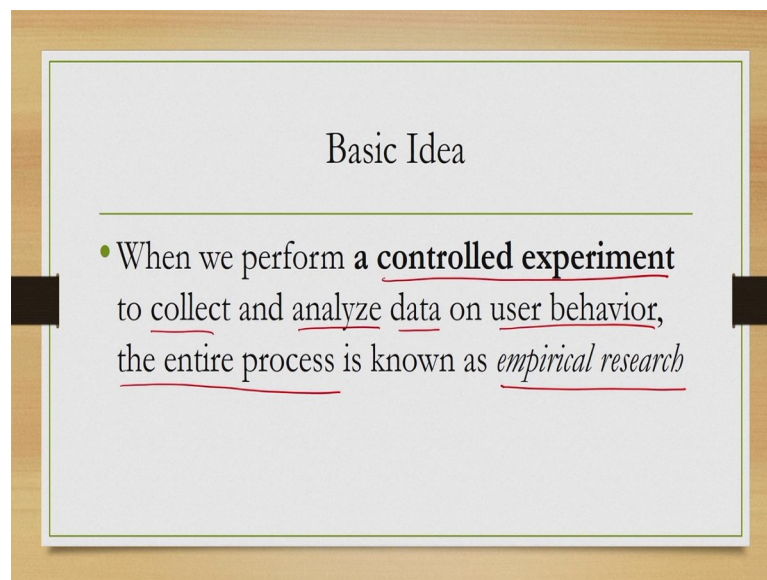
So in order to build the models we required some data collected from human users. This is one of the objective, there are several models also which in this lecture we have not discussed, but which are very much there, which are based on machine learning approaches. Now, in order to build a learning based model, we required training data as well as testing data. These data items refer to data collected from users, data collected from human participants.

So, in order to build any type of model, whether that is in the form of an equation like the Fitts law or the Hick Hyman law or a learning based model, we require some data. Similarly, during our discussion on the lifecycle stages for an interactive system software

development, we mentioned one important stage, that is the empirical evaluation. So, we need to evaluate our system to check for its usability in an empirical study, which means we need to get it checked with the actual users of the system.

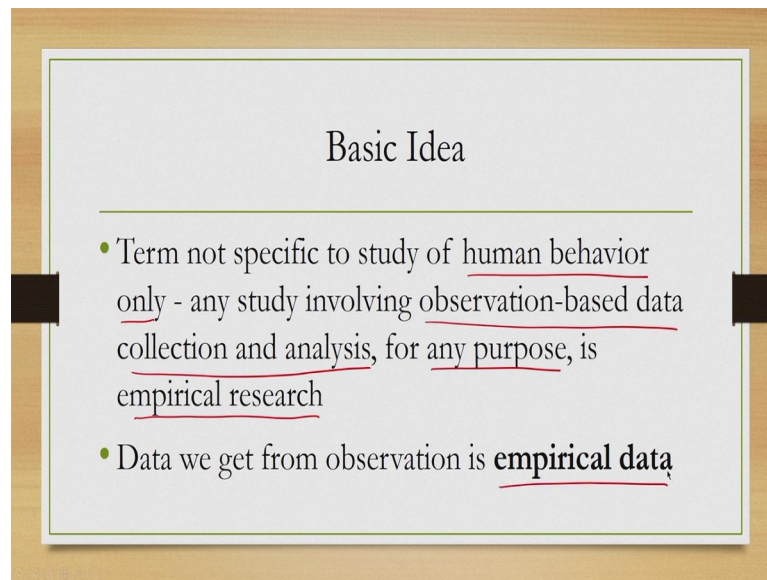
So, there are two purposes; one is to build models, either in the form of an equation or to build a learning based model say for example, a classifier, which classifies a given data into different classes. Also, we need the data to evaluate empirically a system that we have designed. What we want to evaluate? We want to evaluate the usability of the system, but through empirical means. So, for these reasons we need empirical study to collect empirical data.

(Refer Slide Time: 05:09)



Now, when we perform a controlled experiment to collect and analyze data; data on user behavior, this process of collection and analysis of data is known as empirical research. In empirical research what we do is we perform a controlled experiment to collect data and analyze data. In our context this data is related to user behavior; however, that need not be the case always.

(Refer Slide Time: 05:53)



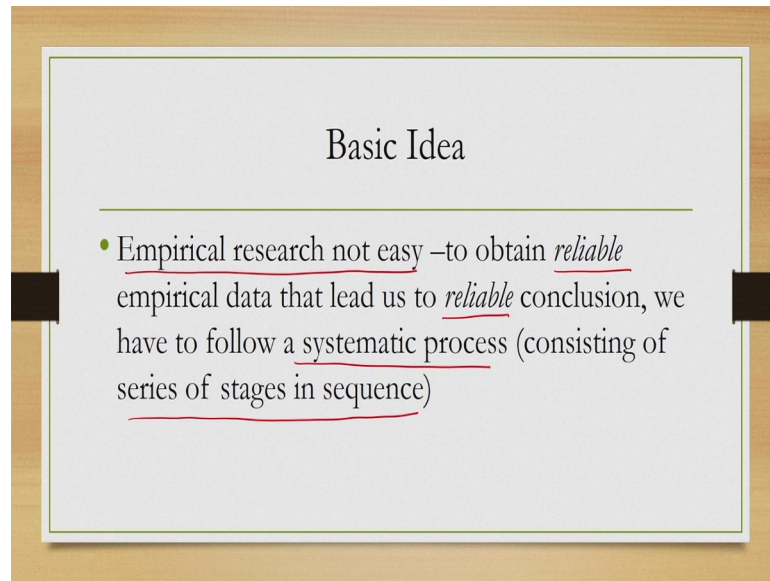
Basic Idea

- Term not specific to study of human behavior only - any study involving observation-based data collection and analysis, for any purpose, is empirical research
- Data we get from observation is empirical data

The term empirical research is not specific to study of human behavior alone. Any study involving observation-based data collection and analysis, for any purpose, is empirical research. And, the data that we collect during this study is called empirical data. So, if somebody is watching the time the sunrises that is also empirical data, it has nothing to do with human activity.

Similarly, if somebody is watching the increase in the water flow in a river and records that observation, that is also empirical data, again this has nothing to do with human activity. So, empirical research means any observations and analysis of the data collected, which may or may not be related to human in our context in the context of user centric system design. The term empirical research refers to observation based collection and analysis of user behavior data, that is one difference in the context of empirical research in user centric system design and empirical research in general.

(Refer Slide Time: 07:23)

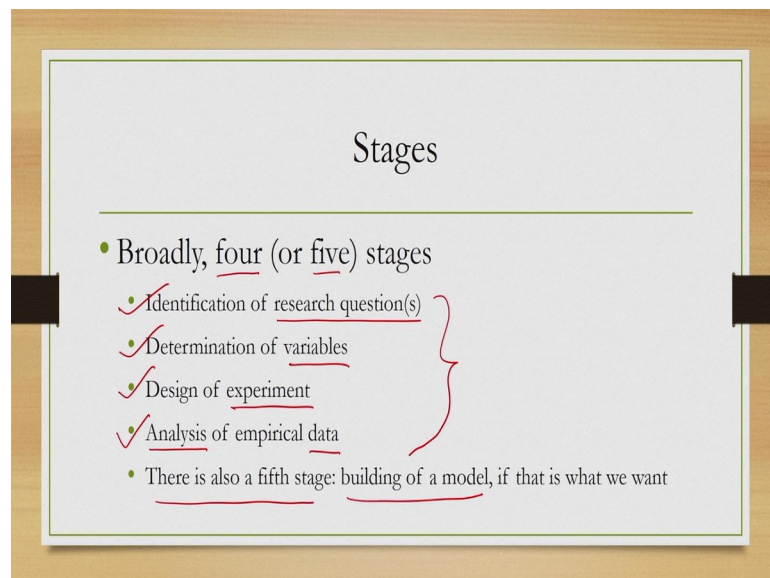


Basic Idea

- Empirical research not easy –to obtain reliable empirical data that lead us to reliable conclusion, we have to follow a systematic process (consisting of series of stages in sequence)

Now, this empirical research is not easy. In order to obtain data that is reliable which can take us to reliable conclusions. We have to follow a very systematic process, which consists of a series of stages in sequence, otherwise we will end up with data that is not reliable and accordingly the conclusion that we may draw from that data will not be reliable as well. So, as I said in order to conduct empirical research you need to perform a series of steps in sequence or series of stages in sequence.

(Refer Slide Time: 08:17)



Stages

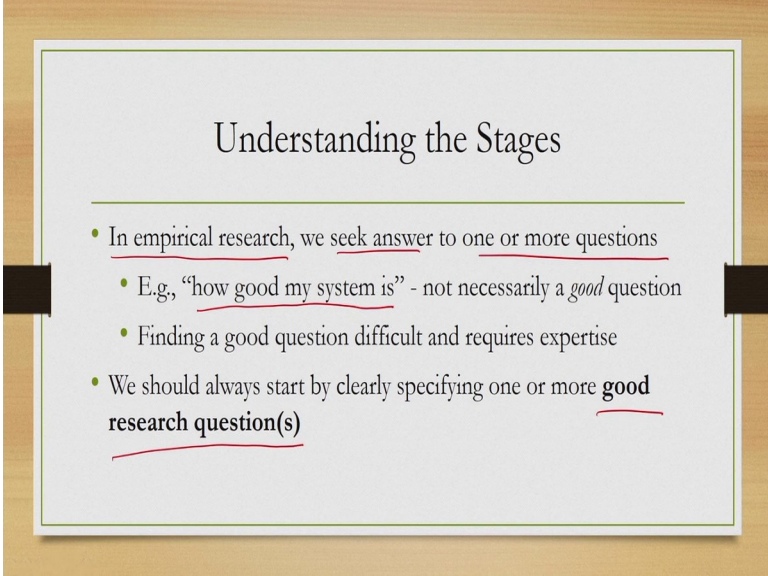
- Broadly, four (or five) stages
 - ✓ Identification of research question(s)
 - ✓ Determination of variables
 - ✓ Design of experiment
 - ✓ Analysis of empirical data
- There is also a fifth stage: building of a model, if that is what we want

So, what are those stages of an empirical research? There are broadly four or five stages, why this confusion we will soon clarify. So, the very first stage is identification of one or more research questions. Second stage is determination of variables, third stage is design of experiment, fourth stage is the analysis of empirical data.

Now, if our objective is to come to a conclusion about validity of certain things then these four stages are sufficient. However, as I said before sometimes we are interested in building some models, some relationship between quantities. Such as the movement model in the Fitts law or the model of choice reaction model of choice reaction time in the hick Hyman law. In that case we also have a fifth stage, where we build a model based on the data.

Now, in this lecture we will learn about these stages and how to perform those stages?

(Refer Slide Time: 09:34)



Understanding the Stages

- In empirical research, we seek answer to one or more questions
 - E.g, “how good my system is” - not necessarily a *good* question
 - Finding a good question difficult and requires expertise
- We should always start by clearly specifying one or more good research question(s)

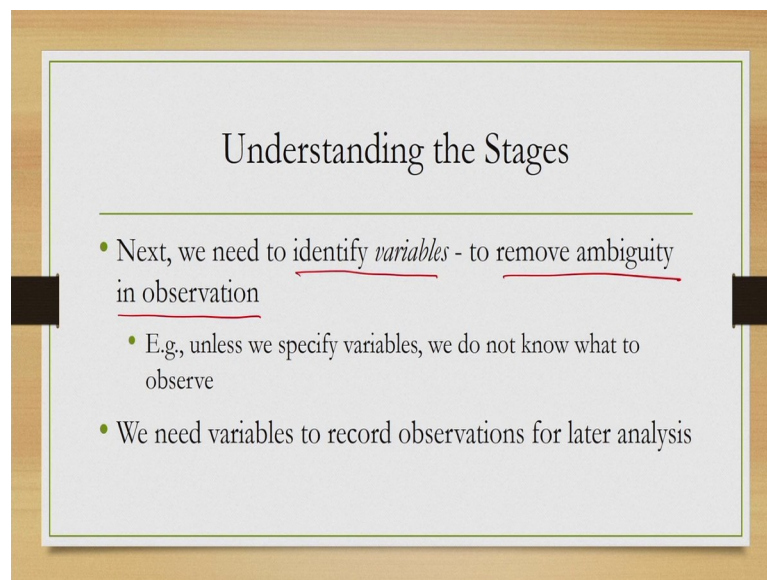
Let us first quickly understand these stages. Now, the primary objective in any empirical research is to seek answer to one or more questions. So, when you are asked to perform a study, there must be some objective, the objective is to seek answer to some questions for which you want to conduct the study. For example, suppose you have built a system and the objective of your study is to determine the quality of the design.

Now, in order to determine the quality in the study what you can do? You can frame some questions. Suppose, I have framed a question like is my system good. Now, if I

frame this question and perform a study to answer this question what typically I will do? I may contact few persons, ask them these questions, get their opinion and try to come to a conclusion, but that need not be a good approach as we shall see in our subsequent discussions.

Is my system good this question can be is my system good or how good my system is these type of questions are not necessarily good questions to start with, why we will soon see. So, the main concern here is to come up with a good research question that will lead us to a good conclusion, which is not a trivial job and there are things that we need to consider to frame such questions.

(Refer Slide Time: 11:18)

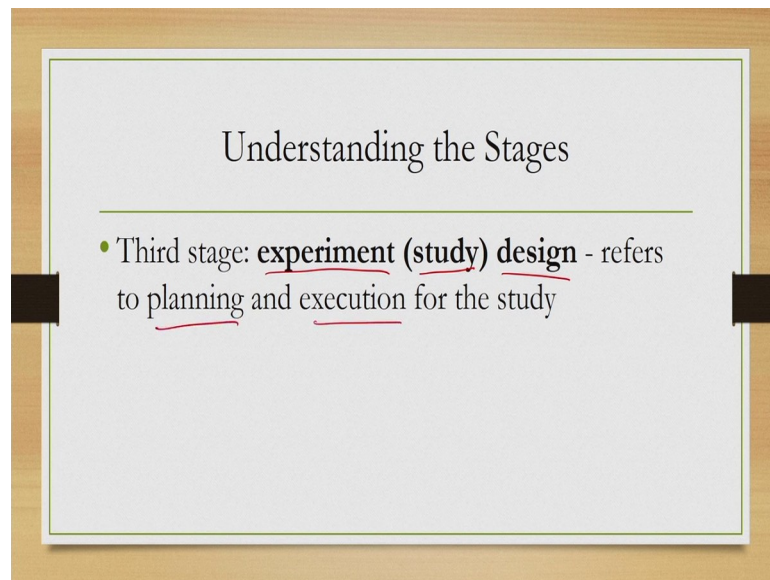


Understanding the Stages

- Next, we need to identify variables - to remove ambiguity in observation
 - E.g, unless we specify variables, we do not know what to observe
- We need variables to record observations for later analysis

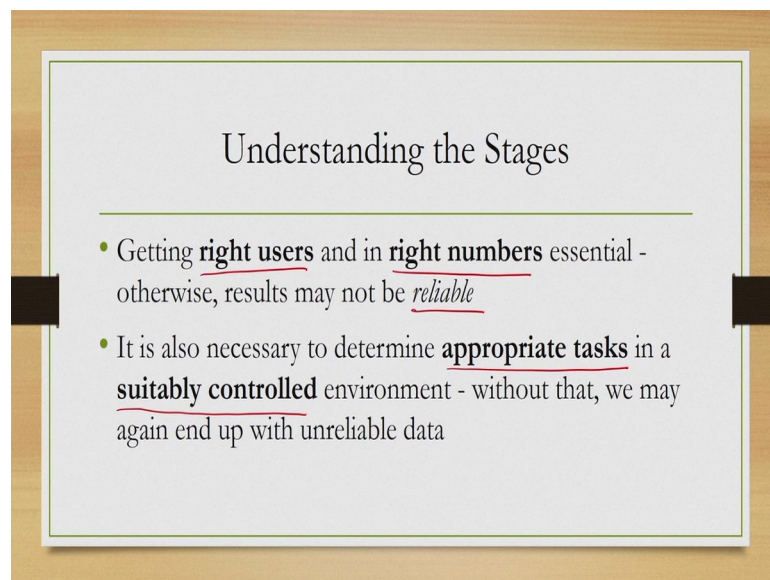
Once a question is formulated, one or more questions are found, what we need to do is identify variables which are required to remove ambiguity in observations. So, unless we know the variables, then we do not know what to observe. Essentially, our objective is to observe something now what those things are will be determined by the variables. So, after identification of one or more good research questions our next stage is to identify variables.

(Refer Slide Time: 12:00)



Then in the third stage, what we do is design the experiment or design the study. It refers to planning and execution of the experiment.

(Refer Slide Time: 12:13)

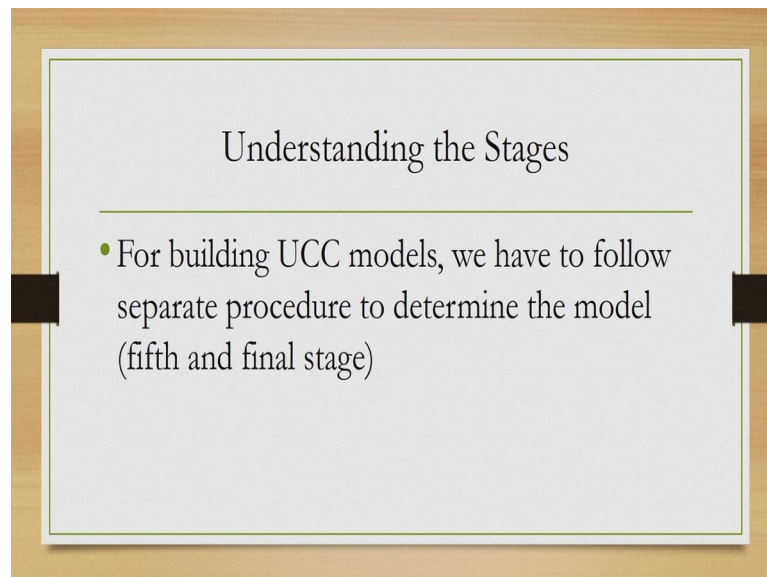


Why this is required, because getting the right users and in the right numbers is essential to observe to collect data. Otherwise, whatever data we collected may lead to conclusions that need not be or may not be reliable. Similarly, it is also necessary to determine appropriate tasks, which we are going to give to the users to perform so, that we can observe their behavior.

And, those tasks are to be given in a suitably controlled environment. So, both are important control the environment suitably and decide on a set of appropriate tasks to be given to the users. Otherwise, we may land up with again unreliable data, which will lead to unreliable conclusion after analysis. And, all these things are taken care of during the third phase that is design of the experiment in the fourth phase we analyze the data.

Now, data analysis is not trivial as you all probably know and we need to identify suitable approach to analyze the particular data that we have collected during our experiment.

(Refer Slide Time: 13:42)



Otherwise, we may not come up with any conclusion or the conclusions may not be reliable. These are the four stages in that sequence, which are essential to perform an empirical research with the sole objective of validating some concerns or trying to figure out if some concerns are justifiable or not.

Now, along with that we also may be interested in building a model, which is nothing, but a relationship between two or more quantities. Now, in order to build the model, we need to take some additional steps and those steps constitute the fifth stage or the final stage of the overall empirical research process. So, these are the five stages. Now, let us start detailed discussion on each of these stages to understand them.

So, we will start with the first stage that is formulation of good research questions.

(Refer Slide Time: 15:02)

Illustrative Case Study

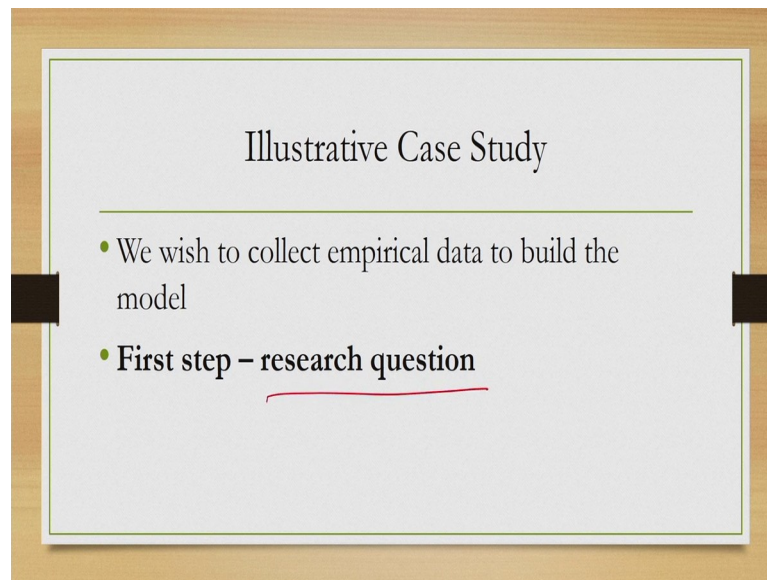
- Suppose we are interested in building a computational model for our aesthetic judgment behavior
- Important topic since there is direct relation with usability (measure of **satisfaction**)

Let us start with an illustrative case study to understand the importance of good research questions. And, what we mean by the term good? Suppose, we are interested in building a computational model for our aesthetic judgment behavior. What is aesthetic judgment behavior? It, essentially the behavior using, which we can tell whether something, which are presented to us is beautiful or not so, it is our sense of beauty and what we want? We want to come up with a computational model to capture this behavior

That means given an interface or interactive system interface, we want to say or predict the quality of its aesthetic and that prediction will be performed using the model this is similar to the Hick Hymans law or the Fitts law, where we are predicting something based on something else. Now, why aesthetic judgment prediction is important, because it is indirectly related to usability or one measure of usability that is satisfaction unless we feel something is beautiful, we may not be satisfied with its performance.

So, if we know that something is beautiful, then there is a there is a likelihood that the resulting interface is likely to be satisfactory to the end users. And, we want to build a computational model to predict the aesthetic quality of an interface.

(Refer Slide Time: 16:56)

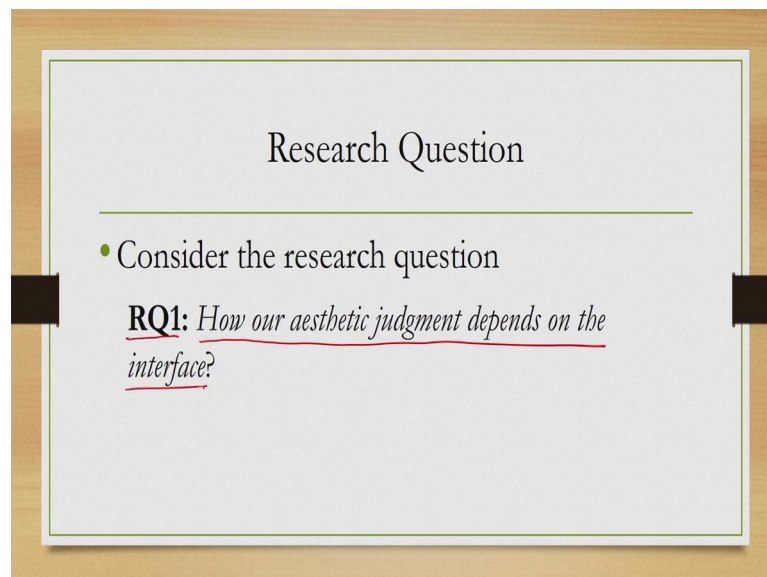


Illustrative Case Study

- We wish to collect empirical data to build the model
- First step – research question

Now, in order to build the model we will follow an empirical approach; that means we will collect data and based on the data we will build a relationship between the interface and our aesthetic judgment. So, the first step as we have already mentioned is to frame suitable research questions, suitable research question or questions.

(Refer Slide Time: 17:17)



Research Question

- Consider the research question
RQ1: How our aesthetic judgment depends on the interface?

Let us start with a simple question. Let us denote it by RQ 1 or research question 1, which is how our aesthetic judgment depends on the interface?

(Refer Slide Time: 17:34)

Research Question

- With RQ1, we are supposed to observe the relationship between the user's aesthetic judgment behavior and the interface

So, with RQ 1, we are supposed to observe the relationship between the user's aesthetic judgment and the interface. So, you are supposed to observe the relationship and using that observation using the data collected during the observation we want to build a model to capture that relationship.

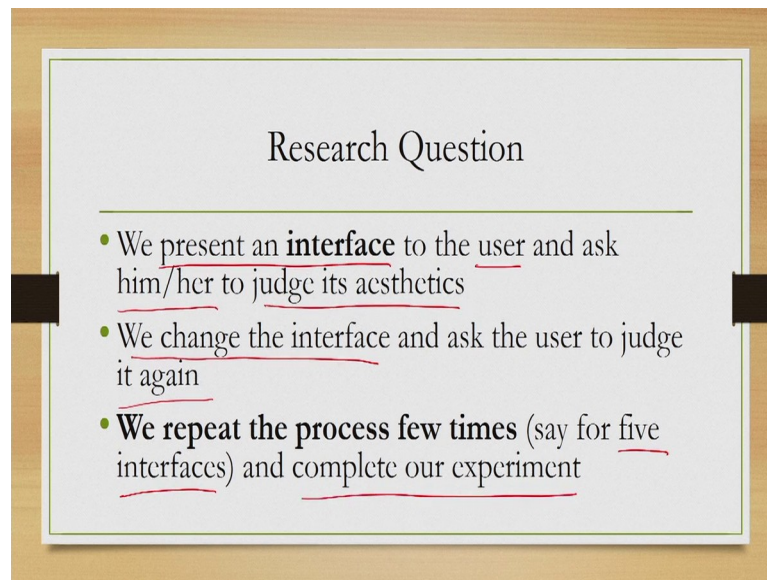
(Refer Slide Time: 17:58)

Research Question

- To observe the relationship, we set up an experiment

So, to observe the relationship we set up an experiment, which is typically, what we do.

(Refer Slide Time: 18:07)

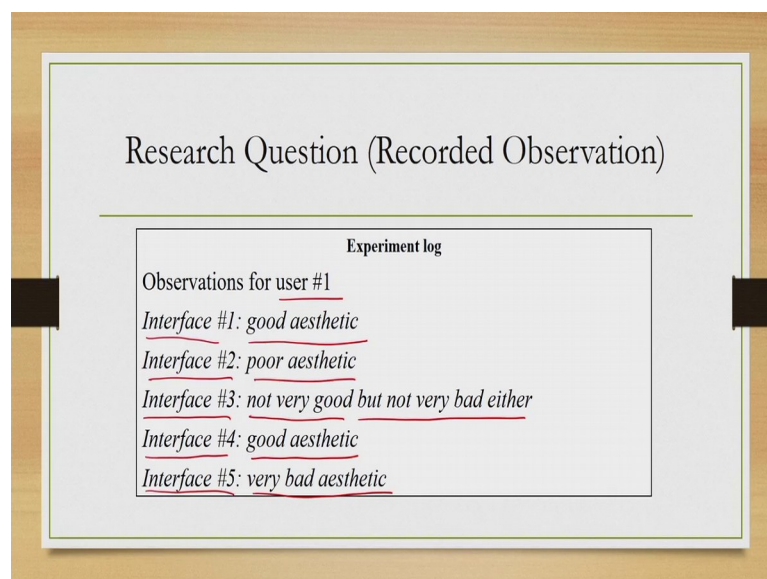


Research Question

- We present an interface to the user and ask him/her to judge its aesthetics
- We change the interface and ask the user to judge it again
- **We repeat the process few times** (say for five interfaces) and complete our experiment

So, what we have done or what we are likely to do in that experiment, we present an interface to a user or a participant and ask him or her to judge it is aesthetics. Then, after the judgment is given, we change the interface bring in a new one and ask the user or the participant to judge it judge the new interface again and we repeat this process few times say for five interfaces. So, for each interface we take the judgment from the participant and we do it for five interfaces and then we complete our experiment.

(Refer Slide Time: 18:55)



Research Question (Recorded Observation)

Experiment log

Observations for user #1

Interface #1: good aesthetic

Interface #2: poor aesthetic

Interface #3: not very good but not very bad either

Interface #4: good aesthetic

Interface #5: very bad aesthetic

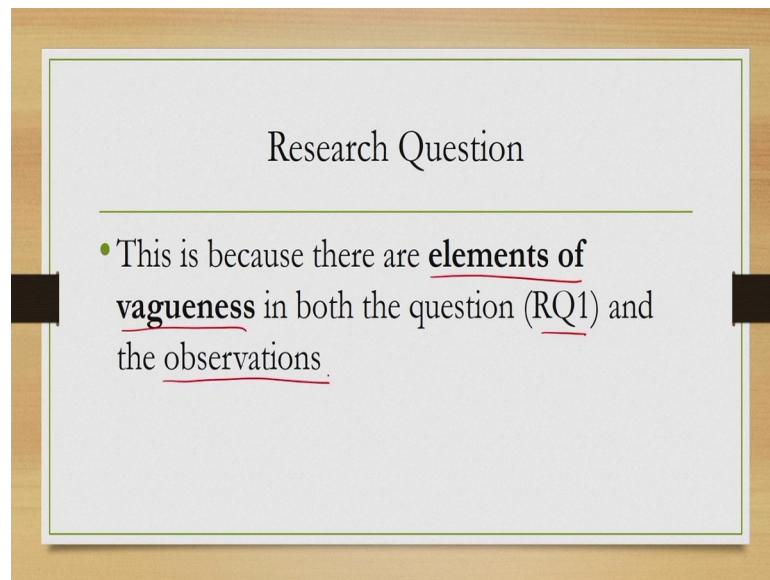
Suppose, this is the observation, that we have recorded, let us assume there are few participants numbered like 1, 2, 3. So, for participant 1 or user 1, we have recorded his or her judgment behavior, in this way for interface 1, the judgment was expressed as good aesthetic for interface 2 it was expressed as poor aesthetic, for interface 3 the user expressed the judgment as not very good, but not very bad either, for interface 4 the judgment was again good aesthetic and for interface 5 the judgment was very bad aesthetic.(Refer Slide Time: 20:03)

Research Question

- Observation reveals *aesthetic judgment depends on the interface*
 - Something well known - study does not reveal anything new
- We are not likely to get answer to RQ1: how they are related

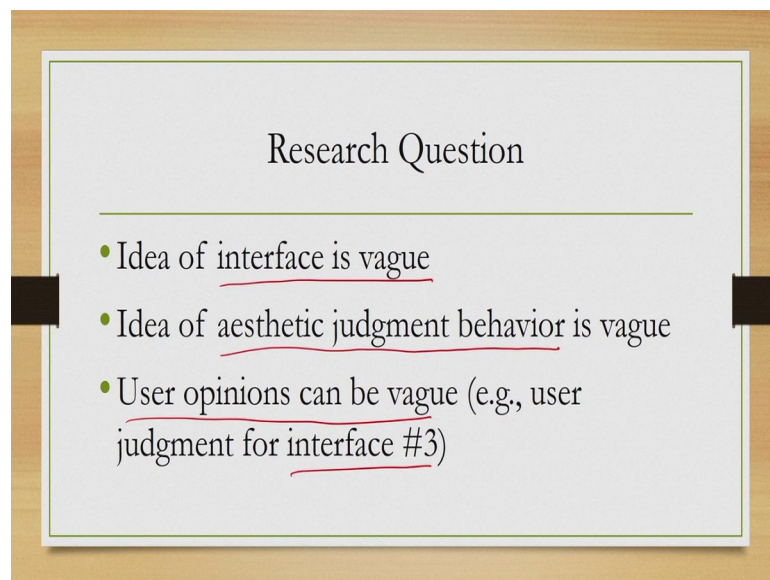
Now, so, this is the observation. So, from this observation what can we do can you do? Anything, unfortunately the answer is no why? Because, the observations reveal that aesthetic judgment depends on the interface, but this is the knowledge that we already have this is nothing new and we want to actually find out the answer to the question, how this judgment is related to the interface, rather than whether there exists some relationship. So, we are not likely to get any answer to our question how they are related?

(Refer Slide Time: 20:38)



Now, why this problem? Because, there are elements of vagueness in both the question and the observations.

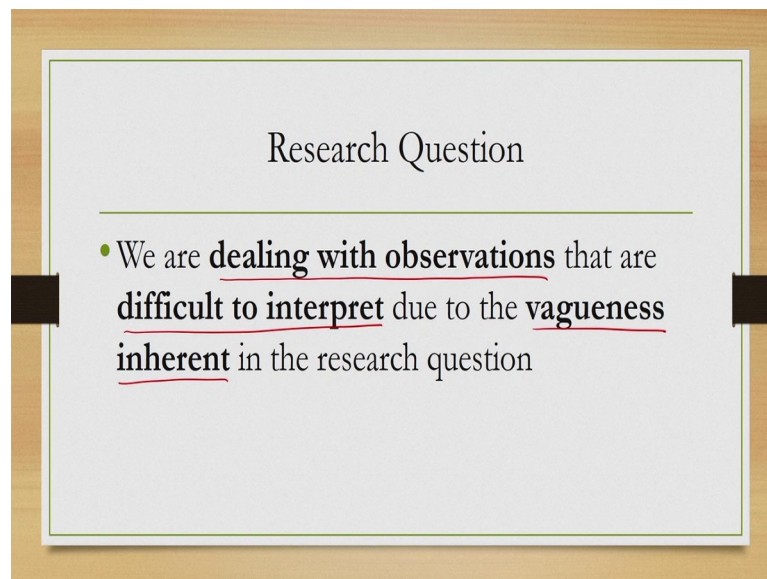
(Refer Slide Time: 20:50)



Let us see what are those elements. First of all the idea of interface is vague, we said that we have used interface 1, interface 2, interface 3, but we did not say, how they are different, what exactly it means that this is interface 1, this is interface 2, how to define an interface, that is not clear.

Secondly, we said that the users have expressed their aesthetic judgment, but it is not clear. What this aesthetic judgment behavior is how to define it? So, users were free to express it the way you are see liked. So, we did not give them clear instruction or clear definition of this behavior. And, since there is no clear definition so, user opinion can be vague. For example, we have seen already that for interface 3, the judgment given was it is not very good, but not very bad either, but this is kind of vague, we cannot conclude anything out of this judgment.

(Refer Slide Time: 21:58)



So, there are vagueness, what it means that we are dealing with observations that are difficult to interpret due to the vagueness inherent in the question. So, we have framed a question, but the question was vague and according to this vagueness inherent in the question, the observations that we have made based on that question are likely to be confusing vague. And, we cannot make clear cut conclusion or we cannot interpret these observations clearly.

(Refer Slide Time: 22:37)

Research Question

- Let us define an interface to be a collection of objects (images, text blocks, headings, tables, animations ...)
- Interface represented in terms of number of objects N

Now, let us modify the definition of an interface. Earlier we did not define anything let us define an interface to be a collection of objects. For example, there can be images, text blocks, headings, tables, animations all these are objects. And, we are defining an interface to be a collection of such objects. Now, we are representing an interface as something which has a number of objects. So, we are now representing interface in terms of number of objects N say for example.

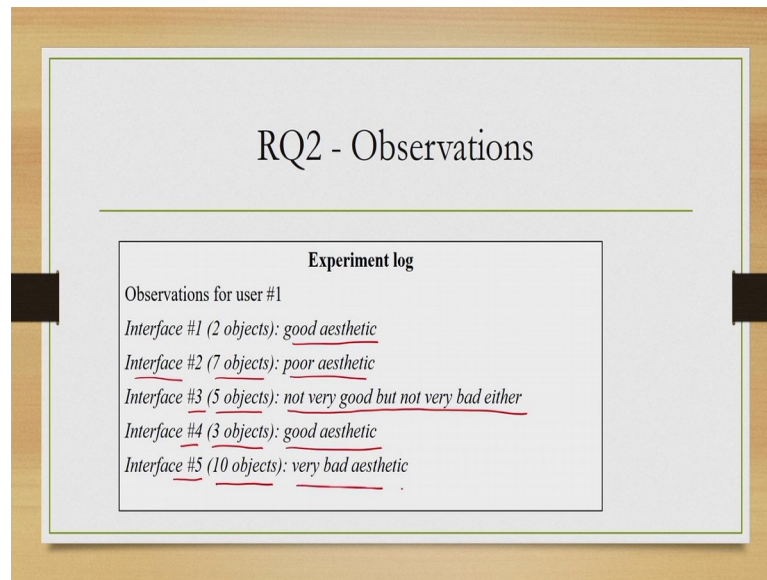
(Refer Slide Time: 23:13)

Research Question

- We reformulate the research question
RQ2: How our aesthetic judgment depends on the number of objects an interface has?

Then, we can reformulate this research question in a different way. We can now frame our question as shown here let us call it RQ 2 or research question 2, it says that how our aesthetic judgment depends on the number of objects an interface has. So, now, we have changed our research question.

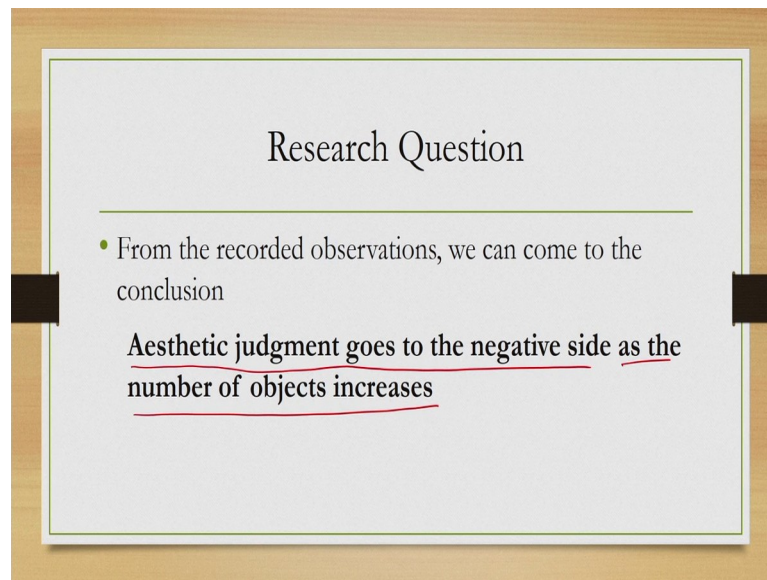
(Refer Slide Time: 23:42)



And, we have again conducted the experiment in a similar way as we have done before and recorded the observations, but now the observations are slightly different as shown here.

Now, each entry indicates interface in terms of number of objects. So, interface 1 has 2 objects and the judgment expressed is good aesthetic interface 2 has 7 objects and the judgment expressed is poor aesthetic. Similarly, interface 3 has 5 objects and the judgment expressed is not very good, but not very bad either interface 4; 4 has 3 objects and the judgment expressed is good aesthetic, interface 5 has 10 objects and the judgment expressed is very bad aesthetic.

(Refer Slide Time: 24:31)



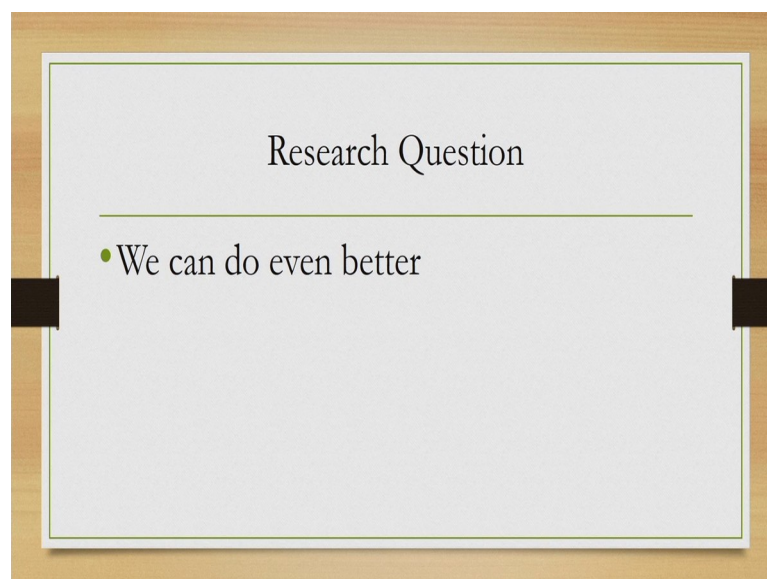
Research Question

- From the recorded observations, we can come to the conclusion

Aesthetic judgment goes to the negative side as the number of objects increases

Now, we can come to some conclusion. Earlier, we are unable to come to any conclusion because of the vagueness everywhere, now the definition of the interface is less vague. Now, we are defining it in terms of number of objects and then we have recorded the aesthetic judgment for those many numbers of objects in an interface. So, we can now conclude based on these observations in the following way, aesthetic judgment goes to the negative side as the number of objects increase.

(Refer Slide Time: 25:14)

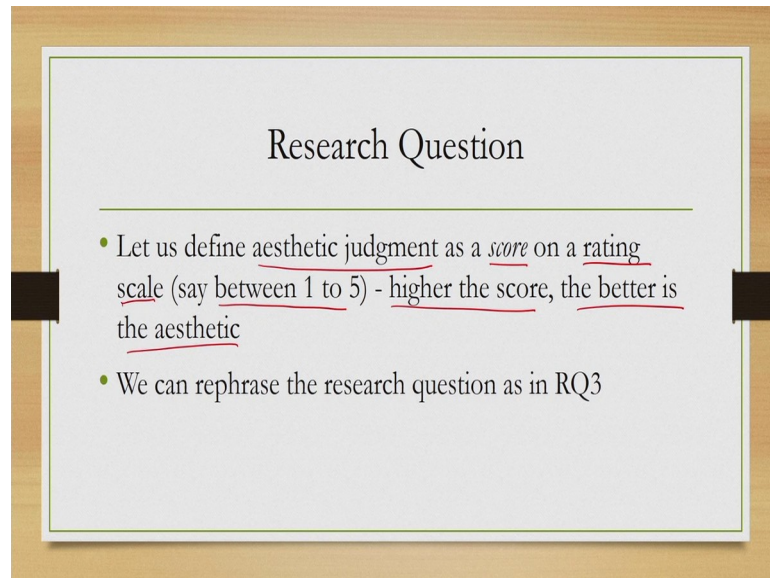


Research Question

- We can do even better

So, in other words it says that if the number of objects increased then a user perceives the interface to have less aesthetic quality. Is that all we can do? Actually we can do even better, let us see how.

(Refer Slide Time: 25:19)

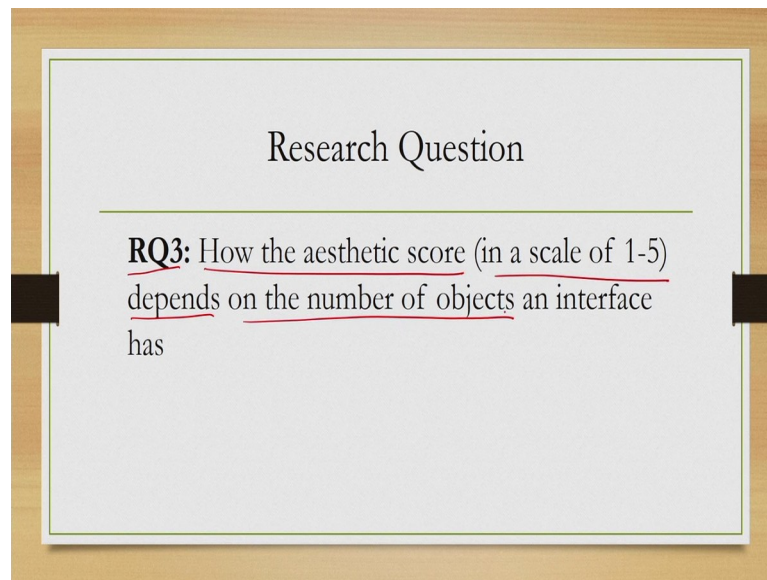


Research Question

- Let us define aesthetic judgment as a score on a rating scale (say between 1 to 5) - higher the score, the better is the aesthetic
- We can rephrase the research question as in RQ3

Let us now define aesthetic judgment as a score on a rating scale say we are using a rating scale of 5 points and the score is between 1 to 5. Now, higher the score, the better is the aesthetic; that means, if the score is 1, it is the worst aesthetic quality and the score is if the score is 5 the quality is best. Now, based on this definition, we can again come up with an even better research question.

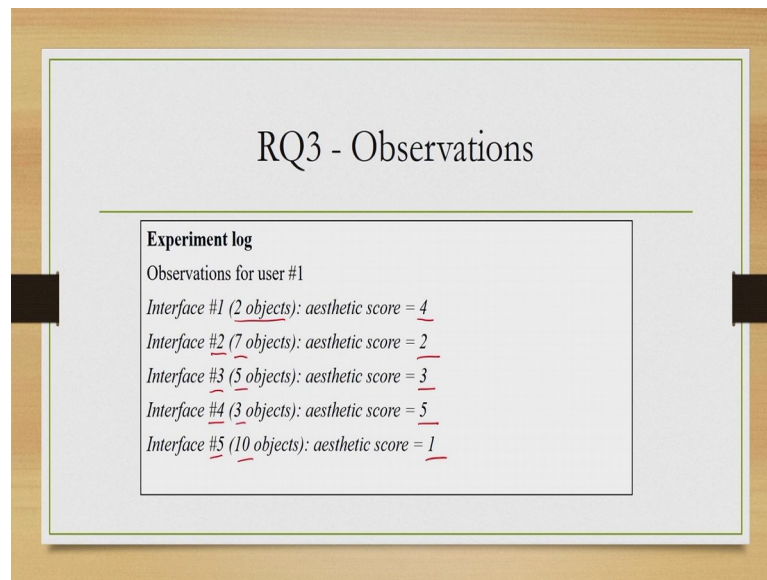
(Refer Slide Time: 25:57)



Let us call it RQ 3 a research question 3, which is how the aesthetic score in a scale of one to 5 depends on the number of objects. So, in RQ one everything was vague, in RQ 2 we have removed vagueness to some extent on the definition of interface, but still the judgment behavior was vaguely defined. Now, in RQ 3 we are removing that ambiguity to some extent, by mentioning a rating scale to define the aesthetic judgment behavior.

So, based on this research question, now we have conducted, suppose now we have conducted the experiment and the recordings from the experiment may look something like this.

(Refer Slide Time: 26:50)



RQ3 - Observations

Experiment log

Observations for user #1

Interface #1 (2 objects): aesthetic score = 4

Interface #2 (7 objects): aesthetic score = 2

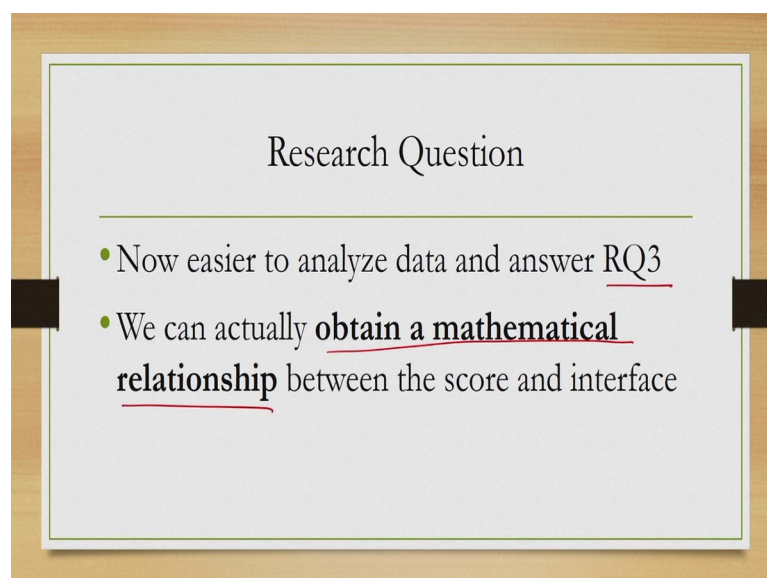
Interface #3 (5 objects): aesthetic score = 3

Interface #4 (3 objects): aesthetic score = 5

Interface #5 (10 objects): aesthetic score = 1

Now, we have interface 1 2 objects with a score of 4, interface 2 7 objects with a score of 2 interface 3 5 objects with a score of 3 interface 4, 3 objects with the score of 5 and interface 5, 10 objects with a score of 1. So, now, on both the sides we have some numbers. Earlier that was not the case, which we have addressed with this definitions of interface and the judgment behavior.

(Refer Slide Time: 27:21)

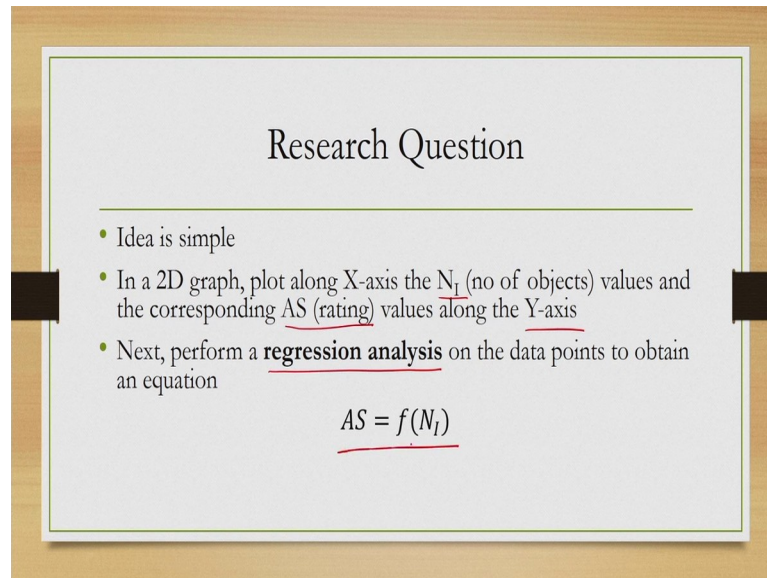


Research Question

- Now easier to analyze data and answer RQ3
- We can actually **obtain a mathematical relationship** between the score and interface

So with numbers it is easier to analyze the data and answer the question how these numbers are related. So, how the score is related to the number of objects on the interface?

(Refer Slide Time: 27:44)



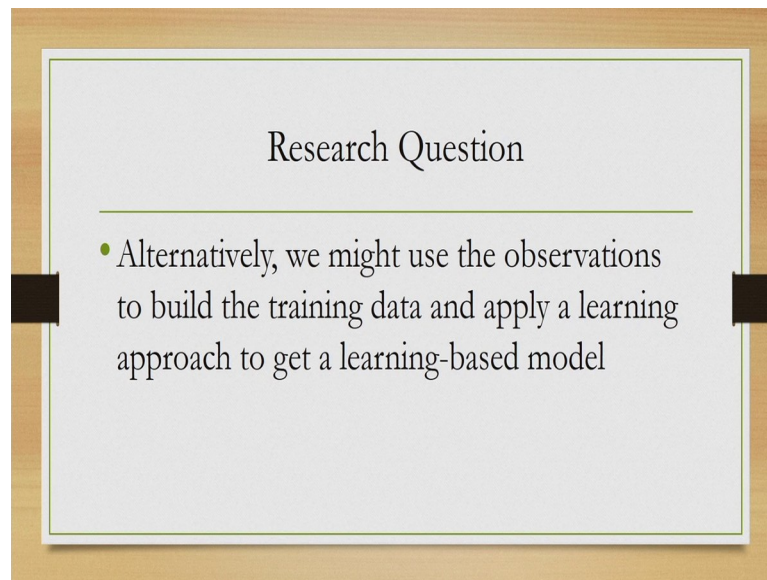
Research Question

- Idea is simple
- In a 2D graph, plot along X-axis the N_I (no of objects) values and the corresponding AS (rating) values along the Y-axis
- Next, perform a regression analysis on the data points to obtain an equation

$$\underline{AS = f(N_I)}$$

We can actually obtain a mathematical relationship between the score and the interface, using a simple approach that is we plot along X-axis the number of objects these values and the rating score is plotted along the Y-axis, and then we perform a regression analysis on the data points in that graph to come up with an equation of the form shown here. A S equal to function of N I, this is how as you may recollect we said that the original formulation of the relationship between index of difficulty and the target distance and width were found out.

(Refer Slide Time: 28:27)

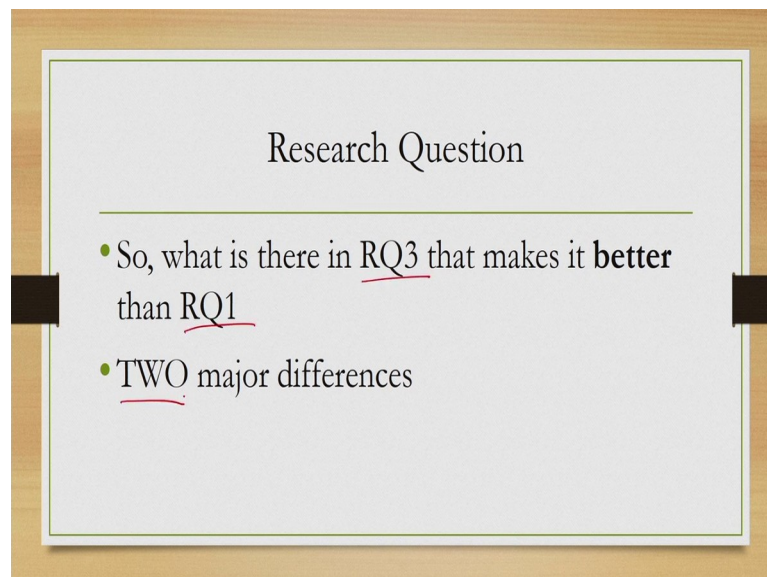


Research Question

- Alternatively, we might use the observations to build the training data and apply a learning approach to get a learning-based model

In fact, if we are not interested in building a specific mathematical relationship, rather we want to go for a classification approach, then we can use the same data to train the classifier as well as test it. So, essentially the same data can be used for a learning based modeling activity.

(Refer Slide Time: 28:48)

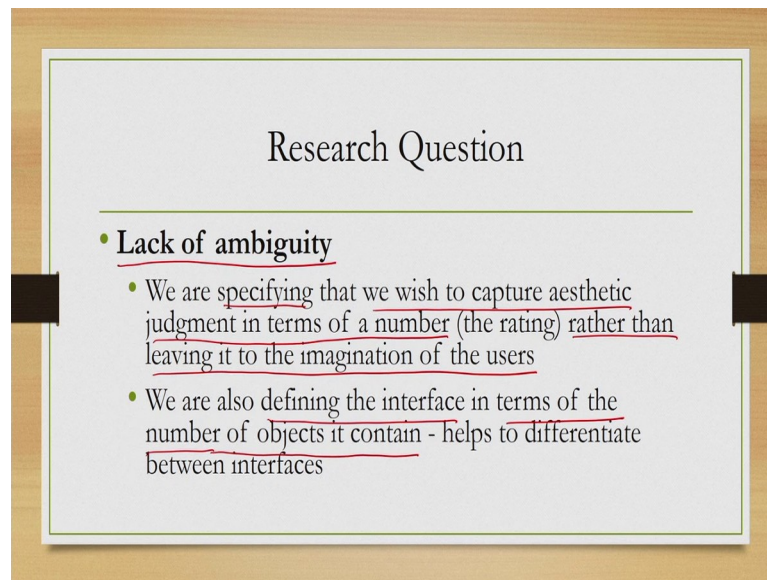


Research Question

- So, what is there in RQ3 that makes it **better** than RQ1
- TWO major differences

So, what is the difference between RQ 3 and RQ 1? So, primarily 2 major differences.

(Refer Slide Time: 28:56)



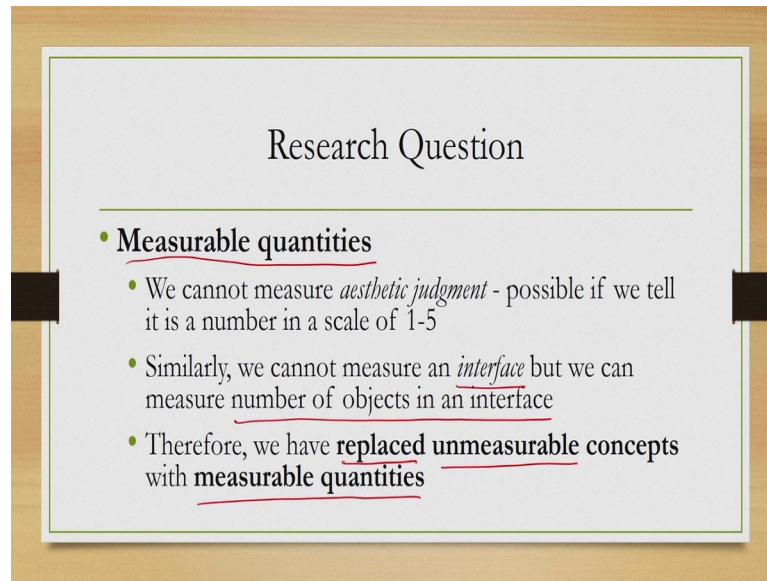
Research Question

- Lack of ambiguity
 - We are specifying that we wish to capture aesthetic judgment in terms of a number (the rating) rather than leaving it to the imagination of the users
 - We are also defining the interface in terms of the number of objects it contain - helps to differentiate between interfaces

First of all RQ 3 is less ambiguous than RQ 1. So, in RQ 3 we are specifying that we wish to capture aesthetic judgment in terms of a number the rating rather than leaving it to the imagination of the users.

So, now we are no longer saying that express your judgment the way you like instead we are saying that you have to express your judgment in terms of the rating. Secondly, we are also defining the interface in terms of a another number, in terms of another number that is the number of objects the interface contains. So, with this definition, it is easier to differentiate between interfaces earlier that was not possible in RQ 1.

(Refer Slide Time: 29:45)



Research Question

- Measurable quantities
 - We cannot measure *aesthetic judgment* - possible if we tell it is a number in a scale of 1-5
 - Similarly, we cannot measure an *interface* but we can measure number of objects in an interface
 - Therefore, we have replaced unmeasurable concepts with measurable quantities

The second difference between RQ 1 and RQ 3 is the use of measurable quantities in RQ 3. If, we simply say aesthetic judgment behavior there is no way we can measure it. However, if we say that aesthetic judgment behavior is expressed in terms of a rating, then we can always measure the rating.

Similarly, we cannot measure an interface, but we can measure number of objects in an interface. So, we have replaced the unmeasurable concepts with measurable quantities, that is the second big difference between RQ one and RQ 3. Because, of these differences these two research questions are treated differently we call RQ 3 as testable research question whereas, we call RQ 1 as non-testable research question.

(Refer Slide Time: 30:40)

Research Question

- We call R3 testable and R1 as non-testable

Now, whatever we have discussed so far indicates that framing of research questions can take place in either of the 2 ways. Either, we can come up with a testable research question or a non-testable research question. Now, there is a tradeoff between the two, let us try to understand that trade off.

(Refer Slide Time: 31:09)

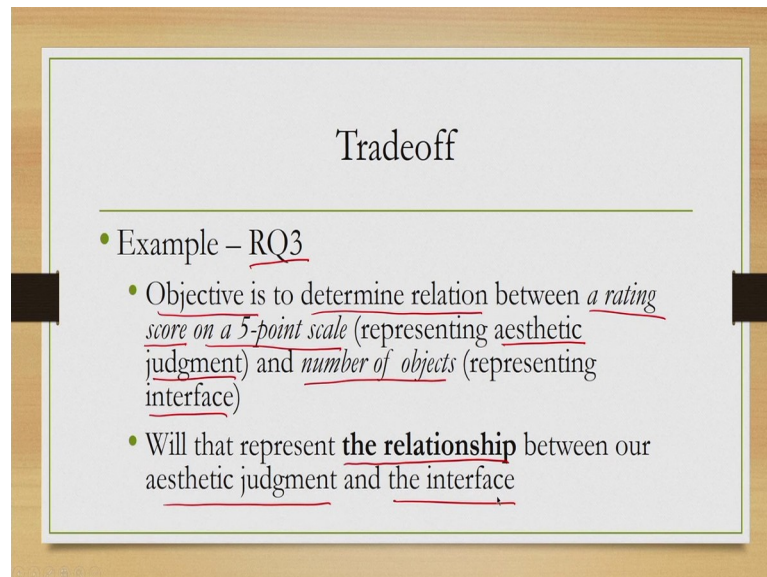
Tradeoff

- Our aim - frame testable questions
- Problem
 - Testable questions designed to seek answer to specific queries
 - Such questions may lack generalizability .

So, what is our aim our aim is to frame testable question why, because with testable question we can record observations unambiguously and come to a conclusion, unlike in the case of untestable or non-testable questions where, we may record something, but it

is difficult to come to a conclusion. However, the testable questions are designed to seek answer to specific queries, that is not the case with non-testable questions. So, testable questions may lack generalizability. Let us try to understand this in terms of our earlier example.

(Refer Slide Time: 31:49)

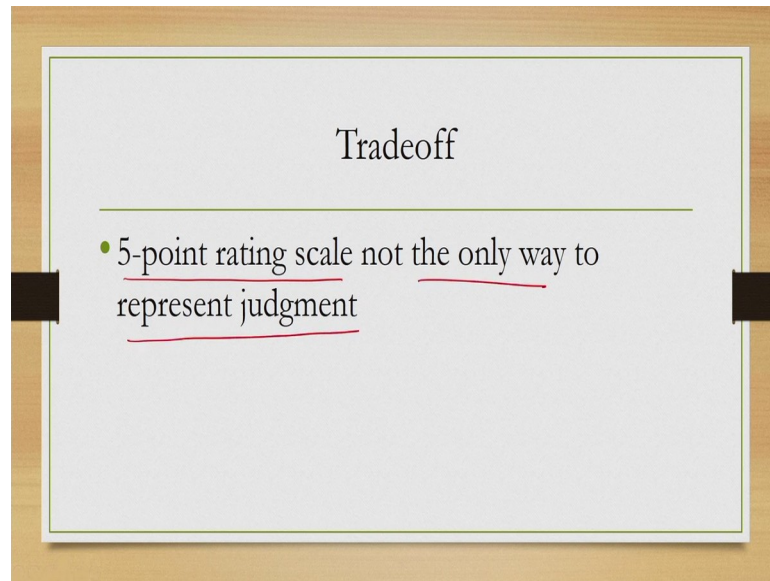


Tradeoff

- Example – RQ3
 - Objective is to determine relation between a rating score on a 5-point scale (representing aesthetic judgment) and number of objects (representing interface)
 - Will that represent the relationship between our aesthetic judgment and the interface

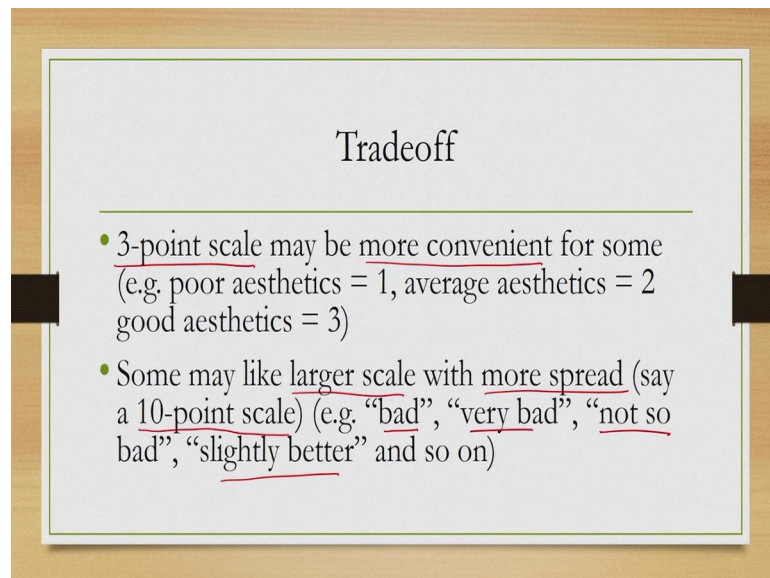
So, in case of RQ 3 or the research question 3, that we have formulated earlier, our objective is to determine a relation between a rating scale, or a rating score on a 5-point scale, which represents the aesthetic judgment behavior and the number of objects which represent the interface. So, this is our specific objective, that establish a relationship between a rating scale and a definition of the interface, which is number of objects. Now, if I get that relationship will that represent the relationship between our aesthetic judgment behavior and the interface, can we say anything on that?

(Refer Slide Time: 32:35)



The 5 point rating scale that we have mentioned need not be the only way to represent judgment. Why?

(Refer Slide Time: 32:44)

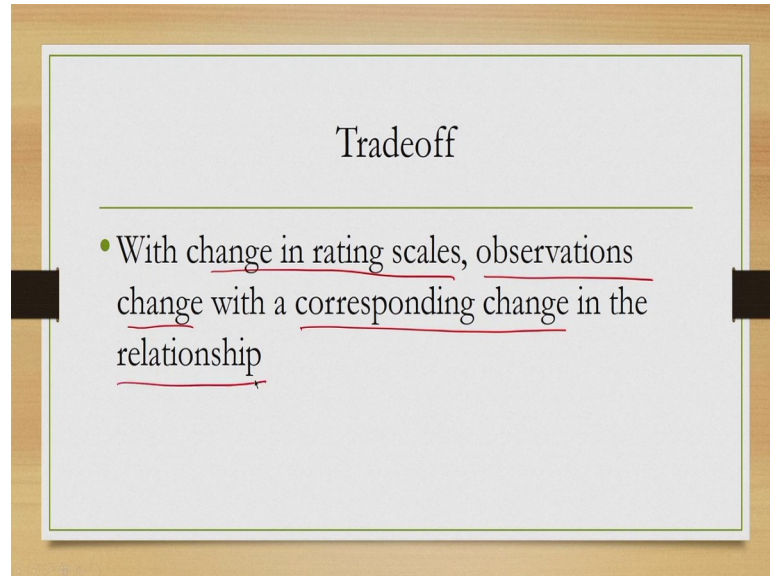


Because, somebody may feel a 3 point scale is better, and more convenient in this scale one will indicate poor aesthetics 2 indicates average aesthetics, and 3 indicates good aesthetics.

Similarly, somebody else miffed, somebody else may find a larger scale is better, where there is more spread. For example, say 10-point scale, which indicates or which can

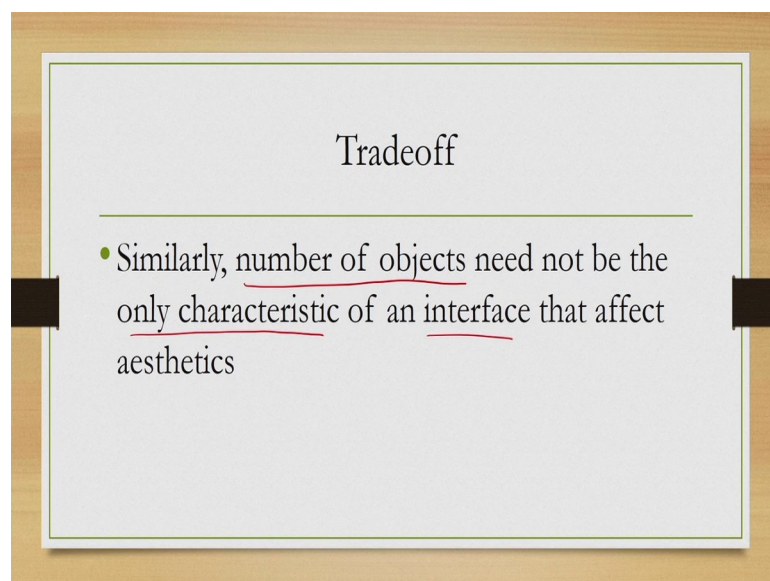
capture judgment's like bad, very bad, not so bad, slightly better and so on at a more coarse level.

(Refer Slide Time: 33:32)



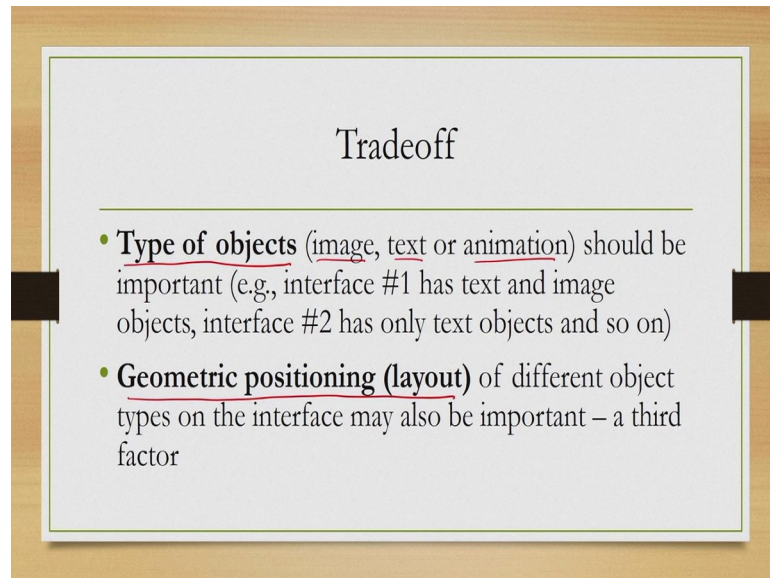
Now, with the change in rating scales definitely our observations are going to change. And, once we use that data to build a model or the relationship between the rating scale and the number of objects in the interface, we are likely to get different relationships. So, there is a corresponding change in the relationship as well.

(Refer Slide Time: 33:57)



Similarly, if the number of objects is used to define an interface that need not be the only characteristic of an interface, which might have an effect on aesthetic judgment behavior. What can be the other characteristics?

(Refer Slide Time: 34:18)



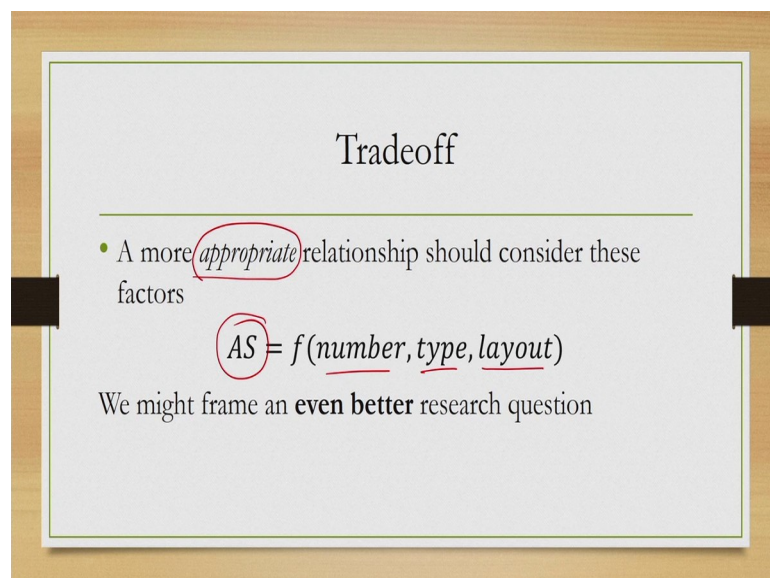
The slide is titled "Tradeoff" and contains two bullet points. The first bullet point discusses the importance of the type of objects (image, text, or animation) on an interface. The second bullet point discusses the importance of the geometric positioning (layout) of different object types on the interface as a third factor.

Tradeoff

- **Type of objects** (image, text or animation) should be important (e.g., interface #1 has text and image objects, interface #2 has only text objects and so on)
- **Geometric positioning (layout)** of different object types on the interface may also be important – a third factor

Intuitively the type of objects that are there on the interface such as image, text, or animation should have some effect on the aesthetic judgment. The geometric positioning or layout of the different object types on the interface may also be important which can potentially be a third factor.

(Refer Slide Time: 34:41)



The slide is titled "Tradeoff" and contains a bullet point and a formula. The bullet point suggests a more appropriate relationship should consider these factors. The formula is AS = f(number, type, layout). The text below the formula suggests framing an even better research question.

Tradeoff

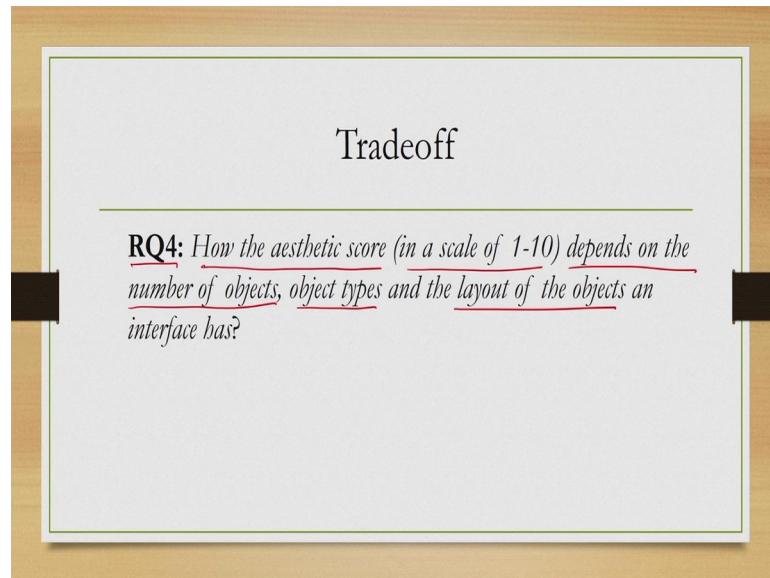
- A more appropriate relationship should consider these factors

$$AS = f(\text{number}, \text{type}, \text{layout})$$

We might frame an **even better** research question

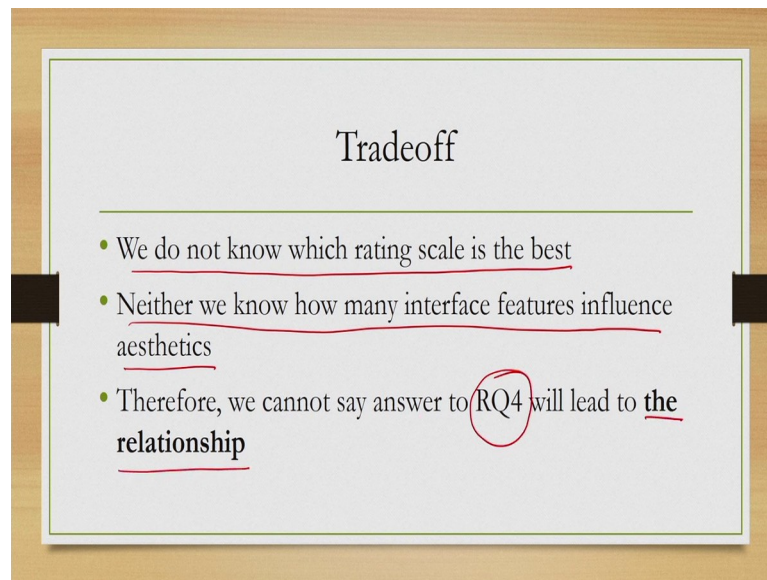
So, we can say that a more appropriate relationship should consider 3 factors instead of one to predict aesthetic judgment behavior. What are those 3 factors number, type, and layout? To predict an aesthetic judgment behavior represented in the form of an aesthetic score, which again can be predicted in either a 5 point scale, or a 3 point scale, or a 10 point scale. Now, with this formulation we can frame an even better research question.

(Refer Slide Time: 35:20)



Let us assume a 10 point scale. So, our research question 4 can be how the aesthetic score in a scale of 1 to 10. Depends on the number of objects, object types and layout of the objects on the interface? Note, here that we are again using measurable unambiguous quantities, but the number of quantities that we are considering here are more compared to RQ 3.

(Refer Slide Time: 35:50)



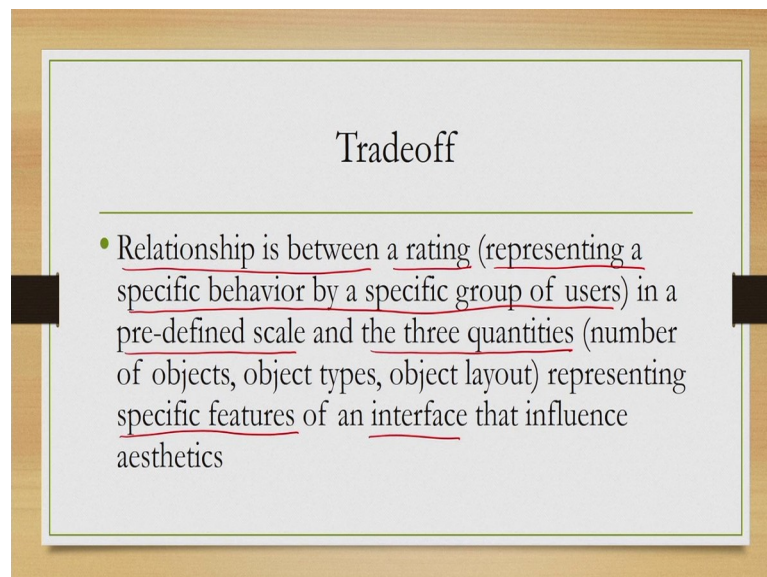
The slide is titled "Tradeoff" and contains three bullet points. The first point is "We do not know which rating scale is the best". The second point is "Neither we know how many interface features influence aesthetics". The third point is "Therefore, we cannot say answer to RQ4 will lead to the relationship". The text "RQ4" is circled in red, and "the relationship" is underlined in red.

Tradeoff

- We do not know which rating scale is the best
- Neither we know how many interface features influence aesthetics
- Therefore, we cannot say answer to RQ4 will lead to the relationship

So, is RQ 4 going to take us to a better relationship, more appropriate relationship, we do not know because, we do not know which rating scale is the best. Neither we know, how many interface features influence aesthetics. So, we cannot say that the answer to RQ 4 is going to lead us to the relationship.

(Refer Slide Time: 36:16)



The slide is titled "Tradeoff" and contains one bullet point. The text is "Relationship is between a rating (representing a specific behavior by a specific group of users) in a pre-defined scale and the three quantities (number of objects, object types, object layout) representing specific features of an interface that influence aesthetics". The text is underlined in red.

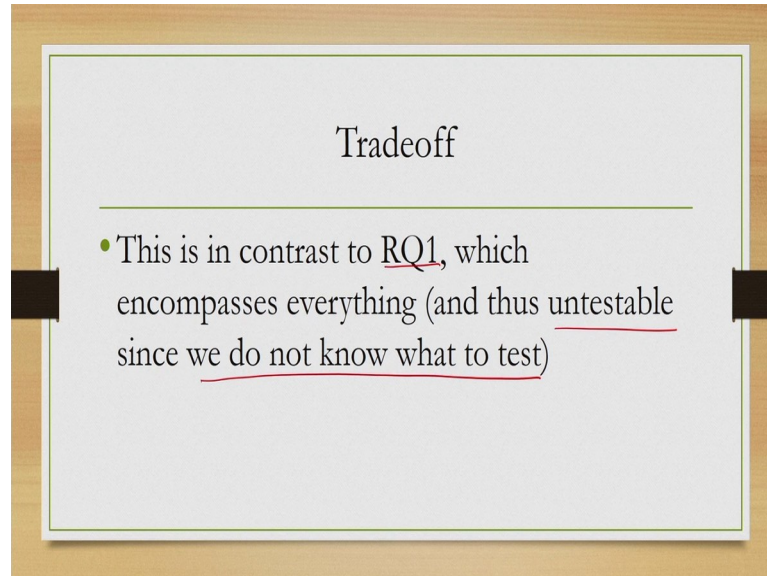
Tradeoff

- Relationship is between a rating (representing a specific behavior by a specific group of users) in a pre-defined scale and the three quantities (number of objects, object types, object layout) representing specific features of an interface that influence aesthetics

And, why this is happening because whatever relationship we are getting by analyzing the empirical data. The relationship is between a rating which represents a specific

behavior by a specific group of users. In a predefined scale and 3 quantities which again represents specific features of an interface, note the emphasis on the terms specific.

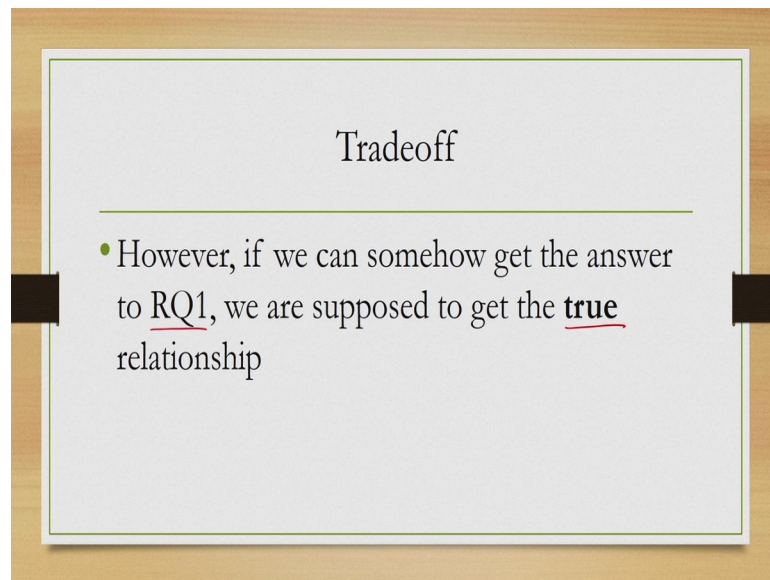
(Refer Slide Time: 36:46)



So, either with RQ 3 or RQ 4 we are going to get some relationship, which relies on specific ways to represent aesthetic judgment and specific ways to represent interface. In contrast in RQ 1, we do not have such considerations; there in that research question we did not define either aesthetic judgment behavior or an interface in a specific way. So, whatever we observed is applicable for any interface or any judgment behavior. So, the scope was broad there whereas, the scope is specific in the case of RQ 3 or RQ 4.

But, because the scope was not specific we have a very generalized concepts there RQ 1 was untestable or non-testable, because it was not clear to us, what to test, which type of interface to test, what type of judgment behavior we are expecting and so on.

(Refer Slide Time: 37:53)

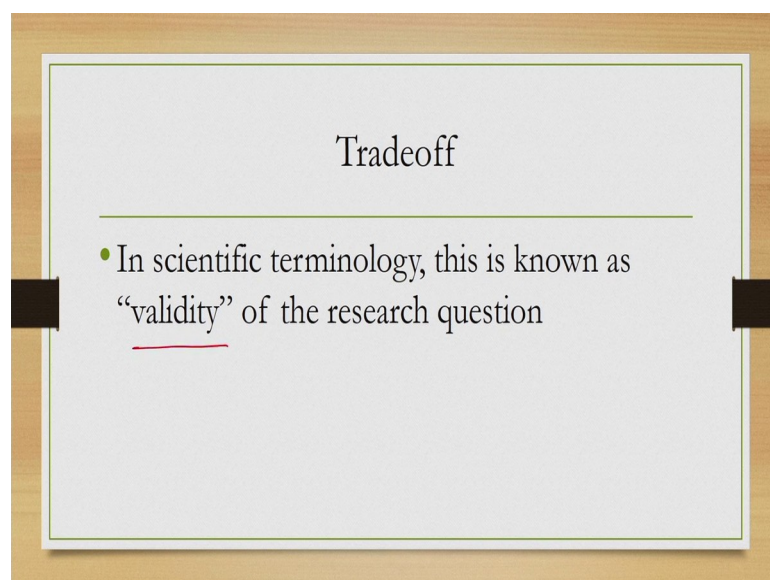


Tradeoff

- However, if we can somehow get the answer to RQ1, we are supposed to get the true relationship

But, if we can get the answer to RQ 1 somehow, then we can certainly say that we have got the true relationship, because RQ 1 is not specific to any interface or any specific judgment behavior representation. So, if we get answer to RQ 1 we get some relationship and this relationship, indicates the true relationship between the judgment behavior and the interface.

(Refer Slide Time: 38:25)

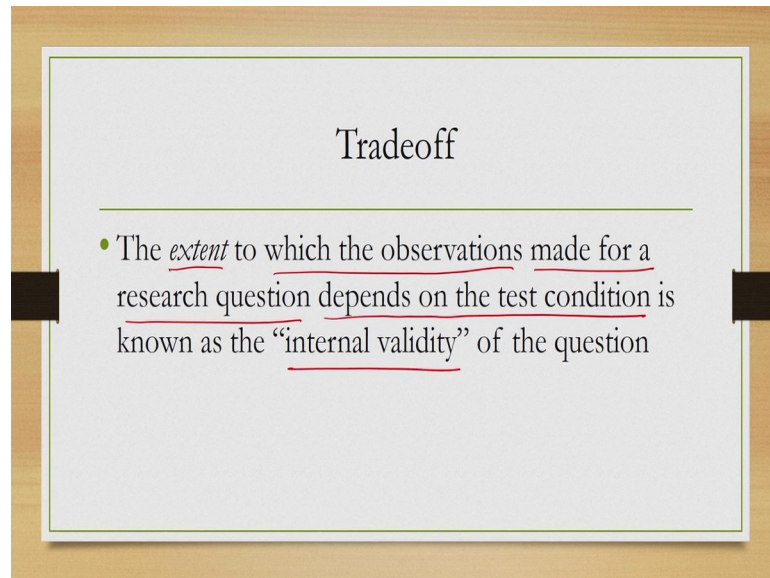


Tradeoff

- In scientific terminology, this is known as “validity” of the research question

Now, this phenomenon, which indicates the quality of a research question the nature of a research question, is called the validity of the question and there are two types of validities.

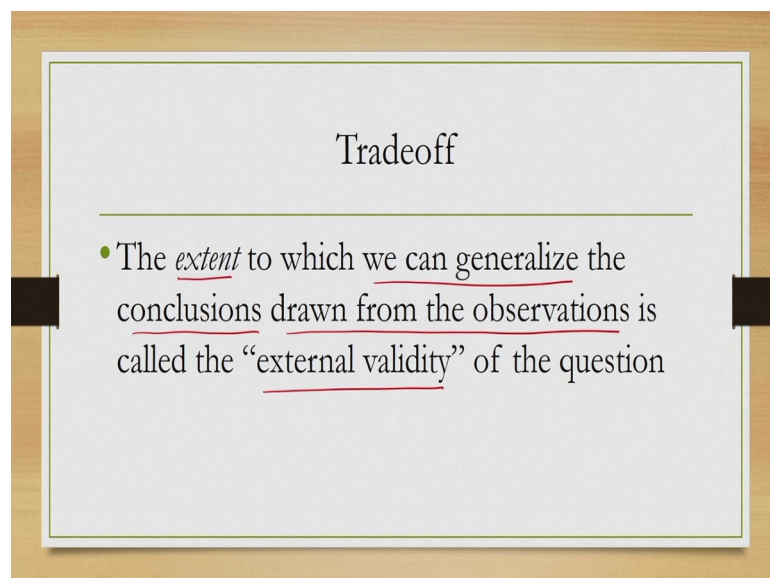
(Refer Slide Time: 38:40)



The slide is titled "Tradeoff" and contains a single bullet point. The text of the bullet point is: "The extent to which the observations made for a research question depends on the test condition is known as the "internal validity" of the question". The slide has a light gray background with a thin green border and is set against a wooden-textured background.

The extent to which the observations made for a research question depends on the test condition is known as the internal validity of the question.

(Refer Slide Time: 38:56)

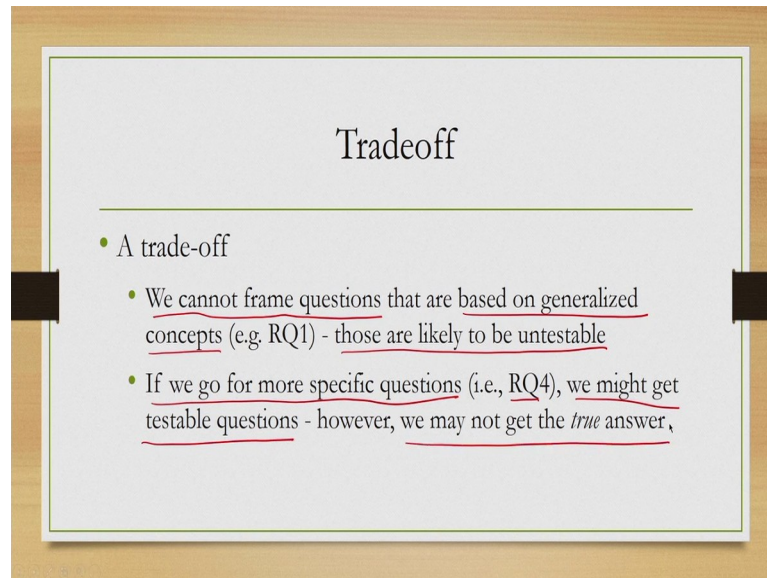


The slide is titled "Tradeoff" and contains a single bullet point. The text of the bullet point is: "The extent to which we can generalize the conclusions drawn from the observations is called the "external validity" of the question". The slide has a light gray background with a thin green border and is set against a wooden-textured background.

And, the extent to which we can generalize the conclusions drawn from the observations is called the external validity of the question. So, the extent to which the observations are

reliant on the test condition is called the internal validity and the extent to which we can generalize the findings, is called the external validity. So, the tradeoff is between these two validities of a research question.

(Refer Slide Time: 39:28)



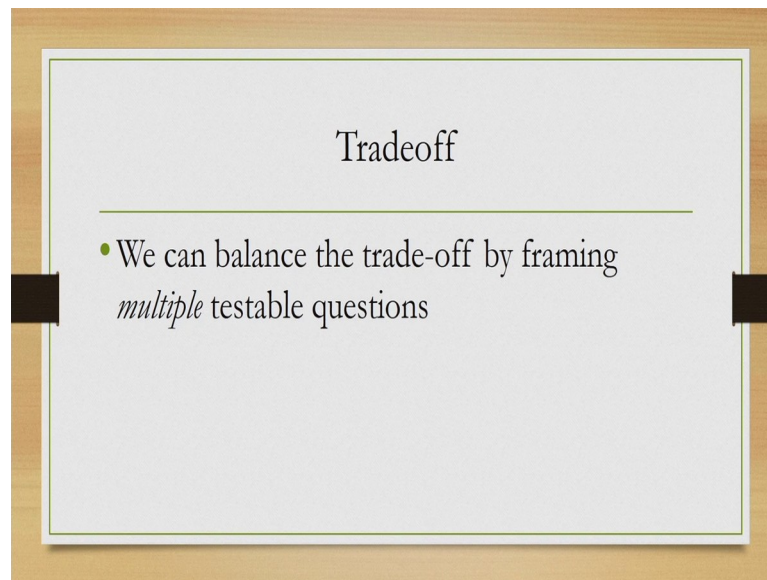
Tradeoff

- A trade-off
 - We cannot frame questions that are based on generalized concepts (e.g. RQ1) - those are likely to be untestable
 - If we go for more specific questions (i.e., RQ4), we might get testable questions - however, we may not get the true answer,

To what extent a question should have internal validity, without affecting the external validity. So, the tradeoff is as follows it is difficult to frame questions that are based on generalized concepts such as RQ 1, those are likely to be non testable or untestable.

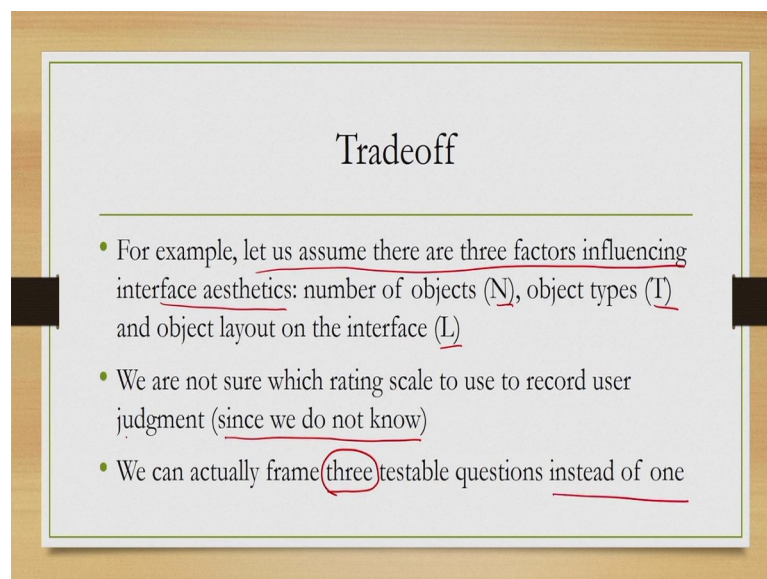
Now, if we go for more specific questions we increase internal validity such as RQ 4, but we might get answers that are not generalizable. So, we might get testable questions in that case, but we may not get true answers.

(Refer Slide Time: 40:10)



How we can balance this trade-off, there is one way to do that that is by posing multiple testable research questions and seeking answers to those questions. From those answers we can come to a conclusion about the generalizability of the observations.

(Refer Slide Time: 40:32)



For example, let us assume that there are three factors; influencing interface aesthetic; the number of objects object types and object layout on the interface. We are not sure which rating scale to use to record our judgment, because we do not know. Now, instead

of having a single research question we can frame three testable research questions instead of one.

(Refer Slide Time: 40:57)



Tradeoff

RQ4 (as before): *How the aesthetic score (in a scale of 1-10) depends on the number of objects, object types and the layout of the objects an interface has?*

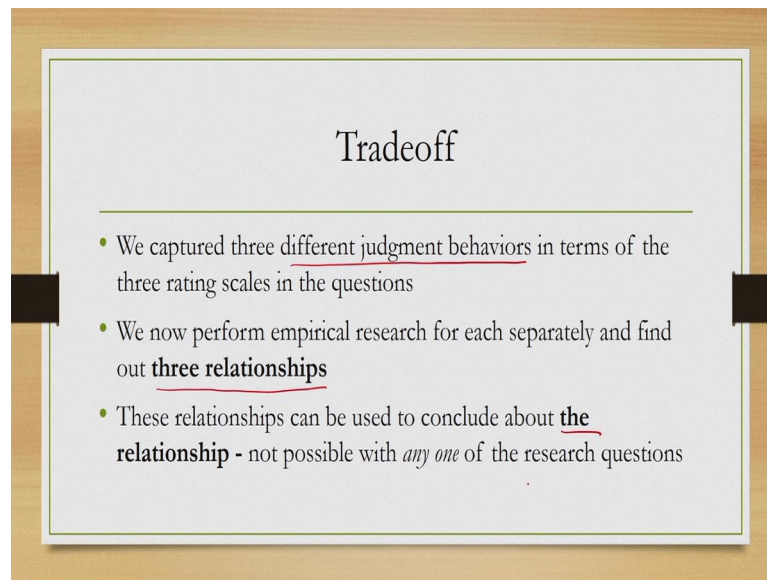
RQ5 (modified form of RQ3): *How the aesthetic score (in a scale of 1-3) depends on the number of objects, object types and the layout of the objects an interface has?*

RQ6: *How the aesthetic score (in a scale of 1-5) depends on the number of objects, object types and the layout of the objects an interface has?*

What are these questions? Let us call them RQ 4, RQ 5 and RQ 6, RQ 4 is as before, how the score in a scale of 1 to 10 depends on the number of objects, object types and layout. RQ 5 is modified from RQ 3, which is how the score in a scale of 1 to 3 depends on number of objects object types and layout. And, RQ 6 says; how the aesthetic score in a scale of 1 to 5 depends on the number of objects object types and layout.

So, in each of these questions interface definition remains the same in terms of 3 variables number of objects object types and layout of the objects, but rating scales are changed. In RQ 4 it is a 10 point rating scale in RQ 5, it is a 3 point rating scale and in RQ 6, it is a 5 point rating scale.

(Refer Slide Time: 41:52)



The slide is titled "Tradeoff" and is set against a light gray background with a thin green border. It features three bullet points, each with underlined text. The first bullet point states that three different judgment behaviors were captured using three rating scales. The second bullet point notes that empirical research on each scale separately yields three relationships. The third bullet point explains that these relationships can be used to conclude about a specific relationship, which is not possible if only one research question were considered.

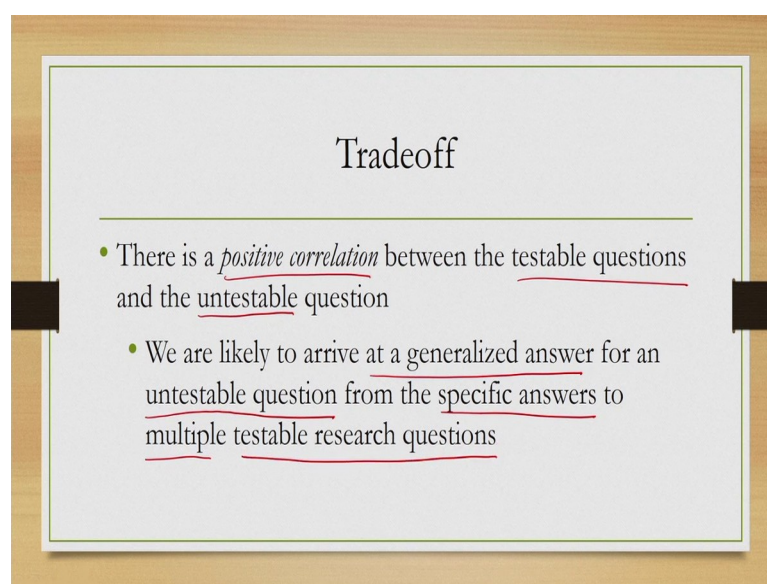
Tradeoff

- We captured three different judgment behaviors in terms of the three rating scales in the questions
- We now perform empirical research for each separately and find out three relationships
- These relationships can be used to conclude about the relationship - not possible with *any one* of the research questions

So, with these 3 questions, we captured different judgment behaviors in terms of 3 different rating scales, and then we perform empirical research as we have done before collect data, and find 3 relationships instead of 1.

Now, we can use these 3 relationships to come to a conclusion about the relationship or the true relationship, which is not possible, if we are considering only one of the research questions.

(Refer Slide Time: 42:21)



The slide is titled "Tradeoff" and is set against a light gray background with a thin green border. It features two bullet points, each with underlined text. The first bullet point states that there is a positive correlation between testable questions and an untestable question. The second bullet point explains that a generalized answer for an untestable question can be derived from specific answers to multiple testable research questions.

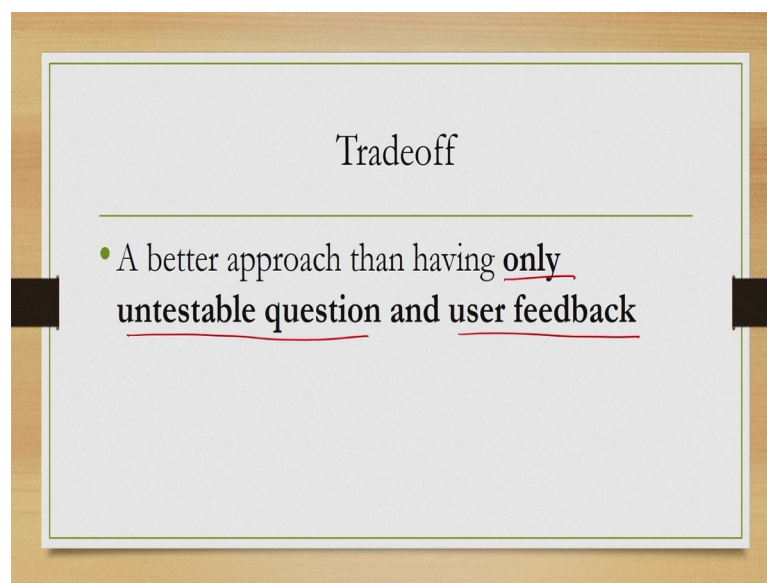
Tradeoff

- There is a positive correlation between the testable questions and the untestable question
- We are likely to arrive at a generalized answer for an untestable question from the specific answers to multiple testable research questions

It has been found, that there is a positive correlation between testable questions and untestable or non-testable questions. We are likely to arrive at a generalized answer for an untestable or non-testable question from the specific answers to multiple testable research questions.

So, if we get answer to multiple testable research questions then from those answers we are likely to find a conclusion, that is generalizable and that is the answer of a non-testable or untestable research question.

(Refer Slide Time: 43:05)

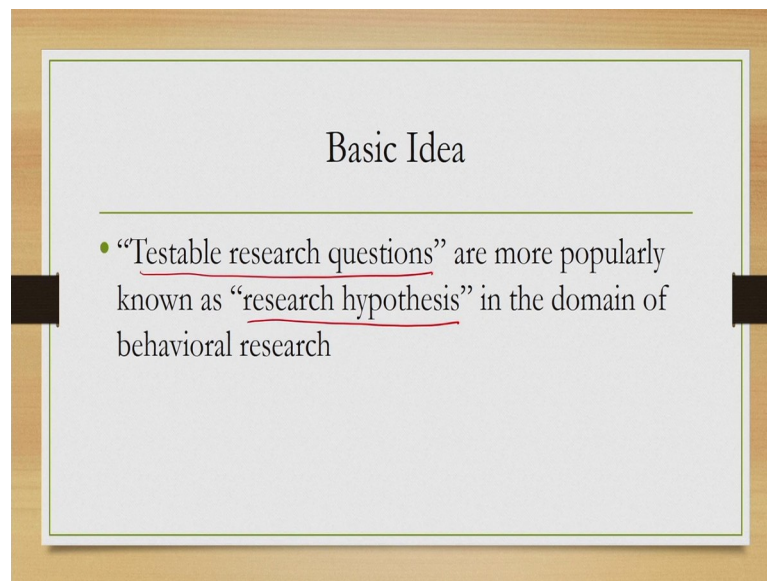


And, this is certainly a better approach than having only untestable question and user feedback. Why this is so? Suppose you are asked to collect data for determining how good your system is and you simply went to the user asked for their opinion. Now, some say it is good some say it is bad. Since, there is a positive correlation between untestable and testable questions; you may think that instead of going for multiple testable questions, we can simply try to get the answer of untestable questions, by going to the users asking for their feedbacks.

Now, the problem is that the feedback that you may get from the users need not be quantifiable, measurable, testable and unbiased. So, the reliability of the feedback is questionable and there is no way we can test the reliability.

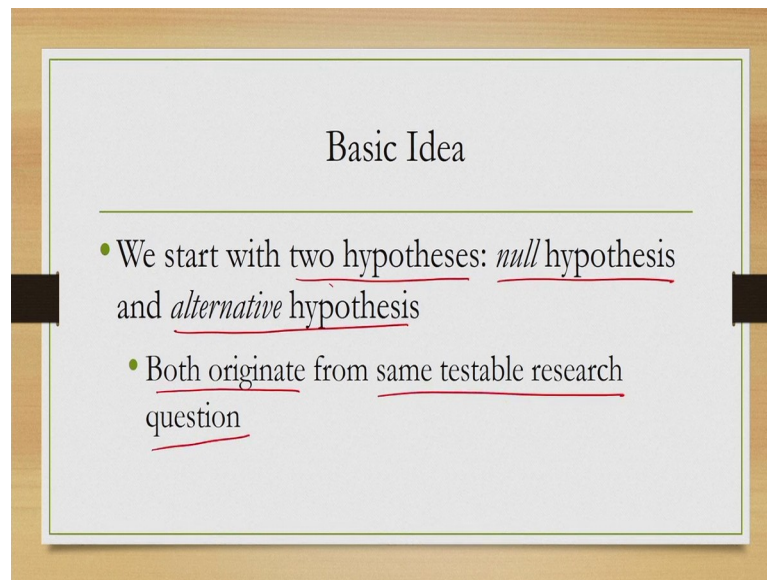
In the case of testable questions whatever data we collect, they are quantifiable and we can test their reliability through statistical means, based on that we can come to a reliable conclusion. And, we can use that conclusion, which are specific in nature to come to a generalized reliable conclusion. So, that is on the idea of testable and non-testable research questions. There is another closely related concepts that are used in empirical research and it is important to know the terminology and the difference between that other concept and this research question concept.

(Refer Slide Time: 44:54)



So, a term which is more popularly used to indicate testable research questions is called the research hypothesis. So, are they same or the term hypothesis and the term research questions are different.

(Refer Slide Time: 45:10)



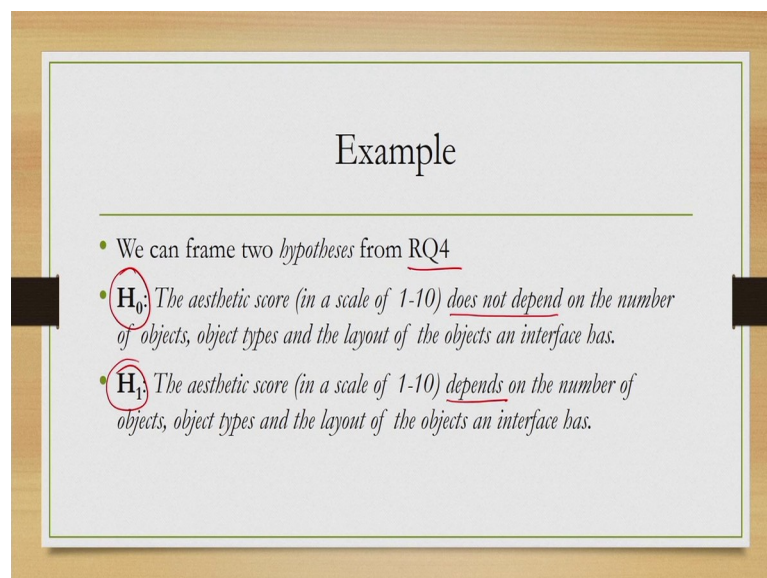
Basic Idea

- We start with two hypotheses: null hypothesis and alternative hypothesis
- Both originate from same testable research question

Let us try to understand when we start our empirical study, we actually are supposed to start with two hypothesis; one is called the null hypothesis and one is called the alternative hypothesis.

Now, both these hypothesis originate from the same testable research question; that means, earlier we said we start with a testable question, now what we are saying we start with a couple of hypothesis, one is the null hypothesis, one is the alternative hypothesis. Now, both these hypothesis come from the same testable question.

(Refer Slide Time: 45:48)



Example

- We can frame two hypotheses from RQ4
- H_0 : *The aesthetic score (in a scale of 1-10) does not depend on the number of objects, object types and the layout of the objects an interface has.*
- H_1 : *The aesthetic score (in a scale of 1-10) depends on the number of objects, object types and the layout of the objects an interface has.*

For example consider RQ 4, which we have already seen before. From this research question we can frame two hypotheses H 0 and H 1, what H 0 states the aesthetic score in a scale of 1 to 10 does not depend on the number of objects object types and the layout of the objects an interface has. And, what H 1 states the aesthetic score in a scale of 1 to 10 depends on the number of objects, object types and the layout of the objects an interface has.

So, H 0 states that the score does not depend and H 1 states that the score depends.

(Refer Slide Time: 46:34)

Example

- We are no longer posing any question - '?' at the end is gone
- Apart from that, there is one important difference - a single question gave rise to two hypotheses.

So, what is the difference between the research question 4 or RQ 4 and these two hypotheses? First of all we are no longer posing any question. So, the question mark at the end is gone apart from that the other difference that we have already mentioned is that the single question RQ 4 gave raise to two hypotheses.

(Refer Slide Time: 46:59)

The slide is titled "Example" and is set against a light gray background with a thin green border. It contains two bullet points. The first bullet point discusses the null hypothesis (H₀), stating it is essentially stating that the test condition is not going to affect the outcome (judgment). A sub-bullet point notes that it is typically opposite to what we set out to establish (effect of the test condition on the observations). The second bullet point discusses the alternative hypothesis (H₁), stating it is just the opposite - we are stating that test condition does affect outcome.

In H₀ we are essentially stating that the test condition is not going to affect the outcome or our judgment behavior, which is called the null hypothesis. So, the null hypothesis is typically an hypothesis which states the opposite to what we set out to establish. And, the H₁ is called the alternative hypothesis, which is the opposite of the null hypothesis, that is here we typically state that the test condition affect the outcome.

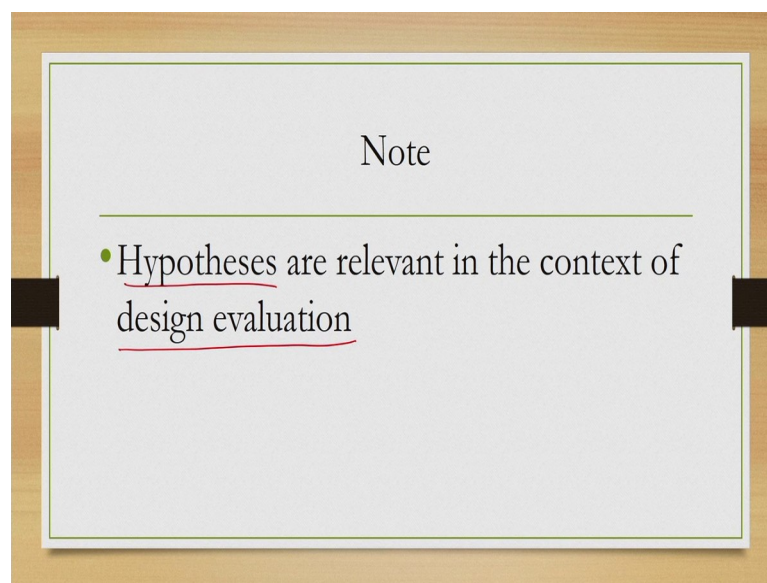
(Refer Slide Time: 47:35)

The slide is titled "Note" and is set against a light gray background with a thin green border. It contains one bullet point stating that in an empirical research, we aim to find statistical evidence to refute or nullify null hypothesis and support alternative hypothesis.

So, in an empirical research, what is our objective to find statistical evidence to refute or nullify the null hypothesis and support alternative hypothesis that is the crux of the matter.

So, when we set out to perform an empirical research we start with two hypotheses, which originate from same testable research questions, our objective is to refute the null hypothesis, because it is the opposite of what we want to establish and support the alternative hypothesis through statistical means, through statistical evidence.

(Refer Slide Time: 48:20)



However, it should be noted that the concept of hypotheses are relevant only in the context of design evaluation. So, when we are concerned about building a model, the idea of hypotheses is not suitable, because of the nature of its formulation. As, we have seen before when we formulate an hypothesis, we are not interested about the relationship between quantities.

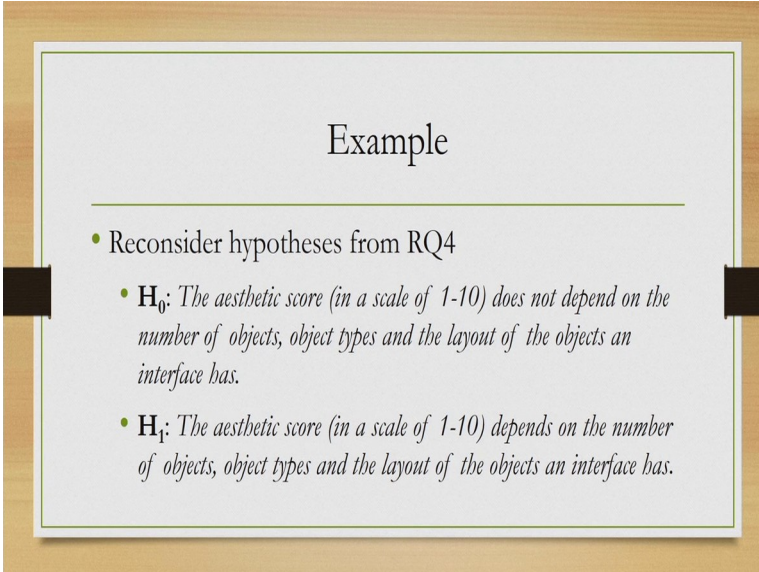
So, the question mark and the use of the words like how they are related these are no longer used. So, then what purpose it serves, it is an essential precondition before we can go for modelling a relationship with hypothesis we establish scientifically. The existence of a relationship unless we establish the existence we cannot build the relationship.

So, it is not that we directly collect some data and build the relationship. Before, that we have to actually establish that this relationship exist and there the hypothesis formulation

and testing plays an important role. So, first we formulate hypothesis on the relationship, which essentially states that there exists some relationship once that existence of the relationship is established by supporting the alternative hypothesis like, in the case of the two hypotheses we framed from RQ 4, we can use the same data to go for building the model.

Once, we do that the scientific basis of the model increases and we can be more certain that the model actually represents the behavior. There exists some relationship which we have established through hypothesis testing and based on that testing outcome we went for building a model which essentially captures that relationship. So, this is a more scientific approach, then directly going for building a model with empirical data. And, that is what we are interested in doing in empirical study.

(Refer Slide Time: 50:27)



Example

- Reconsider hypotheses from RQ4
 - H_0 : The aesthetic score (in a scale of 1-10) does not depend on the number of objects, object types and the layout of the objects an interface has.
 - H_1 : The aesthetic score (in a scale of 1-10) depends on the number of objects, object types and the layout of the objects an interface has.

So, whenever we want to conduct an empirical research to answer some research question or to build some model, first thing is we should start with hypothesis test the validity of alternative hypothesis and then use the data to build the model.

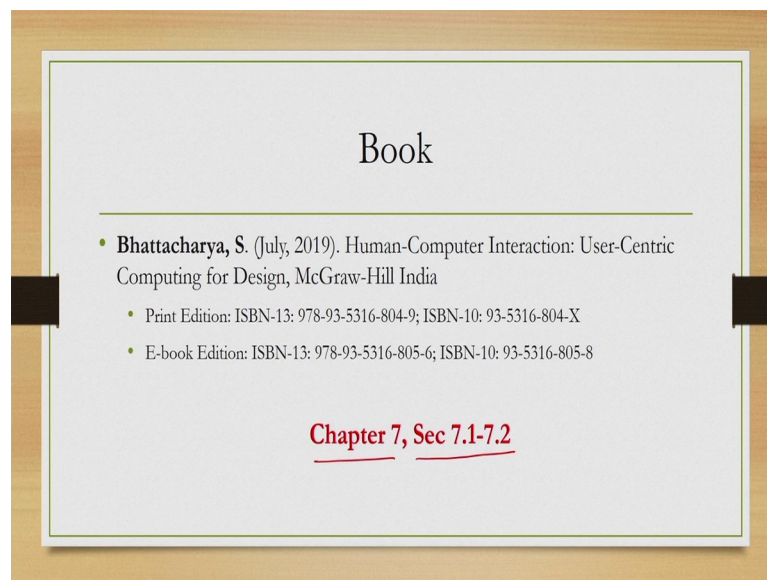
So, today we have discussed the first stage of the empirical research process namely formulation of research question. So, we have gone through the concepts of testable research questions, non testable or untestable research questions, the tradeoff between these 2, the idea of internal and external validities of these questions, and how to balance

this tradeoff namely by posing multiple testable research questions and concluding based on the conclusions of those research questions.

Based on the conclusion drawn from the data, based on the conclusions drawn from the data collected by utilizing those multiple testable research questions. Also, we have clarified on the conceptual difference between the two terms research questions and hypothesis, irrespective of what our objective is for empirical study, whether to evaluate a system or to build a model, we should always start with hypothesis, which results from a testable research question. So, our first objective is to frame a testable research question from there we frame to hypothesis null and alternative hypothesis, then we find statistical evidence to support the alternative hypothesis, based on that evidence we go for building models.

If, our ultimate objective is to build models, using the same empirical data; however, if our objective is just to evaluate a system or to find evidence for supporting alternative hypothesis, then we stop there after the evidence is found.

(Refer Slide Time: 52:54)



The material that is covered today is taken from the same book that we are following you are advised to refer to chapter 7, section 7.1 and 7.2 to find more details on these topics. In the next lecture, we are going to discuss about the other stages of the empirical study till then.

Thank you and good bye.