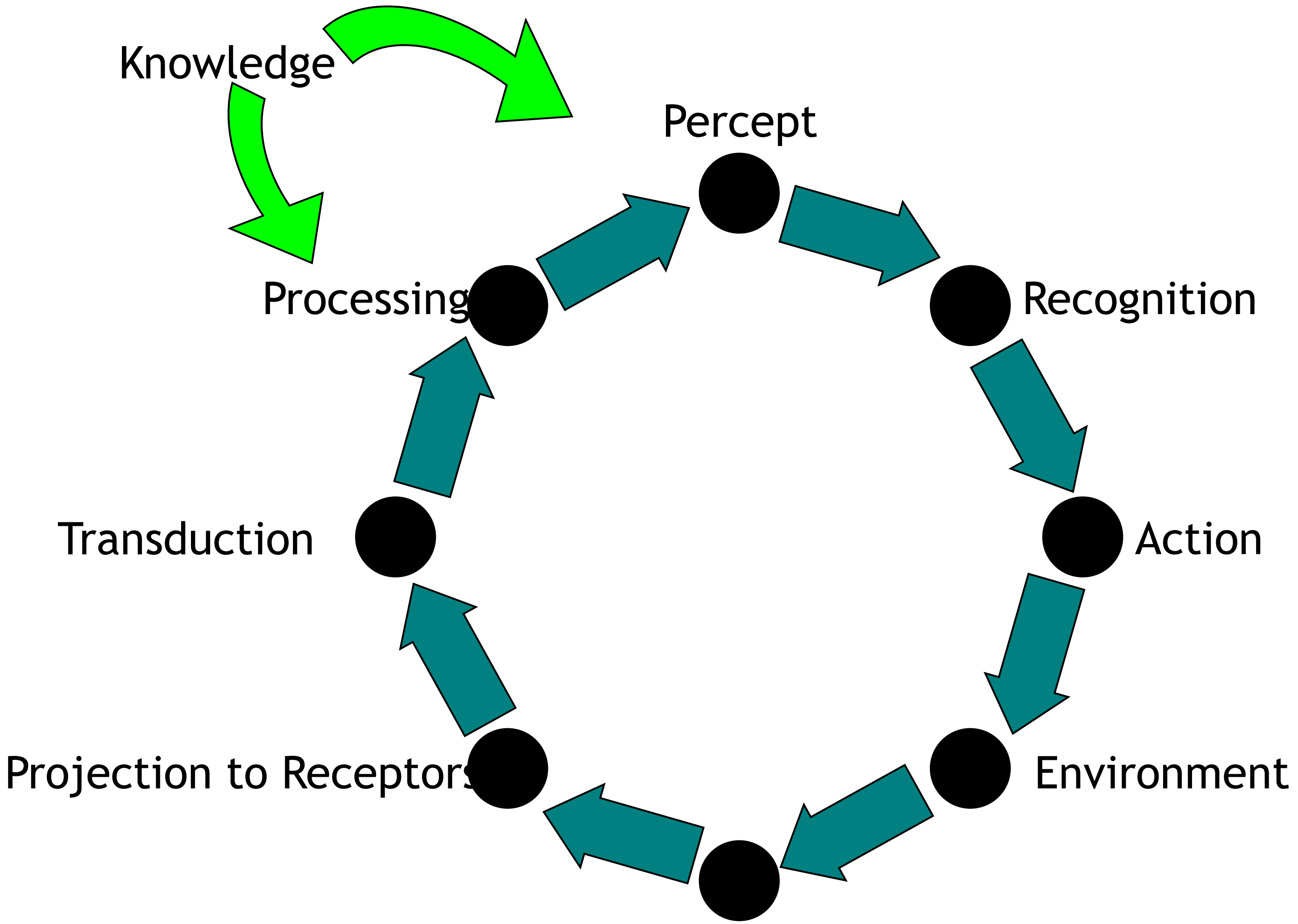
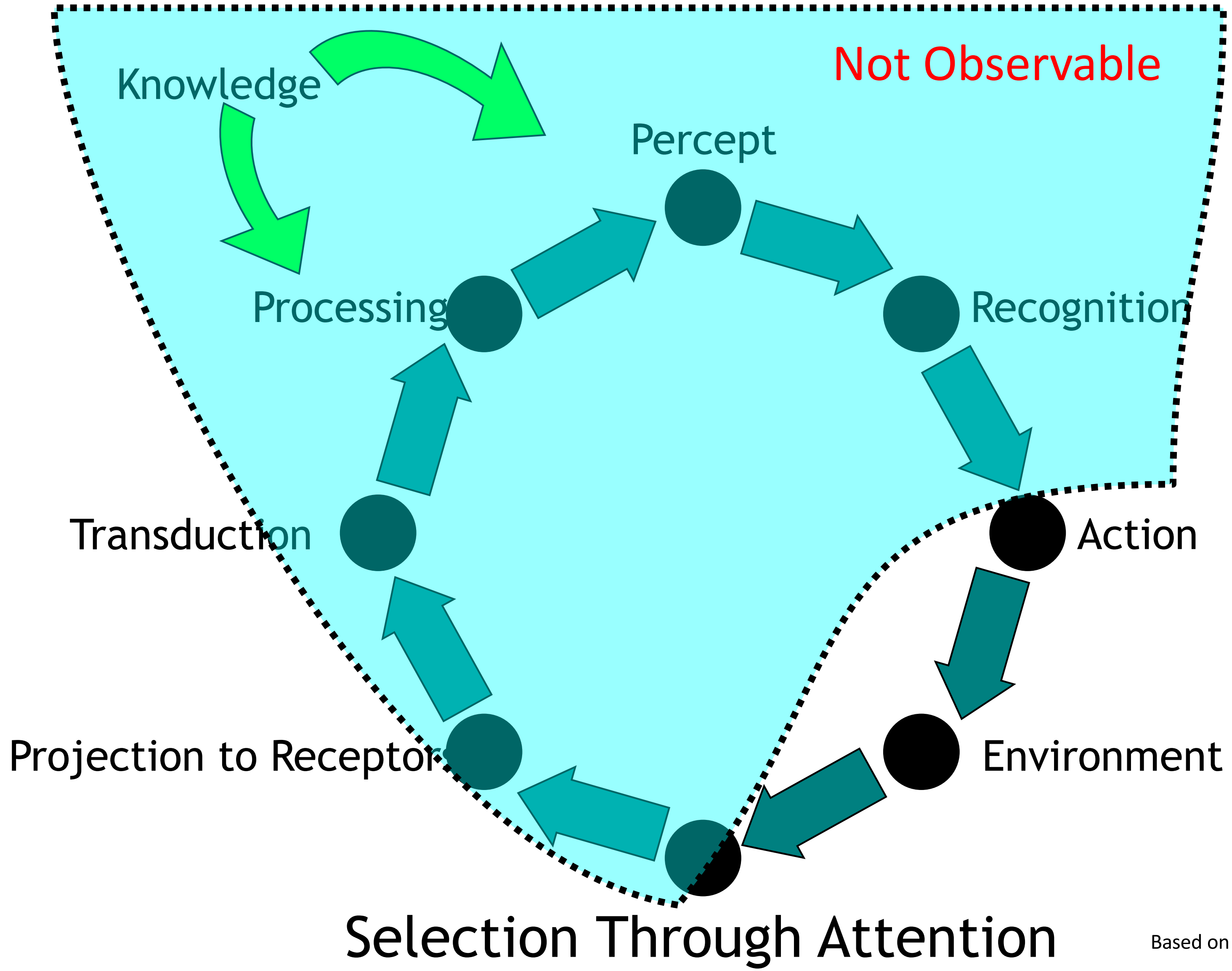
Why is this hard?



Selection Through Attention

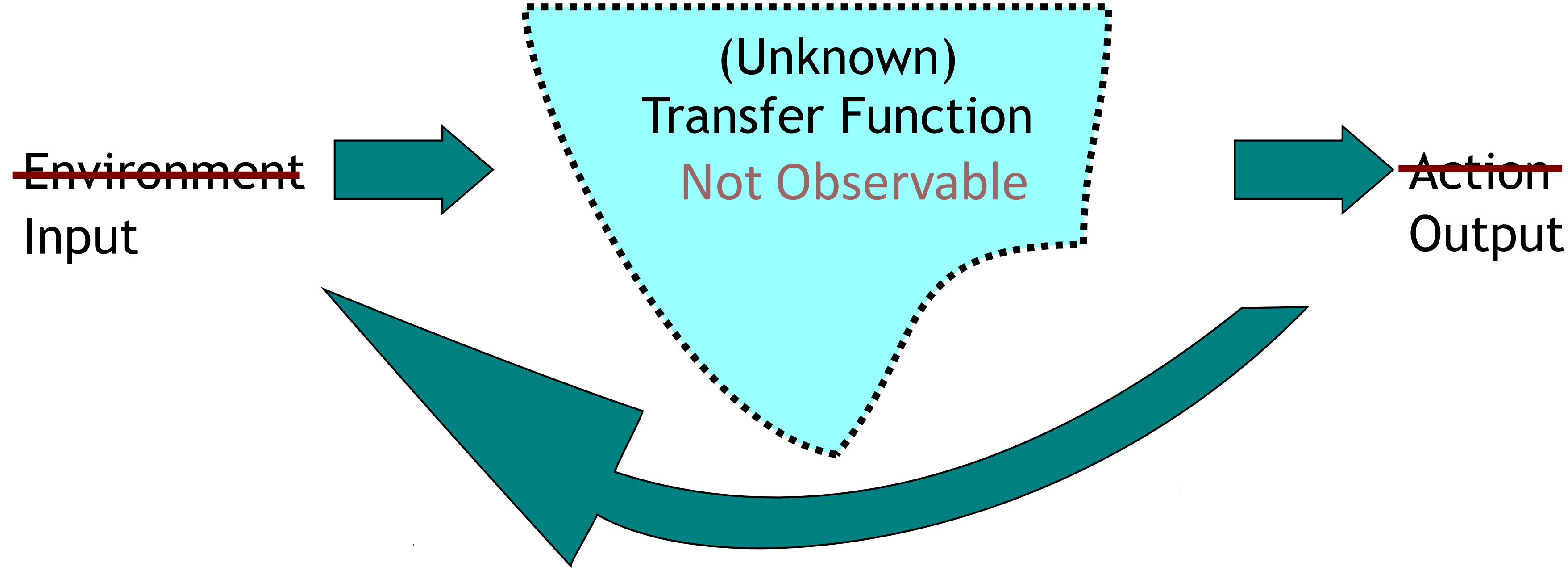
Based on B. Goldstein 2002

Why is this hard?



Based on B. Goldstein 2002

The classic **black** box



Dynamically adaptable, non-linear system
with feedback
(and probably feed-forward...)

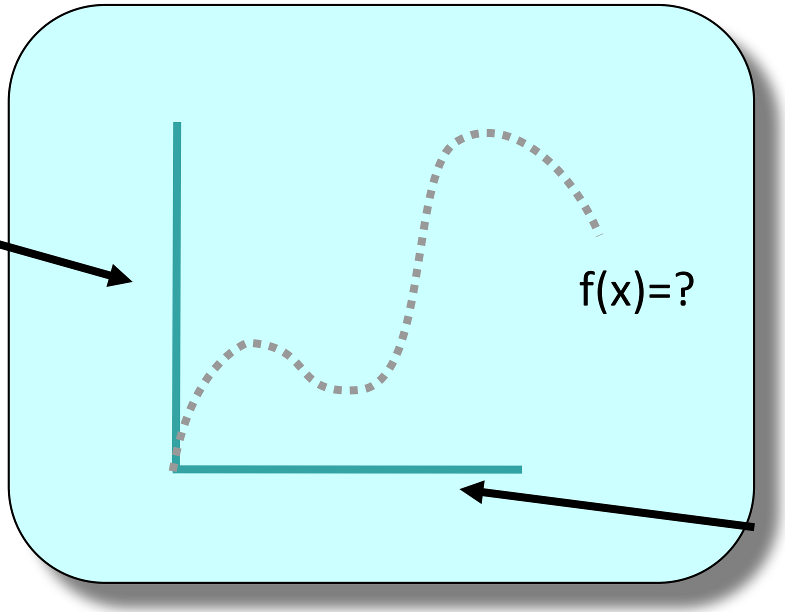
A different (black) box

CognitiveSystems

Goal: Estimate an unknown function

"Action" or
Output or
measured behavior

Dependent Variable



"Environment" or
Input or
Parameter Values or
Stimulus

Independent variable

Method: ??

Based on Cunningham & Wallraven, 2011

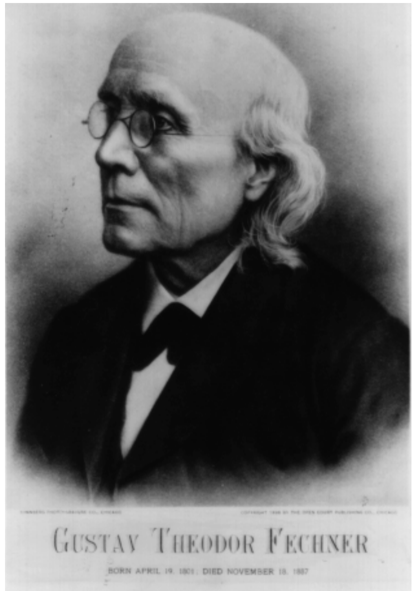
Degree of Control
To interfere or not?

- No control: Observational Research
  - examine things as they happen
  - often uses correlational analyses
  - Examples: Astronomy, Anthropology, Zoology, …
- Complete control: Experimental or controlled research.
  - repeatedly and reliably produce a specific event in order to examine it
  - can talk about causation
  - Examples: Physics, Chemistry, Perceptual Psychology,...

Degree of Control
Type of experimental studies

- Controlled experiment:
  - Examines more general questions (about the underlying parameters of the system and its influence on the participant)
  - Requires control over
    - what is shown (stimuli)
    - how and when the stimuli are shown (experimental procedure)
    - what the participants should do (task)

Degree of Control
Type of experimental studies



- An example of controlled experimental methodology in psychology is Psychophysics:
  - Set of experimental methodology invented by Gustav Fechner in 1860 (and since extended by lots of people).
  - Provides mathematical descriptions of the functional relationship between variations in the physical world and the resulting variations in the psychological (or perceptual) world.
  - Requires very fine control over
    - what is shown (stimuli)
    - how and when the stimuli are shown (experimental procedure)
    - what the participants should do (task)

Degree of Control

- Powerful experiments in psychology (such as those in psychophysics) can be thought of as "exploring the parameter space" of a technique, procedure, or algorithm in order to determine the function relationship between parameter values and perceptual effects (possibly, in order to optimize the technique, etc.)

Control

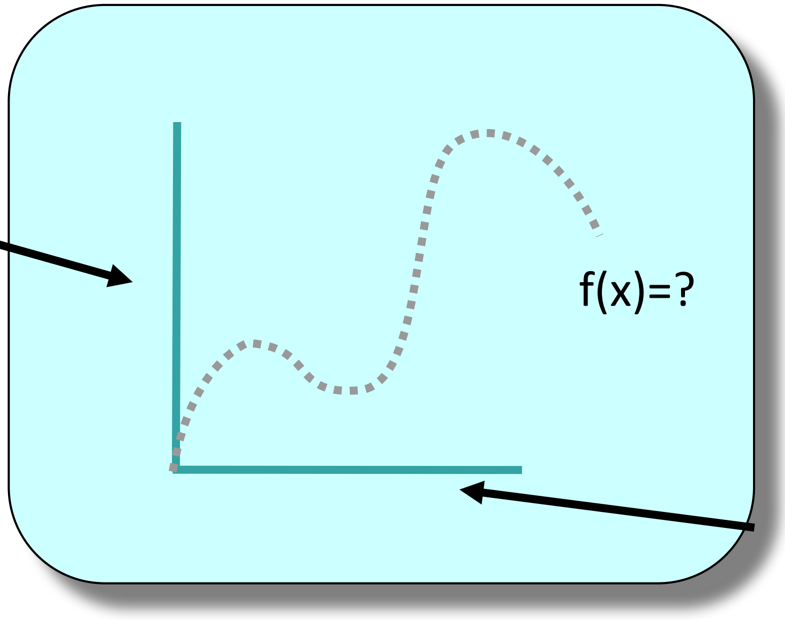
- WARNING: The increased flexibility in answering questions that perceptual and psychophysical experiments offer comes at the cost of a need for increased vigilance, rigor, and expertise.

Flashback: Why is this hard?

Goal: Estimate an unknown function

“Action” or
Output or
measured behavior

Dependent Variable



“Environment” or
Input or
Parameter Values or
Stimulus

Independent variable

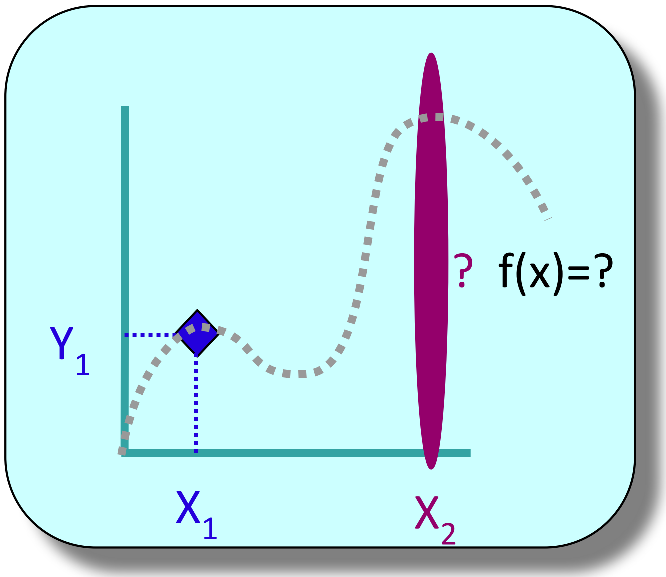
Method: Systematically sample the function

Fundamental compromise

Specificity

- You can only talk about what you measured

Generality

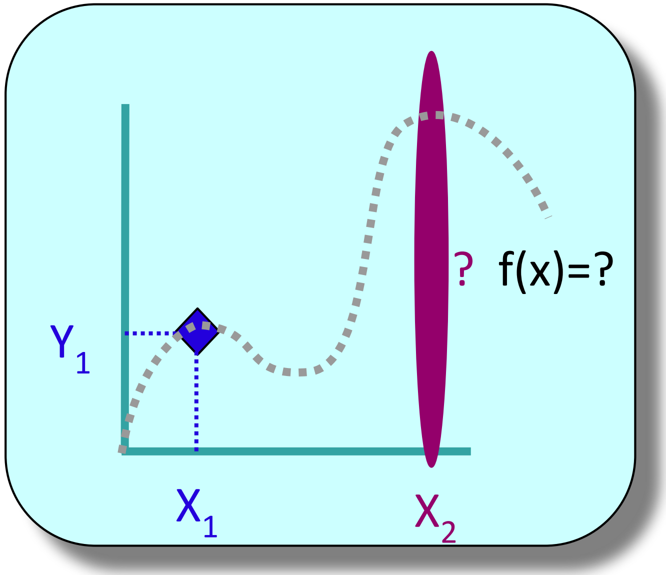


Fundamental compromise

CognitiveSystems

Specificity

•You can only talk about what you measured



Based on Cunningham & Wallraven, 2011

Fundamental compromise

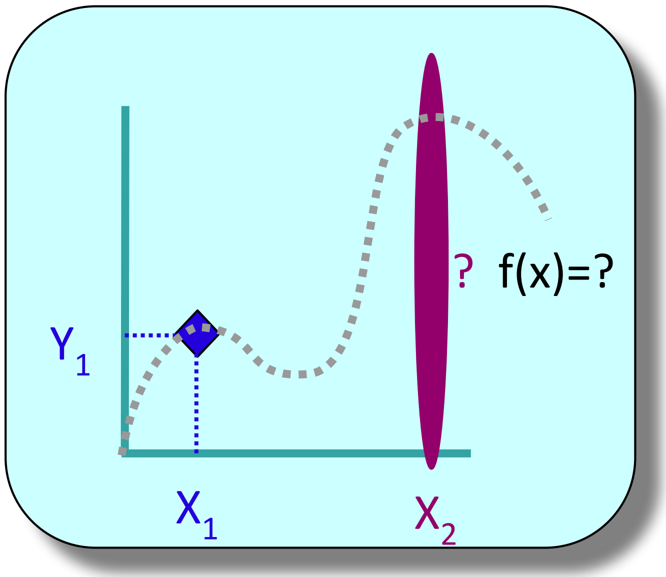
CognitiveSystems

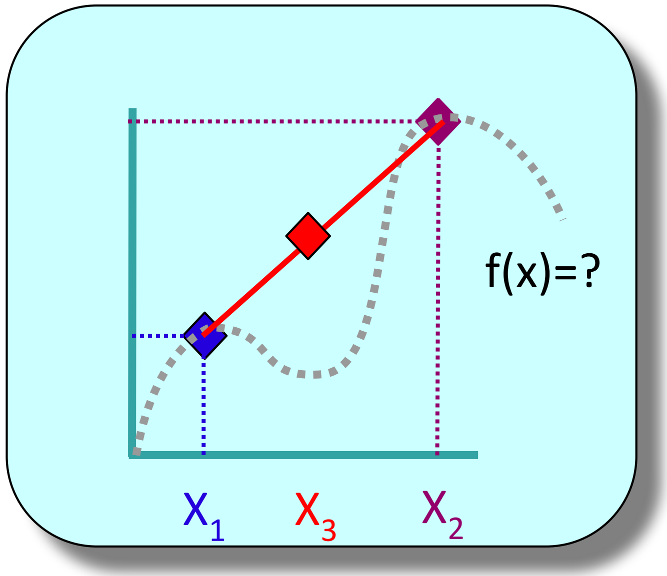
Specificity

•You can only talk about what you measured

Generality

•Make broad statements without measuring every point...**Interpolate!**





Based on Cunningham & Wallraven, 2011

Fundamental Compromise

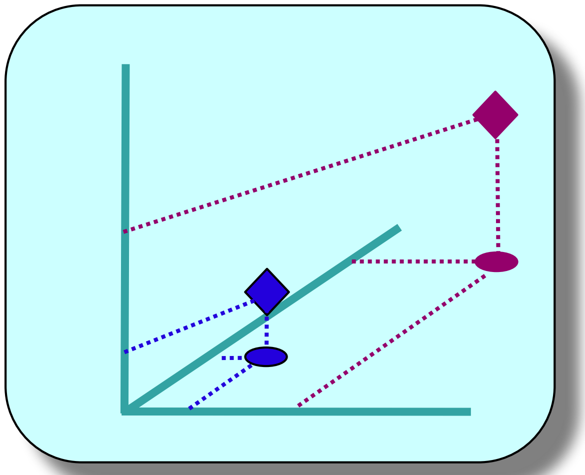
CognitiveSystems

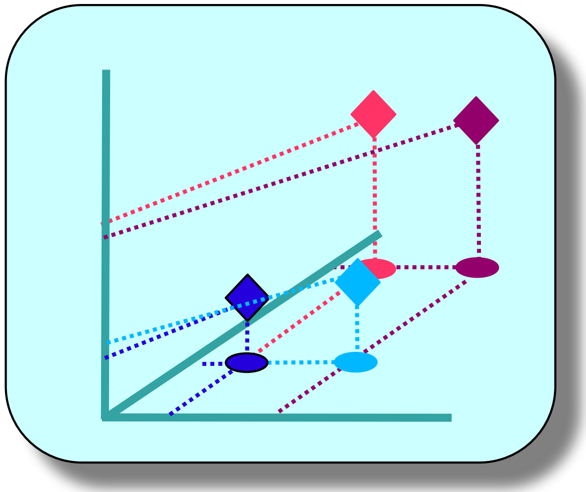
Specificity

- You can only talk about what you measured
- If too much varies at once, you cannot say what **caused** any differences

Generality

- Make broad statements without measuring every point...**Interpolate!**
- Systematically vary dimensions

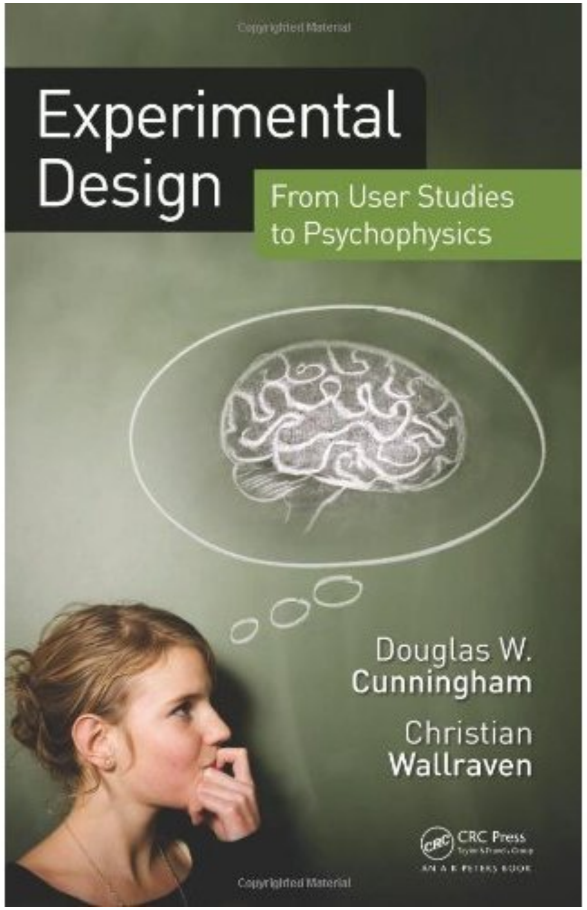




Based on Cunningham & Wallraven, 2011

CognitiveSystems

2: Philosophy of experiments – what is an experiment and **how can we model this mathematically?**



with material from

Douglas Cunningham

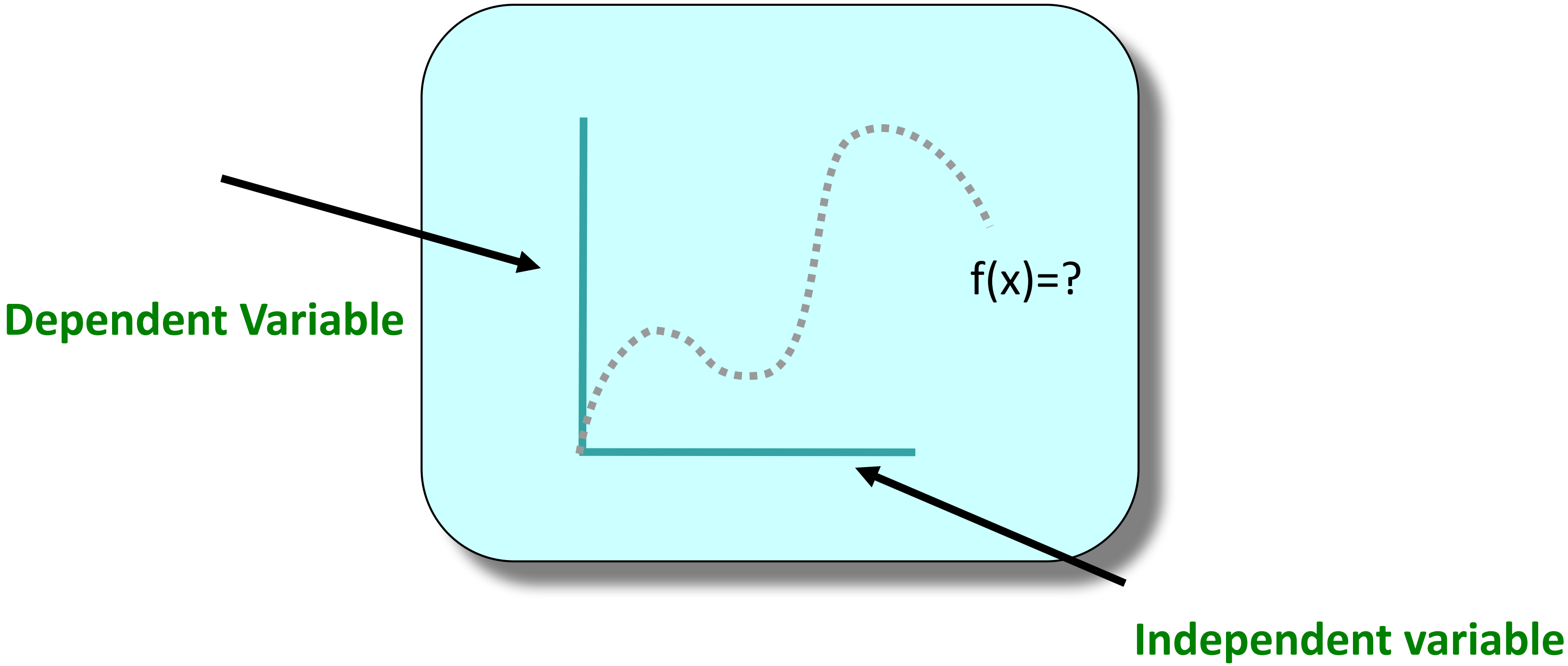
Some terms in Experimental Design

- Each dimension that we manipulate is called a **factor**.
- Each value that we used from a factor is called a **level.**
- Usually, all combinations of factors are used. Any given combination is called a **condition**.
- Each single execution of a condition is a **trial**.
- Since only an examination of what happens under all relevant conditions can answer our question conclusively, the full collection of trials that addresses the current research question is an **experiment**.

What is Experimental Design?

CognitiveSystems

Goal: Estimate an unknown function



Method: Systematically sample the function

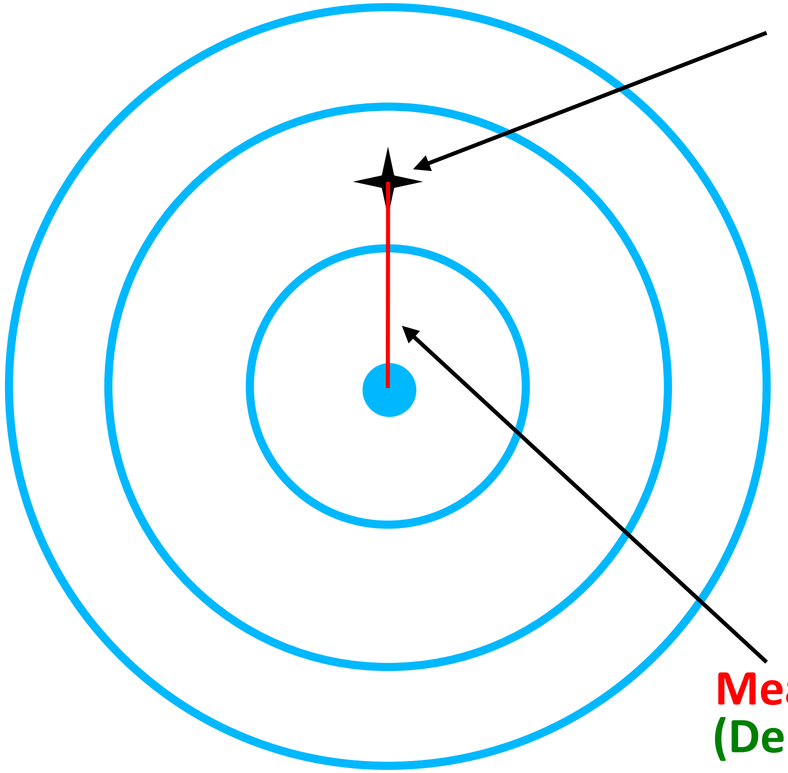
Based on Cunningham & Wallraven, 2011

Experiment I

- **Research Question:**
  - How accurately can people point to a target?
- **Methods:**
  - Stimulus: Blue bullseye target
  - Participants: One
  - Task: Point (once) as quickly and accurately as possible to the center
  - (total number of trials: 1)

Experiment I

Let B(x) be the perception-action loop for pointing accuracy (for this situation)



Response

$$M(x) = B(x)$$

If I repeat the experiment, keeping *x* constant, will I get the same result?

Measurement M(x)
(Dependent Variable)

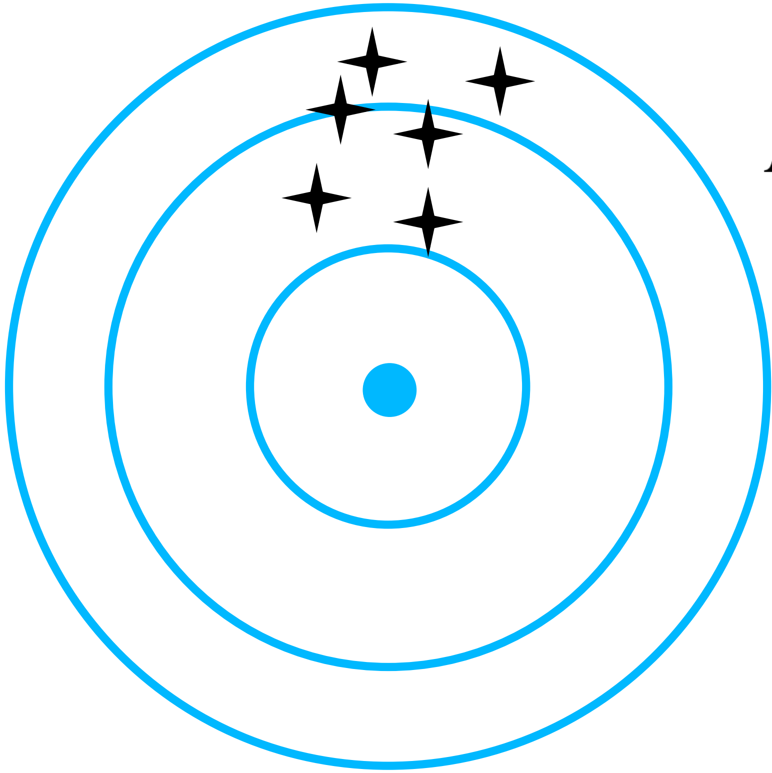
Situation: x (participant, task, stimulus, time of day, ...)
(includes the Independent Variable)

Experiment Ib: Repeated Measures

CognitiveSystems

- **Research Question:**
  - How accurately can people point to a target?
- **Methods:**
  - Stimulus: Blue bullseye target
  - Participants: One
  - Task: Point (n times) as quickly and accurately as possible to the center
  - (total number of trials: 1 x n)

Experiment Ib: Repeated Measures



$$M(x) = B(x) + \varepsilon_W$$

Situation: x (participant, task, stimulus, time of day, ..)

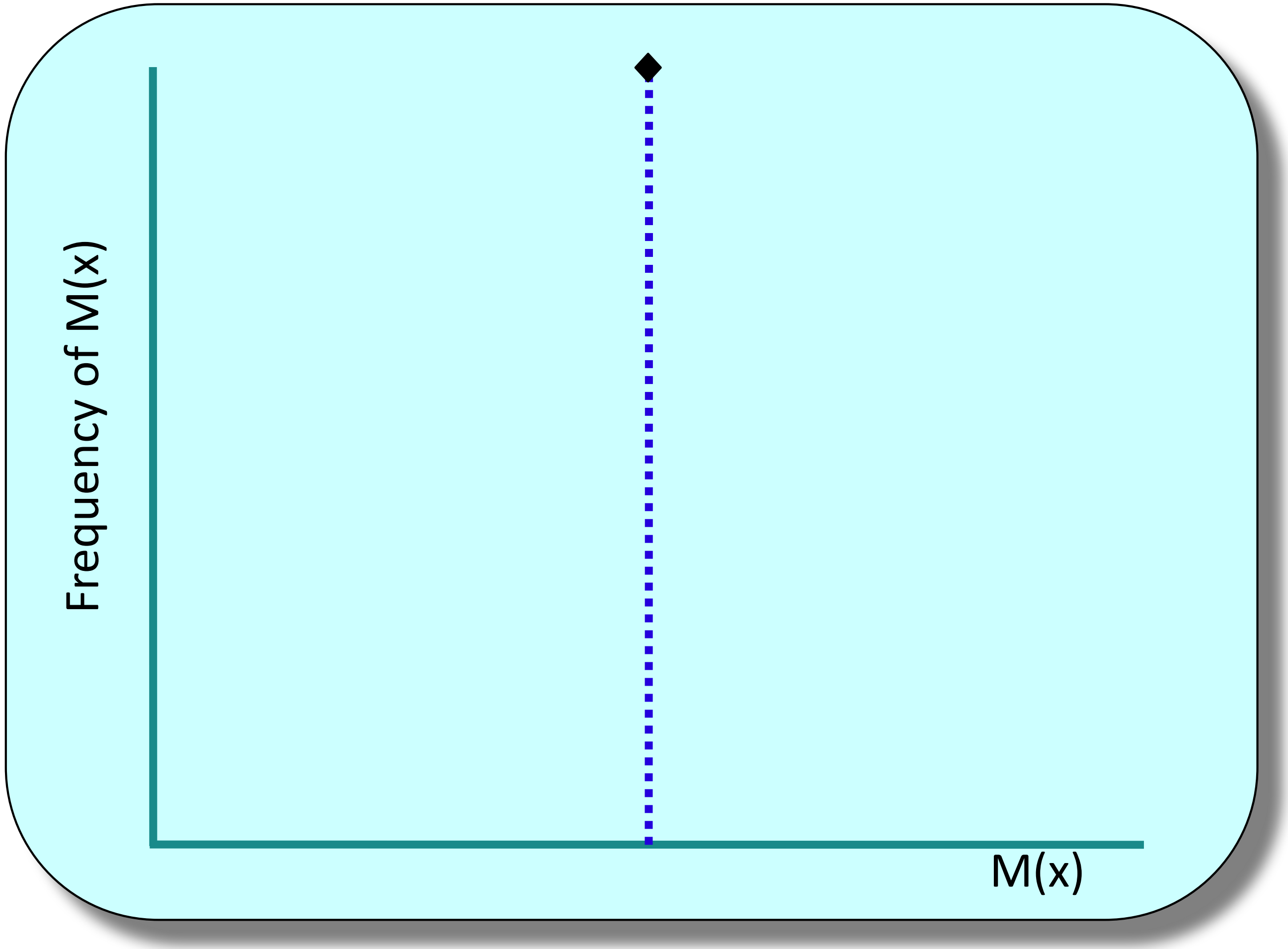
- People cannot exactly repeat any performance
- There is some inherent, unintentional variation/noise

Experiment Ib: Repeated Measures

CognitiveSystems

$$\text{for } \varepsilon_W = 0,\ M(x) = B(x)$$

$$M(x) = B(x) + \varepsilon_W$$

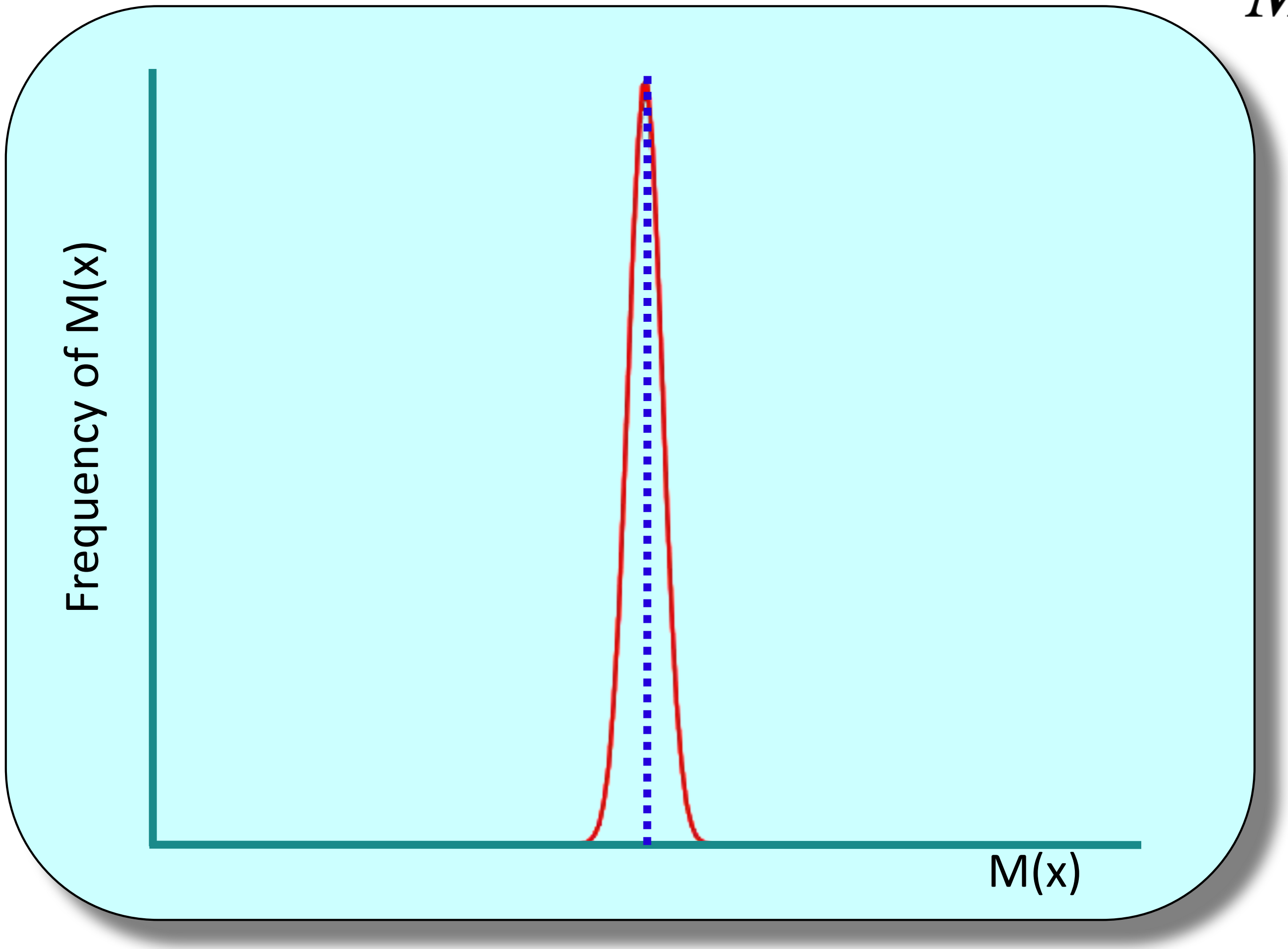


(note: Distributions have been rescaled)

Experiment Ib: Repeated Measures

for small ε_W each measurement is close to $B(x)$

$$M(x) = B(x) + \varepsilon_w$$



in this case, we do many trials and find that most of the time the measurements fall in a very small region around the “true” value of B(x)

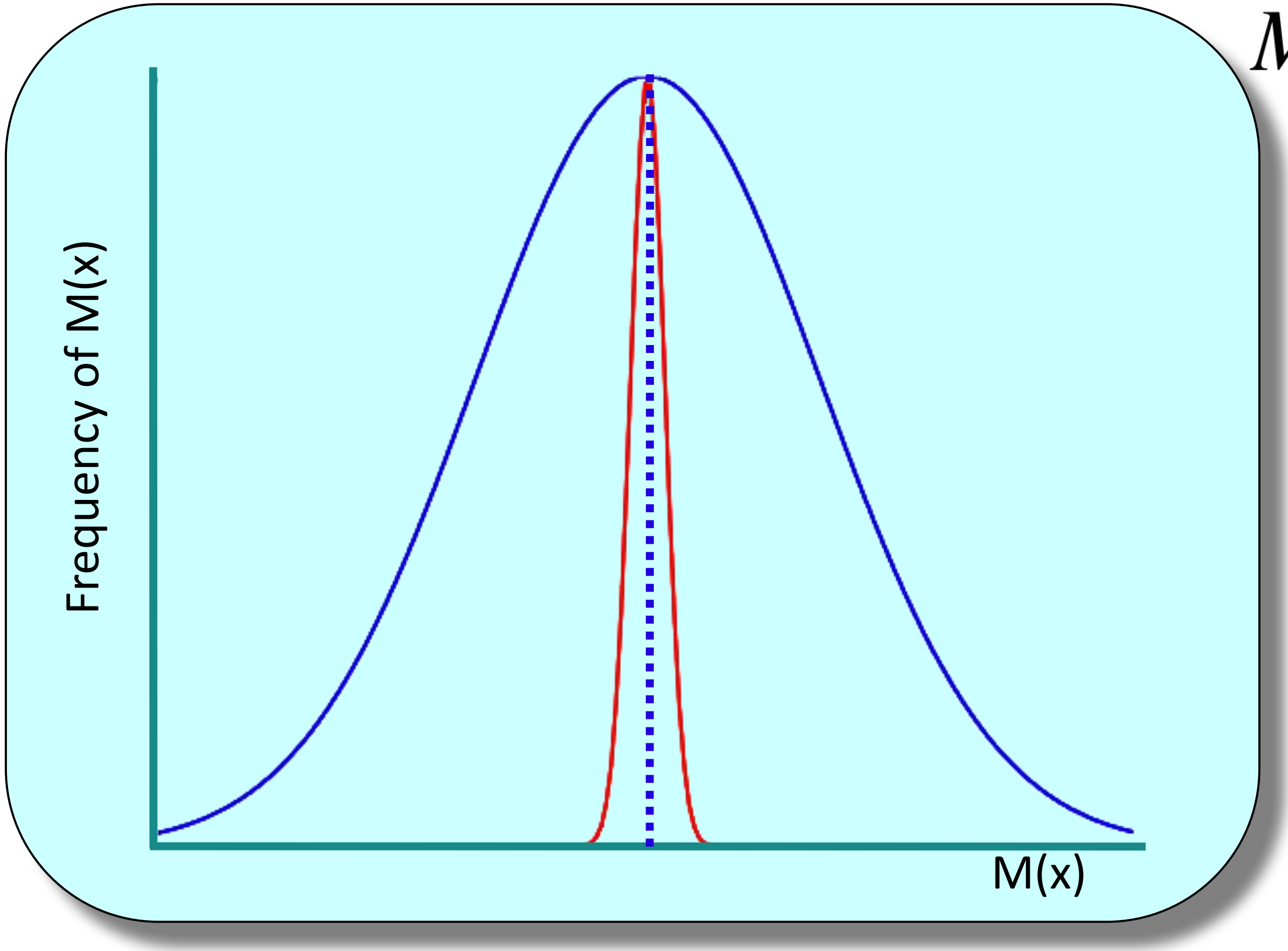
(note: Distributions have been rescaled)

Experiment Ib: Repeated Measures

for small ε_W each measurement is close to $B(x)$

for large ε_W each measurement can be very far from $B(x)$

$$M(x) = B(x) + \varepsilon_w$$

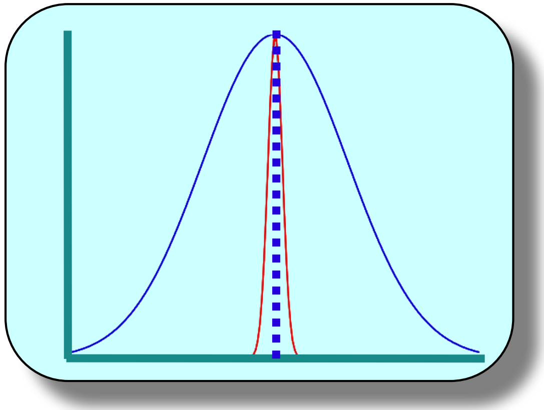


in this case, we do many trials and find that most of the time the measurements fall in a **large** region around the "true" value of B(x)

(note: Distributions have been rescaled)

Experiment Ib: Repeated Measures

CognitiveSystems



- The average approximates B(x)
- B(x) is constant
- ϵ_{w1} is different every time we measure M(x)
- Bias from ϵ_{w1} is sometimes positive, sometimes negative
- With enough trials, we can estimate the error, and factor it out

$$\text{trial 1}: \; m_1(x) = B(x) + \varepsilon_{w1}$$

$$\text{trial 2}: \; m_2(x) = B(x) + \varepsilon_{w2}$$

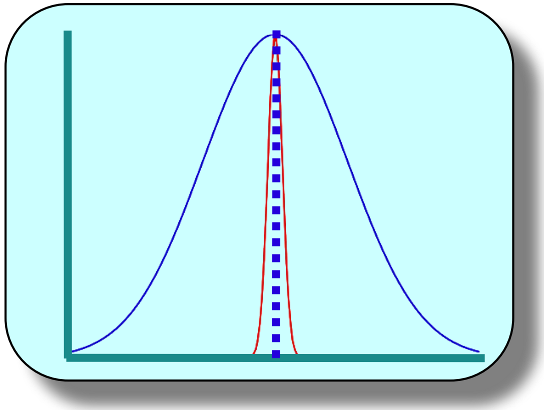
$$\ldots$$

$$\text{trial n}: \; m_n(x) = B(x) + \varepsilon_{wn}$$

$$\text{average}: \; \overline{M}(x) = \frac{\sum_{i=1}^{n} m_i}{n} = \frac{\sum_{i=1}^{n} B(x) + \varepsilon_{wi}}{n} \approx B(x)$$

Experiment Ib: Repeated Measures

CognitiveSystems



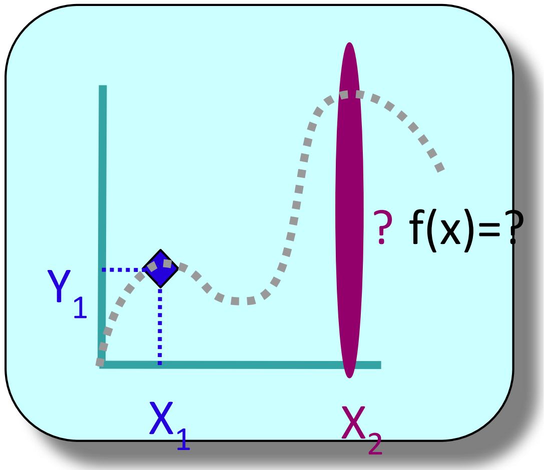
- Anything that can reduce noise improves the approximation
- Increasing number of samples (trials) improves approximation
- How many repetitions?
  - There are equations for calculating this, based in part on
    - expected effect size
    - noise size
- Rule of thumb: More than 5 less than 20

Experiment Ib: Repeated Measures

- **Research Question:**
  - How accurately can people point to a target?
- **Methods:**
  - Stimulus: Blue bullseye target
  - Participants: One
  - Task: Point (**5** times) as quickly and accurately as possible to the center
  - (total number of trials: 1 x **5**)

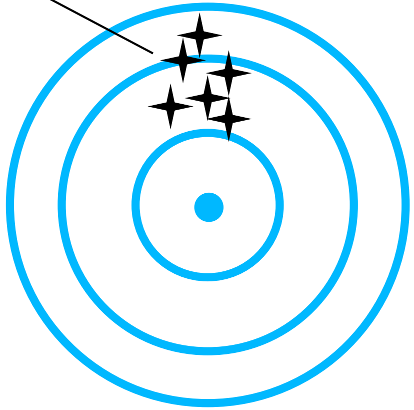
Experiment Ib: Repeated Measures

- **Research Question:**
  - How accurately can people point to a target?
- **Methods:**
  - Stimulus: Blue bullseye target
  - Participants: One
  - Task: Point (**5** times) as quickly and accurately as possible to the center
  - (total number of trials: 1 x **5**)
- The results will be specific for this person
- B(x) might be different for different people
- To generalize to the population, we need more people!

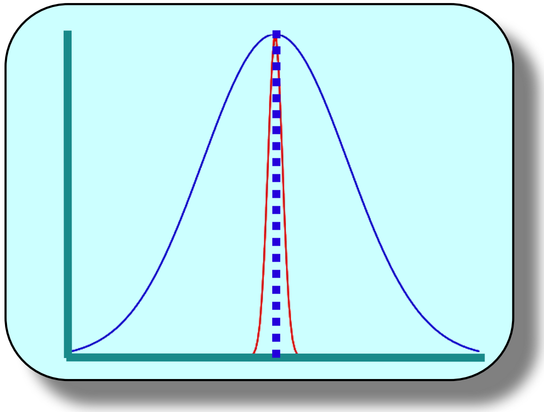


Experiment Ic (Multiple Participants)

- **Research Question:**
  - How accurately can people point to a target?
- **Methods:**
  - Stimulus: Blue bullseye target
  - Participants: Several (*p*)
  - Task: Point (once each) as quickly and accurately as possible to the center
  - (total number of trials: 1 per person)



Experiment Ic (Multiple Participants)



- We can calculate the error as above
- Is measuring many people once really the same as measuring one person multiple times?
- Why are people different?
  - Fundamentally different action-perception loops?
  - Constant (population) action-perception loop with everyone having a minor variation of that (e.g., noise)?

Experiment Ic (Multiple Participants)

- Per person, we assumed a constant effect B(x) plus internal noise ε_w = within person noise
- We can likewise assume a globally constant effect B(x) and additional noise between people ε_b = between person noise

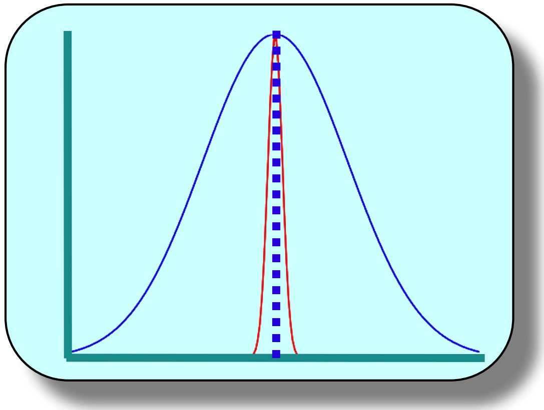
$$\overline{M}(x) = \frac{\sum_{i=1}^{p} m_i}{p} = \frac{\sum_{i=1}^{n} B(x) + \varepsilon_{wi} + \varepsilon_{bi}}{p}$$

- Might be wise to sample each of the two error functions separately! (n trials for each of p participants)

$$\overline{M}(x) = \frac{\sum_{i=1}^{p} \sum_{j=1}^{n} m_{ij}}{pn} = \frac{\sum_{i=1}^{p} \sum_{j=1}^{n} B(x) + \varepsilon_{wj} + \varepsilon_{bi}}{pn}$$

How many samples??

- We are sampling unknown (noise) functions
- Multiple samples are needed per population
- How many participants?
- Again, there are equations for this
- As a rule of thumb,
  - for large effects, 10 is sufficient
  - for smaller effects, more
  - per Person (5+ repetitions per person)
- Two error terms, we need to sample both: Both Population and Person!

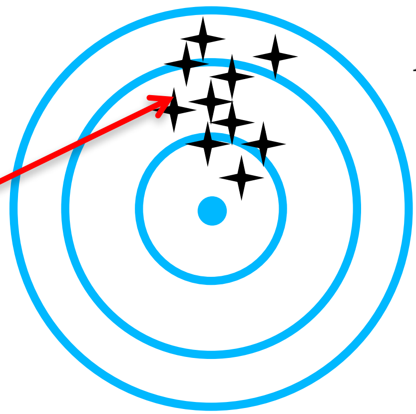


Experiment Id
Multiple Participants and Repetitions

- **Research Question:**
  - How accurately can people point to a target?
- **Methods:**
  - Stimulus: Blue bullseye target
  - Participants: 10
  - Task: Point (5 times) as quickly and accurately as possible to the center
  - (total number of trials: 1 x 5 = 5 per person)

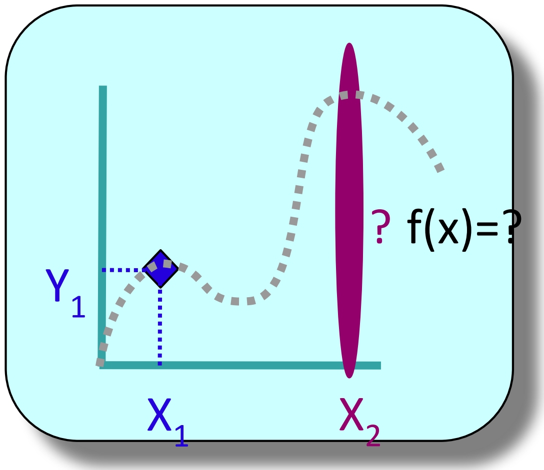
$$M(x) = B(x) + \varepsilon_w + \varepsilon_b$$

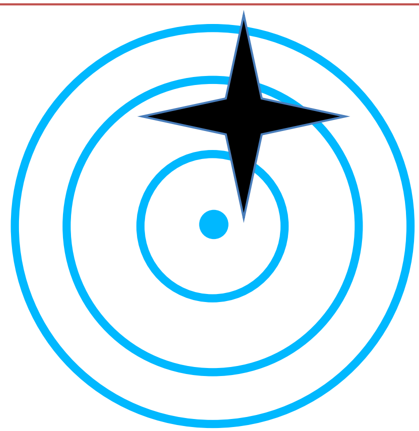
Each point is now the **AVERAGE** of the 5 trials per person!



Experiment Id
Multiple Participants and Repetitions

- Why is performance so bad?





$$M(x) = B(x) + \varepsilon_w + \varepsilon_b$$

Experiment II (Color)

- **Research Question:**
  - How do changes in color affect pointing?
- **Methods:**
  - Stimulus: Red and Blue bullseye targets
  - Participants: 10
  - Task: Point (5 times) as quickly and accurately as possible to the center
  - (total number of trials: 2 x 5 = 10 per person)

Experiment II (Color)

CognitiveSystems

$$M(x) = B(x) + \varepsilon_w + \varepsilon_b$$

Effect of color change

$$M(x+\Delta c) = B(x+\Delta c) + \varepsilon_w + \varepsilon_b$$

$$M(x+\Delta c) \approx B(x) + B(\Delta c) + \varepsilon_w + \varepsilon_b$$

Change in stimulus color

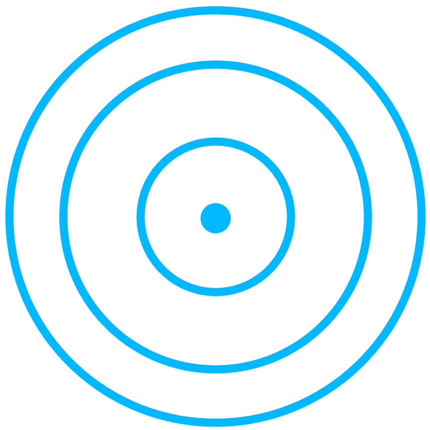
$$M(x+\Delta c) - M(x) = \{B(x) + B(\Delta c) + \varepsilon_w + \varepsilon_b\} - \{B(x) + \varepsilon_w + \varepsilon_b\}$$

$$M(x+\Delta c) - M(x) = \{B(x) - B(x)\} + \{\varepsilon_w + \varepsilon_b - \varepsilon_w - \varepsilon_b\} + B(\Delta c)$$

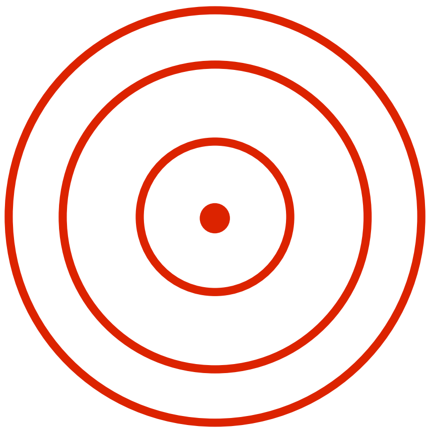
$$M(x+\Delta c) - M(x) = B(\Delta c)$$

*The difference between conditions is the effect of changing color **from blue to red***

specificity



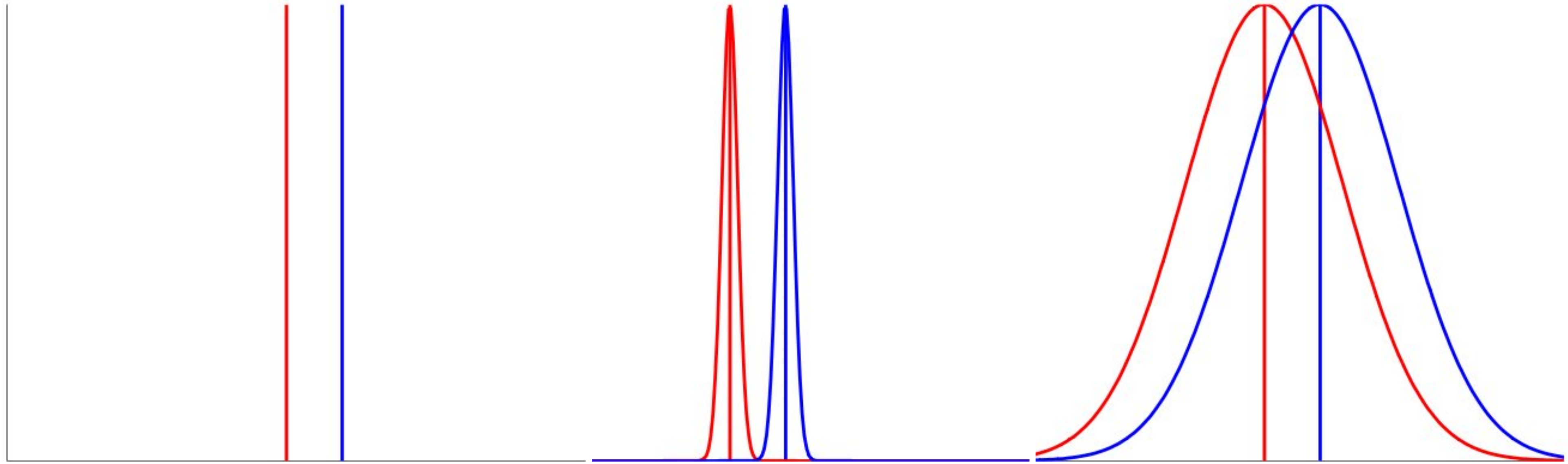
Situation: x



Situation: $x + \Delta c$

Note that splitting the function B(x+Δc) into its component parts (B(x) and B(Δc)) requires that the function B be homomorphic. Linear functions satisfy this property. Since we have assumed that the elements of x are independent of each other and can be modeled with as a linear, weighted sum, B is homomorphic.

Experiment II (Color)

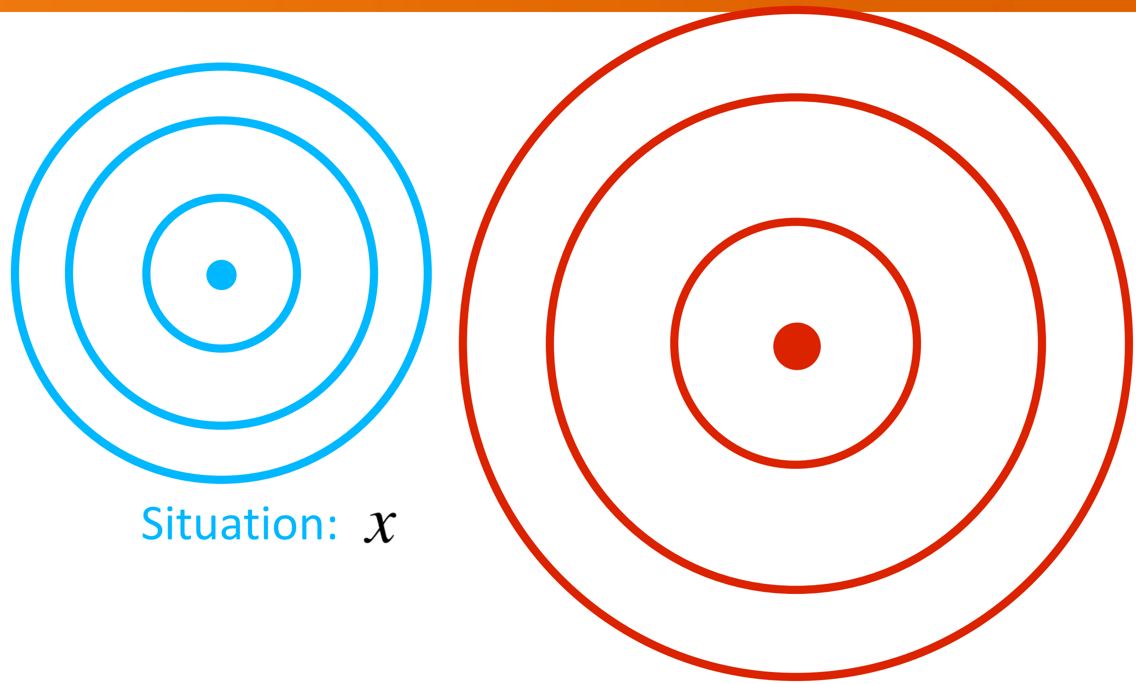


- Are the means *really* different?
- Noise may "swamp" effect!
  - Control noise
  - Identify and remove unwanted variance
  - Run (complicated) statistics

Experiment III (Color and Size)

- **Research Question:**
  - How do changes in color and size affect pointing?
- **Methods:**
  - Stimulus: Large and small, Red and Blue bullseye targets
  - Participants: 10
  - Task: Point (5 times) as quickly and accurately as possible to the center
  - (total number of trials: 2 x 5 = 10 per person)

Experiment IIIa (Color and Size)



Situation: x

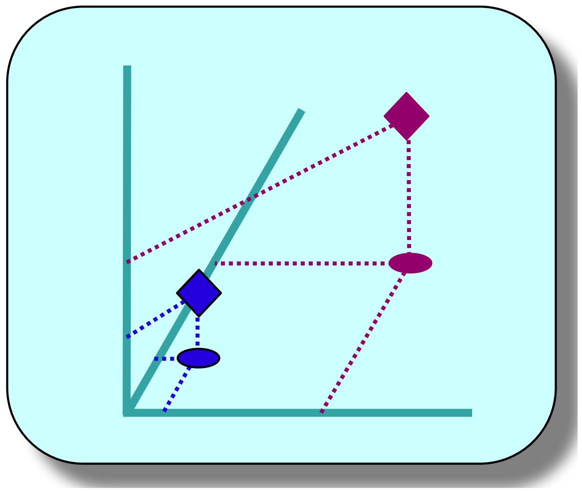
Situation: $x + \Delta c + \Delta s$

$$M(x) = B(x) + \varepsilon_w + \varepsilon_b$$

$$M(x + \Delta c + \Delta s) = B(x) + B(\Delta c) + B(\Delta s) + \varepsilon_w + \varepsilon_b$$

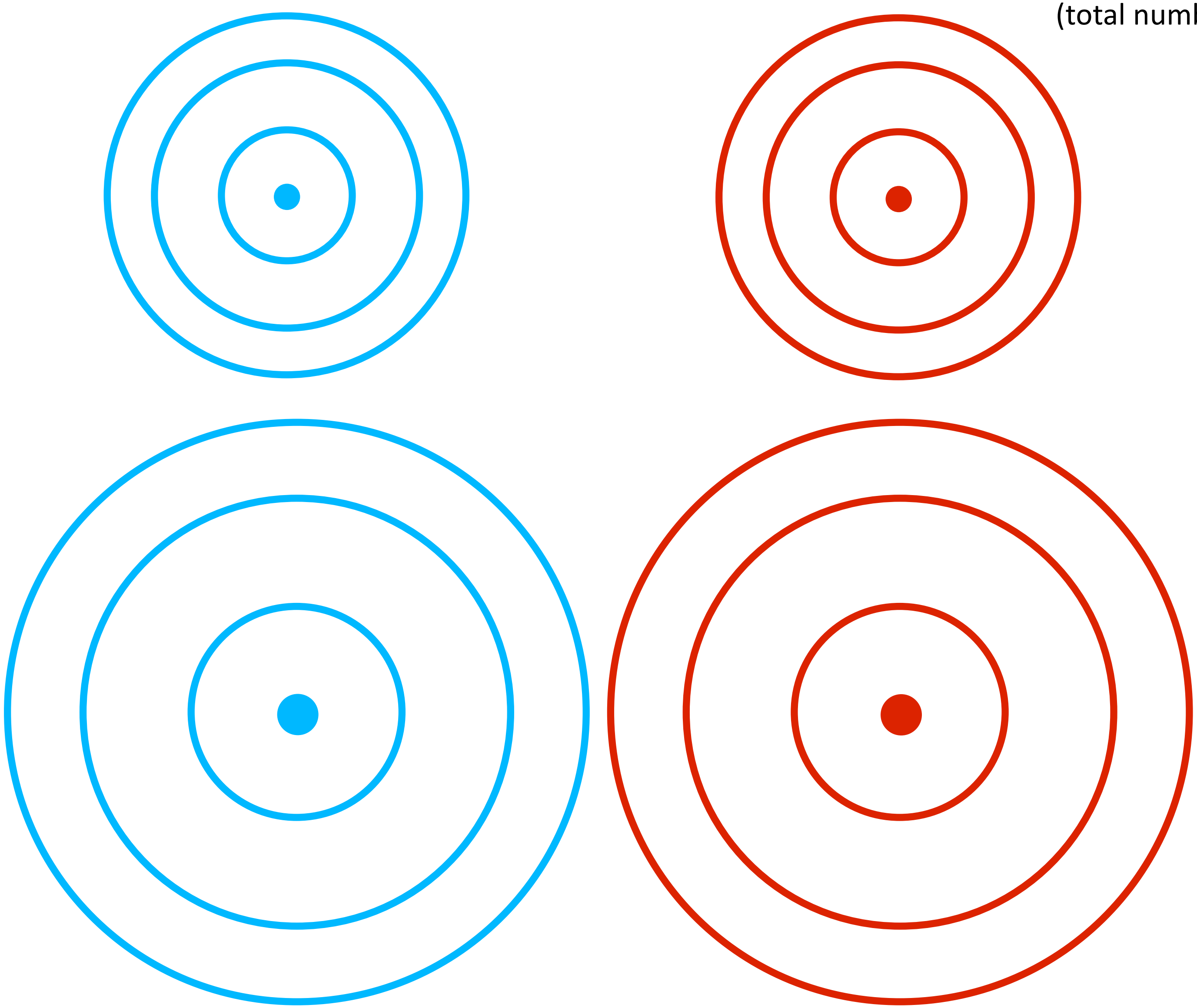
$$M(x + \Delta c + \Delta s) - M(x) = B(\Delta c) + B(\Delta s)$$

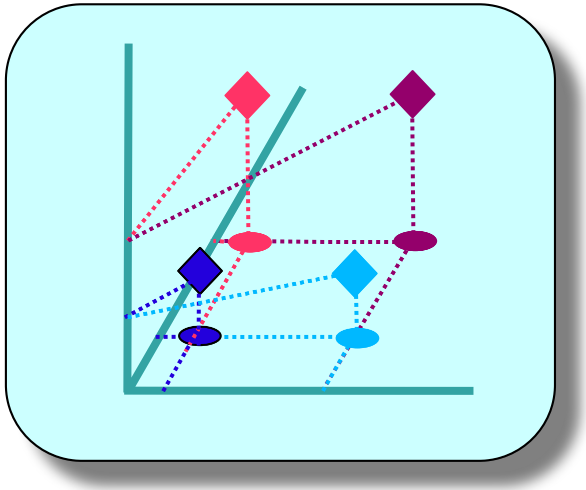
When changing two variables at the same time, you can no longer **conclusively** say what caused the difference between conditions



Experiment IIIb (Color and Size)

(total number of trials: 4 x 5 = 20 per person)



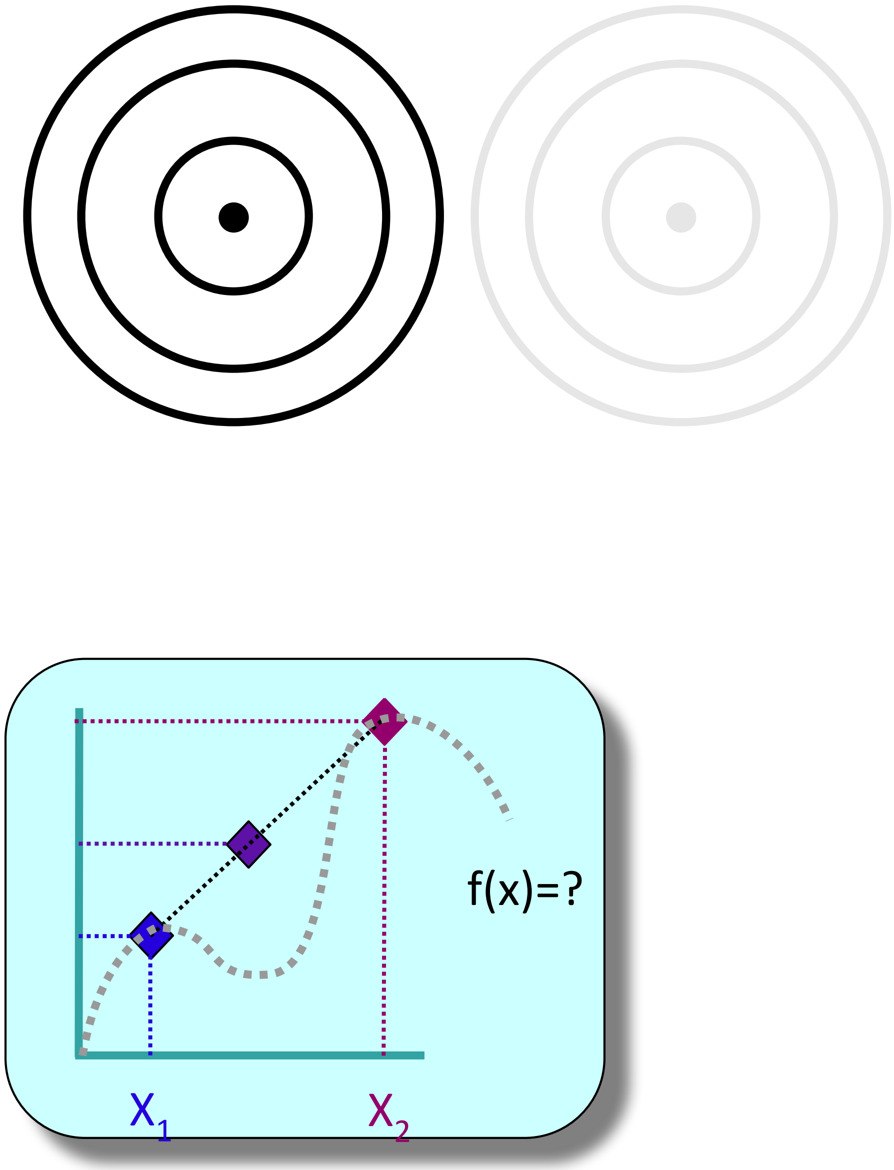


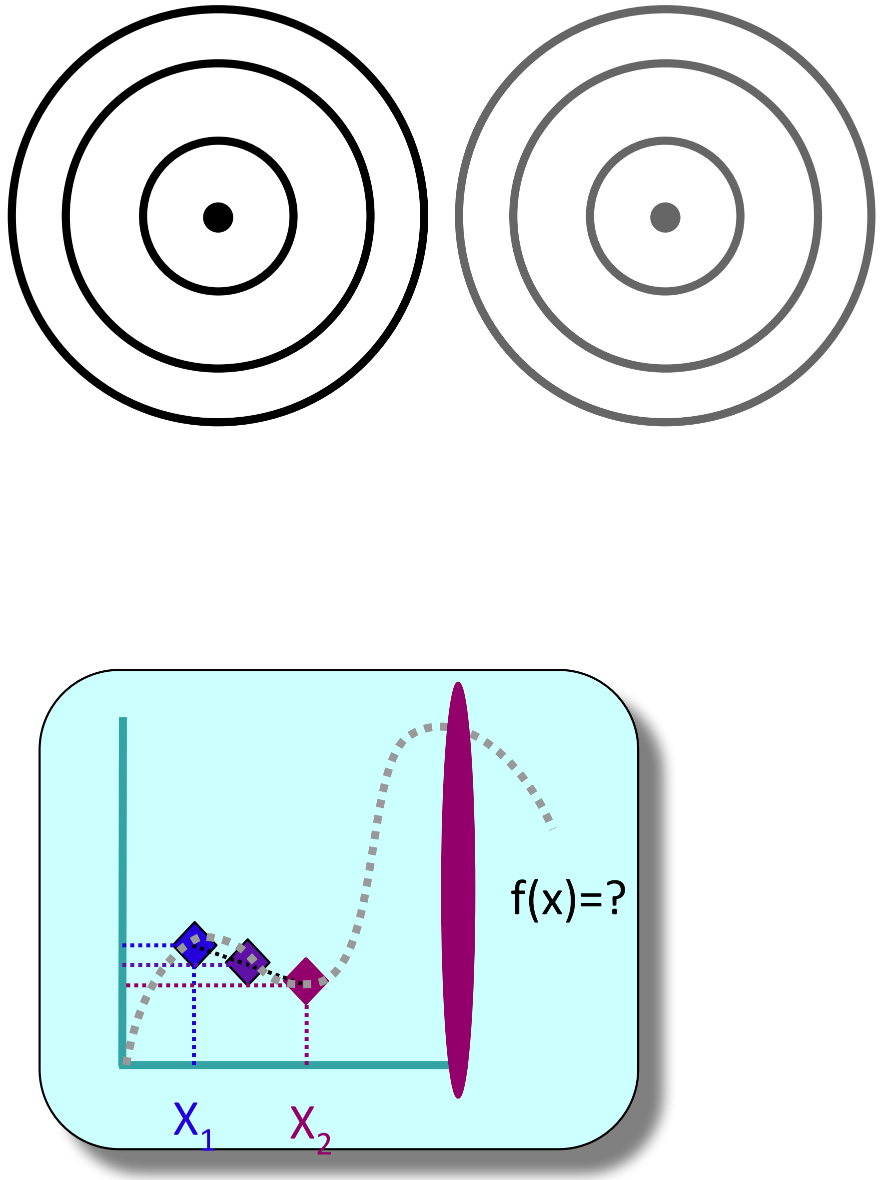
Experiment IV (Contrast: Trial Order)

CognitiveSystems

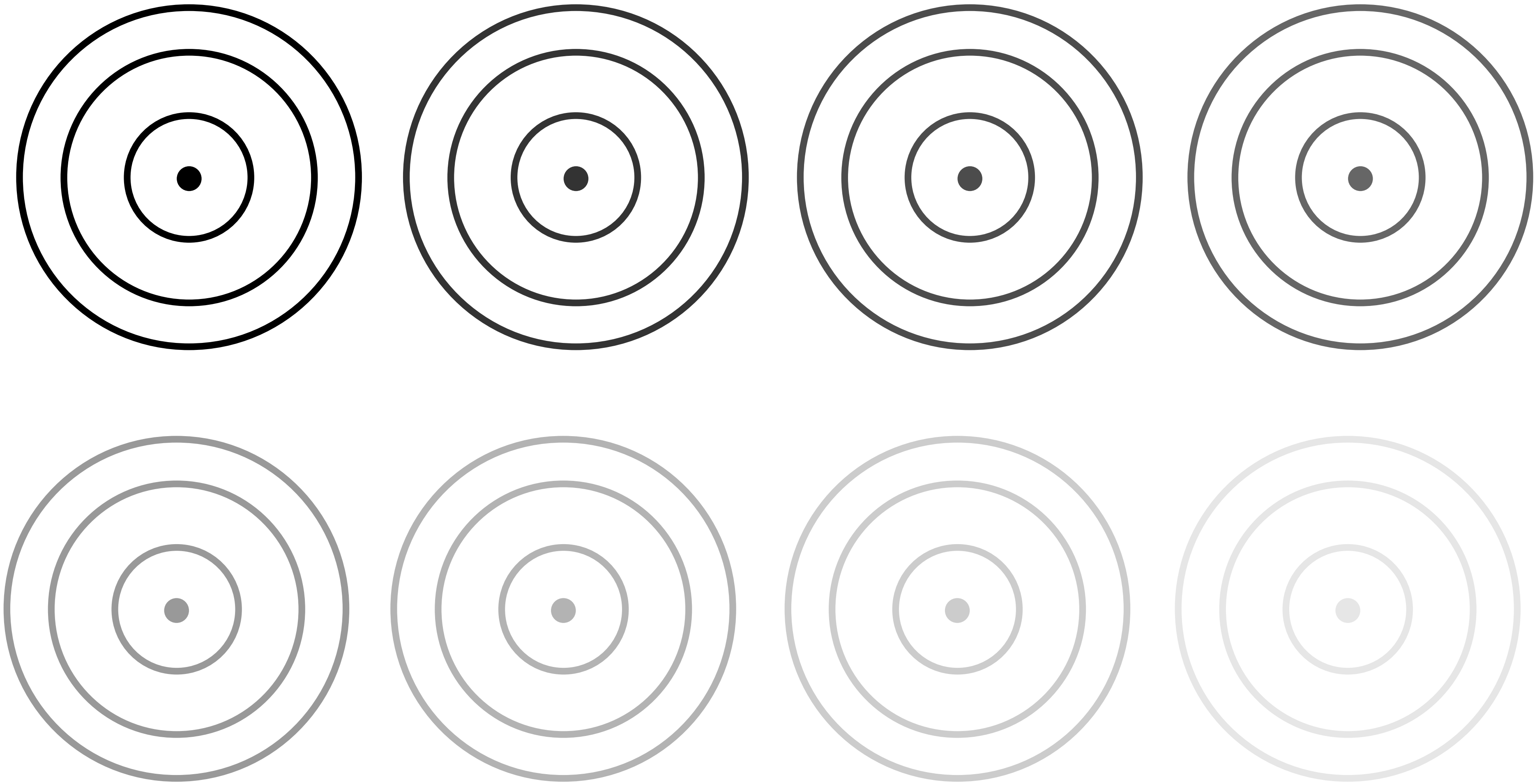
- **Research Question:**
  - How do changes in contrast affect pointing?
- **Methods:**
  - Stimulus: high and low contrast bullseye targets
  - Participants: 10
  - Task: Point (5 times) as quickly and accurately as possible to the center
  - (total number of trials: 2 x 5 = 10 per person)

Experiment IV (Contrast: Trial Order)





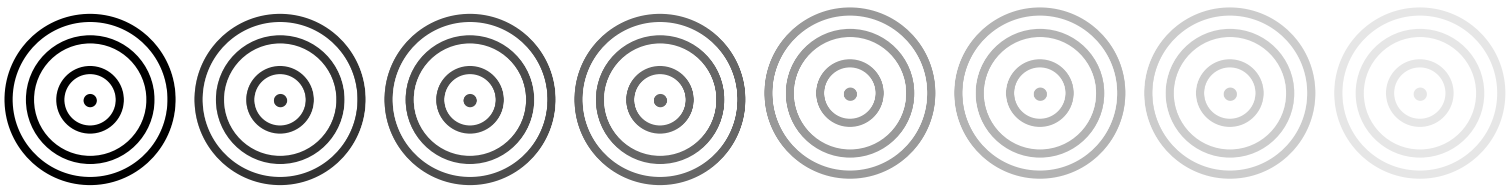
Experiment IV (Contrast: Trial Order)



Experiment IV (Contrast: Trial Order)

CognitiveSystems

- **Research Question:**
  - How do changes in contrast affect pointing?
- **Methods:**
  - Stimulus: 8 bullseye targets, systematically varying contrast in equal steps
  - Participants: 10
  - Task: Point (5 times) as quickly and accurately as possible to the center
  - (total number of trials: 8 x 5 = 40 per person)



Experiment IV (Contrast: Trial Order)

$$M(x) = B(x) + \varepsilon_w + \varepsilon_b$$

$$M(x - p) = B(x - a) + \varepsilon_w + \varepsilon_b$$

$$M(x - 2p) = B(x - b) + \varepsilon_w + \varepsilon_b$$

$$M(x - 3p) = B(x - c) + \varepsilon_w + \varepsilon_b$$

$$\ldots$$

$$M(x - 7p) = B(x - g) + \varepsilon_w + \varepsilon_b$$

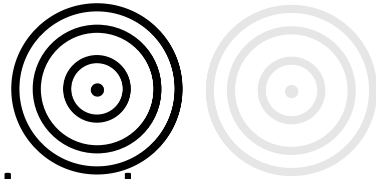
Notes:
Measurements are in terms of a base condition x and multiples of 10% contrast change.

The underlying perception-action loop **does not** use this periodic representation.

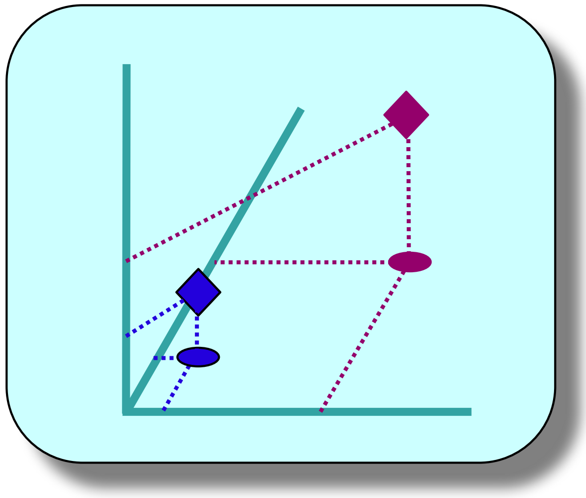
Experiment IV (Contrast: Trial Order)

Experiment IV (Contrast: Trial Order)

- Is trial 1 identical to trial 8?



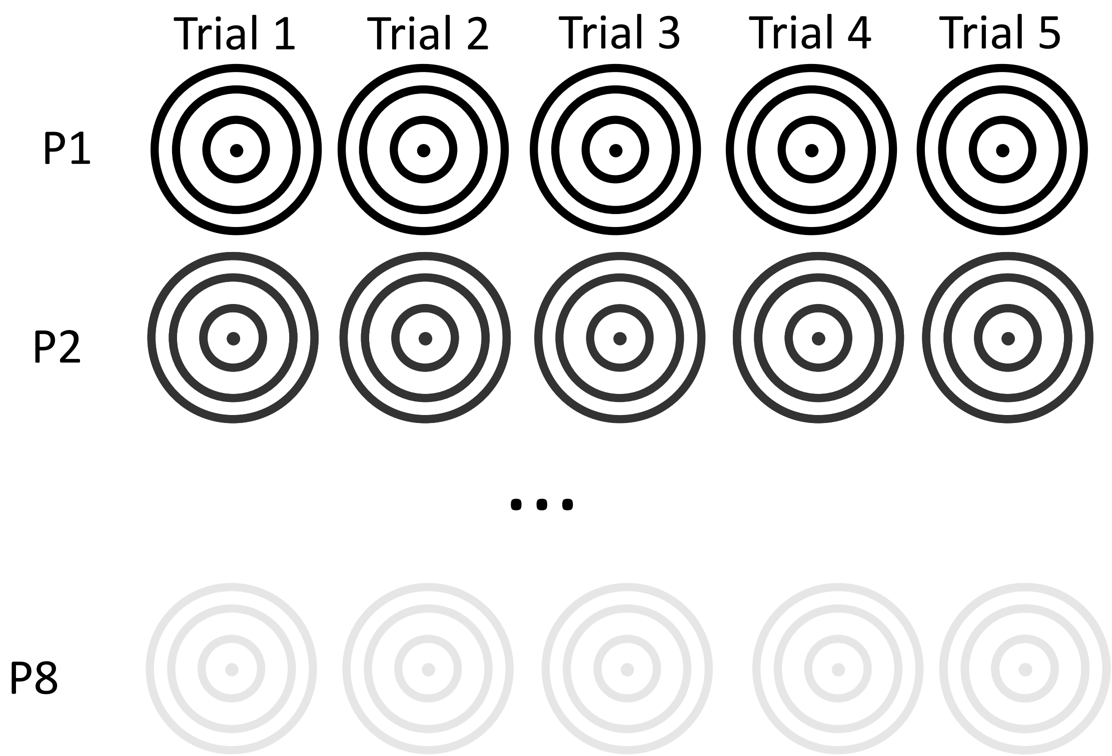
- Different contrasts
- Difference in practice!
- So, any difference between trial 1 and trial 8 might be due to
  - Contrast
  - order/practice effect



Experiment IV (Contrast: Trial Order)

(total number of trials: 1 x 5 = 5 per person)

- Solution 1: Eliminate order!
  - Everyone sees only one contrast
  - Each contrast is seen by one person
  - Between-participants design



- Difference between high and low might be due to
  - contrast
  - participant (P1 vs P8)
- So, multiple people per contrast.
- The higher ε_b is, the more participants one will need for **each** contrast condition
- 10 people for 8 groups= 80 people!

Experiment IV (Contrast: Trial Order)

- Solution 2: Fully control for order
- We let everybody see every contrast, so we know that they act as their own control in terms of performance
- To explicitly check for order effects, we need to investigate how c1, c2, c3 may be different from c2, c3, c1 – for example
  - so we would need 10 people per possible order
- How many possible orders are there?
  - n! for n conditions
  - 6 for 3 conditions
  - 8! = 40320 for our 8 contrast levels → >400000 participants!

Experiment IV (Contrast: Trial Order)

- Solution 3: Compromise by controlling only for some orders
- Maybe we are just interested to see, whether there IS an order effect, not WHICH order it has
- So, we don't test all possible orders (n!), but only a subset (n) of orders
- A popular selection is the latin square design, which for three conditions A,B,C looks like this
- We recruit 10 people per row, and compare their performance – if any row is different from the others, we know that there is an order effect and that we have to be careful!
  - 8 * 10 participants total

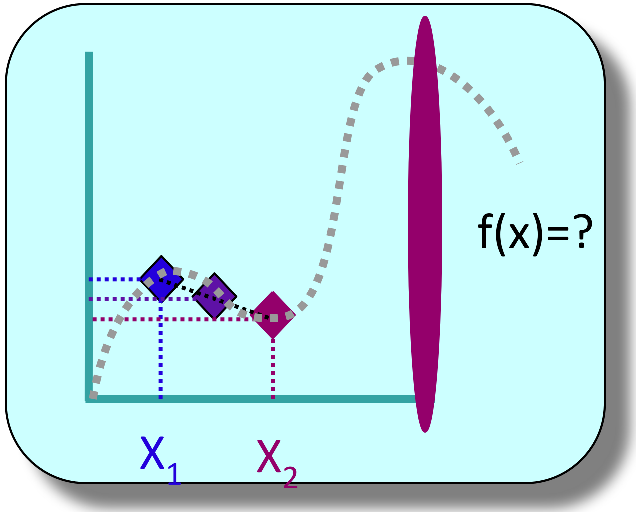
	1	2	3
1	A	B	C
2	B	C	A
3	C	A	B

Experiment IV (Contrast: Trial Order)

- Solution 4: Compromise by accepting some noise
- We let every participant see a different, random order and hope that order noise is washed out over many participants
- This is a good assumption if individual differences in performance are smaller than order effects (which is often the case)
- Again, more participants are needed than 8, but perhaps one may be able to spend fewer than 80 – let's say 40 participants

Special Factors: Participants

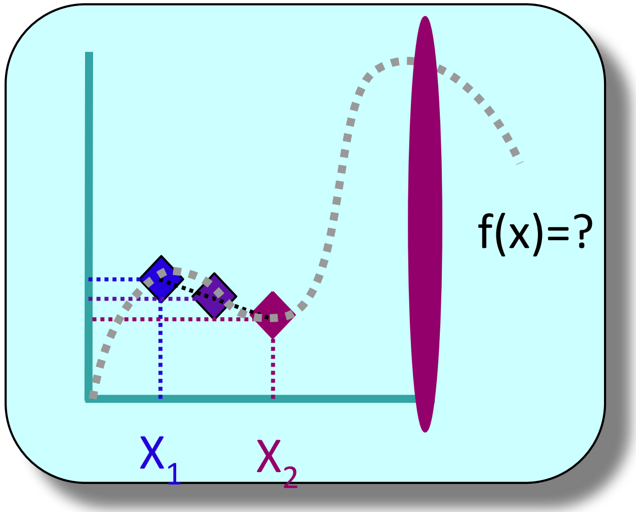
- Why do people differ?
  - Natural talent
  - Expectations
  - Motivation
  - Fatigue
  - Physical differences (e.g., eye sight)
  - Experience



- What if we only used the authors?
  - Authors are not naive, can Bias
- What if we only used experts?
  - Experts/Novices have different skill levels, can Bias

Special Factors: Participants

- Sampling a distribution!
- Participants should be representative
- If you want to understand people in general, use people in general
- Use naive participants, unless
  - sure knowledge cannot affect results
  - desired population consists solely of experts
- If you want to understand how a technique will affect surgery, use surgeons!



What do you think?

- Going abroad, what are the differences that you have observed in other cultures?
  - …

People are WEIRD

- Arnett (2008) surveyed top journals in six sub-disciplines of psychology and found that
  - 68% of subjects were from the US
  - 96% from ‘Western’ industrialized nations (European, North American, Australian or Israeli) = 12% of the world’s population
- You’re 4000 times more likely to be studied by a psychologist if you’re a university undergraduate at a Western university than a randomly selected individual

People are WEIRD

- The weirdest people in the world? Joseph Henrich, Steven J. Heine and Ara Norenzayan (2010). Behavioral and Brain Sciences, Volume 33, Issue 2-3, 2010; 61-83.
- Claim:
  - psychologists test people from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies
  - this population shows considerable variation among visual perception, fairness, cooperation, spatial reasoning, categorization and inferential induction, moral reasoning, reasoning styles, self-concepts and related motivations, and the heritability of IQ
  - least representative population!
- Solution:
  - cross-cultural research agenda!
  - cautionary words in publications

People are WEIRD


Volume 40 Number 1 February 2014

Published bimonthly by the American Psychological Association

ISSN 0096-1523

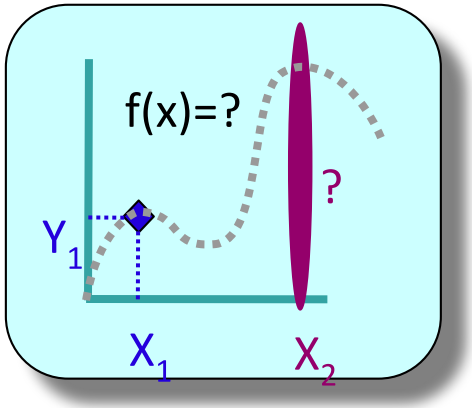
Journal of Experimental Psychology: Human Perception and Performance

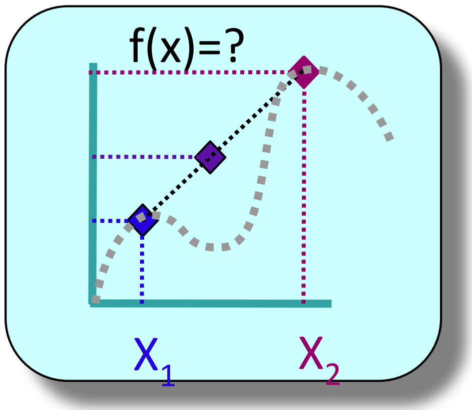
Editor
James T. Enns

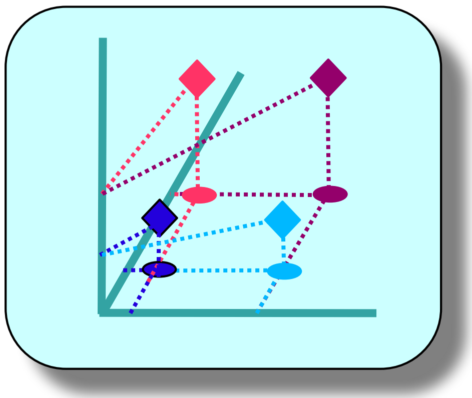
Associate Editors
Sarah Creem-Regehr
Michael D. Dodd
Marc O. Ernst
Brett Fajen
Susanne Ferber
Rob Gray
Kathleen Rastle
Geoffrey Woodman

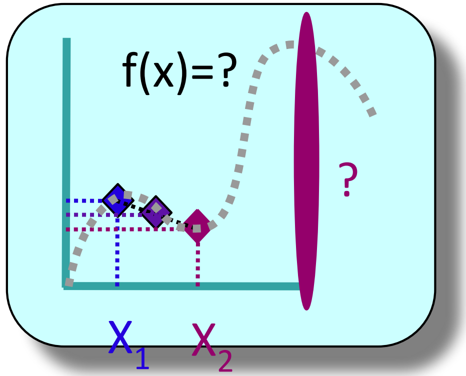
www.apa.org/pubs/journals/xhp

Summary so far









Summary so far

Summary so far

- Try to remove variance through proper experimental design, and not through complex statistics
- The more complex your statistics become, the fewer the number of people that understand what you did or why becomes (i.e., you loose your audience)

General summary

- Once we are clear about our research question, we need to go about trying to answer it. In general, that means:
  - Show something (stimuli)
  - somehow (stimulus presentation)
  - to someone (participants)
  - And ask them to do something (the task)
  - (and then analyze the results)