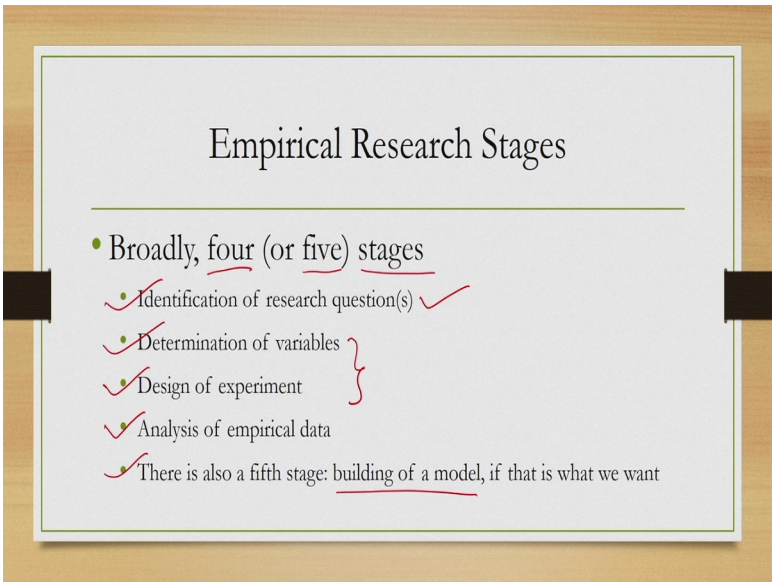
User-Centric Computing for Human-Computer Interaction
Prof. Samit Bhattacharya
Department of Computer Science and Engineering
Indian Institute of Technology, Guwahati

Lecture - 27
Variables determination and experiment design

Hello and welcome to lecture number 27 the course User-Centric Computing for Human-Computer Interaction.

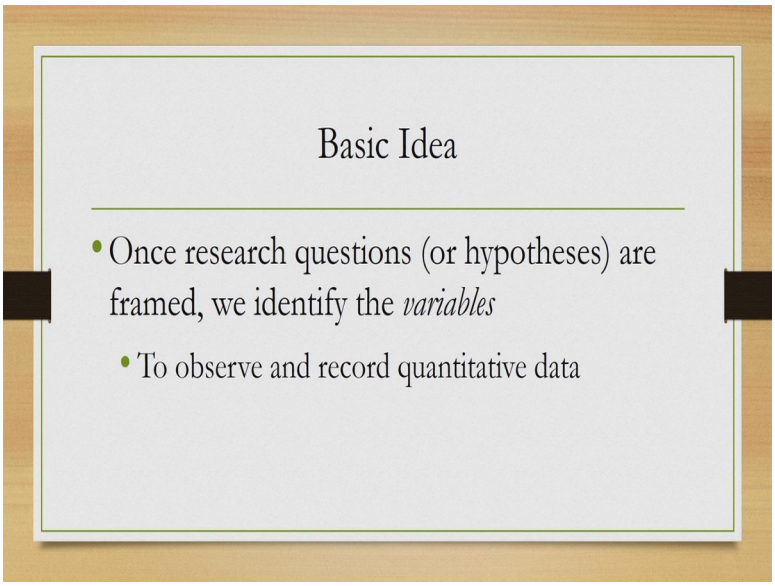
(Refer Slide Time: 00:53)



So, we started our discussion in the last lecture on the topic of empirical research. We have seen there that there are several stages in the process of empirical research. Broadly there are four stages and if we are concerned about building a model out of the empirical data then, we can consider there to be five stages.

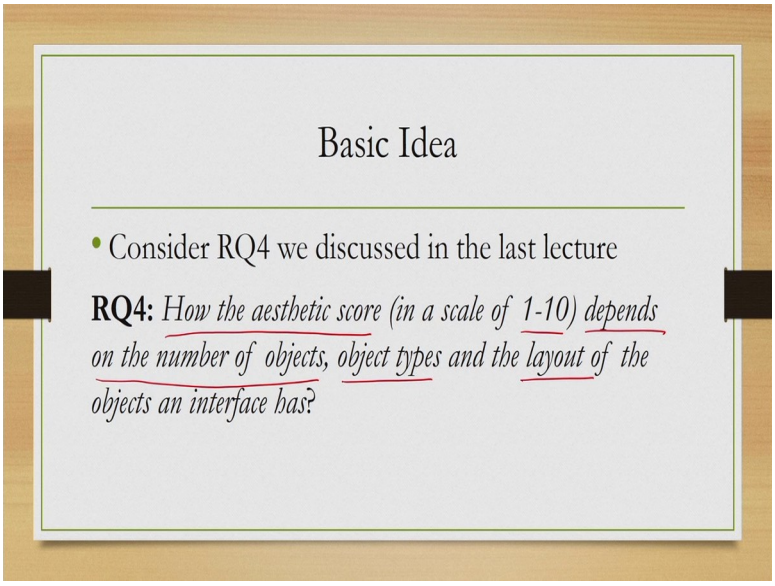
First stage is identification of research question, second stage is determination of variables, third stage is experiment design, fourth stage analysis of the data to come to a conclusion, and the fifth stage is about model building; building of a model. Among them in the last lecture we have covered the first stage that is identification of research questions. In this lecture we are going to cover the two stages; second and third namely variable determination and experiment design.

(Refer Slide Time: 01:52)



So, in order to understand the idea of variables and determination of variables, in an empirical research, let us start with an example. If you may recollect in the previous lecture, we framed few research questions for the study on building a relationship between our aesthetic judgment behavior and an interface.

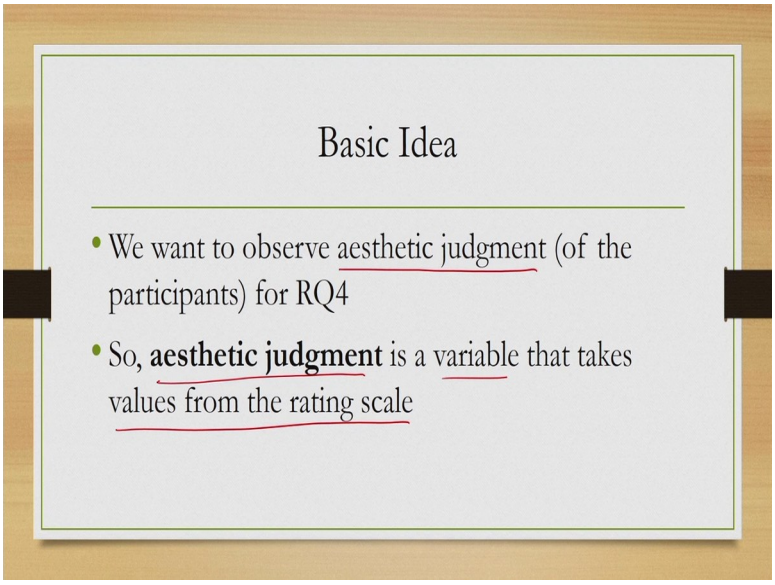
(Refer Slide Time: 02:26)



So, we have framed one research question that we termed RQ 4 or research question 4 for the benefit of recollection; we will reproduce the question here, but with respect to that question let us try to understand the concept of variables and how to identify them.

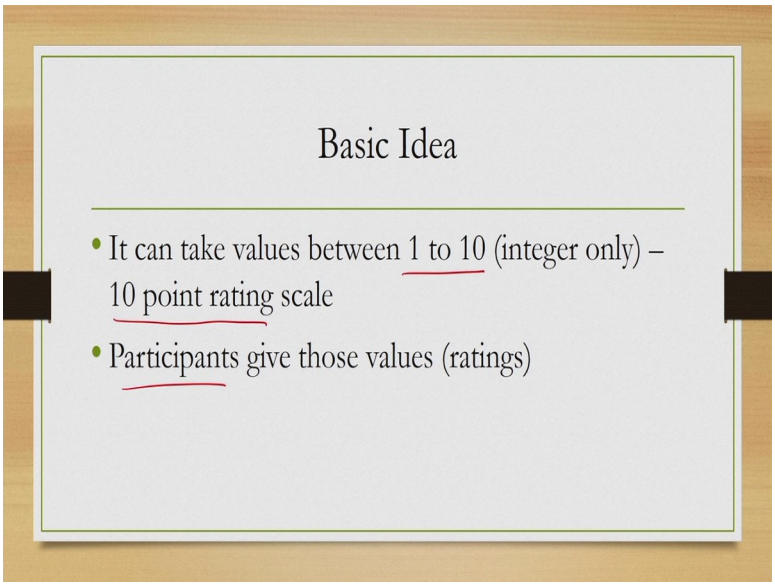
So, what was the research question? How the aesthetic score in a scale of 1 to 10 depends on the number of objects, object types, and the layout of the objects on the interface?.

(Refer Slide Time: 03:06)



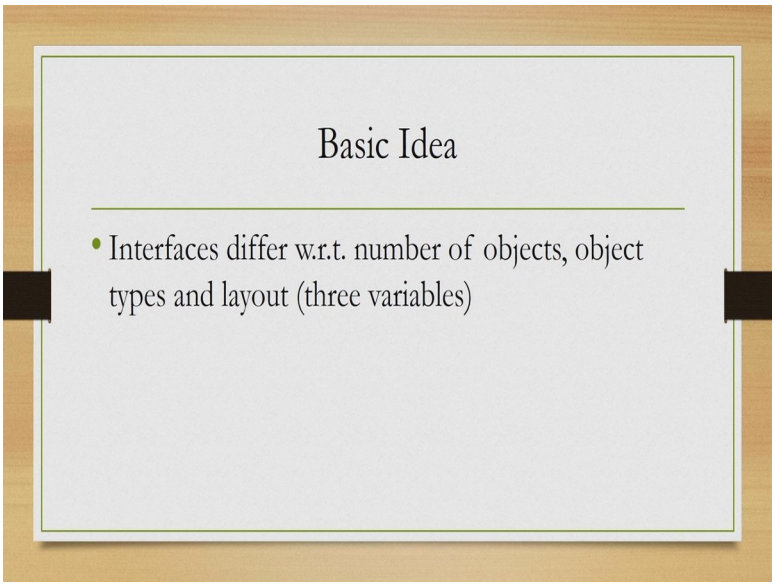
Now, if we have this research question, then what this question tells us. First of all we want to observe ratings. Why we want to observe rating, because we want to observe aesthetic judgment and we are assuming that ratings give us that judgment behavior. So, in other words aesthetic judgment is a variable that takes values from the rating scale. So, this is one variable which is there in the research question.

(Refer Slide Time: 03:45)



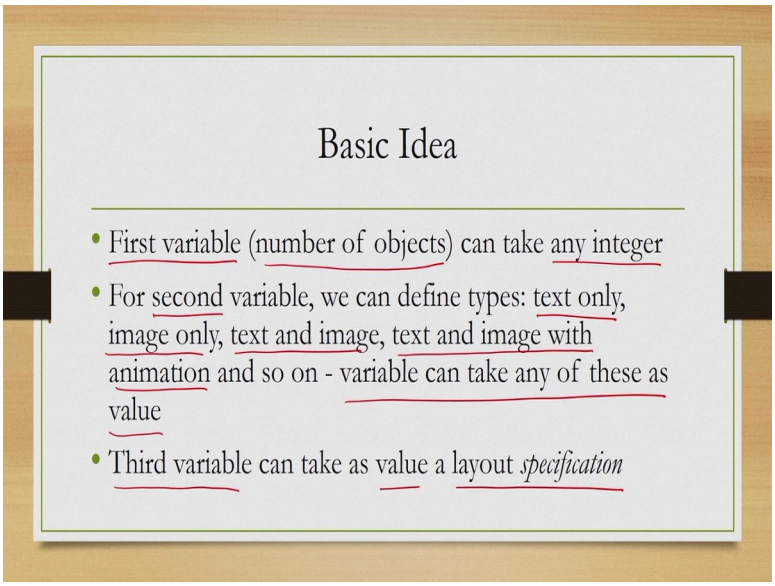
And, this variable aesthetic judgment can take values between 1 to 10, because we are using a 10 point rating scale and from where these values come from the participants. So, this is one variable.

(Refer Slide Time: 04:08)



Let us see, what other variables are there in the research question? Now, the interface we have defined in terms of three quantities; number of objects, object types and interface layout.

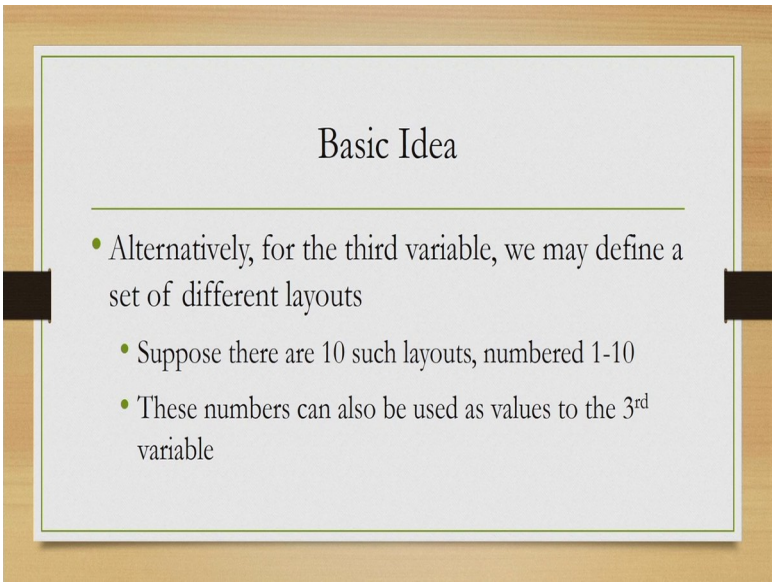
(Refer Slide Time: 04:29)



Now, the three quantities represent three variables. The first variable number of objects, which can take any integer as it is value because the number can be any integer. For the second variable, we can define object types. For example, text only, image only, text and image, text and image and animation and so on.

Each of these represent type. So, for example, one interface can have only texts. So, the object types are texts, an interface can have texts and images both and so on. So, each of these types can work as a value assigned to the second variable. So, the variable can take any of these as value. And, in the third variable namely the layout the value can be a specification of the layout in terms of the position and size of the objects present on the interface.

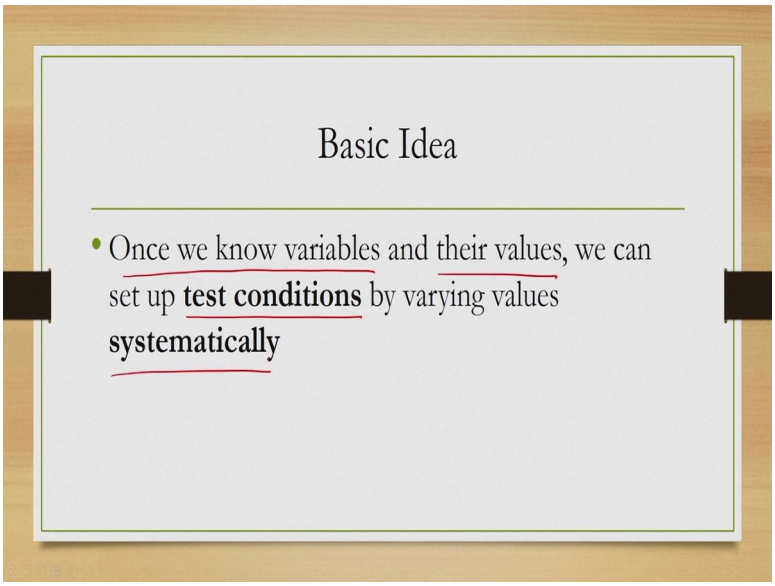
(Refer Slide Time: 05:45)



Now, the specification is a complex thing in order to specify a layout, we require to specify the object dimensions, object positions, maybe object colors and many such things, which makes the specification complex. Instead what we can do is we can define a set of layouts, each having different specification, assign them some numbers, suppose we work with 10 such layouts.

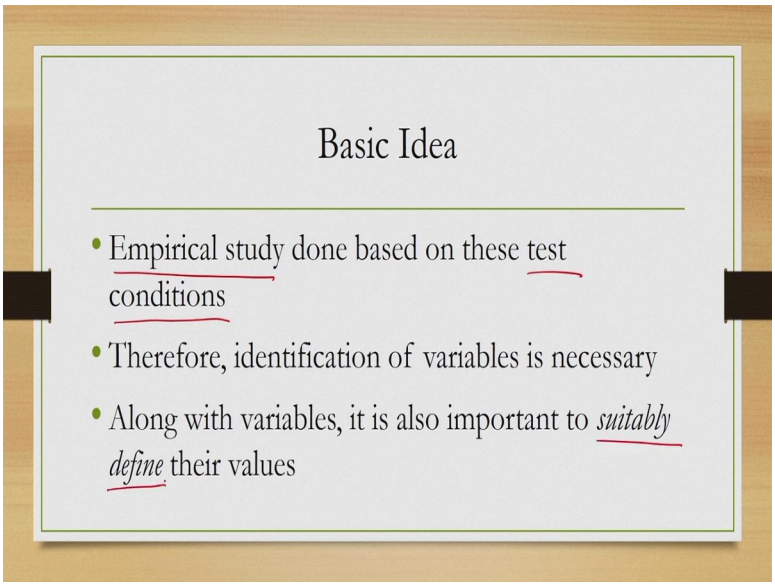
So, in that case layout 1 indicates as particular specification, layout 2 indicates another specification, which is different from 1 and so on up to layout 10. So, in that case the 3rd variable can be assigned a value between 1 to 10 or the layout numbers. Each number is indirectly pointing to a specific layout, this is another way of specifying the values.

(Refer Slide Time: 06:48)



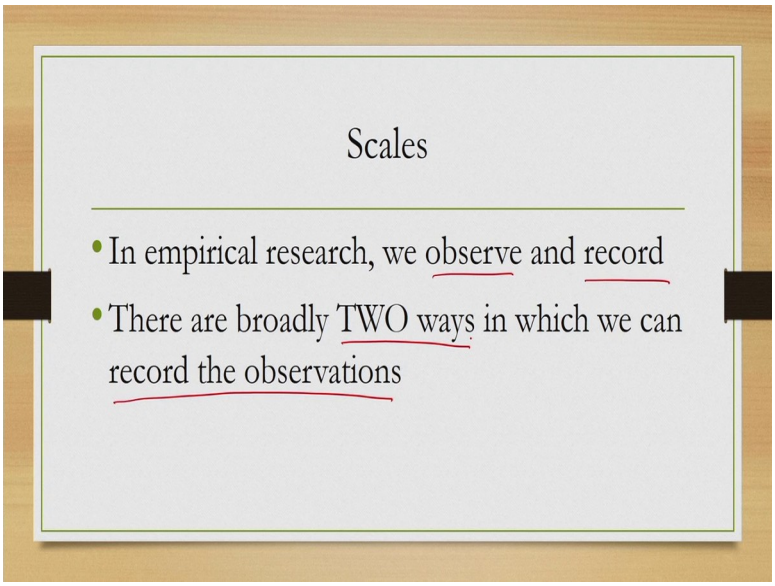
So, once we know what are the variables and what values they can take, then we can set up test conditions by varying these values systematically. This is very important in empirical research.

(Refer Slide Time: 07:04)



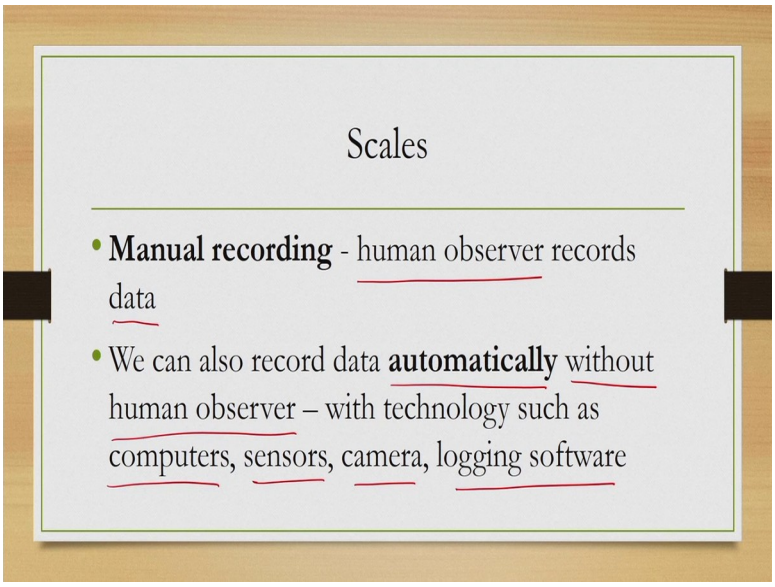
Because, empirical study is done based on these test conditions only so, we have to first define the test conditions otherwise we will not be able to conduct the study. Along with that it is also important to suitably define their values. So, one is identification of the variables second is defining their values so, that we can set up proper test conditions.

(Refer Slide Time: 07:33)



Now, what we do in empirical study, we observe and record. In order to record our observations, we can take recourse to either of two ways; one is manual recording, other one is automatic recording.

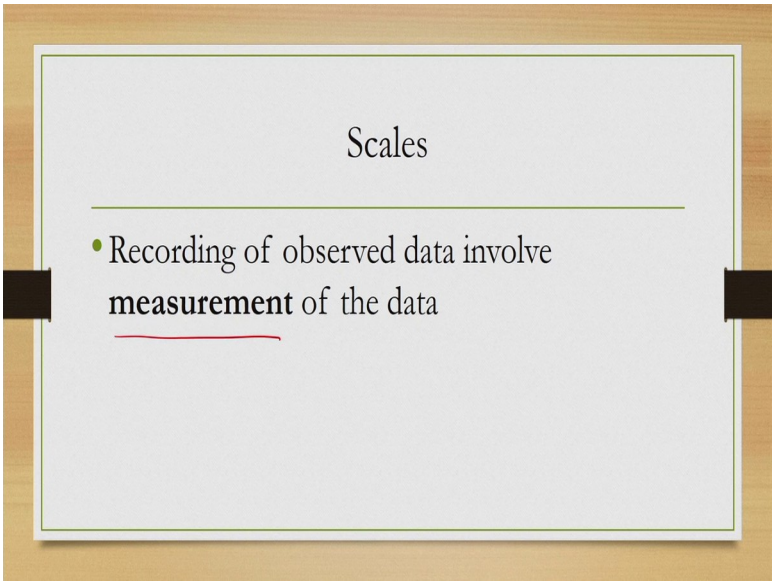
(Refer Slide Time: 07:50)



Now, in manual recording a human observer is present and records the data, the observation whereas, in case of automatic recording there is no human observer. So, we perform the recording without human observer and with the aid of technological tools

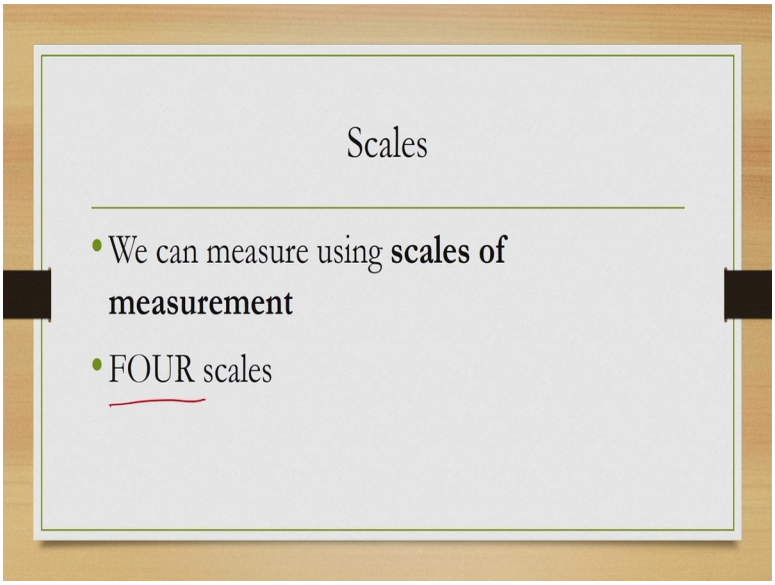
such as computers, sensors, camera, logging software and so on. These are utilized to record the data that we observe.

(Refer Slide Time: 08:25)



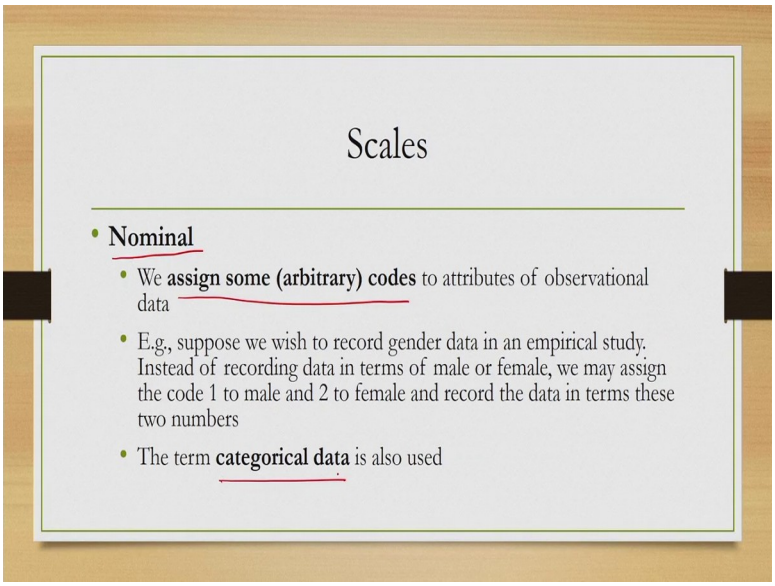
Now, recording of observed data involve measurement unless we are able to measure our data we will not be able to record it. So, measurement plays a very important role in empirical study, our ability to measure is very important.

(Refer Slide Time: 08:48)



And, we use scales of measurements for the purpose. There are broadly four such scales available which you can make use of during your study.

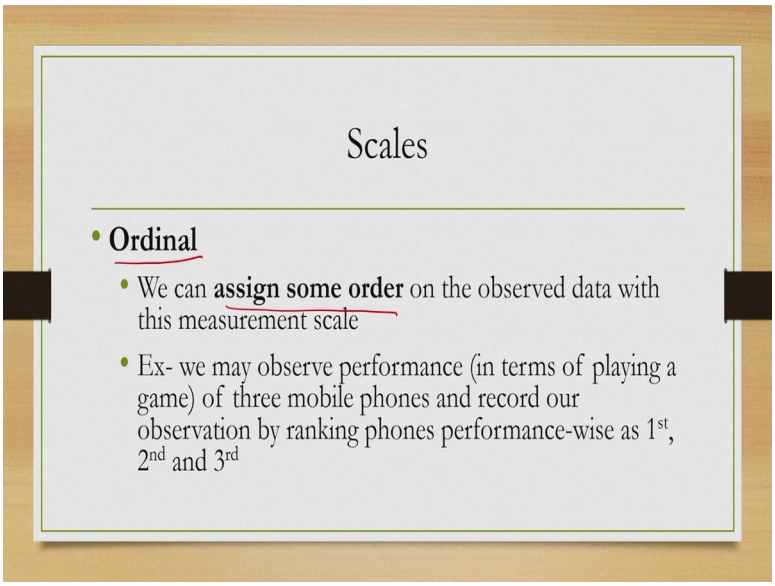
(Refer Slide Time: 09:02)



Let us see those scales one by one the first scale is nominal. So, in this case when we are using a nominal scale of measurement, what we do? We assign some arbitrary codes to attributes of the observational data. For example, suppose we wish to record gender data in an empirical study. Now, instead of recording data in terms of male or female, we may simply assign the code 1 to male and 2 to female and record the data in terms of these two numbers.

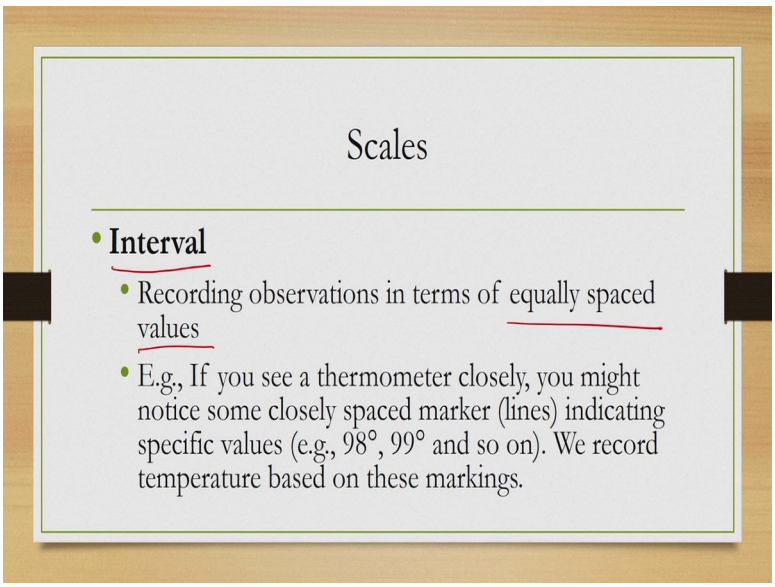
So, what we are doing here instead of using the actual gender, we have assigned some code, that is 1 to male and 2 to female and then record data against these codes, this is some arbitrary assignment of codes to the attributes of data items. So, this is a nominal scale of measurement we are using. Often we use the term categorical data also to represent this nominal measurement scales.

(Refer Slide Time: 10:07)



Second scale of measurement is ordinal scale. Here, we assign some order to the data items we observe. For example, we may observe performance of a set of phones in terms of playing of a video game and then we can order their performance as 1st, 2nd, 3rd and so on. So, this is putting some order on the observations and this is called ordinal scale of measurement.

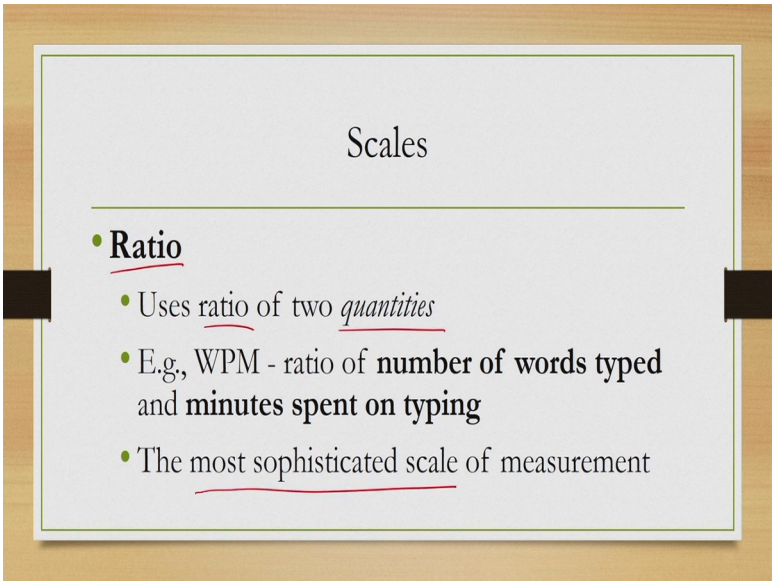
(Refer Slide Time: 10:42)



The third scale is interval scale. Here, what we do we record observations in terms of equally spaced values. For example, consider a thermometer. We all probably have seen

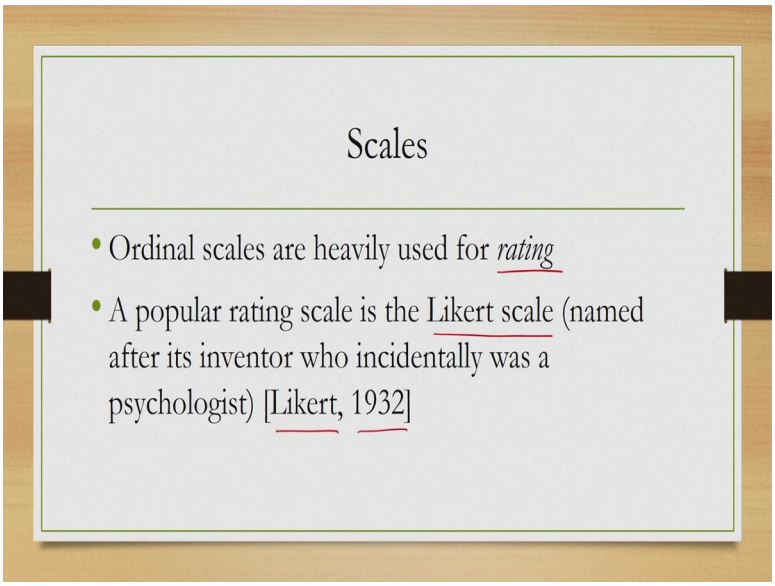
it or made use of it, in order to record body temperature. Now, in thermometer there are markings, which are separated by a value which is equal for all the markings. This is an example of a interval scale of measurement with equal spacing and we use it to record our body temperature, using that particular scale of measurement. This type of scale is this type of scales is known as interval scale.

(Refer Slide Time: 11:33)



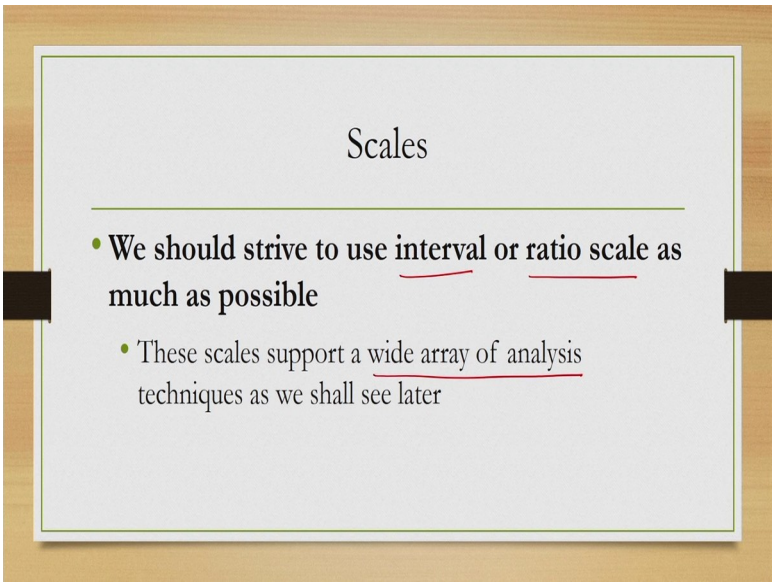
And, finally, there is the fourth scale that is a ratio scale, which uses a ratio of two quantities. For example, consider the measurement of typing speed in terms of Watts Per Minute or WPM. Now, this is a ratio scale of measurement. Here, we are representing the observed data in terms of ratio of 2 numbers namely the number of words typed and the amount of time elapsed. It is often considered the most sophisticated scale of measurement among all the four scales.

(Refer Slide Time: 12:10)



Now, among these scales namely the nominal, ordinal, interval and ratio, the ordinal scales are heavily used for rating purpose. And, one very well-known rating scale is the Likert scale, which is named after it is inventor Likert who invented it in 1932.

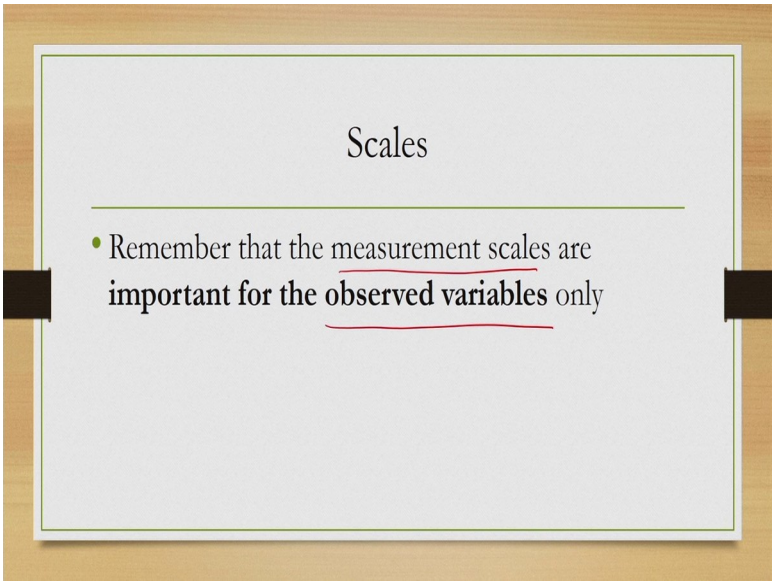
(Refer Slide Time: 12:35)



Now, our objective in any recording of data is to use either interval or ratio scales, because when we record data using those scales? The data actually can be used for performing a wide array of analysis which is not possible, if we record data in terms of

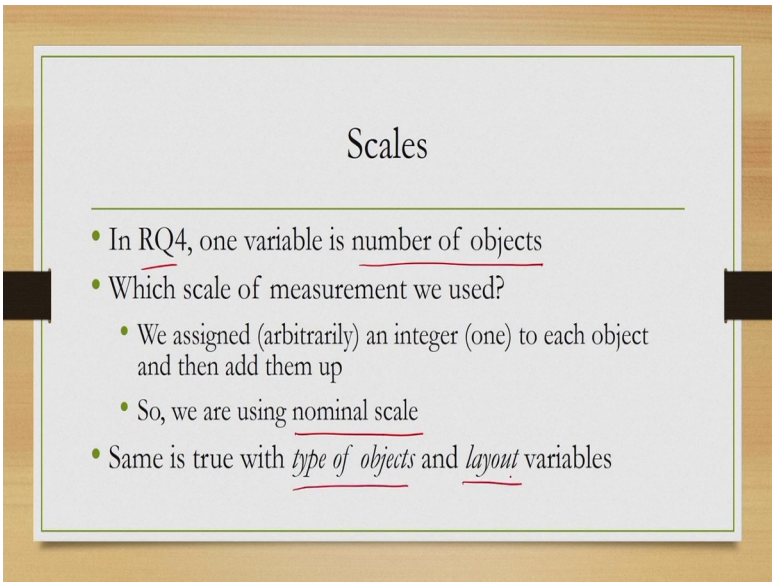
ordinal or nominal scales. We will soon see how to use these data values for analysis in subsequent lecture.

(Refer Slide Time: 13:13)



One thing we should note here. So, when we are talking of this measurement scales and our intention to use this particular scale of measurement, that is actually important when we are concerned about variables, for which we observed their values or for the observed variables only. Let us try to understand this in terms of an example.

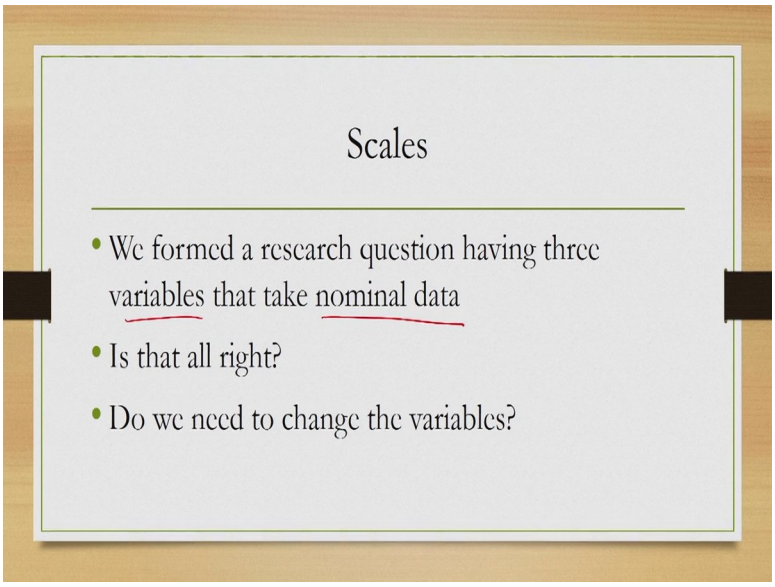
(Refer Slide Time: 13:39)



In RQ 4 that we have just seen there is one variable, which is the number of objects it is related to the definition of the interface. So, for number of objects which scale of measurement we used, remember that we assigned arbitrarily an integer to each object and then add them up. So, if an interface contains say two text items, one image items, then we arbitrarily assigned number one to one text item, number 2 to another text item.

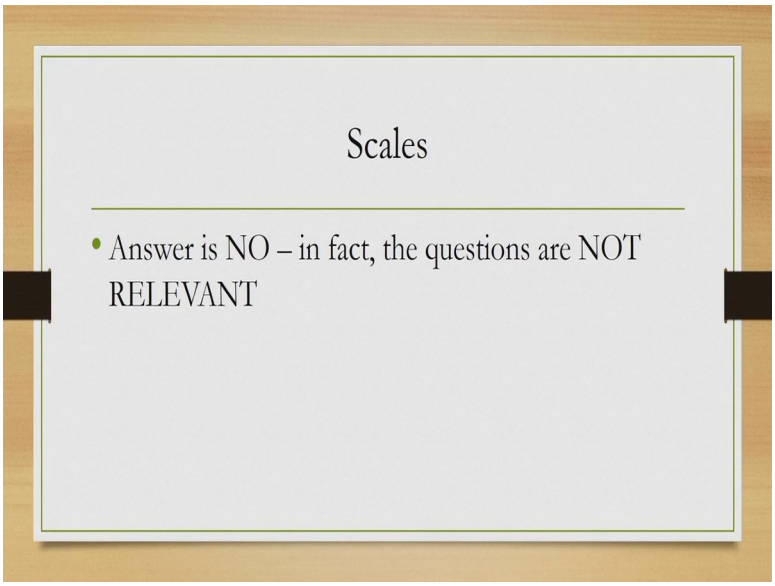
We arbitrarily assigned number one to one text item, number one to another text item, number one to the image item, and then added them up to come to the value of 3 objects in that interface. So, this assignment is arbitrary each object irrespective of its type has been assigned the same integer value and then we added them up, clearly we have used a nominal scale. In a similar line of argument we can find out that the variables type of objects and layout have also used nominal scales.

(Refer Slide Time: 14:55)



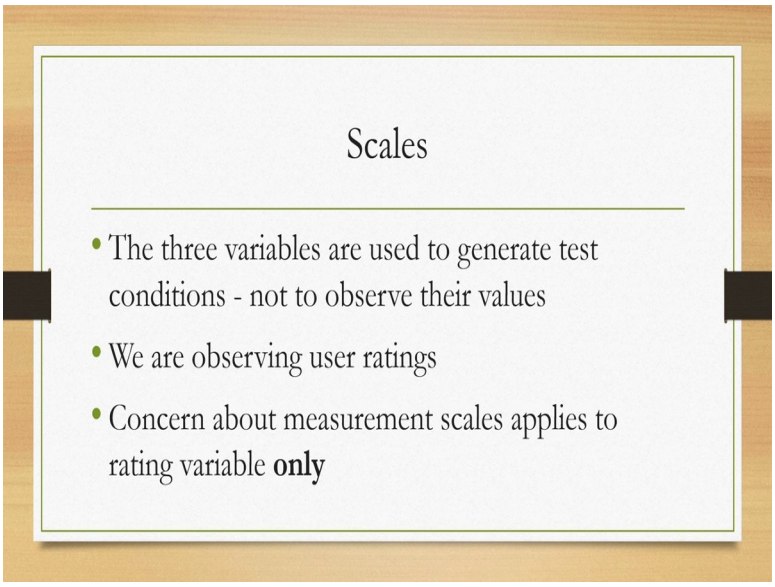
So, we formed a research question having three variables that take nominal data. Now, you may wonder whether that is the right way to proceed, whether we are doing the right thing.

(Refer Slide Time: 15:08)



In fact, this question is not at all very relevant.

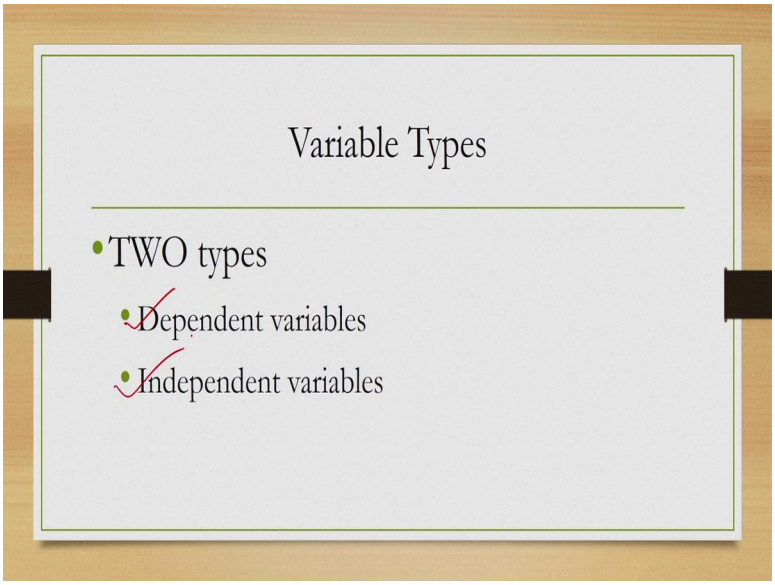
(Refer Slide Time: 15:15)



Why, because when we are talking of variables that we should be able to define in terms of interval or ratio scale of measurement, we are actually referring to variables whose values we observed. Now, in the case of the three variables we just discussed namely number of object types and layout. We do not observe their values instead we set their values to observe the overall value of the aesthetic judgment that is the rating. So, the rating variable is actually what we should be bothered about, rather than these other three

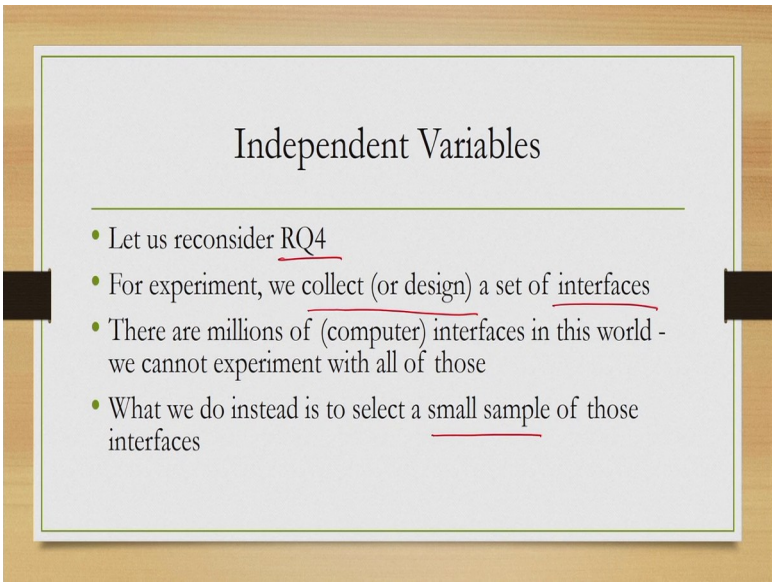
variables. So, this brings us to the concept of dependent and independent variables. Let us try to understand, what are these variables and what is the difference between them.

(Refer Slide Time: 16:05)



So, when we are talking of variables broadly there are two types of variables; dependent variable and independent variable.

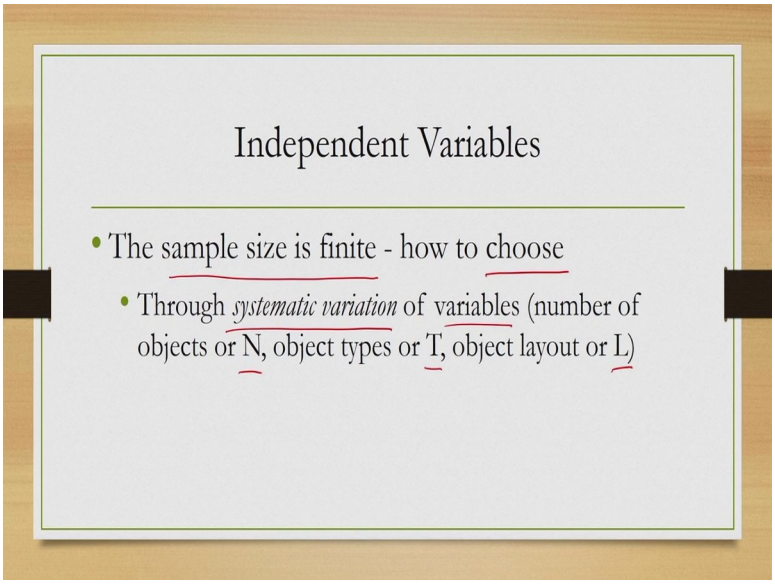
(Refer Slide Time: 16:14)



In order to understand these variable types, let us reconsider RQ 4 or research question four, that we have mentioned sometimes back. So, now, in order to perform the experiment, we have to decide or design on a set of interfaces. Now, the point is the

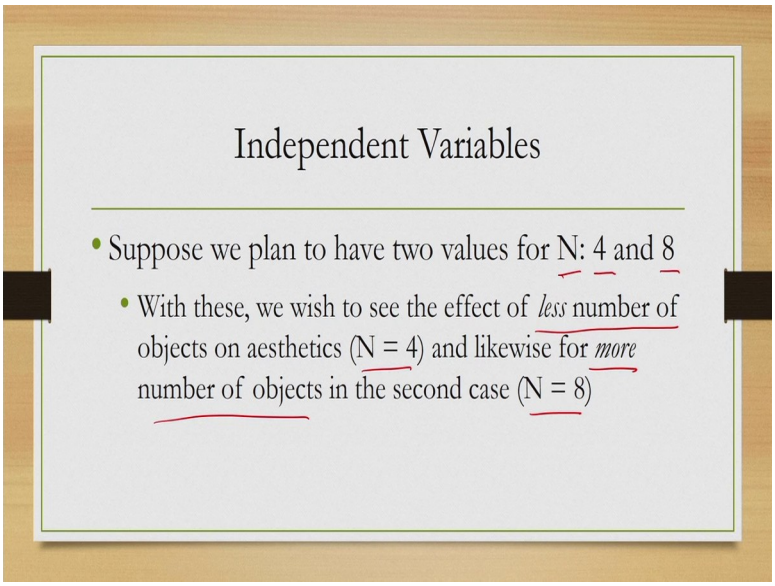
number of possible interfaces can be very large. And, of course, we cannot experiment with all possible interfaces. So, what we can do is we can select a small sample of that large number of interfaces possible and we work with that small sample only.

(Refer Slide Time: 16:55)



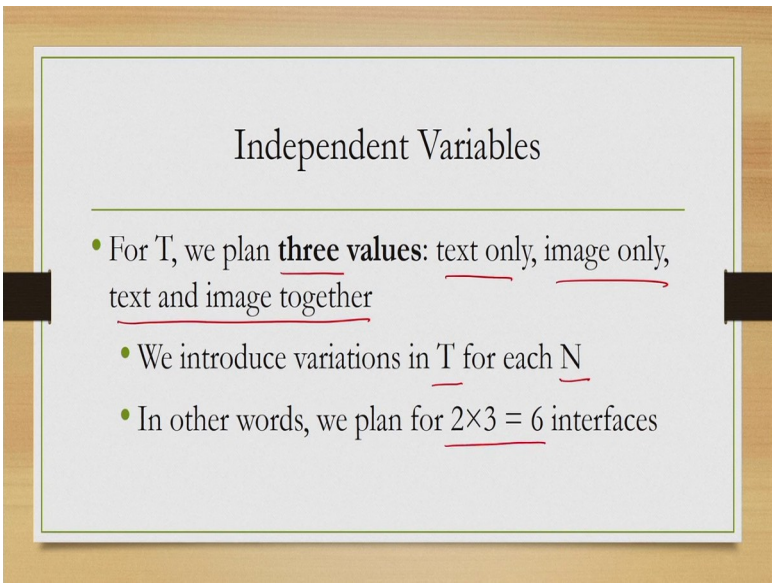
Now, this sample size is finite. So, you are asked to choose a set of interfaces from a very large set and how to choose, how to make the choice that is the question. So, what we do is we systematically vary the variables to make the choice. Now, what are the variables here number of objects N, object type or T, and object layout or L. So, these variables we systematically vary to choose a small set of interfaces as sample.

(Refer Slide Time: 17:35)



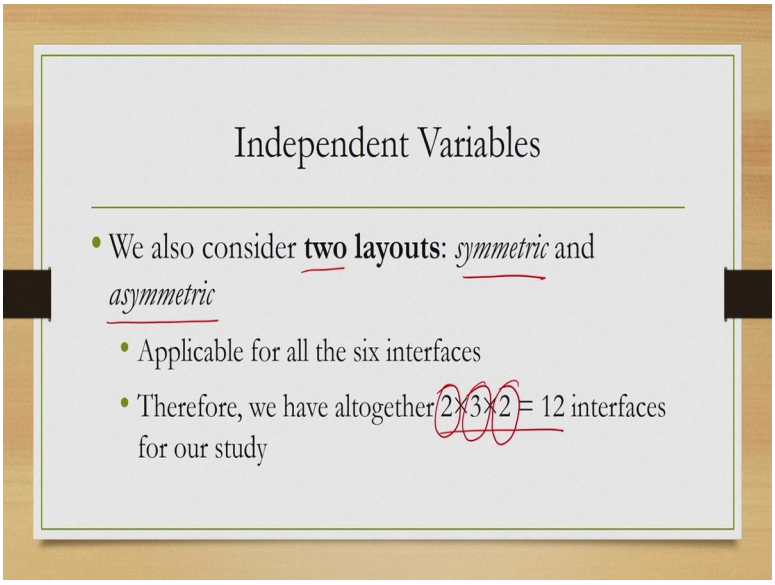
Now, let us consider two values for N; 4 and 8; that means, one interface with 4 objects and another interface with 8 objects. Now, with these we used to see the effect of less number of objects on aesthetics when N equal to 4 and likewise we wish to see the effect of more number of objects on aesthetics when N equal to 8. So, our logic for choosing these two values is that one value serves as an indicator of a less number of objects and the other value serves as an indicator of more number of objects. And, we want to check the effect of these more or less numbers on the overall aesthetic judgment.

(Refer Slide Time: 18:22)



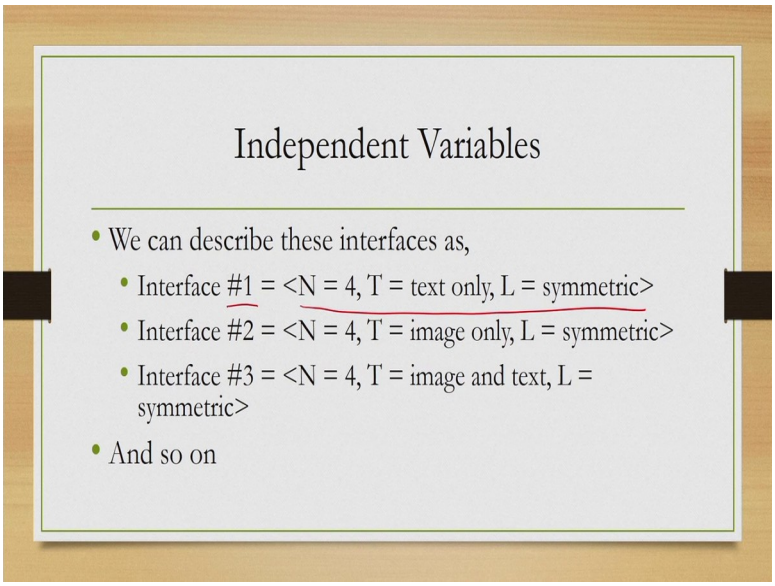
Similarly, for T, let us assume that we are dealing with three values text only, image only, and text and image together. So, for each value of N, we have three values of T. So, total number of interfaces, if we consider these variations of N and T becomes 6.

(Refer Slide Time: 18:50)



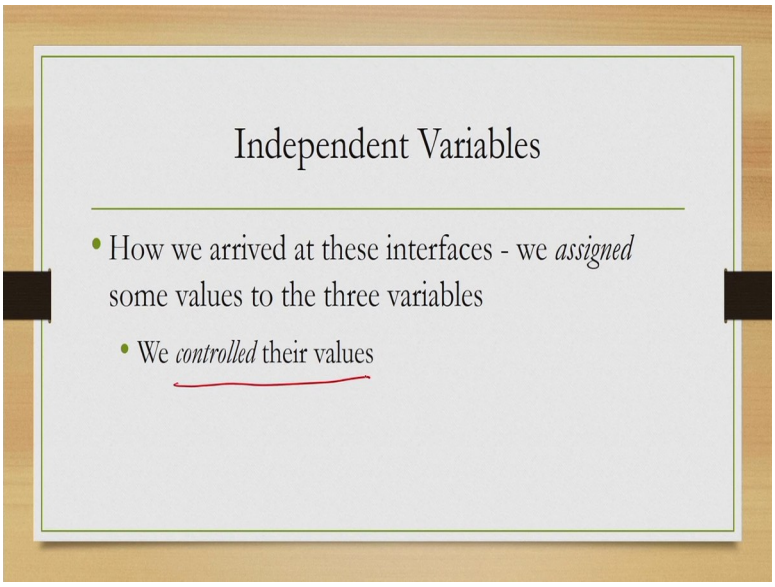
We still have one variable left that is L; now let us assume two layouts for this variable. So, the value it can take is from either of these two values symmetric and asymmetric. So, either it the layout can be symmetric with respect to some access or asymmetric with respect to the same access. So, then we have total 12 interfaces; 2 for N for each of these 2 3 values for T and for each of these T value, we have 2 layout values.

(Refer Slide Time: 19:33)



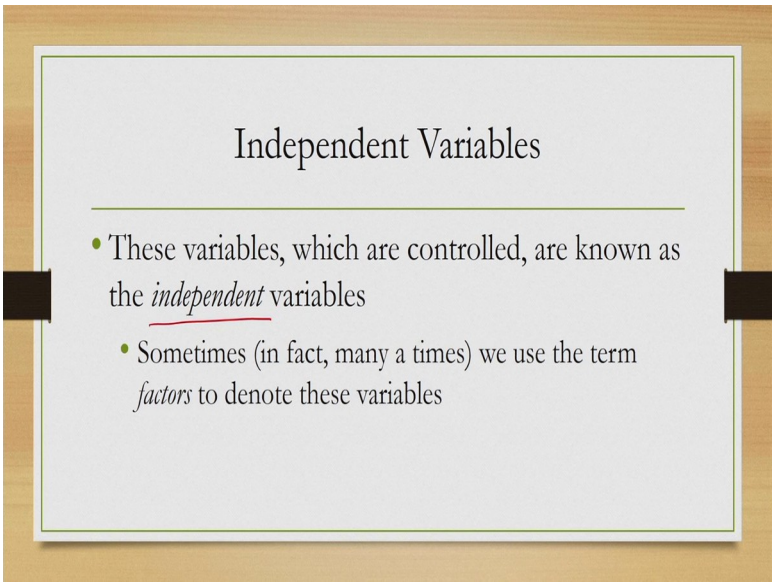
So, total 12 interfaces. And, we can use a convention to represent these different interfaces. Namely, number one in terms of this convention N equal to 4 T equal to text only and L equal to symmetric. Similarly, we can define all the 12 interfaces.

(Refer Slide Time: 19:51)



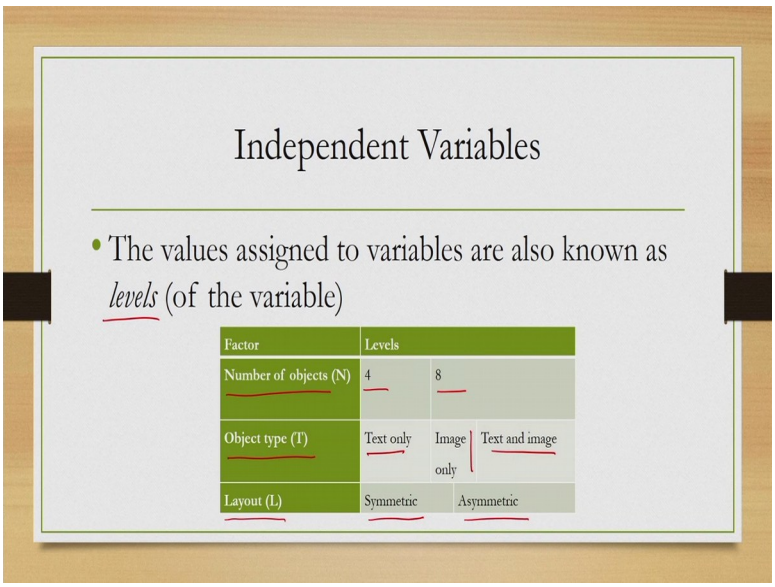
So, how we arrived at these values by putting some systematic effort in varying the variables? Now, when you say we varied the variables; that means, the variation is in our hand. How to vary these variables or in other words we controlled these variation or we control these variables, we controlled the values that these variables take.

(Refer Slide Time: 20:18)



So, when we control the values of a variable that variable is called independent variable. Another term we use to indicate independent variable is factor.

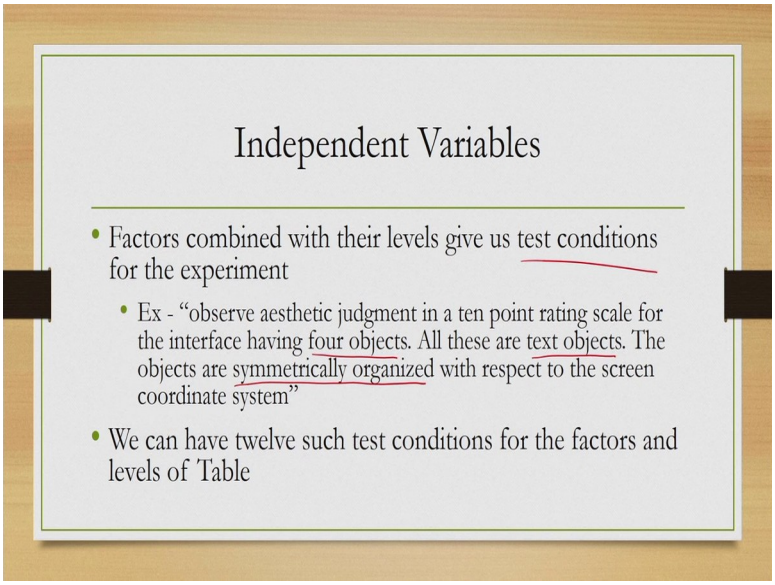
(Refer Slide Time: 20:31)



And, the values that are assigned to independent variables are also known as levels. So, in our example our factors are number of objects with 2 levels, 4 and 8 object type is another factor having 3 levels text only, image only and text and image. Layout is the third factor which has 2 levels symmetric and asymmetric. So, when we talk off variables; one is independent variable, that is the variables for which we control their

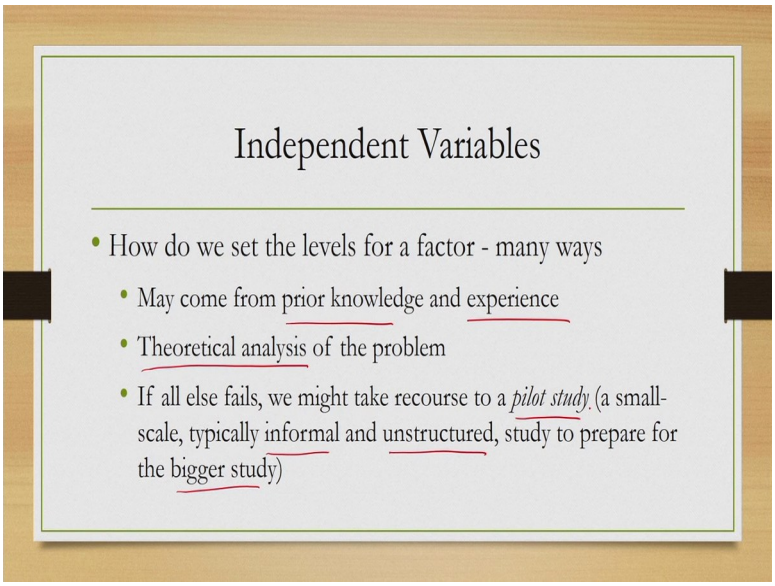
values, these are also known as factors and the values that they take are sometimes called levels.

(Refer Slide Time: 21:22)



Now, when we combine factors with their levels we get test conditions. Example, "observe aesthetic judgment in a ten point scale for the interface having four objects. All of these are test objects. And, the objects are symmetrically organized. So, this is one text condition example. And, this test condition you obtained by combining the values of the independent variables. And, as we have said if we combine all the values of the variables, that we have set, then you will get 12 such test conditions.

(Refer Slide Time: 22:10)

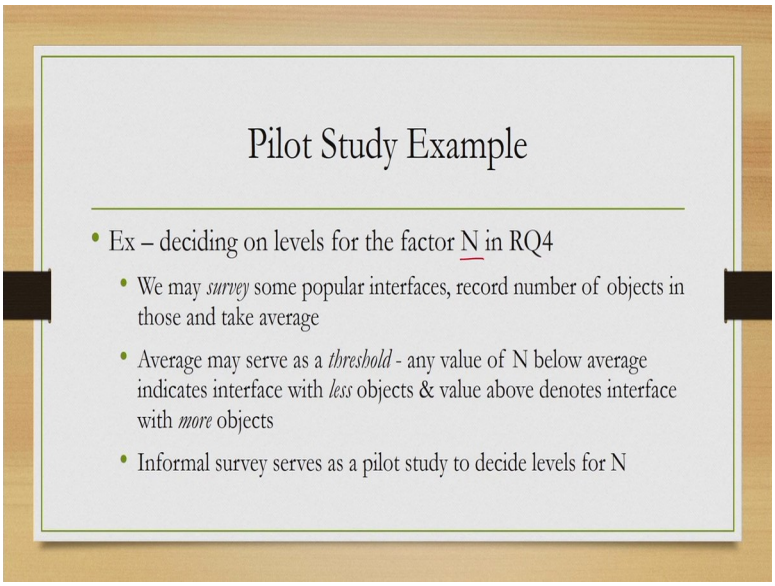


The problem is as we have already said we want to vary the values in a systematic way. So, sometimes it is easy to set the values as we have just seen, if the problem size is small. So, our idea of systematic variation can come from prior knowledge and experience or a theoretical analysis of the problem.

If these are not sufficient, even with our experience or intuition or theoretical analysis of the problem we are unable to figure out, how to set the values, then what we can do is we can go for small study called a pilot test. In a pilot test we can obtain some insight about the nature of the data and we can use that idea to set the values.

Now, the study is typically informal and unstructured with only the purpose to get some idea of the data so, that we can prepare for the bigger study. So, such small studies are called the pilot studies which we typically conduct before we perform the actual study.

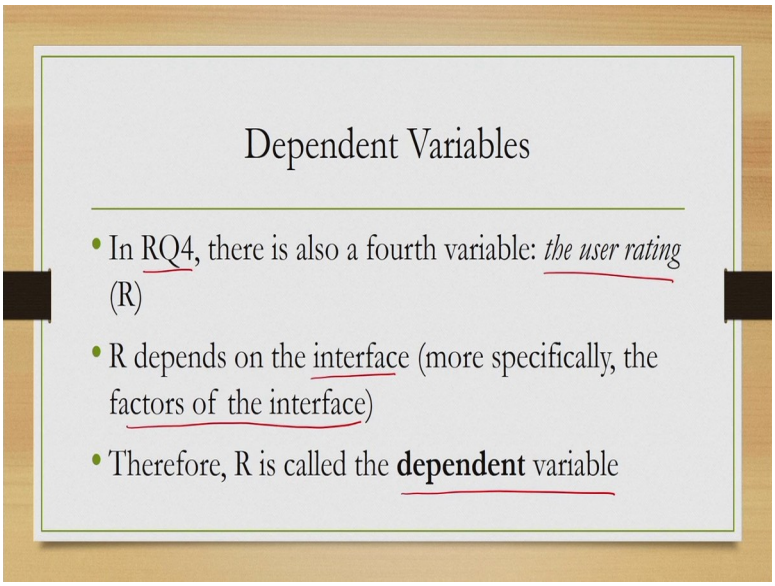
(Refer Slide Time: 23:38)



As an example consider the case of deciding on the levels for the factor N in RQ 4, in other words in RQ 4 we want to decide on different values of N the number of objects. So, how to do that? Suppose, we surveyed a set of popular interfaces and in those interfaces we recorded the number of objects present. So, for each interface we will get some value. Now, we take an average of all those numbers. Now, this average may serve as a threshold value.

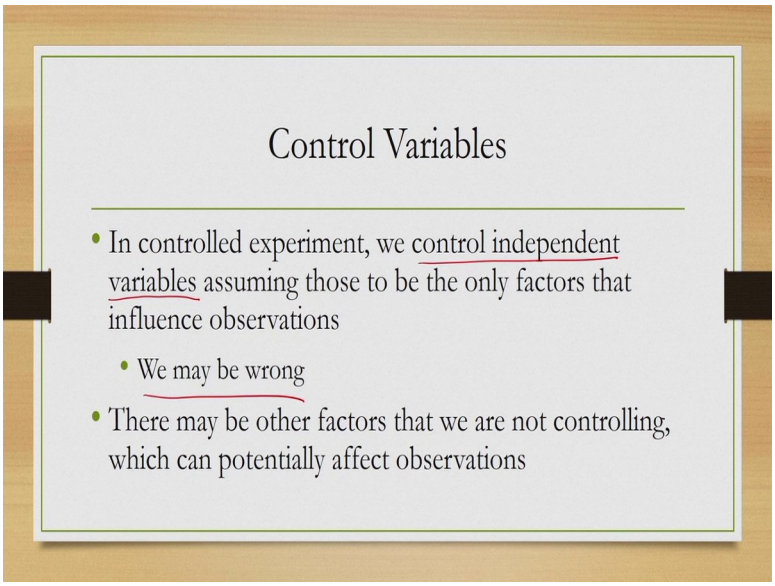
Any interface having less number of objects than this threshold, maybe categorized as interface with less number of objects. Similarly, any interface having more number of objects than this threshold maybe categorized as an interface with more number of objects which we have just discussed before. So, this pilot study of surveying popular interfaces serves our purpose of defining interfaces, that are having less number of objects and interfaces, that are having more number of objects in terms of some threshold value which we have obtained based on the pilot study.

(Refer Slide Time: 25:04)



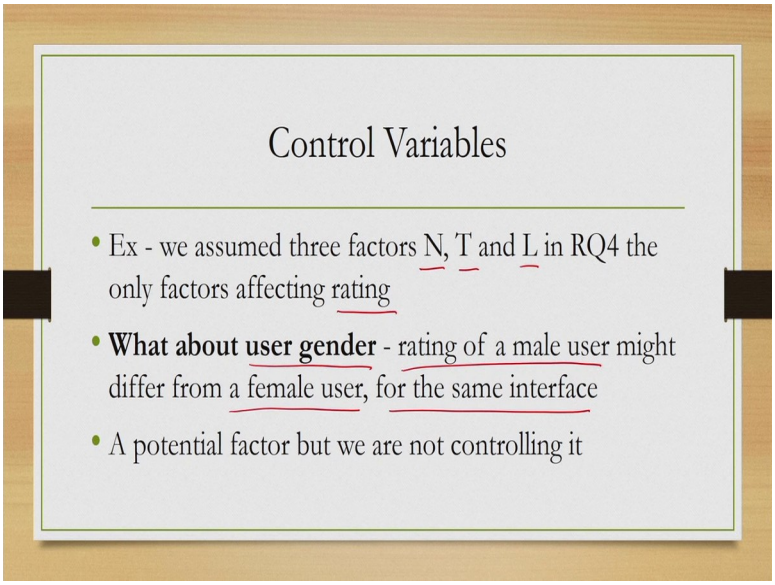
Now, in the research question as we have noted there is also a fourth variable the rating which indicates the aesthetic judgment behavior. So, the rating value depends on the interface or more specifically the factors of the interface. Based on these factors the users judge the interface and provide the rating. So, such variables are called dependent variables; that means, dependent variables are those which depends on the independent variables or the factor. Now, apart from these two broad types of variables there are two more types which we should know one is control variable, other one is confounding variables.

(Refer Slide Time: 25:51)



Let us try to understand these variables. So, in a controlled experiment so, every empirical study is considered to be a controlled experiment and in a controlled experiment, we are supposed to control independent variables by setting predetermined values to those variables. Now, our assumption here is that these are the only variables that affect, the outcome, that affect the dependent variable. However, this assumption may be wrong. There may be other factors or independent variables, which affect dependent variable, but which we are not controlling.

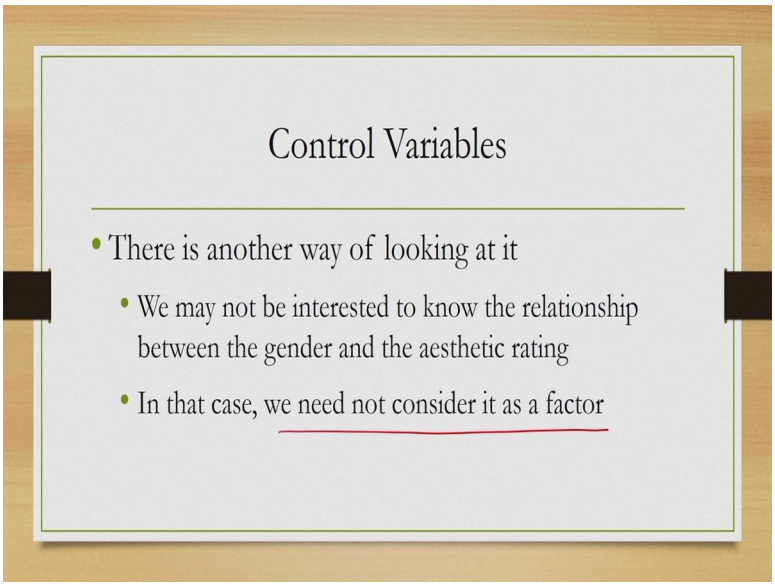
(Refer Slide Time: 26:37)



Let us considered one example. So, in RQ 4 we assumed that there are three factors N T and L, which affect the rating. It may be the case that user gender is another factor. So, the rating of a male user may be different, than the rating of a female user even if the levels of the three factor remain the same or for. In other words for the same interface, we can have a different rating from a male than a female user.

So, in that case clearly gender plays a role in affecting the rating behavior or the aesthetic judgment behavior, which we are not explicitly controlling unlike the other variables namely N T and L.

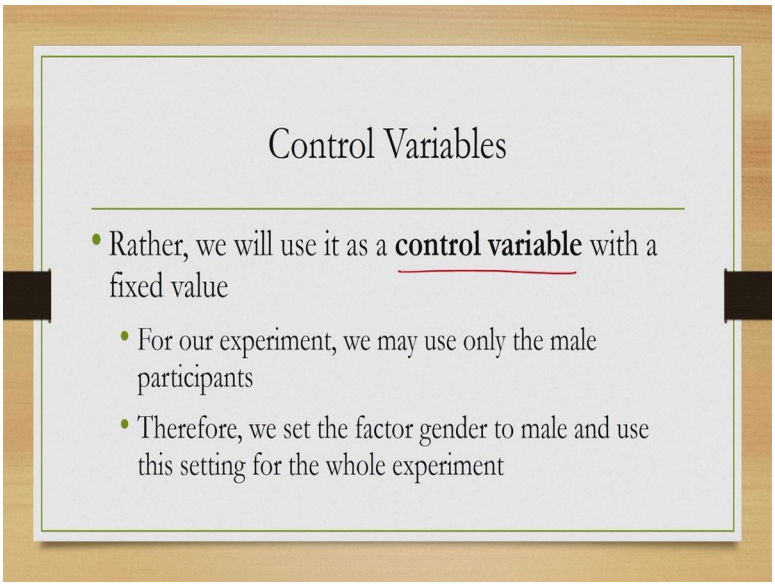
(Refer Slide Time: 27:40)



So, what to do with this, how to deal with this situation? If we know that gender is a factor, then the most straightforward way is to add a fourth factor, namely gender, to our list of factors. However, there can be another way of looking at it. We may not be interested to know the relationship between gender and the judgment. Our main objective is to know the relationship between factors that are characterizing the interface only rather than other factors.

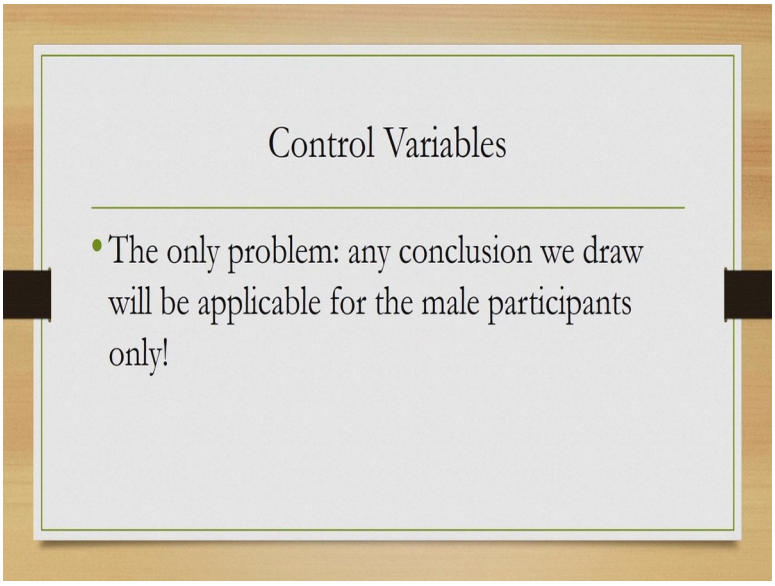
So, in that case how gender is related to the judgment behavior may not be something which we are concerned about. So, what we can do? We need not consider it as a factor. So, we need not included it in our overall formulation of the relationship.

(Refer Slide Time: 28:32)



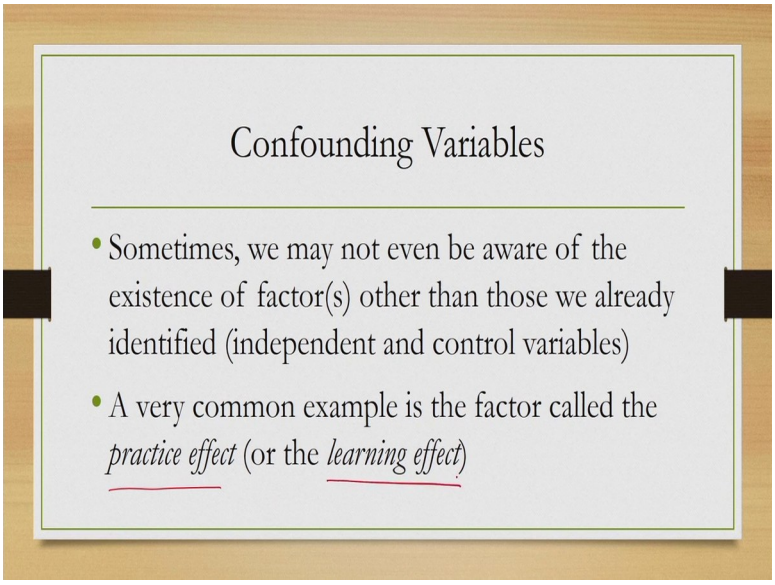
Instead we can use it as a control variable although we are using the term variable what we do here is essentially fix a single value to this variable. In other words in a study we make it a constant unlike the independent variables where we are varying the values. So, in this example what we can do is during the study we can fix the gender of the user. Say, we collect data from only male users or only female users, then we are treating this variable as a control variable by making it a constant with fixed value.

(Refer Slide Time: 29:12)



Is there a downside of it? Yes of course, what we can conclude from the data becomes applicable only for the male participants if we are fixing the gender as male. So, the generalizability of our findings may reduce.

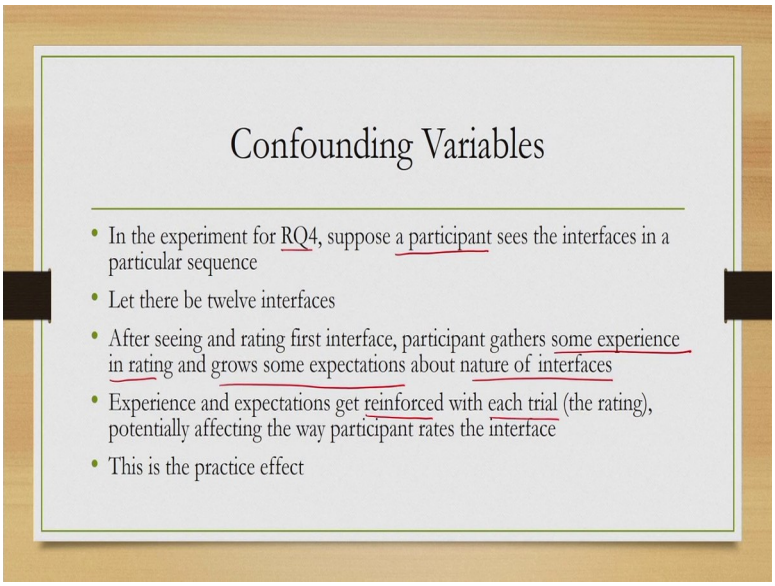
(Refer Slide Time: 29:34)



The other type of variable is the confounding variable. So, far we are assuming that we are aware of all the variables that affect the dependent variable, whether it is independent variable or controlled variable. Sometimes it may so; happen that we are not even aware of the existence of a variable that affects the outcome.

So, we are not explicitly controlling it either by treating it as a control variable or by treating it as an independent variable. However, the result that we get actually gets affected by the values of this variable. And, the conclusion that we can draw from the observations may lead us to a wrong conclusion, because we are not considering this variable. Now, a common example is what is called a practice effect or the learning effect. Let us try to understand this effect.

(Refer Slide Time: 30:27)

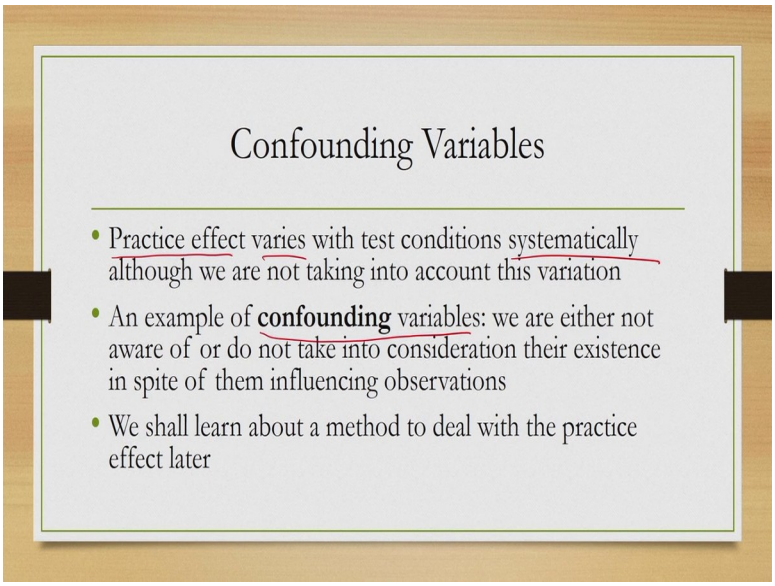


Suppose, in the experiment that we are conducting to answer the research question RQ 4, a participant gets to see the interfaces in a particular sequence. Remember there are twelve interfaces and the participant is asked to see all the interfaces to give their give his or her rating on each of the interface. Now, after seeing and rating the first interface, the participant gathers some experience in rating and also grows some expectation about the nature of the interfaces. And, these experiences or expectations get reinforced with each trial or rating of each interface.

What it leads to it leads to some biased judgment, if there is some mismatch between the experience and expectations that the participant has grown and the design of the interfaces that we have used in our study. So, some biasness may creep in into the overall observation. Although, the aesthetic quality may not be bad, but because the particular participant is expecting a particular way of organizing elements on the interface, the participant feels the aesthetic quality is bad. So, the judgment is actually clouded or biased by the expectations.

When that happens we call it a practice effect or a learning effect that is an effect that comes due to practice or an effect that comes due to learning of the prior conditions.

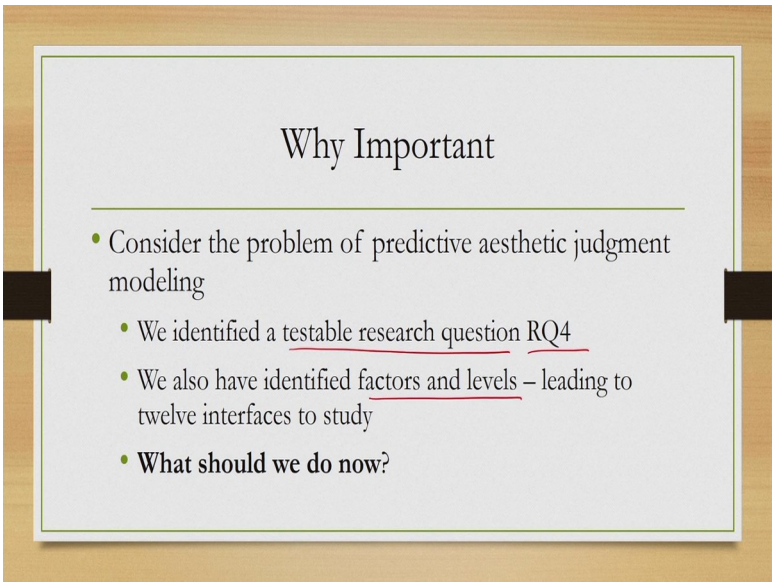
(Refer Slide Time: 32:04)



Now, this practice effect varies with the test conditions systematically, but we are not taking care of it in our observations and recording, these type of variables. Where we are explicitly not taking into account their presence or their effect, but they are affecting the outcome is known as confounding variables. Later on we will learn how to take care of the practice effect. So, essentially there we are trying to eliminate the effect of this confounding variable by explicitly taking care of practice effect in a specific way which we will learn in a later lecture, but our objective should be to detect presence of such variables. So, that the data we get are reliable data.

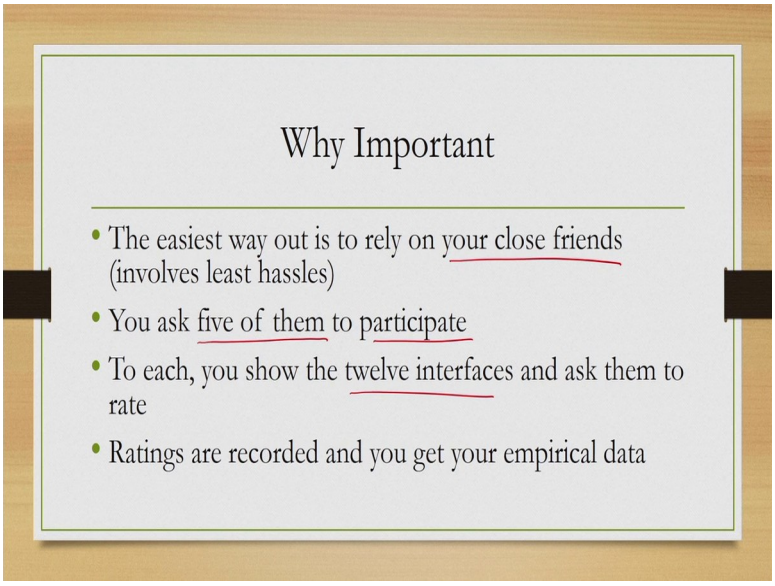
So, in summary we have mentioned four types of variables, independent variables and dependent variables, these are the basic two types along with that we should be aware of control variables and confounding variables in order to make our observations as good as possible. The next stage once we are able to detect or once we are able to identify the variables, the next stage in empirical research is the design of the experiment itself.

(Refer Slide Time: 33:25)



So, the problem is like this. So, we identified a testable research question namely RQ 4 and we have also identified factors and levels. Next, what should we do how to collect data? Now, if I give this problem to most of you what you should do next, then we have a tendency to follow a very simple approach.

(Refer Slide Time: 33:52)

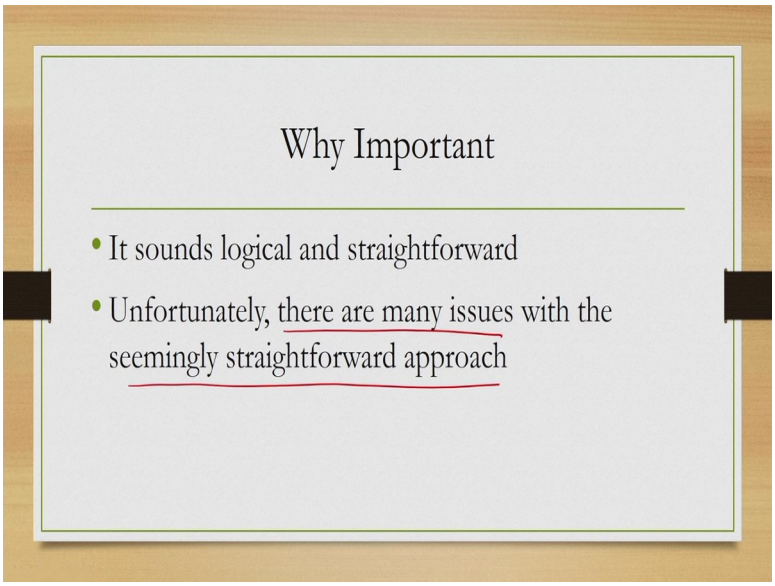


Most of us probably will do the easiest way possible that is we will ask a set of our friends, close friends, for helping us in data collection. For example, suppose we are asking for help to five of our friends. So, we ask them to participate. To each of these

friends we show the twelve interfaces and ask them to rate. Now, ratings are recorded and you get your data. So, you asked your friends to rate twelve interfaces, you asked five of them and you get your data.

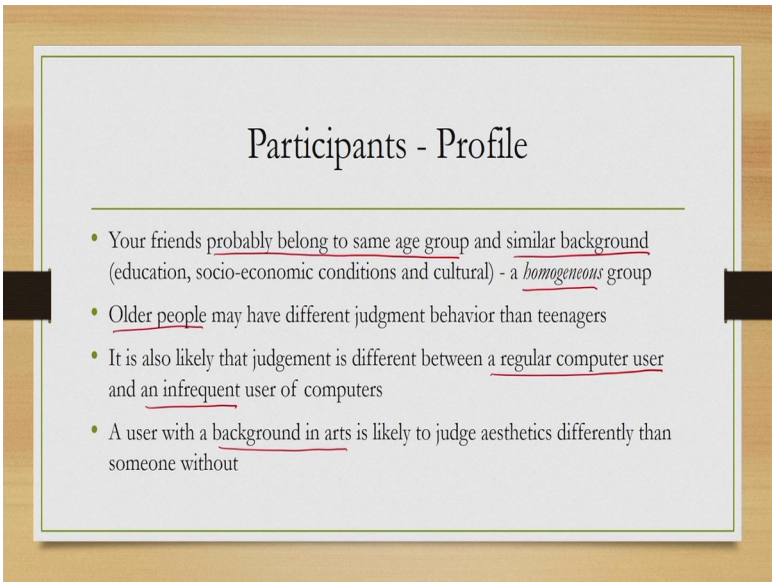
Can we rely on this data, can we simply go for building, a model or conclude something about the nature of the relationship based on this data. And, even if we are able to do that what will be the nature of that conclusion.

(Refer Slide Time: 34:55)



In fact, there are many issues with this seemingly straight forward approach. What are those issues, let us try to see one by one.

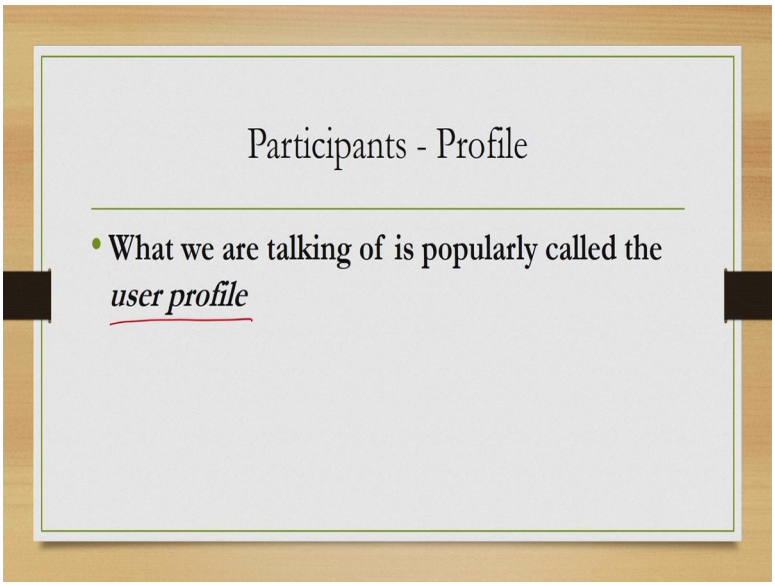
(Refer Slide Time: 35:05)



First of all we need to bother about the participants. Whom we are engaging for data collection, when you ask your friends now, your friends probably belong to the same age group and similar background in terms of education and socioeconomic conditions culture and so on. In other words they essentially refer to a homogeneous group.

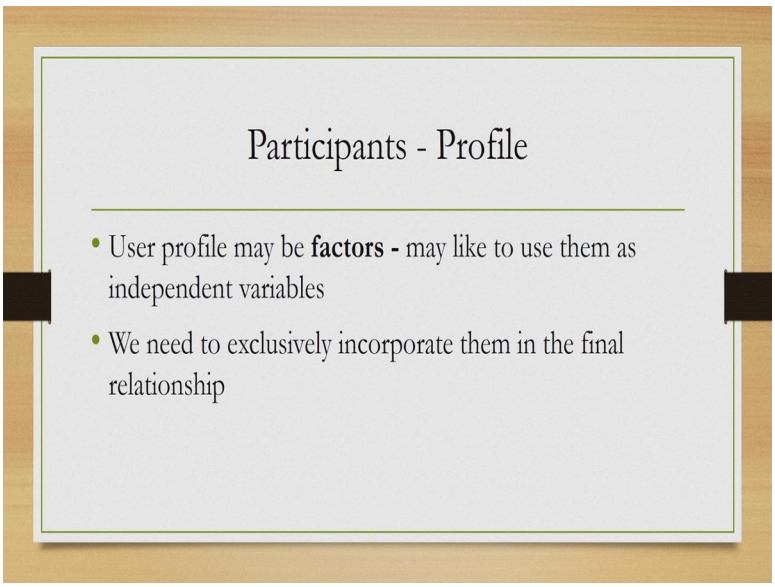
Now, why this is important? Because if we take data from some other groups such as the older persons or somebody who is a regular computer user vis a vis, somebody who is not a frequent user of computers, or someone who has a background in arts vis a vis, someone who does not have a background then we are likely to get different judgments. So, what we are referring to with these terms that homogeneous group, older people, regular computer users in frequent users, person with background in arts, these facts what we are referring to is essentially known as user profile.

(Refer Slide Time: 36:26)



So, we should bother about the profile of the users whom we are planning to engage in our data collection.

(Refer Slide Time: 36:41)

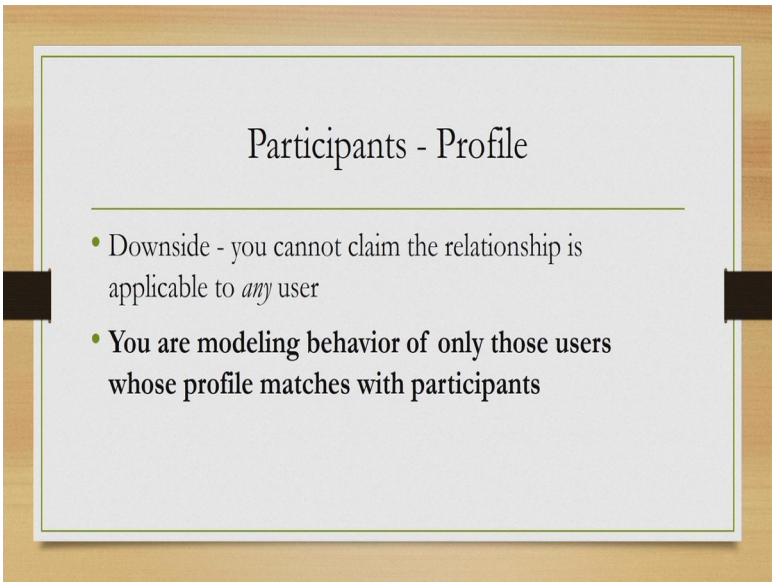


Now, there are many components in a profile as we have seen educational background maybe a component of the profile, culture socioeconomic background, gender many things may be part of the profile. And, as we have just discussed the profile affects the judgment behavior or is likely to affect the judgment behavior.

So, ideally we should take into account this profile information in our study. How to do that? As, we have just seen before one way to do it is to include this profile information or components of the profile explicitly in the relationship that we used to build, but that is likely to make things very complicated. What else we can do, we can treat them as control variable, we can fix the values for these components of the profile and accordingly we collect data as we have discussed earlier.

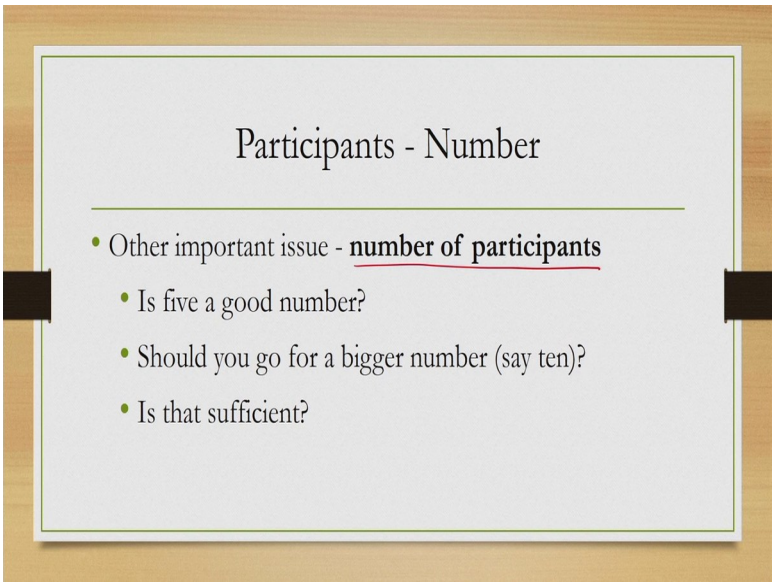
This approach comes with the same pitfall of the process that is whatever data we collect is actually representative of the particular user belonging to that profile, we cannot generalize it. So, if we fix the user profile and collect data, then whatever conclusion that we can draw can be only for users of that particular profile. And, we cannot say that the particular relationship that we have found after the study is applicable for any user with any profile.

(Refer Slide Time: 38:29)



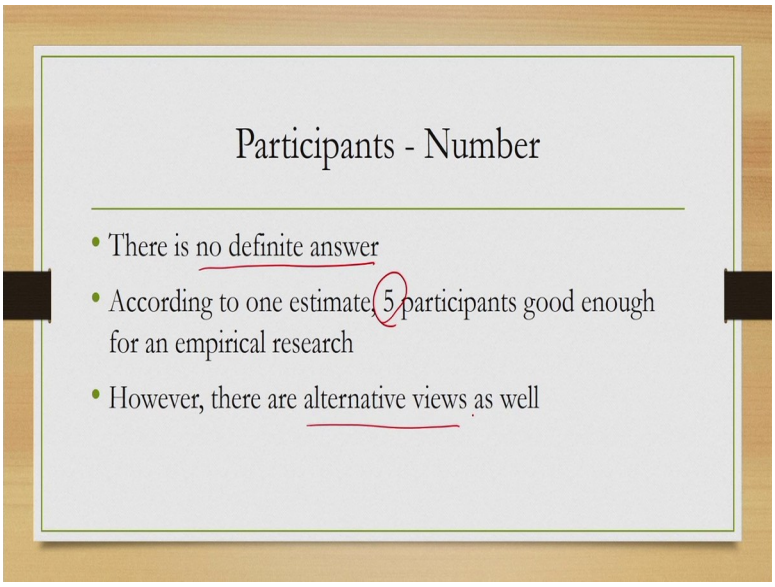
So, essentially your model or conclusion becomes somewhat less generalizable. If, you treat the profile information as control variables, that is one issue the other issue is number of participants.

(Refer Slide Time: 38:42)



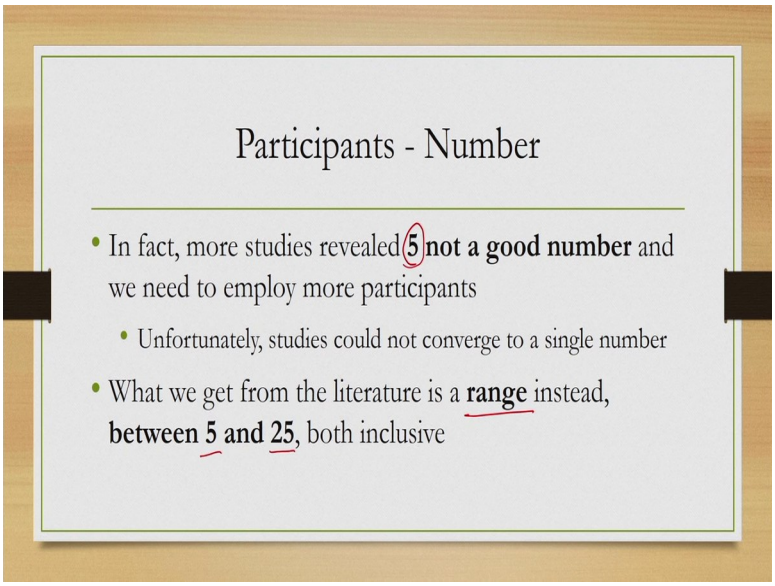
So, when we want to collect data, how many participants we should involve? Should you involve five is that a good number or should we involve more, how many more, what is a right number? That is another issue that often confuses the designer of the experiment.

(Refer Slide Time: 39:15)



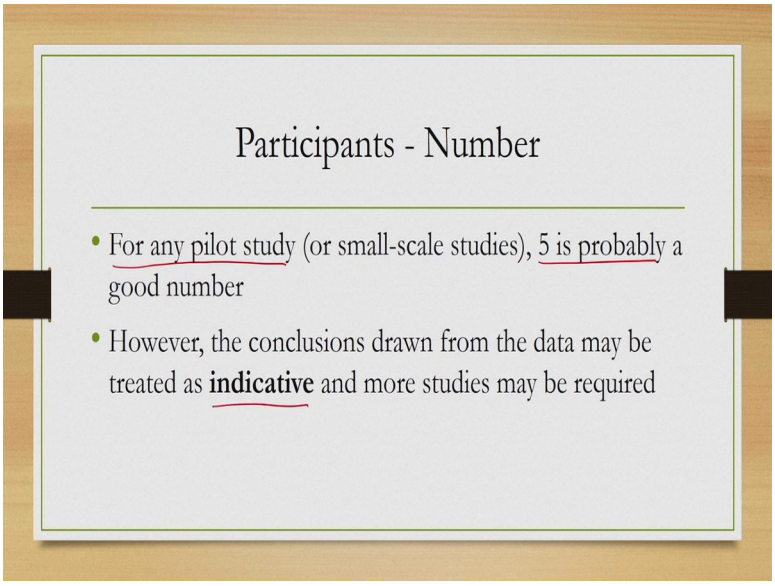
In fact there is no definite answer. According to ones study 5 is a good enough number. So, with 5 participants we can get sufficient data for generalized conclusion, but there are other studies which give some alternative views, which give some alternative views.

(Refer Slide Time: 39:40)



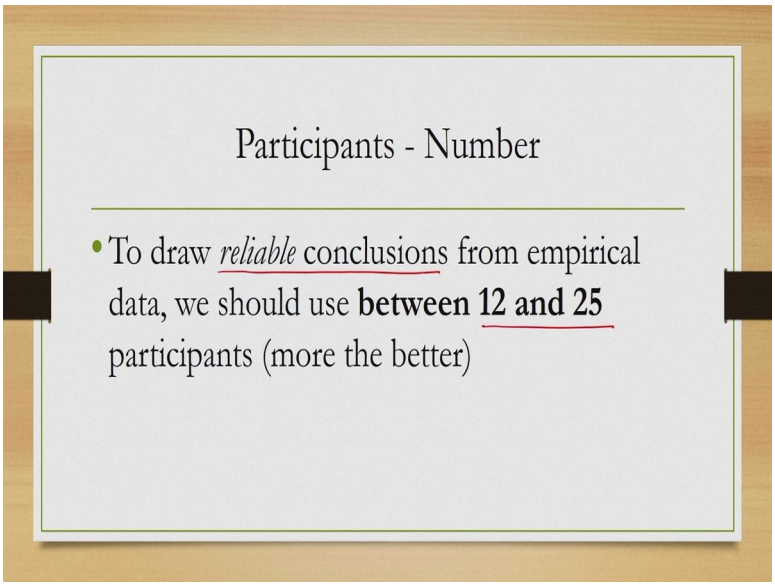
Other studies have revealed that 5 is not necessarily a good number, but no study could convincingly come to a conclusion about the right number. In fact, from different studies in the literature, what we can find out is that there is a range instead of a single number of participants to be engaged in a study and the range varies the range is between 5 and 25, where both the numbers are included. So, we can engage between 5 to 25 both included number of participants in our study.

(Refer Slide Time: 40:32)



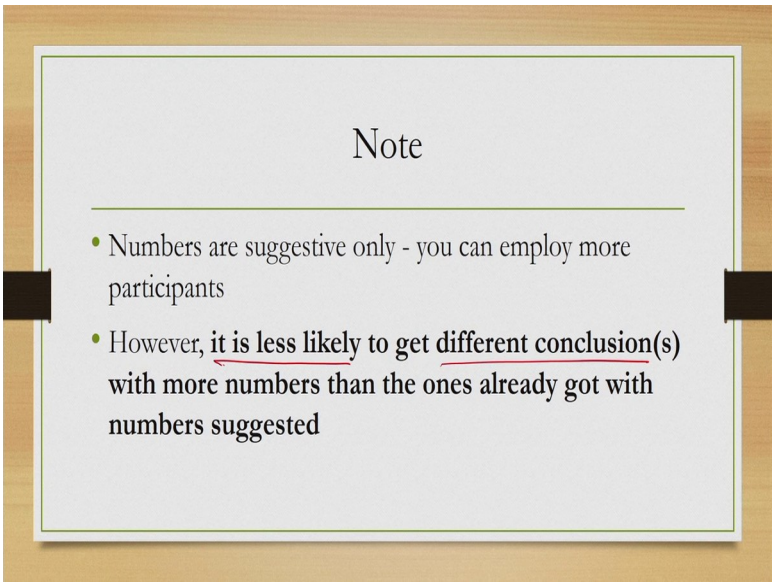
So, for any pilot study as we have mentioned earlier these are small-scale studies, 5 is probably a good number, but for main studies we require more participants, but when we are conducting pilot study with 5 participants, we should not conclude conclusively instead whatever results we get can be termed at best as indicative of the result.

(Refer Slide Time: 41:04)



And, for main studies it is advisable to use between 12 and 25 participants to get reliable data for reliable conclusions.

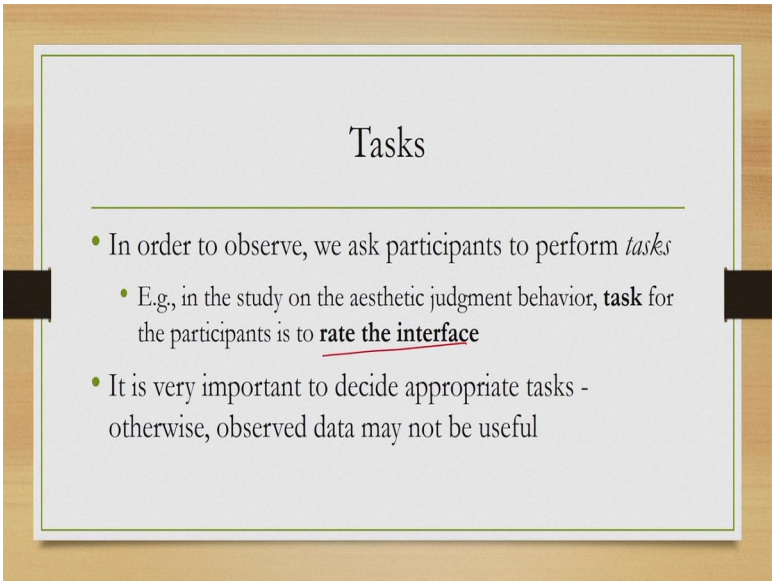
(Refer Slide Time: 41:21)



You may engage more, if you feel and if you have access to the resources namely money and other resources time, but it is less likely that you will get a different conclusion, then the conclusion than you have already got or you would have got if you would have used between 12 to 25 participants.

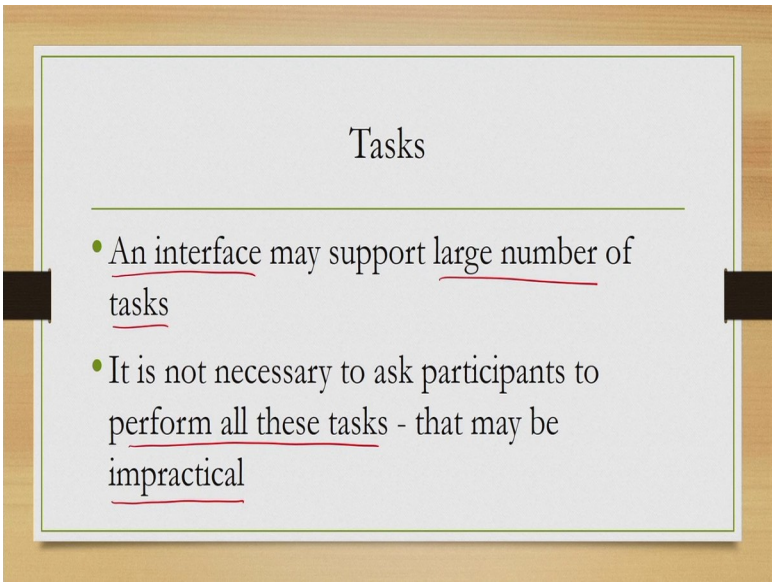
So, with more participants it is not likely that you will get different conclusions or too many different conclusions one or two different conclusions you may get, but not significantly. So, it is advisable to use between 12 to 25 participants in a study, if it is main study. For pilot studies 5 participants may be sufficient.

(Refer Slide Time: 42:26)



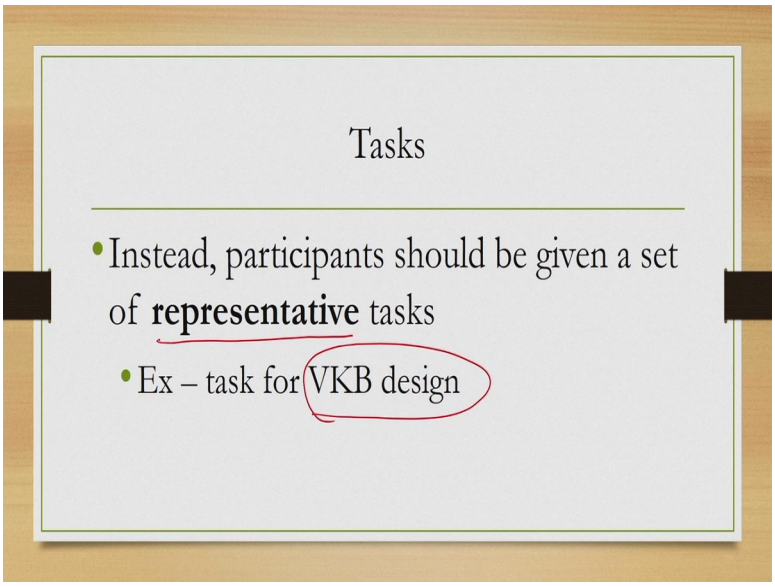
The other component is the task. So, each participant is asked to carry out a task. And, during that process you record, you observe the participant behavior and record. Now, the question is what task to be given to a participant and in which way. In the example, that we have mentioned with respect to the research question RQ 4 for aesthetic judgment behavior ratings. The task is simple that is to rate the interface, but in some cases it may not be that simple and it is very important to decide the appropriate tasks. Otherwise, the data that you get the observations that you make may not yield to reliable conclusion.

(Refer Slide Time: 43:22)



In fact, an interface may support a large number of tasks, but it is not necessary to ask participants to perform all these tasks. It may be impractical in many situations. So, which tasks to choose for a participant is a very challenging problem.

(Refer Slide Time: 43:50)



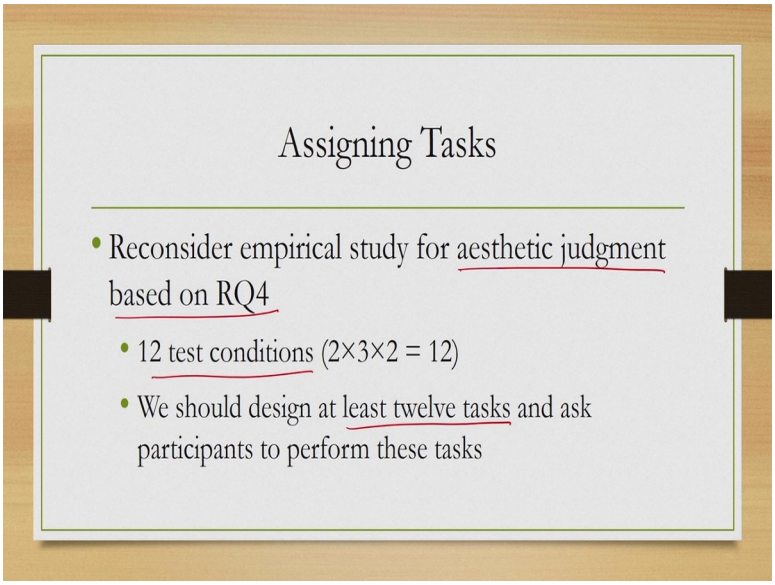
Let us consider one example, earlier we have discussed about the virtual keyboard design, one objective of empirical study is to evaluate the design. Now, in order to evaluate we need to employ participants and give them some tasks to perform on the keyboard. What tasks to be given? We can give them some text to type. Now, these text

should not be any arbitrarily chosen text instead the text should be representative of the underlying language. Typically, the text that we type when we use a keyboard, otherwise if we test our keyboard on the basis of any arbitrary text, then may be the drawbacks of the keyboard during typing will not be visible or identifiable during evaluation phase.

So, that type of text is called representative of the underlying language or more specifically the underlying language, which is text not spoken language. So, when we type we type the text. So, essentially the language is text language and we pick up a representative text from this language to evaluate a design. So, in a nutshell what we want to say is that, when you are asked to choose a task the tasks should be representative.

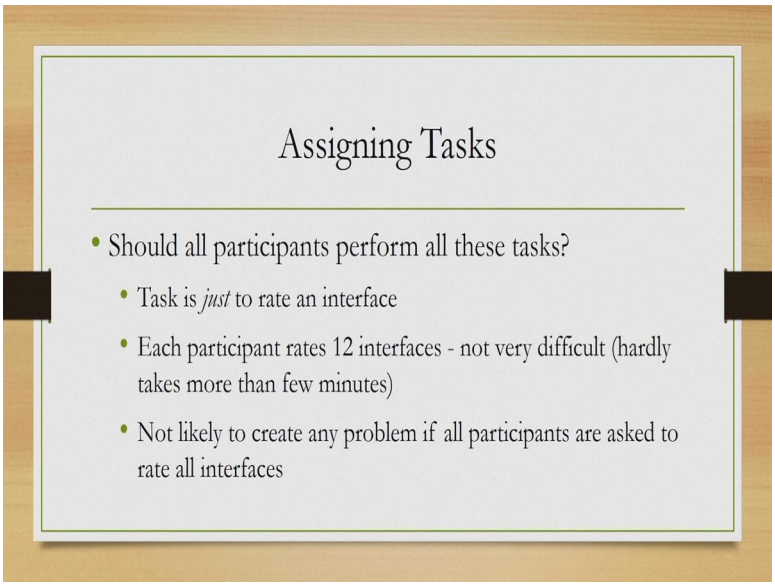
In other words the task should represent the situation that is likely to happen during regular usage of the system. Otherwise, the task and the corresponding data will not be helpful in coming to a useful and reliable conclusion about the system.

(Refer Slide Time: 45:57)



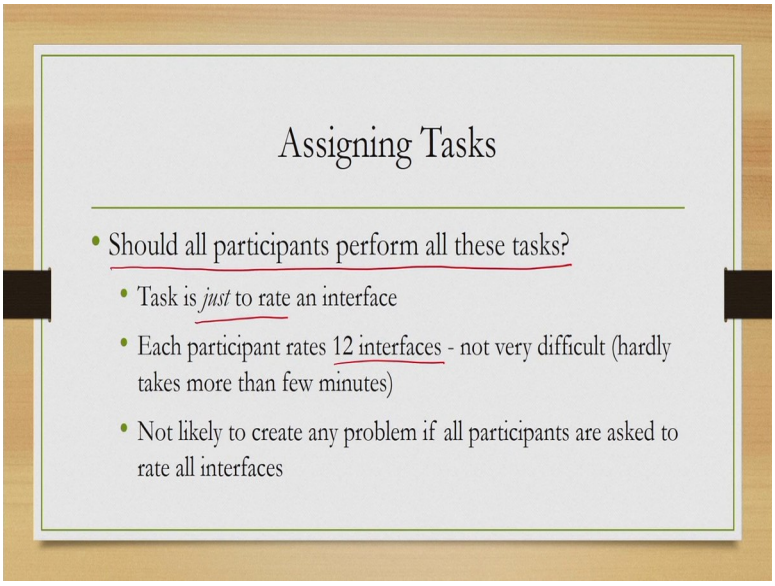
For example, let us reconsider the aesthetic judgment study based on RQ 4. So, we have 12 test conditions and we should design at least twelve tasks to perform the study.

(Refer Slide Time: 46:12)



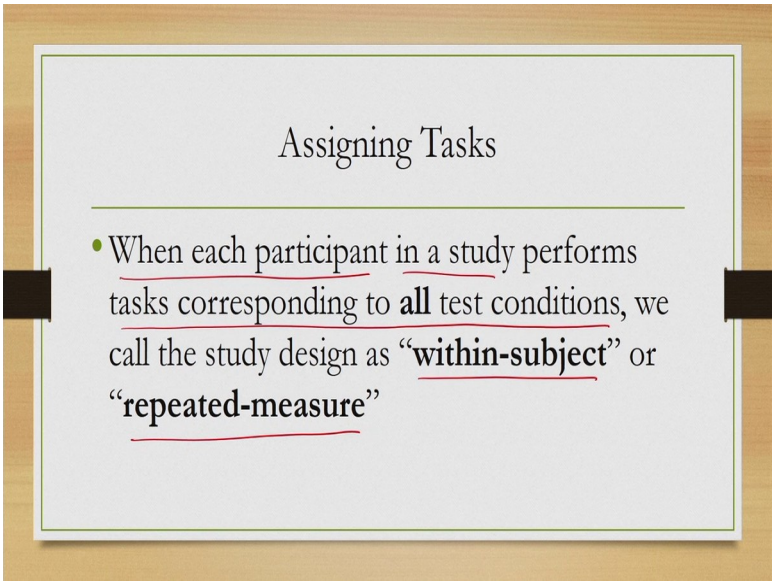
Another issue that comes up when we are trying to assign these tasks is how to assign the tasks? So, determination of task is one issue other issue is once the tasks are determined how to assign the tasks? Let us try to understand this in terms of an example. Assume that, we are dealing with the empirical study related to the research question 4 on aesthetic judgment and there are twelve test conditions and we have decided on twelve tasks that is rating by the participants.

(Refer Slide Time: 46:54)



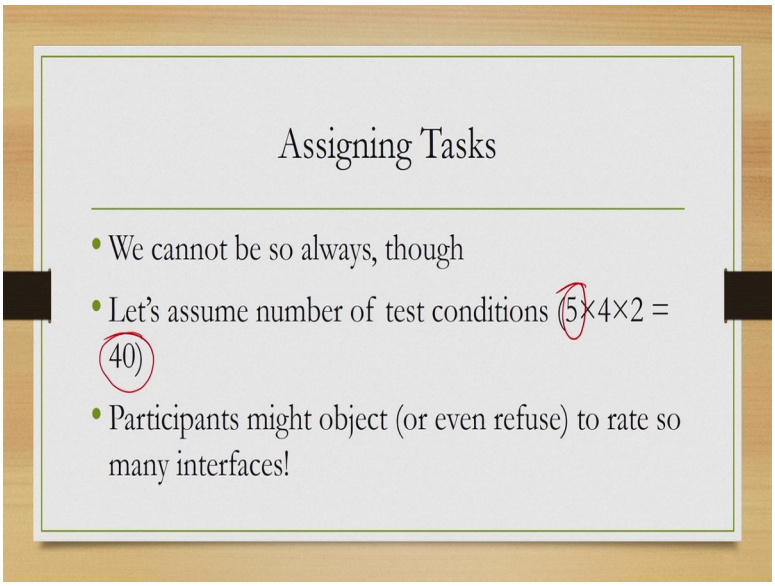
The question is should all participants perform all these tasks, in other words should I ask a participant to perform ratings for all 12 interfaces. In our example the task is very simple that is just to rate and each participant rate 12 interfaces, which is not likely to take too much time and not very difficult as well. So, if we ask our participants to read 12 interfaces they may agree they may not create any problem.

(Refer Slide Time: 47:33)



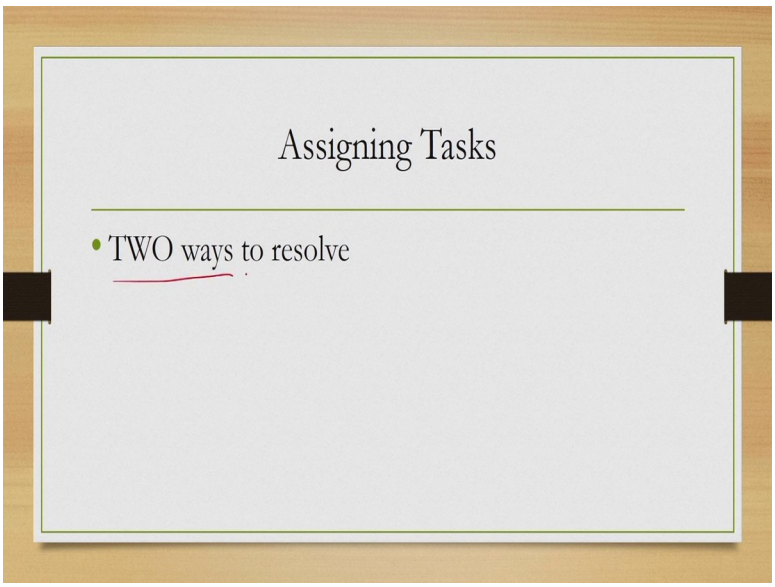
What happens? If, we increase the number of tasks now when each participant in a study performs tasks corresponding to all the test conditions, then we call the particular design as within subject or repeated measure design. So, in a repeated measure or within subject experiment design we ask each participant to perform all the tasks.

(Refer Slide Time: 48:00)



Now, let us assume that we have many more test conditions for our aesthetic judgment study. So, instead of two now we have 5 values for N. So, total number of interfaces to rate becomes 40. Now, with respect to 12 40 is reasonably large number. And, if we ask our participants to rate all the 40 interfaces they may refuse, because that may be very tiring to see the interfaces and rate.

(Refer Slide Time: 48:40)

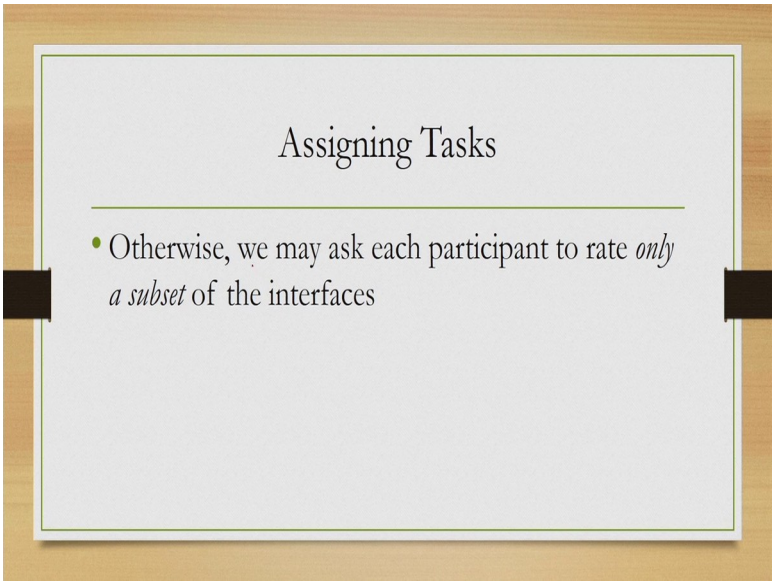


Now, we can resolve this in two ways, what is the first way? This is most straightforward what we can do is we can divide the data collection sessions. So, we can ask the

participants to rate few interfaces in one session and ask them to rate again in another session maybe next day or maybe after a gap of few hours.

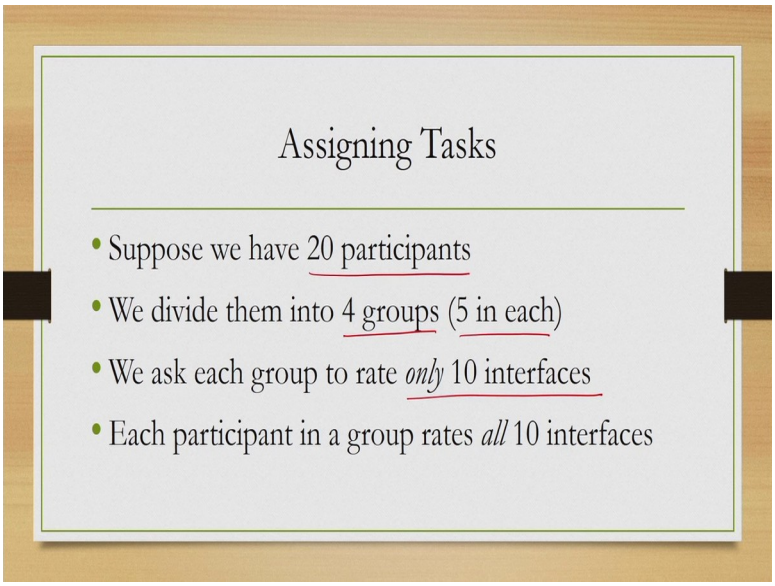
So, that they get sufficient rest and then they are asked to rate a less number of interfaces. In more specific terms, let us assume that on one day one participant is asked to rate 20 interfaces and on the next day the remaining 20 interfaces out of the 40 interfaces. Now, in that way the chances that the participants refuse to rate, because it is too tiring may reduce.

(Refer Slide Time: 49:48)



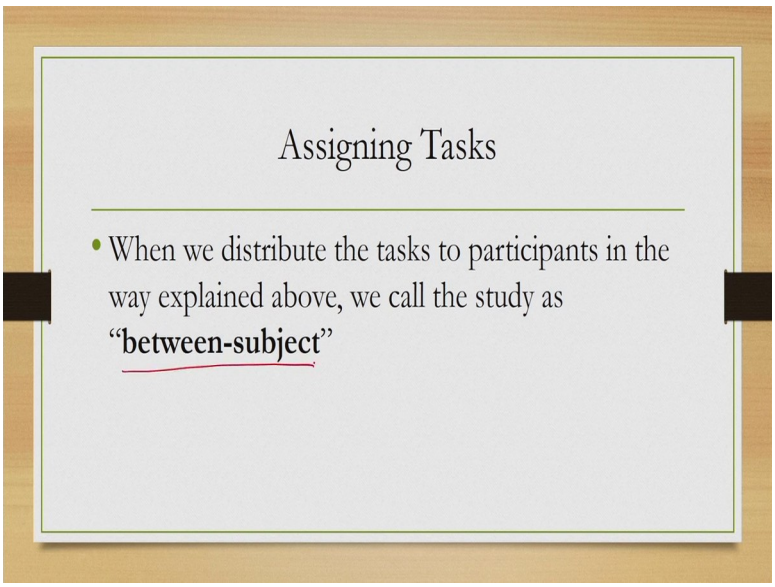
There is another way in the second way we may ask each participant to rate only a subset of the interfaces; that means, all participants are not rating all the interfaces, instead each participant is asked to rate only a subset of all the interfaces.

(Refer Slide Time: 50:07)



For example, suppose there are 20 participants we divide them into 4 groups with 5 in each group and then we ask each group to rate only 10 interfaces. So, total 40 interfaces we get the rating from the 4 groups, but each group rates only 10 interfaces. So, the members of each group are not likely to complain, because they are given a very manageable task.

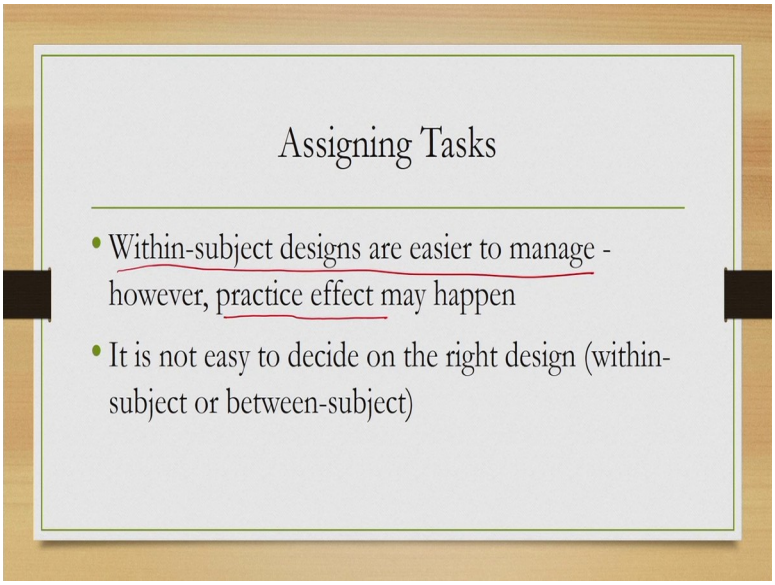
(Refer Slide Time: 50:40)



So, when we assign tasks in this way where we ask a participant to perform a subset of all the tasks, then this type of experiment design is called between subject design. So, we

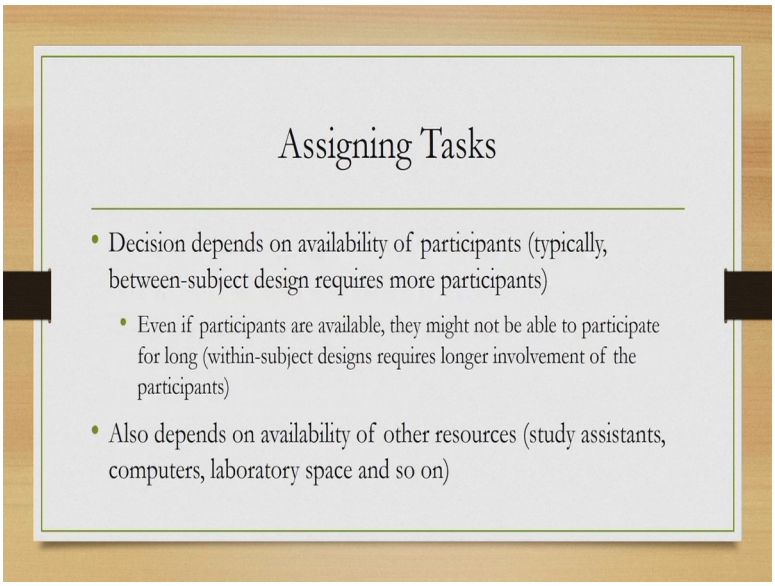
have two types of experiment designs depending on the way we distribute the tasks to the participants; one is repeated measure or within subject design, where each participant performs all the possible tasks or the all the tasks that are designed for the experiment. And, in between subject design we divide the tasks between participants. So, each participant is not required to perform all the tasks.

(Refer Slide Time: 51:26)



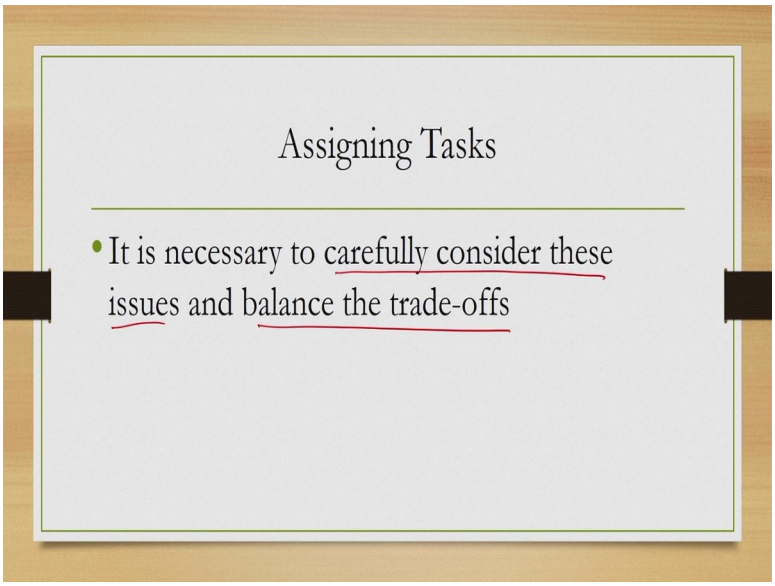
Now, there is always a confusion which approach to choose, which assignment approach, which task assignment approach to choose. It may be noted that within subject designs are easier to manage, because we are dealing with same set of participants, but when we are working with within subject design there may be practice effect as we have explained earlier.

(Refer Slide Time: 51:51)



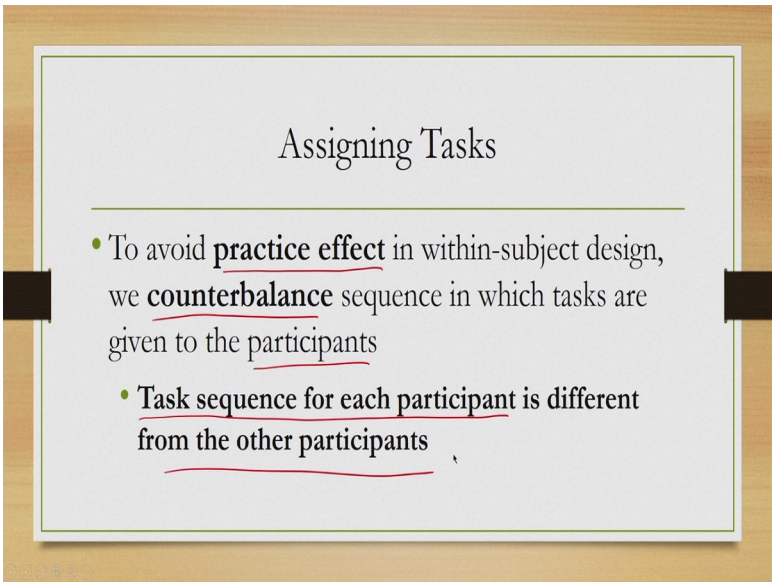
On the other hand the problem with between subject design is that we require more participants, because we are distributing the tasks that may not always be available. So, within subject design requires less participant, but may get affected due to practice effect. Between subject design requires more resources, more participants, more participant times, more cost, but may avoid practice effect. And, it is not necessarily easy to decide on which approach to choose, because even if we have more participants the participants may not be available for longer durations or we may not have resources to recruit more participants.

(Refer Slide Time: 52:43)



So, on a case by case basis we have to make a decision. So, it is necessary to carefully consider these issues and balance the trade-off there is no thumb rule of course, and you have to make your own decision depending on the availability of resources and the kind of participants you are planning to recruit.

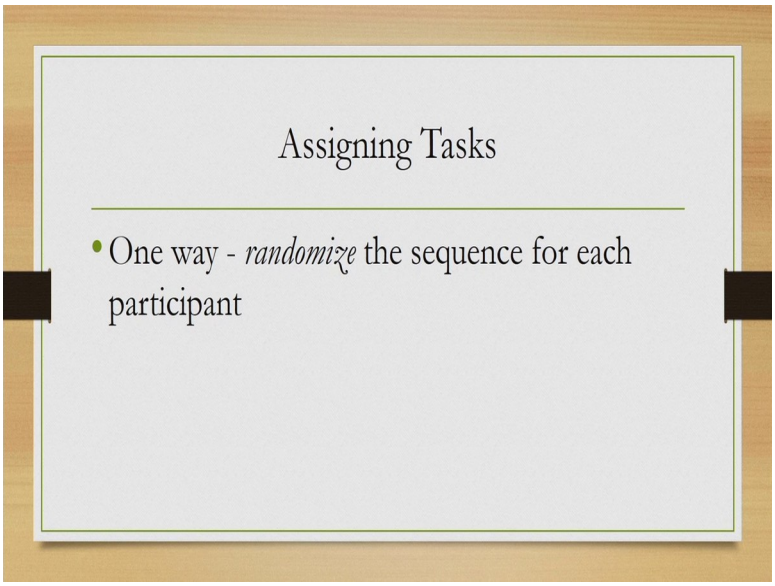
(Refer Slide Time: 53:01)



Now, earlier we said that if we are going for a repeated measure or within subject design it may lead to practice effect. Can we try to avoid practice effect? There is one way to do that. What we can do is we can counterbalance the sequence in which the tasks are given to the participant. What it means? It means that, the task sequence for each participant is different from the other participants.

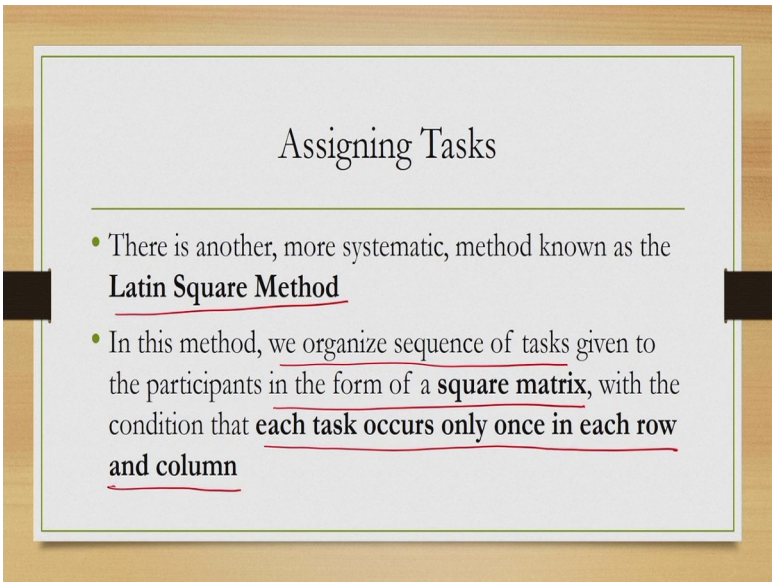
So, in an experiment suppose there are 5 participants, each participant has to perform 5 tasks. So, the sequence in which participant one gets the task is different than the sequence in which participant two gets the tasks and so on for all the participants.

(Refer Slide Time: 53:46)



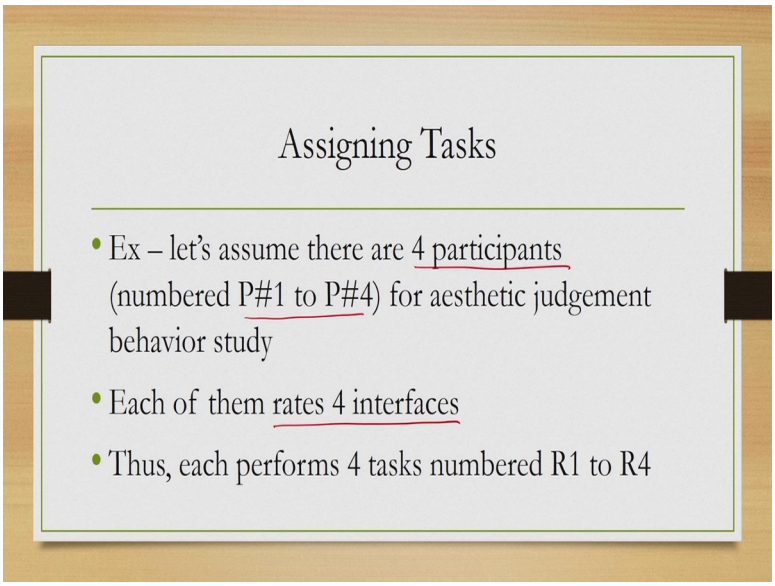
Now, how do we counterbalance? One way is to randomize the task sequence. So, for each participant we randomly generate the sequence now with randomization of course, there is a possibility that sometimes the sequences may become same.

(Refer Slide Time: 54:01)



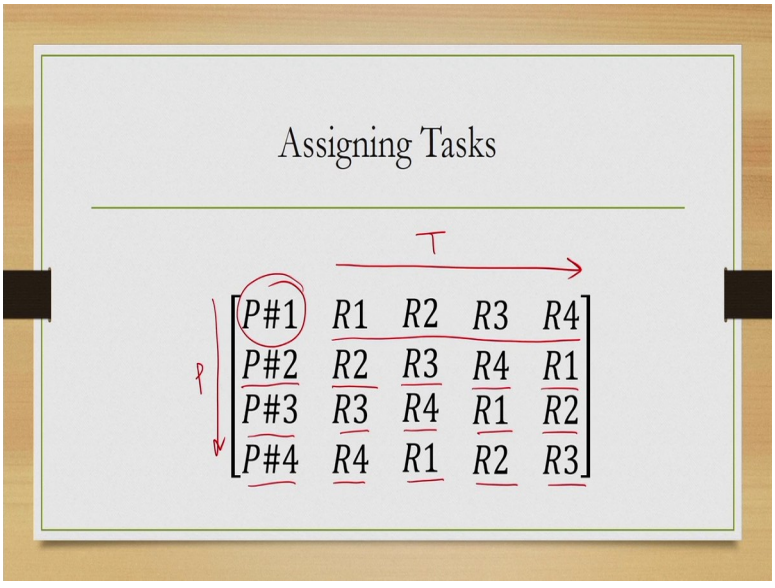
There is one more systematic approach which is called the Latin Square Method. So, in this method, what we do is that we organize the sequences of tasks in the form of a square matrix with the condition that each task occurs only once in each row and column of the matrix. So, this square is called Latin square.

(Refer Slide Time: 54:26)



As an example, let us assume that there are 4 participants number P 1 to P 4, who took part in our aesthetic judgement behavior study and each of them rates 4 interfaces. So, there are 4 interfaces to rate or 4 tasks now this situation, we can represent in the form of a Latin square to indicate how to assign the tasks to the participant.

(Refer Slide Time: 54:51)



So, this side indicates participant this side indicates task let us denote the tasks by R 1 R 2 R 3 R 4. So, for participant 1 we have assigned task in this sequence R 1 followed by R 2 followed by R 3 followed by R 4, where each R represents 1 rating for a particular

interface namely R 1 represents rating of interface 1 by the participant. For P 2 we change the sequence here, we make it R 2 followed by R 3 followed by R 4 followed by R 1.

For P 3 again the sequence changes R 3 followed by R 4 followed by R 1 followed by R 2 and for P 4, it is changed again R 4 followed by R 1 followed by R 2 followed by R 3. So, if you organize it in this way where the same task is not repeated in a row or column then we get a Latin square and that is one systematic way of generating the sequence of tasks to be given to a particular participant.

Now, why it helps counterbalancing helps, because here we are changing the sequences? So, statistically it indicates that whatever practice effect maybe there for a participant for a particular participant we will get nullified by the counter effect of changing the sequence for different participants. So, participant 1 will get a practice effect for this sequence, participant 2 will get a practice effect for some other sequence, participant 3 will get a practice effect for yet another sequence and so on. And, all the sequences are different. So, the practice effects are different. So, statistically they will nullify each other.

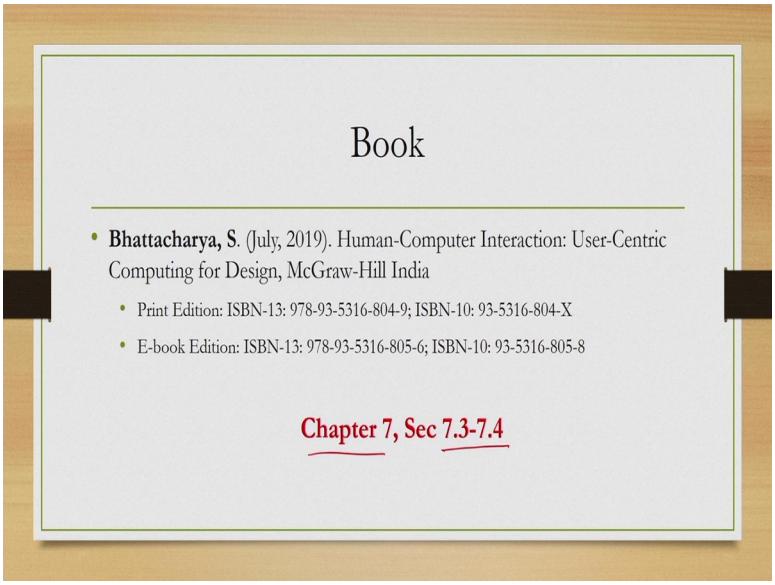
And, what we will get at the end is observations that are not affected by or biased by participants practice effect. So, this is one way of taking care of the practice effect. So, let us summarize what we have learn today. So, two of the stages of an empirical study we have discussed namely the identification of variables and the design of experiment. So, we talked about 4 variables, independent variables, dependent variables, these two are the main variables and then there are control variables and confounding variables.

And, in experiment design we talked about two major issues one is participant and one is task. The issue with participant is that we need to know how many participants to use and how to decide those participants? So, essentially we talked about profiling the target users and then choose participants accordingly. And, secondly, decide on the number which should be at least five for a pilot study and between 12 to 25 for the main study the more number is better. However, it is not likely to lead to significant new findings, then what we can get with the prescribed number.

In the context of task we have touched upon two issues; one is how to design the tasks, we talked about representative tasks, that is the tasks that are likely to occur in practice

and the other issue that we discussed is how to assign the tasks to the participants? There we talked about within subject and between subject designs and in case of within subject design, we mentioned that the chances of practice effect is there to avoid that we can counterbalance the sequence by either randomizing the sequence or using a Latin square method.

(Refer Slide Time: 58:52)



So, whatever we have discussed today can be found in this book you are adviced to refer to chapter 7 section 7.3 and 7.4 for more details on the topics that we have covered today. In the next lecture, we will cover the remaining parts of the empirical study process namely analysis of data and using the data for building predictive models.

Thank you and goodbye.