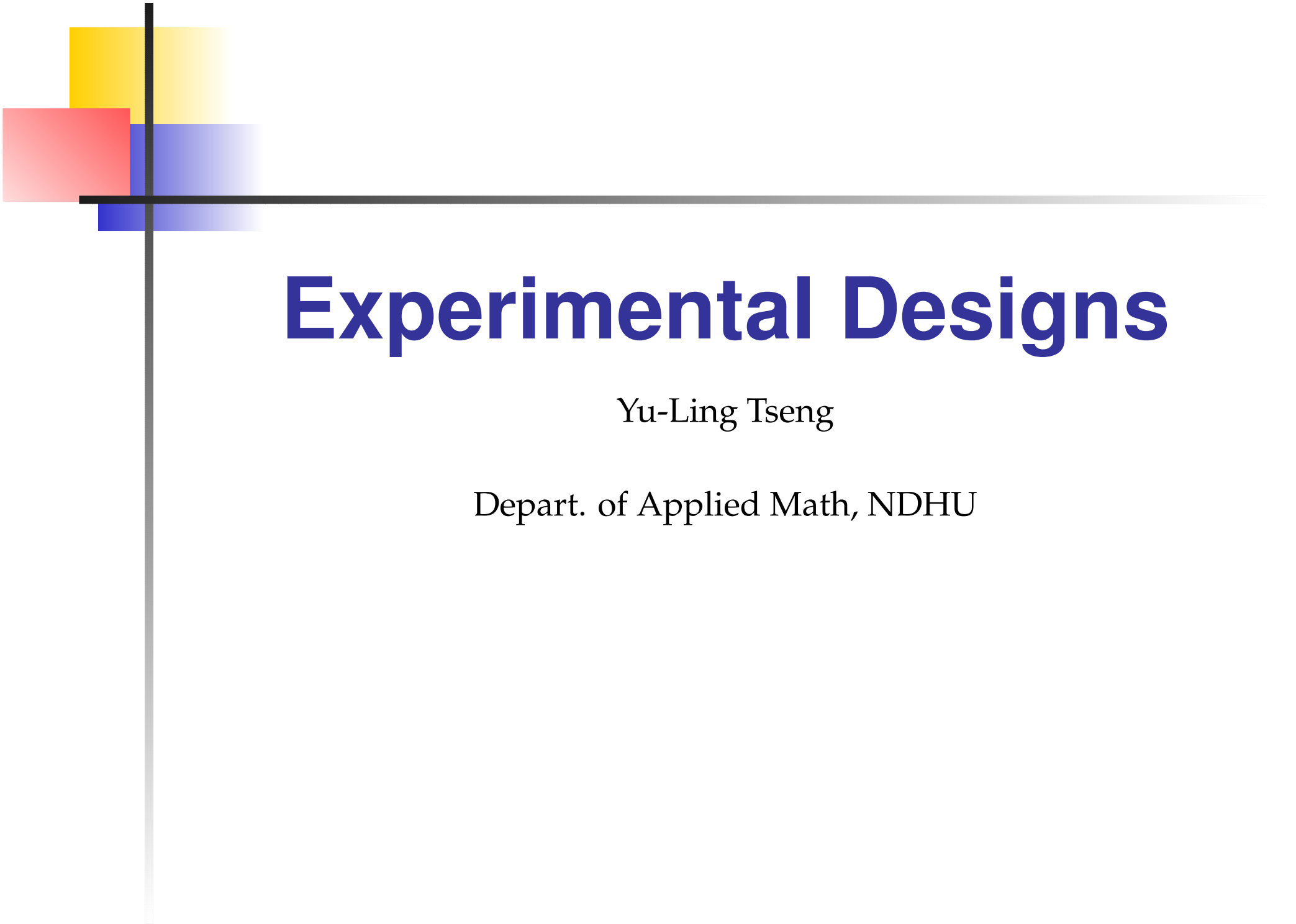


Experimental Designs

Yu-Ling Tseng

Depart. of Applied Math, NDHU

Introduction

- Observational v.s. Experimental studies
  Passive data collection / Active data production
  Descriptive / Inferential (cause-effect) conclusions
- The contribution of Stat. to experimentation
  Problem of interpretation, Stat. inferences, Function of randomization, Concept of local control.

Problem of interpretation

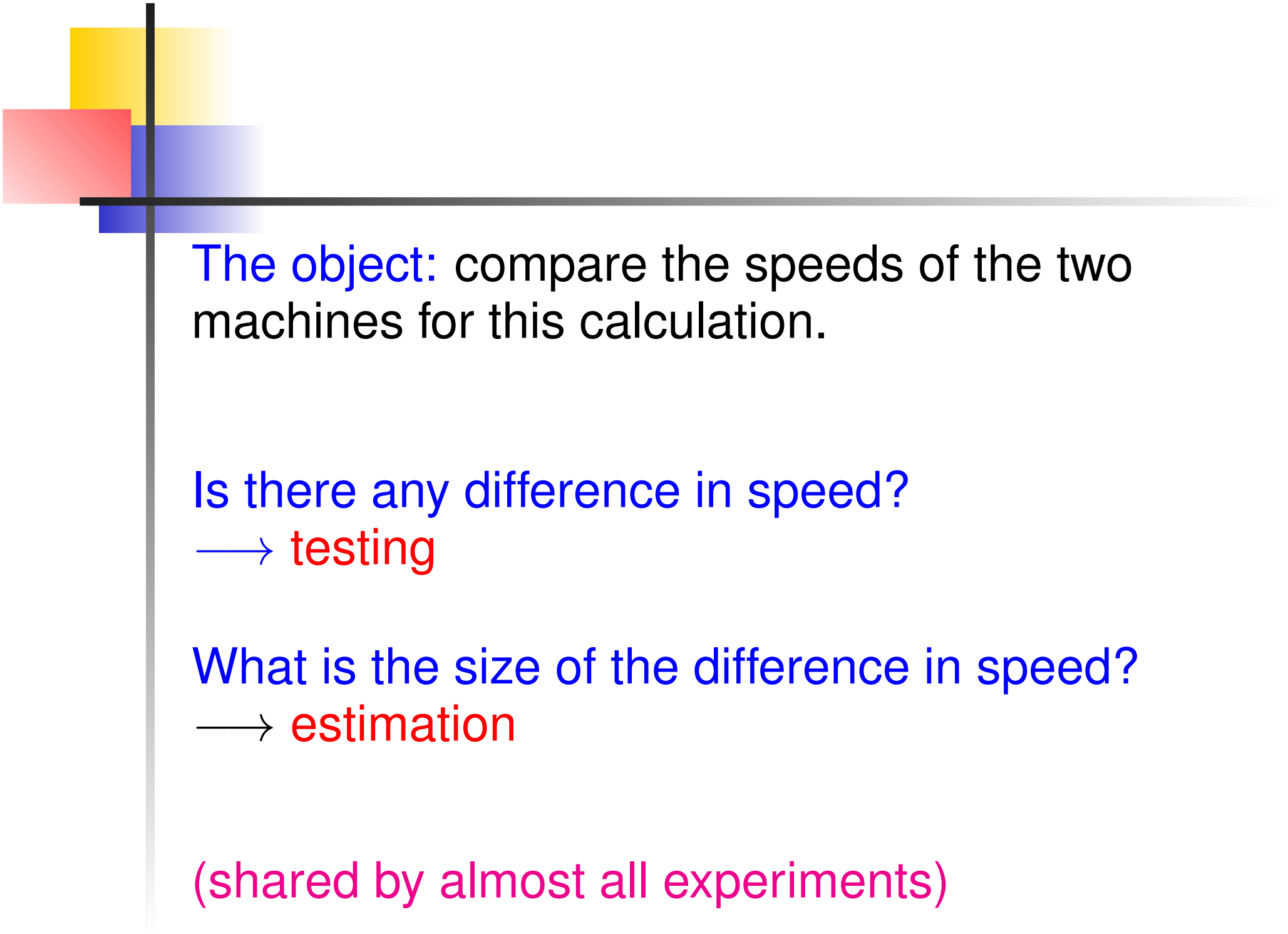
Common characteristic of experiments:
Treatment effects vary from trial to trial

$\longrightarrow$ uncertainty into any conclusions that are drawn from the results

Successive trials may be so discrepant in their results that it is doubtful which treatment would turn out better in the long run.

Ex 1: Time in seconds (minus 2 minutes) required for computing S^2 of 27 observations using two machines A, B

Replication	A	B	$(A-B)$
1	30	14	16
2	21	21	0
3	22	5	17
4	22	13	9
5	18	13	5
6	29	17	12
7	16	7	9
8	12	14	-2
9	23	8	15
10	23	24	-1
Means	21.6	13.6	8.0

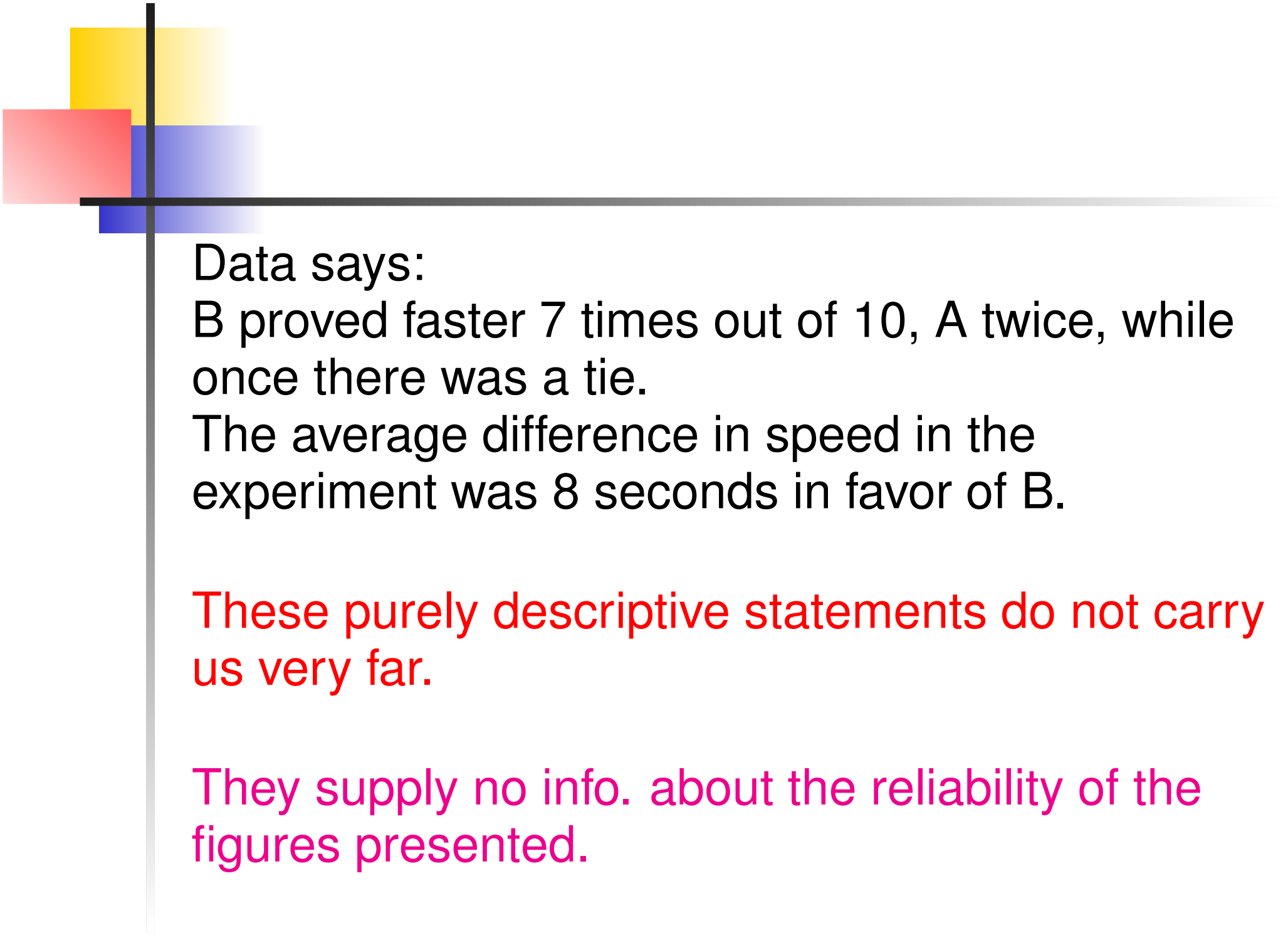


The object: compare the speeds of the two machines for this calculation.

Is there any difference in speed?
$\longrightarrow$ testing

What is the size of the difference in speed?
$\longrightarrow$ estimation

(shared by almost all experiments)



Data says:
B proved faster 7 times out of 10, A twice, while once there was a tie.
The average difference in speed in the experiment was 8 seconds in favor of B.

These purely descriptive statements do not carry us very far.

They supply no info. about the reliability of the figures presented.

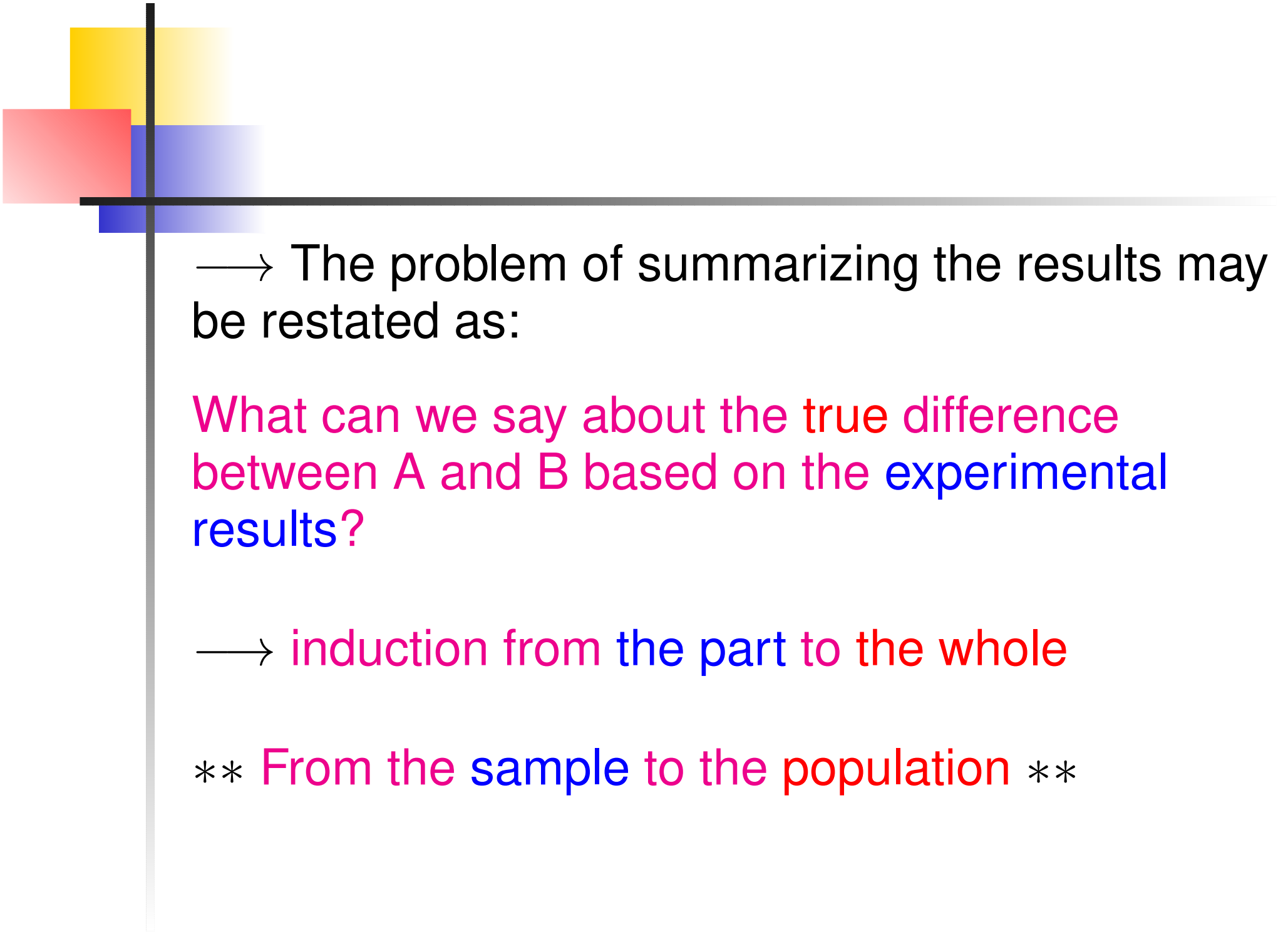
Q: If the experiment were continued for another set of trials, would the advantage at the end still be close to 8 seconds in favor of B?

A point of view

Suppose it were feasible to continue the experiment indefinitely under the same conditions . . . . . the average difference in speed between A, B would presumably settle down to some fixed value (call it the true difference) which is independent of the size of the experiment that was actually carried out.

Stat. point of view in Ex 1

Replication	A	B	$(A-B)$
1	30	14	16
2	21	21	0
3	22	5	17
.	.	.	.
9	23	8	15
10	23	24	-1
.	.	.	.
.	.	.	.
.	.	.	.
$\approx \infty$	.	.	.
Means	?	?	? = true difference θ



$\longrightarrow$ The problem of summarizing the results may be restated as:

What can we say about the true difference between A and B based on the experimental results?

$\longrightarrow$ induction from the part to the whole

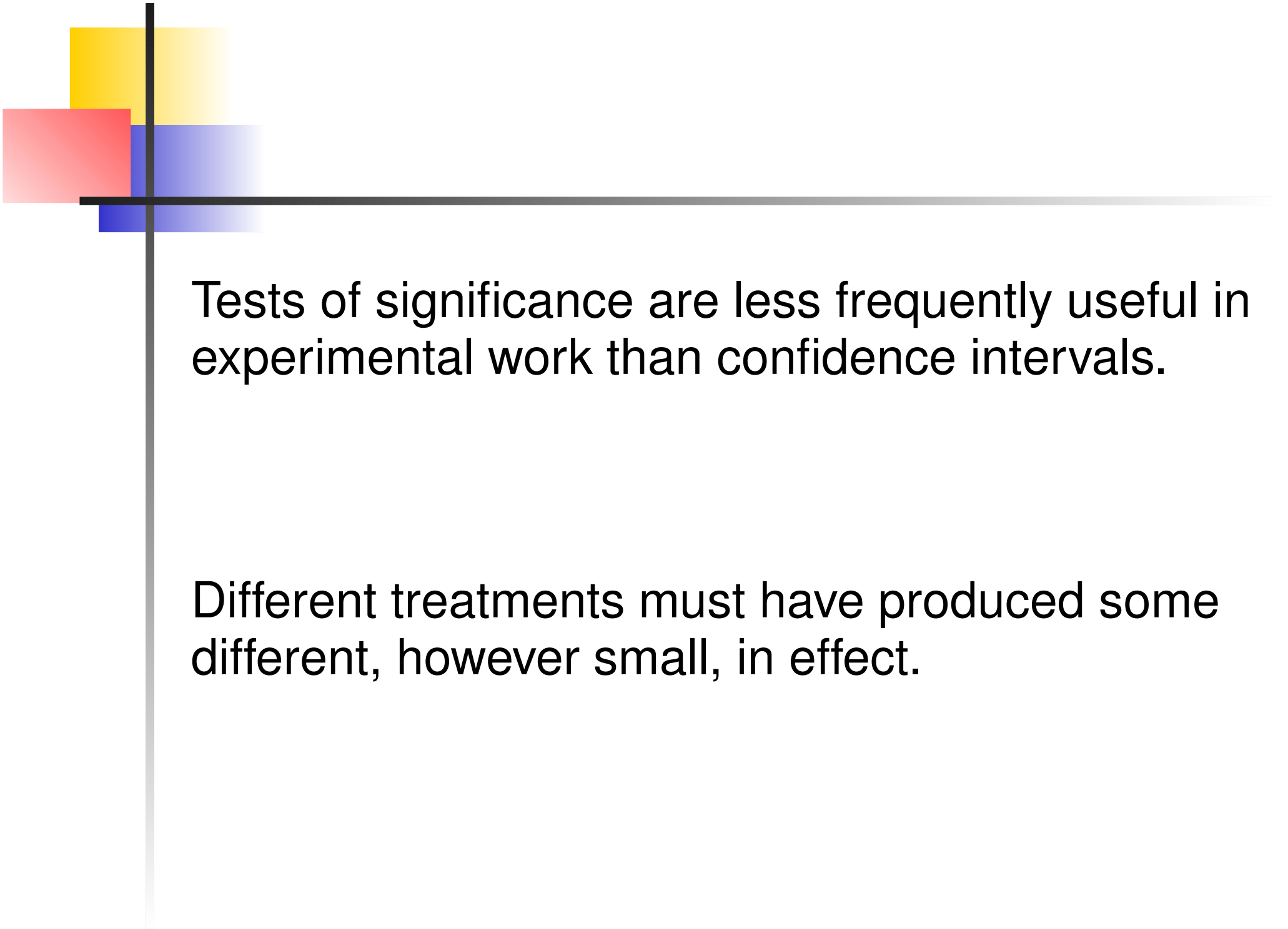
$**$ From the sample to the population $**$

Statistical inference

It cannot be expected that the solution will provide the exact value of the unknown true difference.

Let θ be the true difference
Base on data in Ex 1:

- point est. of $\theta = 8$
- 95% (80%, 99%)conf. int. for θ is $[3.3, 12.7]$ ($[5.1, 10.9]$, $[1.1, 14.9]$)
- the testing: $H_0 : \theta = 0$ is rejected at sig. levels 5%, 20% and 1%, resp'ly.



Tests of significance are less frequently useful in experimental work than confidence intervals.

Different treatments must have produced some different, however small, in effect.

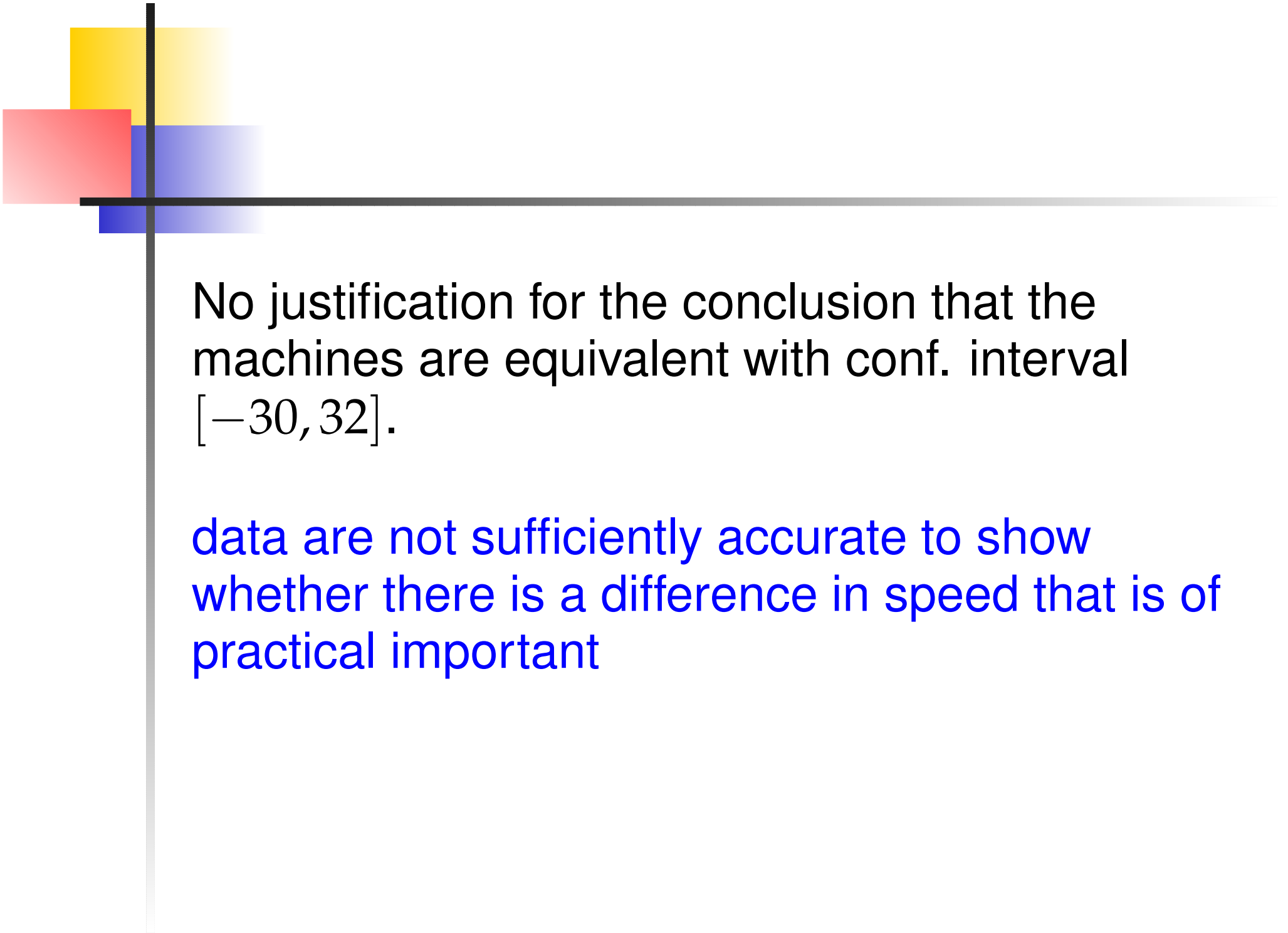
Statistically significant v.s. Practically significant

e.g. 95% intervals: $[-2, 4]$ v.s. $[-30, 32]$

H_0 is not rejected, i.e. the two machines are not sig. different.
$\not\Rightarrow$ they are identical in speed

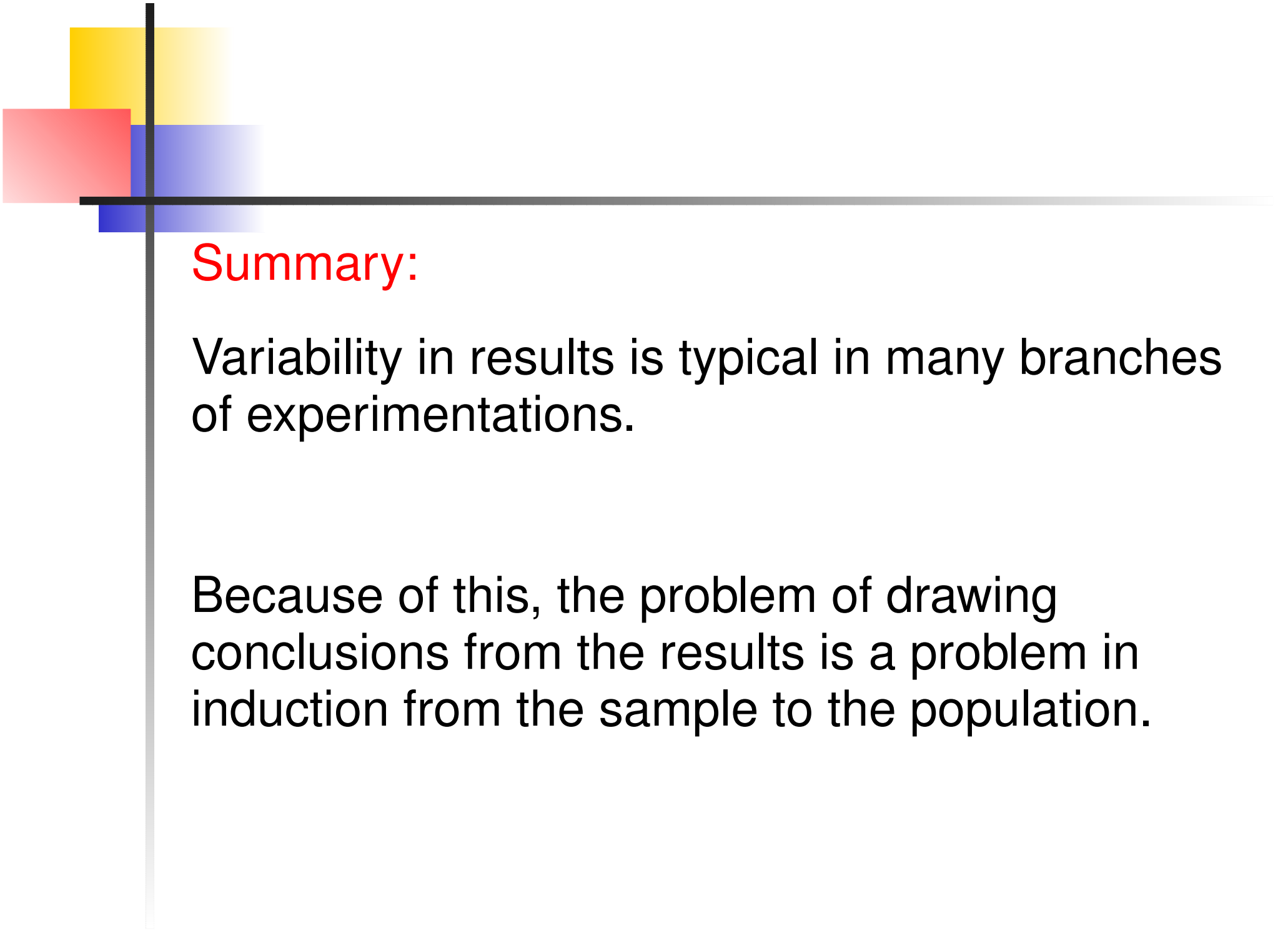
A true difference of 4 seconds, even if it existed, would be of no practical significance.

For all practical purposes the two machines are identical in speed.



No justification for the conclusion that the machines are equivalent with conf. interval $[-30, 32]$.

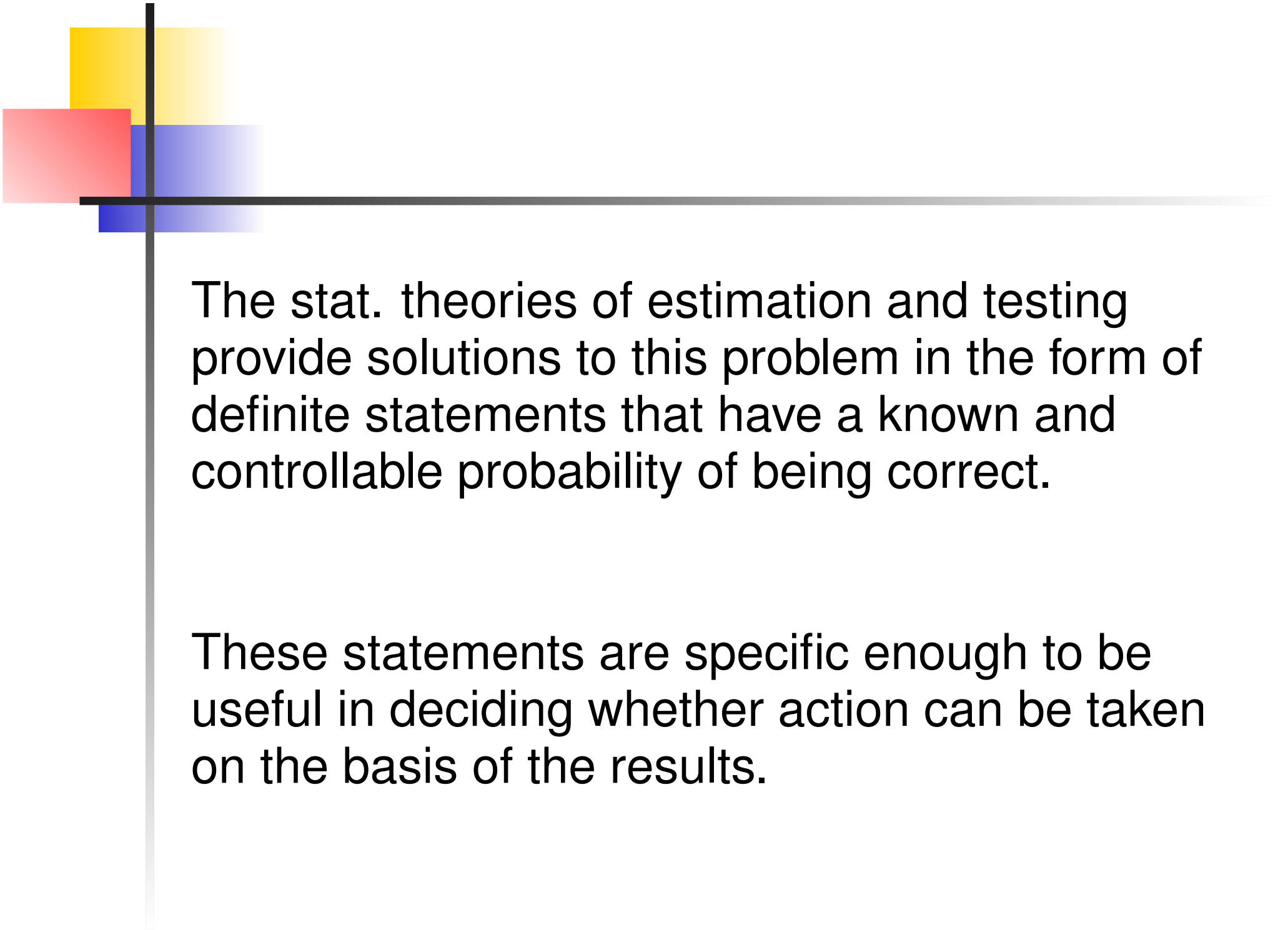
data are not sufficiently accurate to show whether there is a difference in speed that is of practical important



Summary:

Variability in results is typical in many branches of experimentations.

Because of this, the problem of drawing conclusions from the results is a problem in induction from the sample to the population.



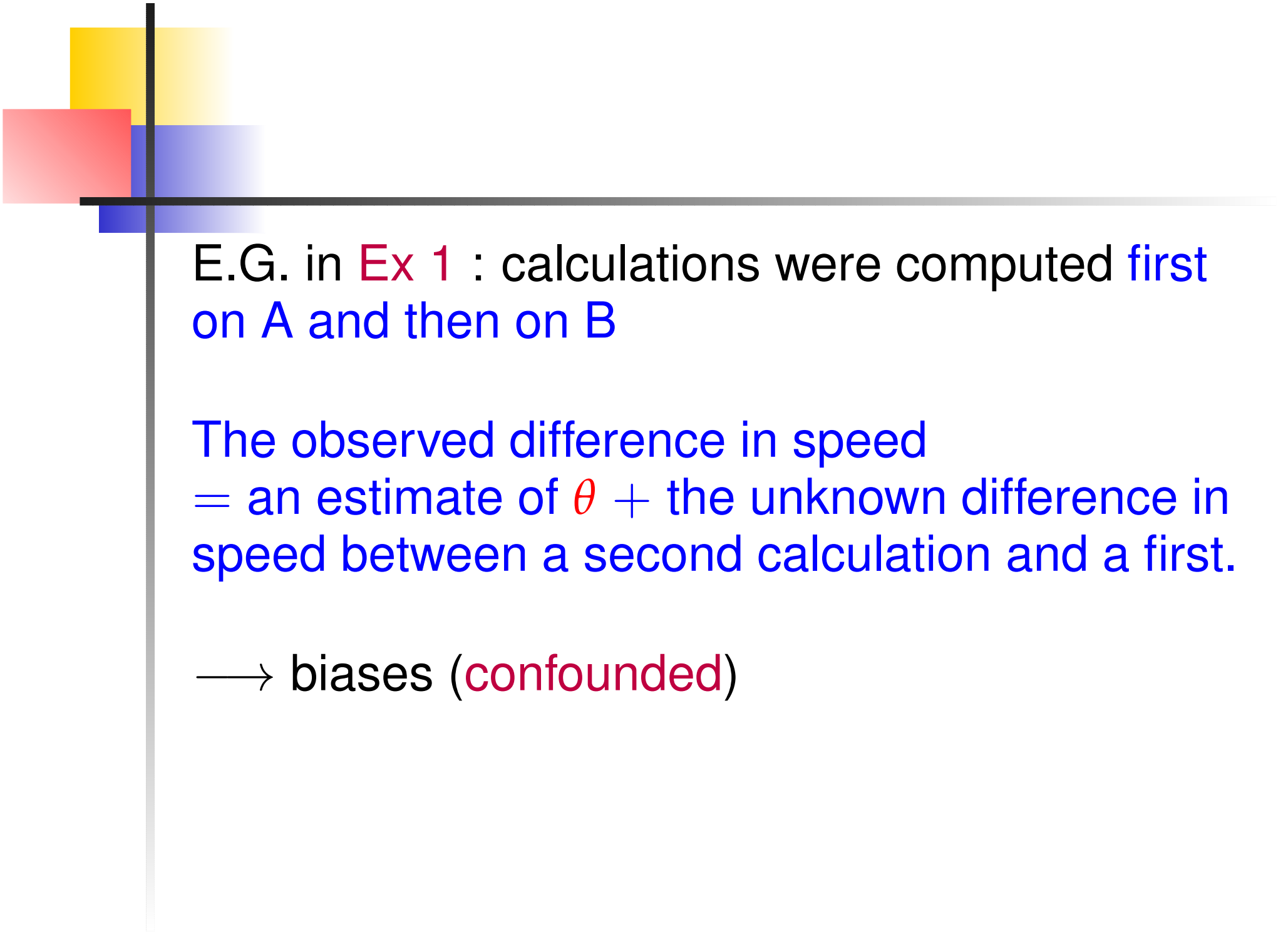
The stat. theories of estimation and testing provide solutions to this problem in the form of definite statements that have a known and controllable probability of being correct.

These statements are specific enough to be useful in deciding whether action can be taken on the basis of the results.

Function of Randomization

The type of statistical inference that can be made from a data set depends on the nature of the data

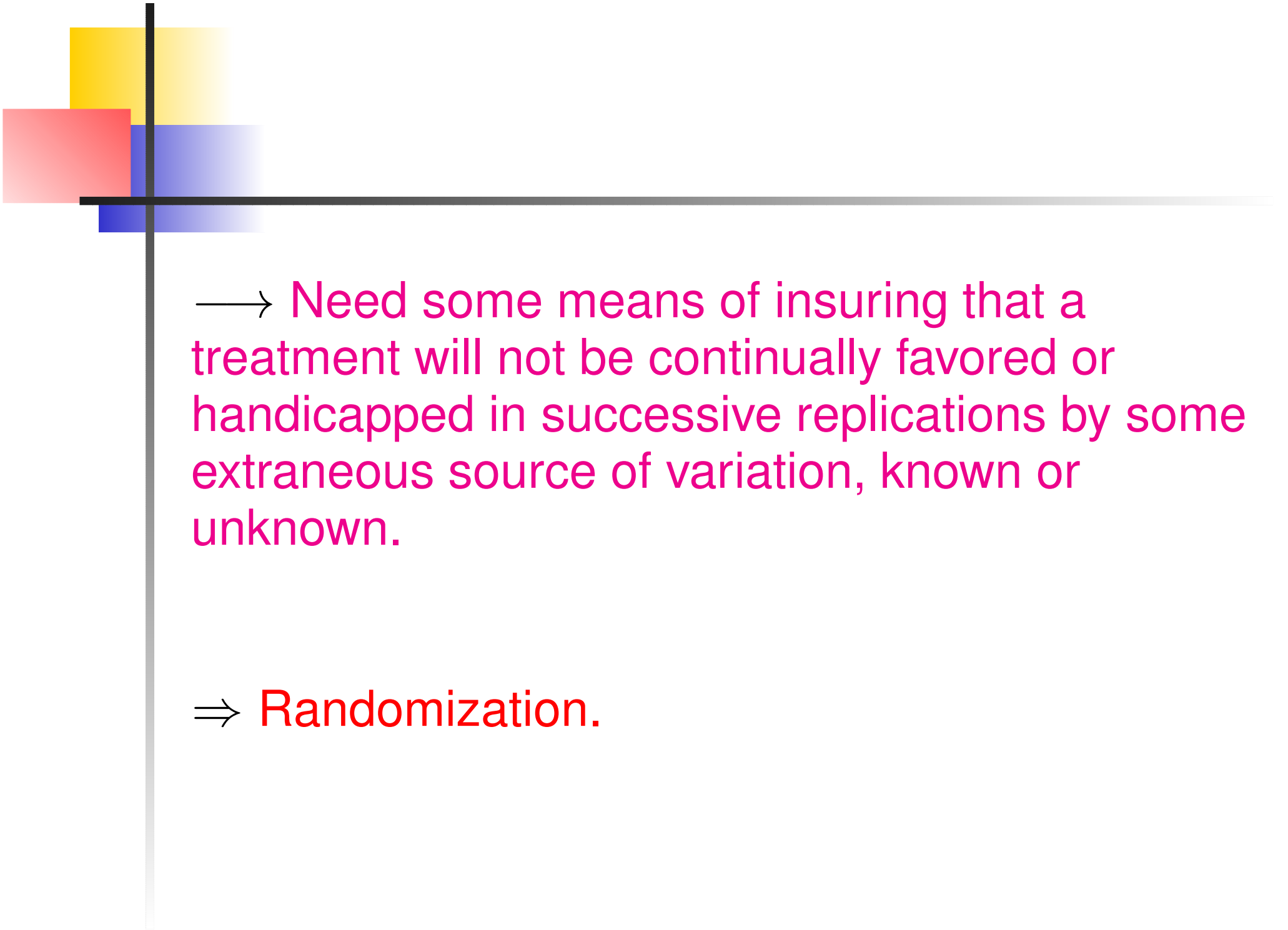
It is easy to conduct an experiment in such a way that no useful inferences can be made.



E.G. in Ex 1 : calculations were computed first on A and then on B

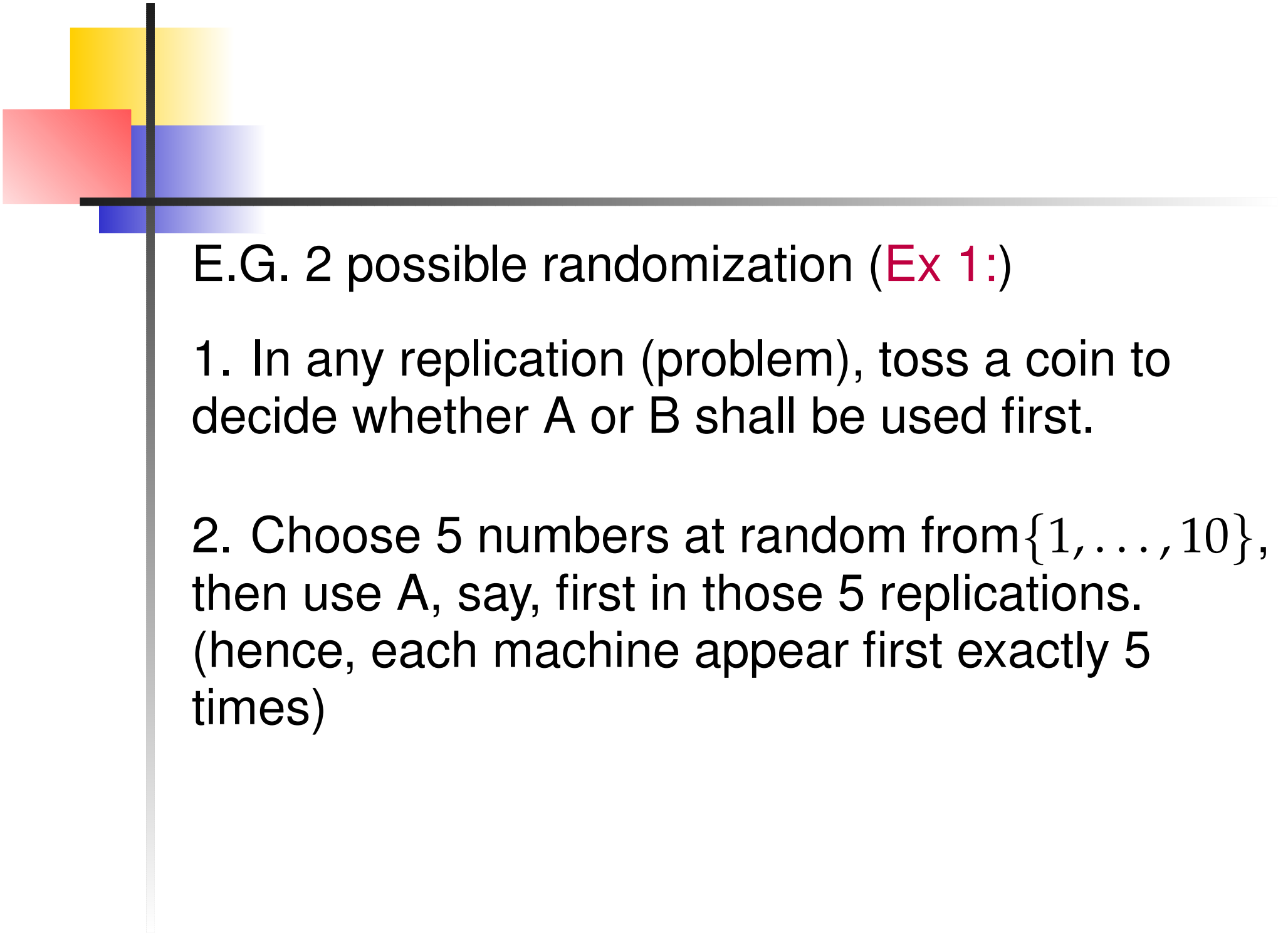
The observed difference in speed
$=$ an estimate of θ $+$ the unknown difference in speed between a second calculation and a first.

$\longrightarrow$ biases (confounded)



$\longrightarrow$ Need some means of insuring that a treatment will not be continually favored or handicapped in successive replications by some extraneous source of variation, known or unknown.

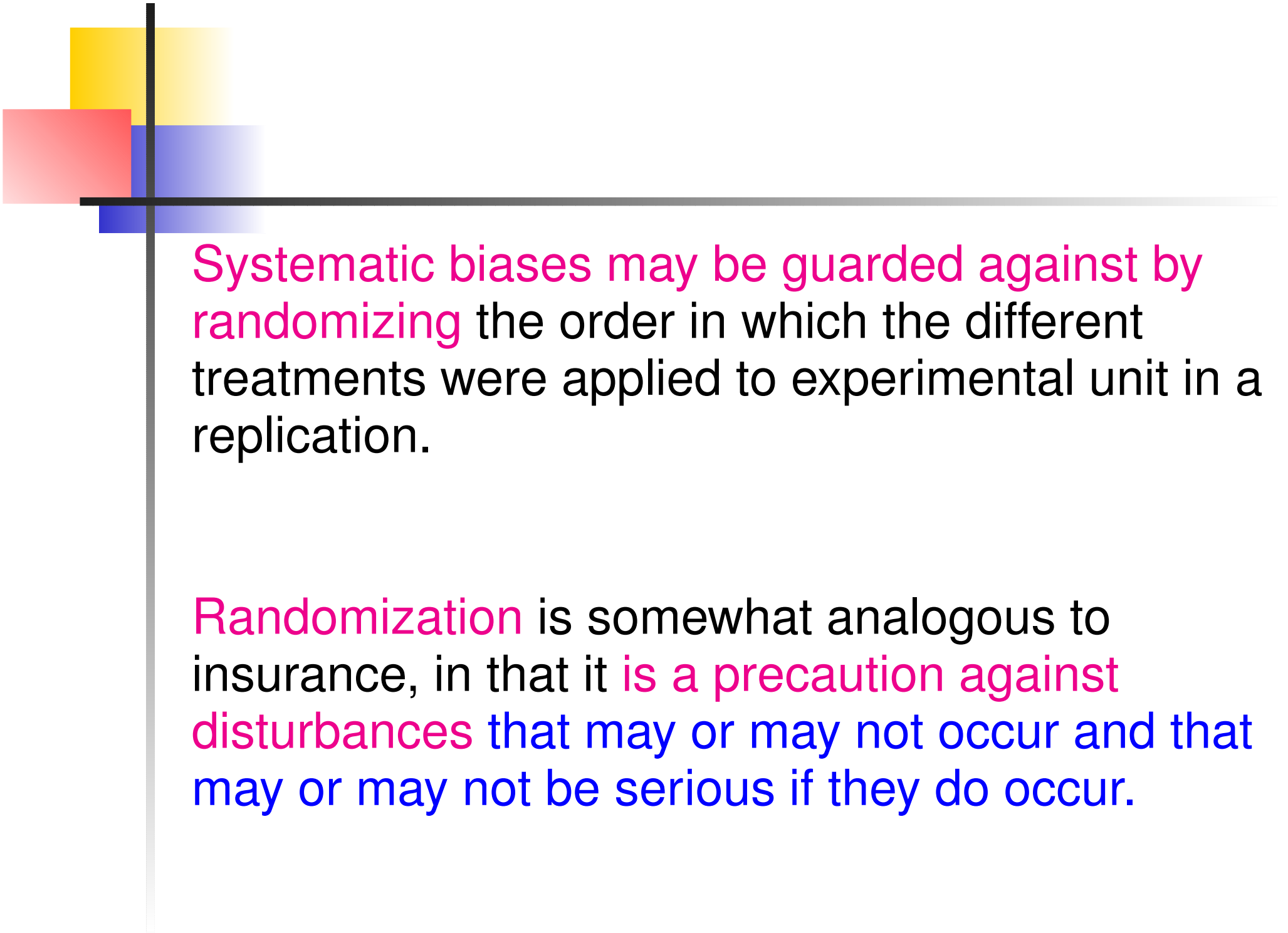
$\Rightarrow$ Randomization.



E.G. 2 possible randomization (Ex 1:)

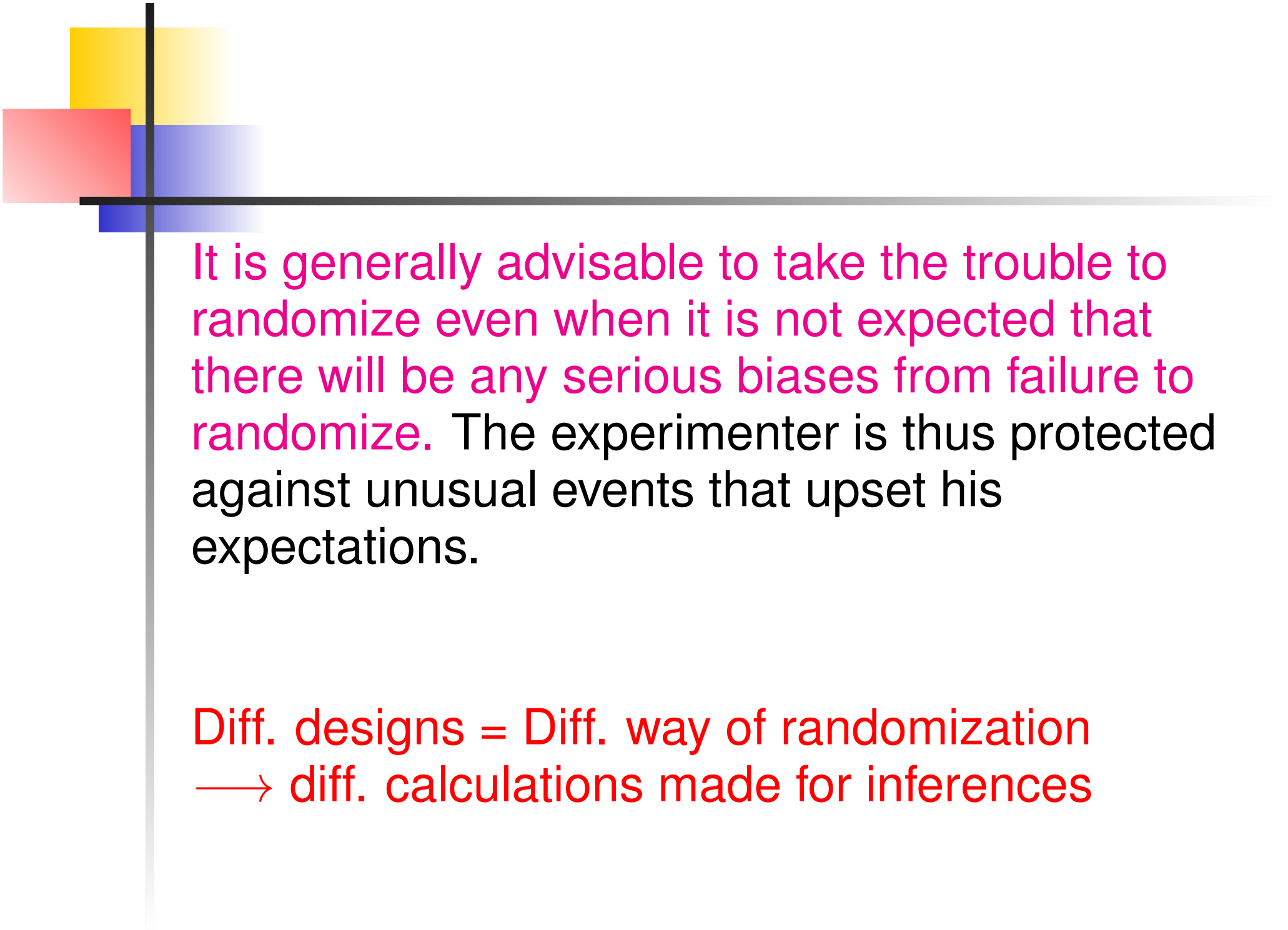
1. In any replication (problem), toss a coin to decide whether A or B shall be used first.

2. Choose 5 numbers at random from $\{1, \ldots, 10\}$, then use A, say, first in those 5 replications. (hence, each machine appear first exactly 5 times)



Systematic biases may be guarded against by randomizing the order in which the different treatments were applied to experimental unit in a replication.

Randomization is somewhat analogous to insurance, in that it is a precaution against disturbances that may or may not occur and that may or may not be serious if they do occur.



It is generally advisable to take the trouble to randomize even when it is not expected that there will be any serious biases from failure to randomize. The experimenter is thus protected against unusual events that upset his expectations.

Diff. designs = Diff. way of randomization
$\longrightarrow$ diff. calculations made for inferences

Local control – Blocking

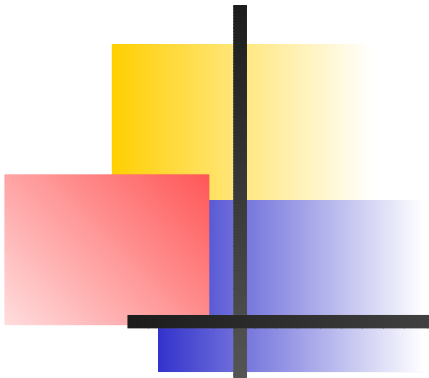
to reduce experimental errors and make the experiment more powerful

by suitable restrictions on the randomization of treatments to E.U.'s.

E.G.(Ex 1:)

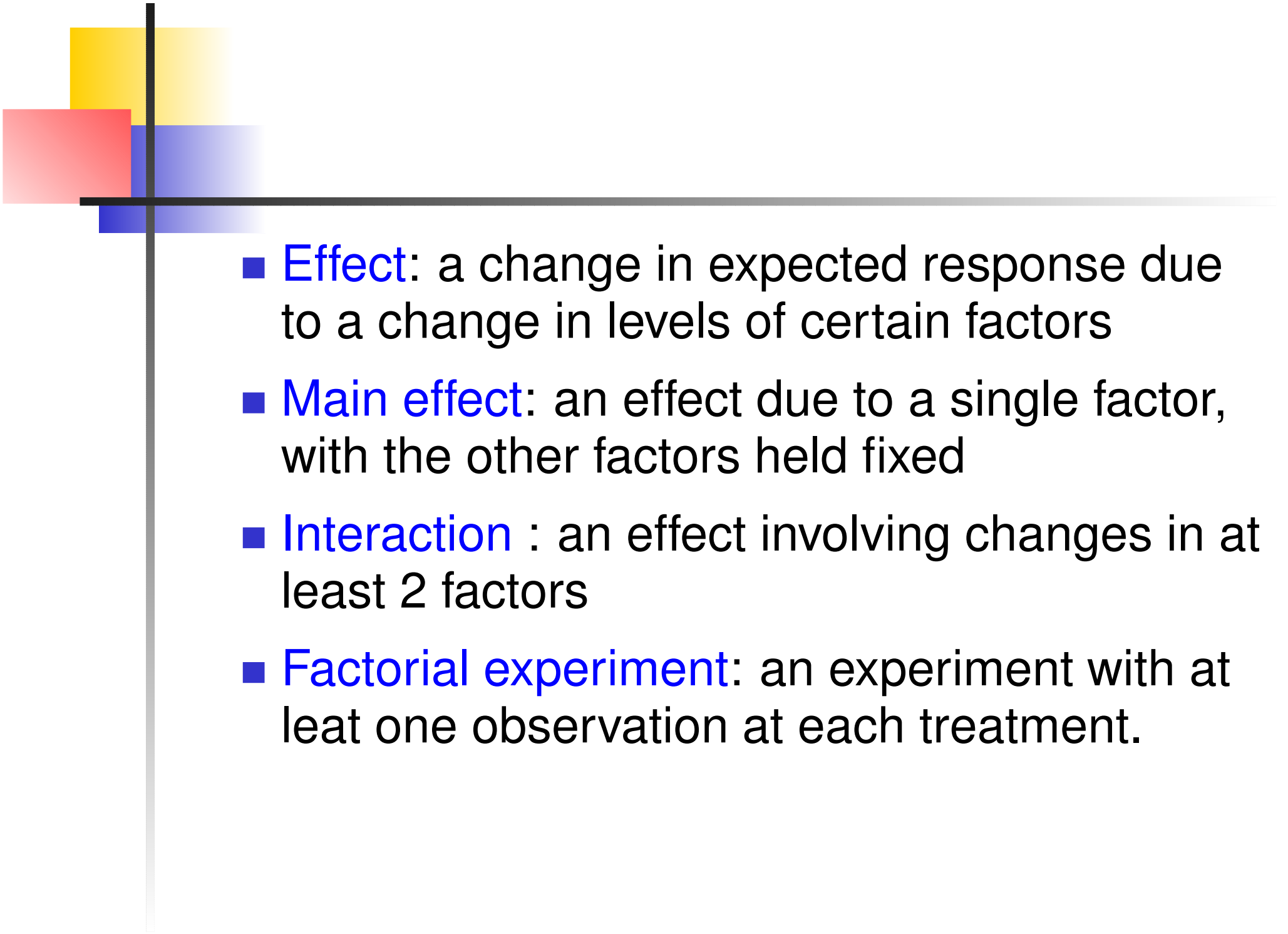
1. In any replication, toss a coin to decide whether A or B shall be used first.
$\longrightarrow$ 1 factor (machine), 1 block (problem at 10 levels)

2. Choose 5 numbers at random from $\{1, \ldots, 10\}$, then use A, say, first in those 5 replications.
$\longrightarrow$ 2 factors (machine at 2 levels; order at 2 levels), 1 block (problem at 10 levels)

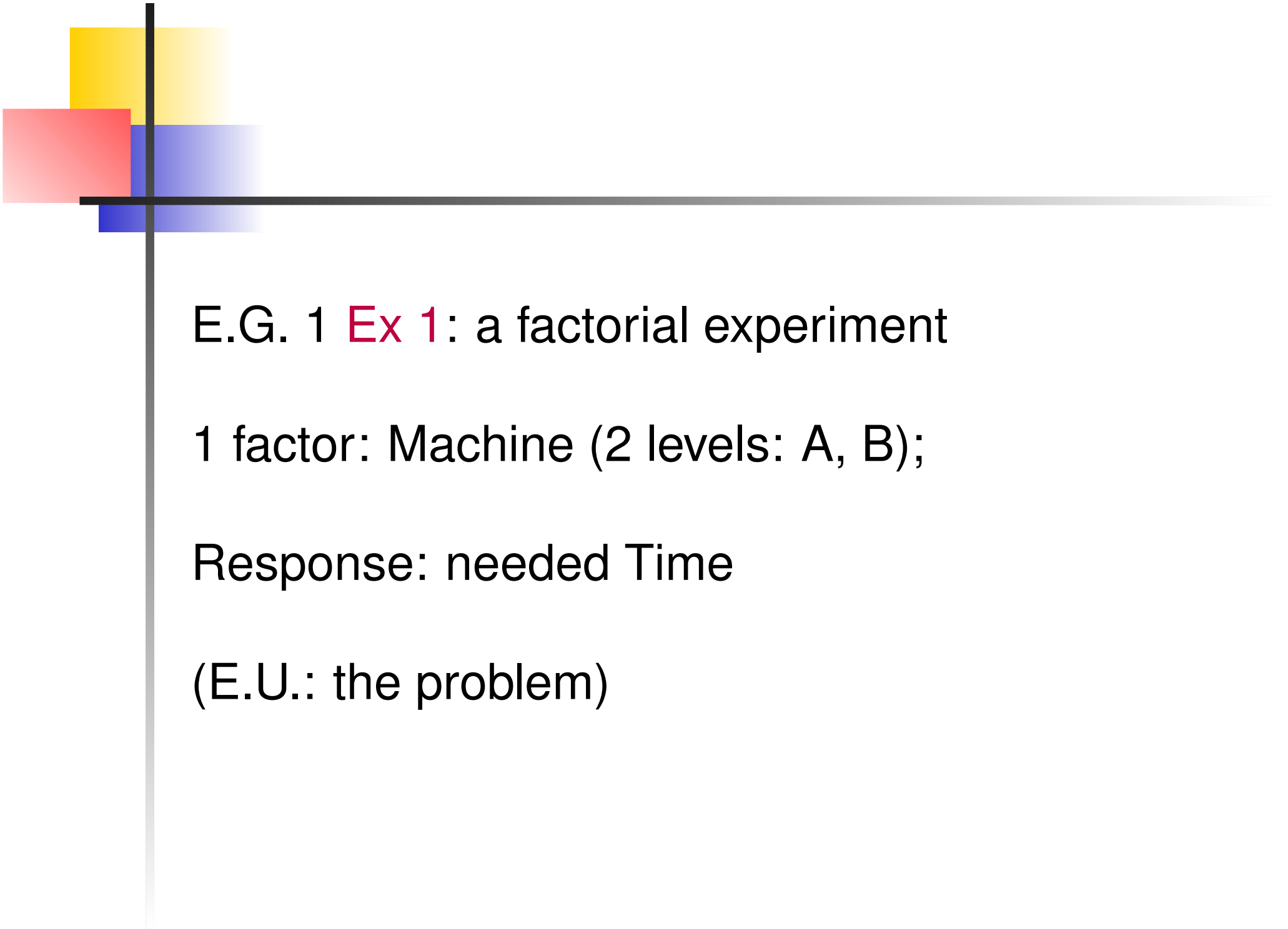


Terminology

- Factor: a variable which can be controlled
  Fixed effect /Random effect
- Factor level: a possible value of the factor
- Experimental unit: an item on which one can run an experiment (c.f. observational unit)
- Treatment: factor-level combination
- Response variable: the characteristic of an experimental unit that is measured.



- Effect: a change in expected response due to a change in levels of certain factors
- Main effect: an effect due to a single factor, with the other factors held fixed
- Interaction : an effect involving changes in at least 2 factors
- Factorial experiment: an experiment with at leat one observation at each treatment.

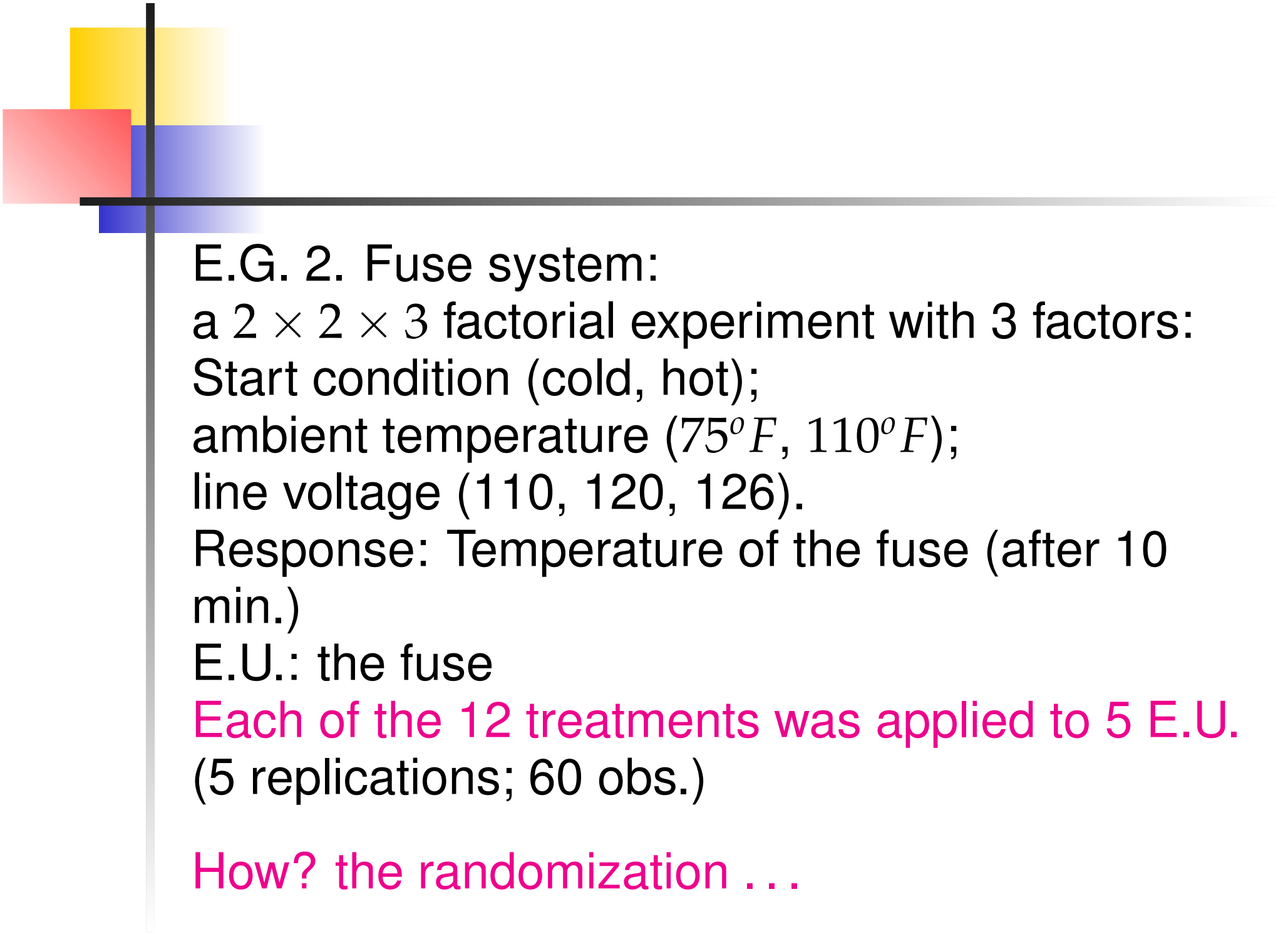


E.G. 1 Ex 1: a factorial experiment

1 factor: Machine (2 levels: A, B);

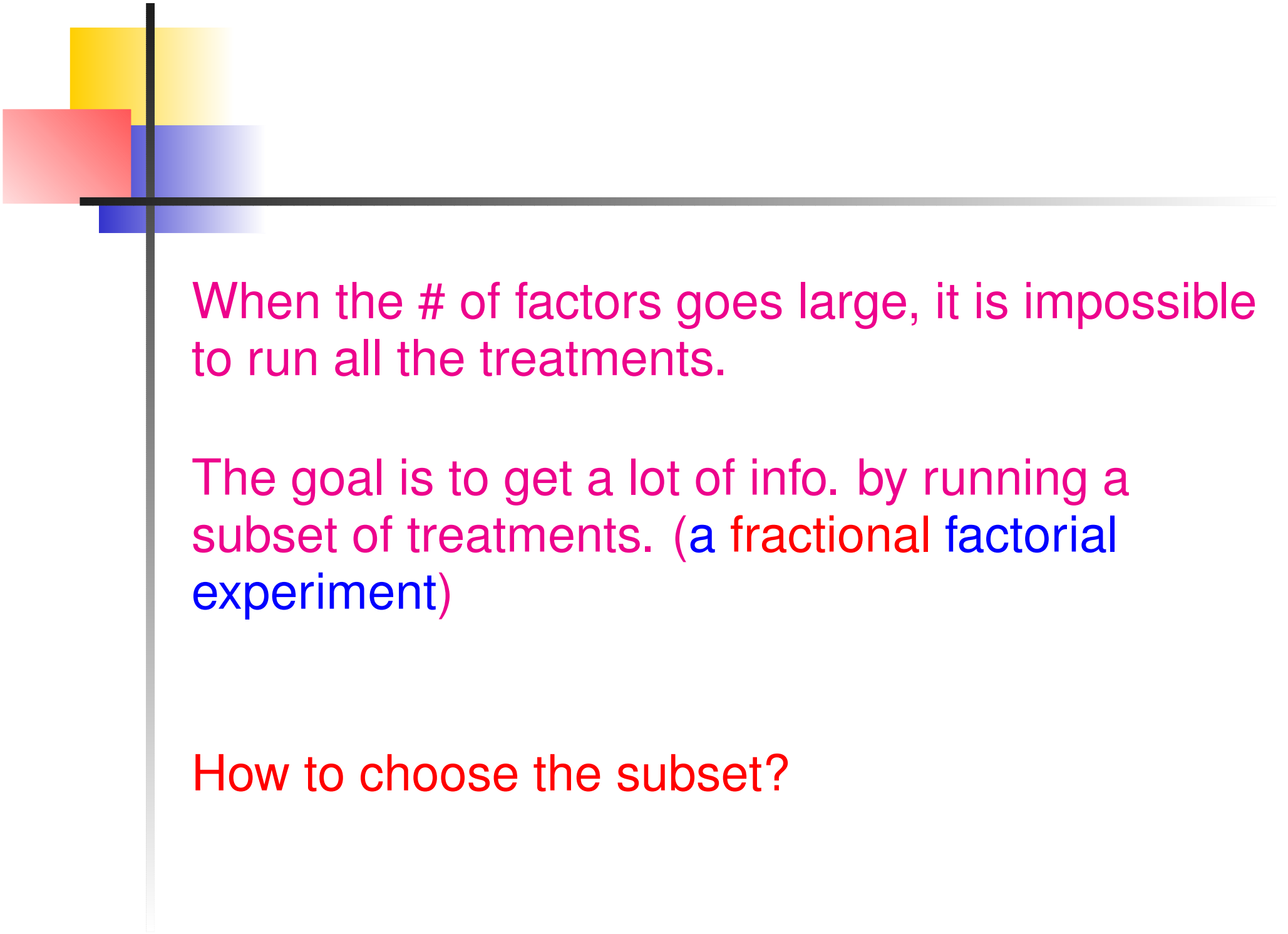
Response: needed Time

(E.U.: the problem)



E.G. 2. Fuse system:
a $2 \times 2 \times 3$ factorial experiment with 3 factors:
Start condition (cold, hot);
ambient temperature (75^oF, 110^oF);
line voltage (110, 120, 126).
Response: Temperature of the fuse (after 10 min.)
E.U.: the fuse
Each of the 12 treatments was applied to 5 E.U.
(5 replications; 60 obs.)

How? the randomization . . .



When the # of factors goes large, it is impossible to run all the treatments.

The goal is to get a lot of info. by running a subset of treatments. (a fractional factorial experiment)

How to choose the subset?

Experimental errors

The results of experiments are affected not only by the action of treatments, but also by extraneous variations (experimental errors) which tend to mask the effects of the treatments.

- inherent variability in E.U.'s
- lack of uniformity in the physical conduct

To reduce experimental errors

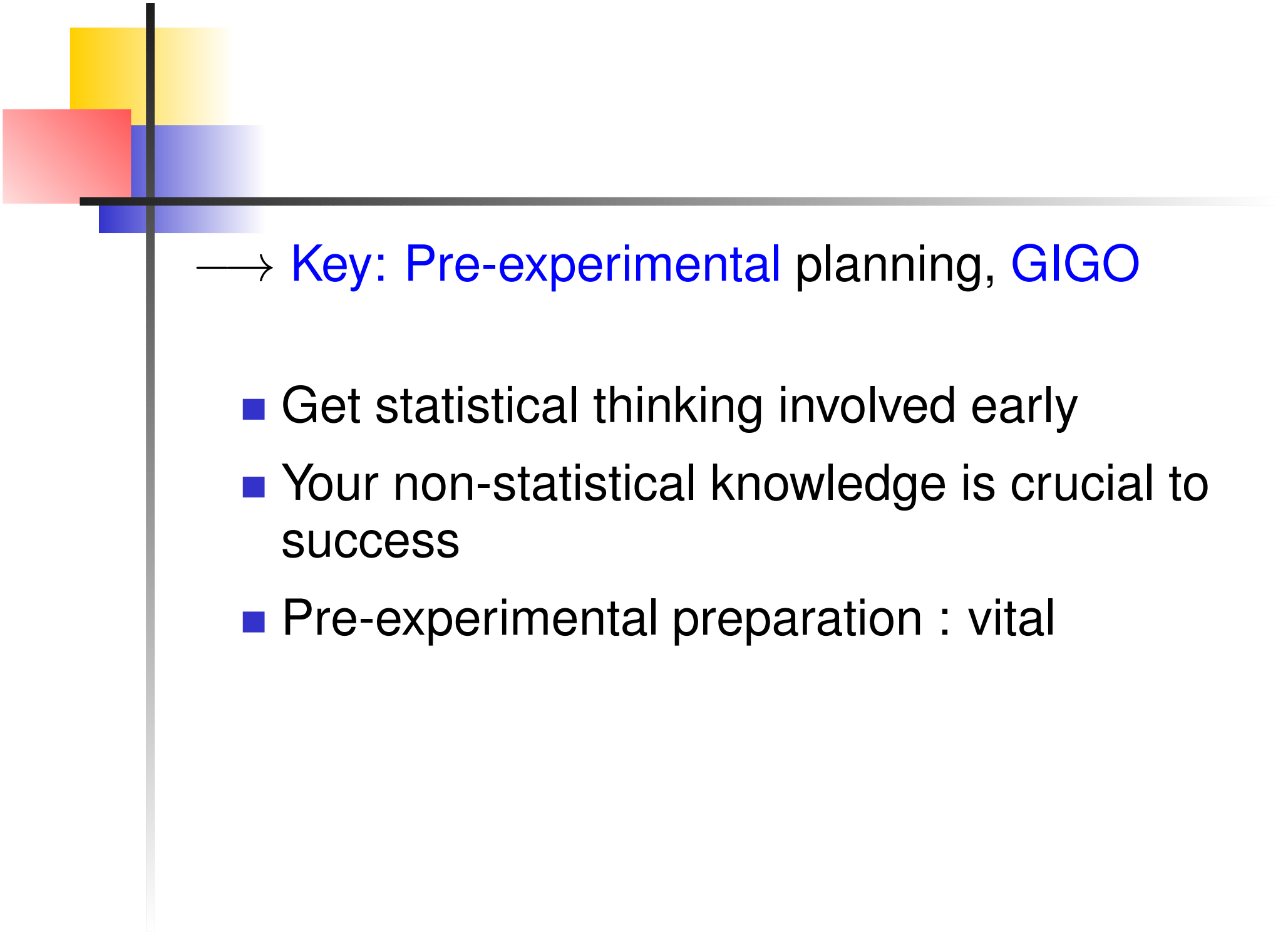
- Properly handle independent variables
  1. rigidly controlled – the factors remain fixed throughout the experiment
  2. manipulated or set at levels of interest
  3. randomization – to average out the errors that cannot be controlled
- Carefully select the E.U.'s so that they are closely comparable $\rightarrow$ Blocking, sometimes
- Refine the experimental technique
- Increase the size of the experiment

Planning the experiment

Inferences can be made depend on the way the experiment was carried out . . .

Need: a detailed description of the experiment and it objectives; A written proposal with

1. a statement of the objectives
2. description of the experiment including treatments, size, E.U., randomization . . .
3. an outline of the method of analysis of the results



$\longrightarrow$ Key: Pre-experimental planning, GIGO

- Get statistical thinking involved early
- Your non-statistical knowledge is crucial to success
- Pre-experimental preparation : vital

Conducting the experiment(s)

In practice, cost-control, effectiveness . . .

$**$ Think and experiment sequentially

Experiments:
pilot (screening, exploratory) / confirmatory

Scale: small / large

$\hookrightarrow$ factorsss. . ., few; 2, 3 levels / few factors, levelsss. . .

$\hookrightarrow$ fractional; ANOVA / factorial; ANOVA, Reg, . . .

$*$ Several pilots (now to then) $+$ 1 Confirmatory