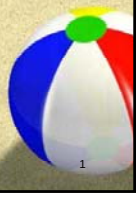
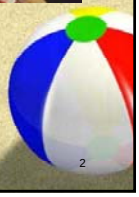
Agenda 2/28/08

- Review of key concepts from recent material
- New lecture material:
  - Causality and experiments



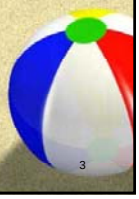
Warming up…

- What if I told you that studying with a partner would help you do better on your final exams, and you were a little skeptical about my claim, and interested in better understanding the how this is supposed to work?
  - What kinds of Qs about the causal relationship would help you figure out whether or not you think this is something that you'd like to try?



What sorts of things would you be curious about?

- Is the ultimate payoff big enough to be worth going through all that trouble?
- I'm a little skeptical about the connection here: does this help mainly by forcing me to schedule in study *time*, or is there something about talking to another *person* that is special?
- How much does this study method help, all other things being equal?
- How much will it help as compared to pulling an intense all-nighter or any OTHER special efforts I make - - how much MORE powerful is this method than others?
- How focused and consistent could I really expect myself to be at this, in normal life?



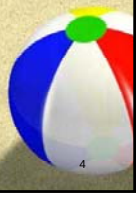
Quick Story

Younger me…yikes…

AMYLIN

AMYLIN PHARMACEUTICALS is a biopharmaceutical company committed to improving lives through the discovery, development and commercialization of innovative medicines.

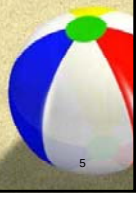
Challenging Science. Changing Lives.



2/28/08 4

New Topics to be Covered

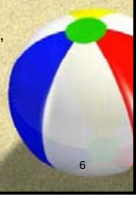
- Experiments
- Confounding Variables
  - Subject
  - Procedural
  - Other
- Controlling for Confounds
  - Strategies
  - Study design features
- Experimental Validity
  - Internal Validity
  - External Validity



2/28/08 5

Quick Review (1 of 7)

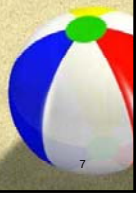
- Just to prime your memory for these things, but let's not get too hung up on them for now …
- Testing for differences between sample means (what is this? How are the variables treated? What question do you ultimately want to answer?)
  - Type 1 error and Type 2 error - what are they, how do researchers try to reduce the risk of them?
  - Statistical significance, P-value, Alpha levels
  - Power. How can researchers increase power?



2/28/08 6

Quick Review (2 of 7)

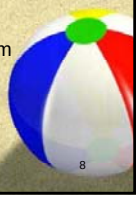
- Independent variables, dependent variables
- Operational definitions (a.k.a. "operationalization")
- Necessary and sufficient causes
  - Why are they useful to identify when possible?
  - How can you evaluate claims about necessary and sufficient causes?
- Deterministic vs. indeterministic causes
- Partial or contributory causes
  - How can you evaluate claims about these, when they're concerning types of events, and concerning particular individuals?
- Counterfactual analysis of causation –
  What is it?
  When might you try to use it?



2/28/08 7

Quick Review (3 of 7)

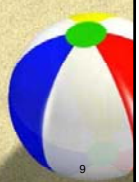
- Proximal vs Ultimate causes
- Causal overdetermination, what is it, how it relates to the counterfactual analysis of causation
- Mill's methods for forming causal hypotheses
- Be able to read and draw diagrams of causal relations, according to the system articulated in the course reader



2/28/08 8

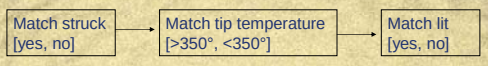
Quick Review (4 of 7)

- Be able to identify examples of some of the common errors of reasoning about causal relations
  - wrongly assuming or neglecting common causes,
  - mistaking causes for effects, neglecting circular causal relationships,
  - post hoc ergo propter hoc

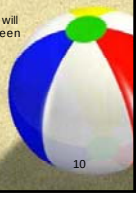


2/28/08 9

Quick Review (5 of 7)

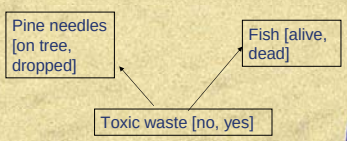


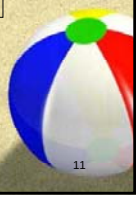
- Diagramming causal relations
  - Variables as nodes (boxes)
  - Causal relations as arrows
- Not tracing the flow of activity, but causal relations
  - If there are conditions under which changing one variable will result in change of another variable, include a arrow between the variables
- Sometimes there are important intermediate causes such that a more ultimate cause only produces its effect through a more proximate cause



Quick Review (6 of 7)

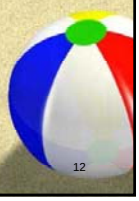
- Common cause
  - A positive correlation between two variables may be the result of a common cause for both





Quick Review (7 of 7)

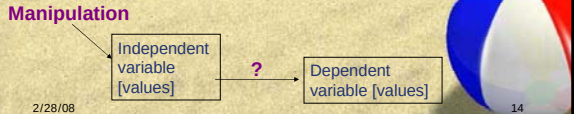
- If a causal relation is direct, there should be no way to screen off the effect from the cause.



Experiments

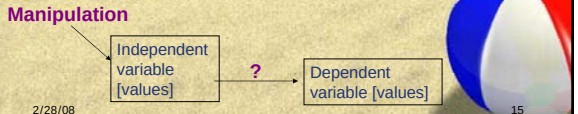
The basic idea of an experiment

- If the independent variable is a cause of the dependent variable, *manipulating* the independent variable should *change* the value of the dependent variable
- If it weren't a cause, we wouldn't expect such a result from manipulation



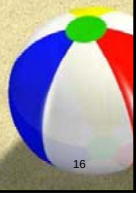
The basic idea of an experiment - 2

- Just to keep you on your toes:
  - What if the dependent variable is causally overdetermined?
  - Is it always easy to manipulate JUST the independent variable you're interested in, and nothing else?



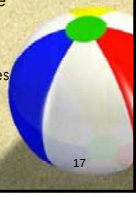
Experiments on regular deterministic systems

- When there is no **variance** in the population being studied, statistical analysis isn't necessary
  - The main danger is affirming the consequent
- The key is to test a causal hypothesis in which it is *unlikely* for the effect to occur *unless* you were right about the cause.



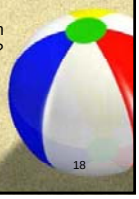
Variability in nondeterministic systems

- Different systems of the same type, or the same system/different times, will **vary** in their responses to a manipulation, depending on
  - Their particular composition and history
  - Effects of a prior manipulation
  - Interaction of the manipulation with other relevent variables
- You might also see variability in your data due to
  - Imprecision in the manipulation, or in your data collection
  - Unknown extraneous variables affecting responses
- Challenge: how to detect and learn about causal relations in the face of background variability



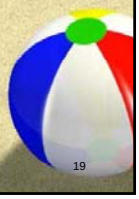
Pretend we tried our study-buddy experiment on everyone in the class

- Randomly, we assign 1/2 the students to study in groups, 2 hrs/wk, the other half don't do this.
  - No other manipulations or instructions.
  - We will compare the Midterm Score with the Final Exam Score
- Even *assuming* that studying in groups has some some effect on the Finals scores:
  - would you expect *everyone's* scores in each group to change *exactly* the same amount?
    - (e.g. everyone in the "study buddy" group improving by 10 points, everyone in the other group only improving by 2 points)
  - **what sorts of things help explain why *that* wouldn't be likely to happen?**



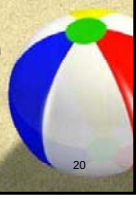
Variability can be your friend!

- On one hand, it can make things complicated
- On the *other* hand, you can LEVERAGE variability in data to get a lot of information
  - (In my opinion) the only kind of variability you "shouldn't like" is the kind that's just "noise" or error due to sloppy experimental design and procedure
    - In fact, decreasing this is another way to increase "power"
  - But variability that's due to "real" differences between individuals *can* be incredibly informative



Using variability for information

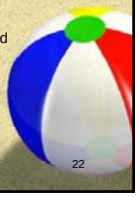
- After manipulating the independent variable, and measuring the dependent variable, you can *compare:*
  - the differences in the average values of the dependent variable *between* different samples, to
  - How much variability on the dependent variable there is *within* each sample
- *Based* on this information:
  - assess how *likely* it is that you would have seen such differences *between* the samples by chance alone
  - If it's unlikely, perhaps the independent variable you're investigating is what caused the difference
    - ...But could it have been something else?



Confounds/ Confounding Variables

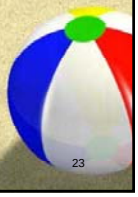
Confounding variables

- **Extraneous variables** are ones you're not trying to investigate.
- Some might be related to the effect of interest.
- These are often called ***confounds.*** 2 kinds are **particularly** important:
  - **Subject variable *confounds***: Systematic differences between the subjects in different groups
  - **Procedural variable *confounds***: Systematic differences in the way different groups are treated



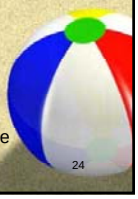
Confounding variables

- If extraneous variables are *correlated* with the independent variable and are *also causes* of the dependent variable, the experiment may be ***confounded***



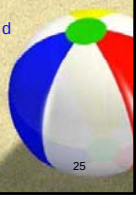
Subject variable confounds

- Subjects in an experiment may be differently affected by different values of other variables, e.g.
  - People of different ages sleep different amounts
  - Women might be affected differently than men
- These variables are extraneous to what you want to test
- If there is a correlation between these variables and the independent variable, *they*, rather than the variable you are *focusing* on, may be what produce the change in the dependent variable, or may alter the degree or character of the change



Procedural variable confounds

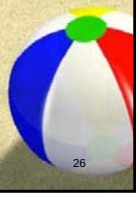
- When you conduct a manipulation, generally more than one thing will be changed
  - These variables will then be correlated with the independent variable—extraneous
    - (like the influence of having study buddies on budgeting your time)
  - If one of the other variables is causally related to the effect of interest, it rather than the variable you are considering may either be the cause, or at least alter the magnitude or character of the effect--confound



A general, intuitive way to think of confounds

- Roughly, you might think of them as *any* sort of factor at all that could interfere with or in any way skew your ability to accurately see, characterize, understand, isolate the effect of your manipulation of the independent variable on the dependent variable.

Things are interfering with my ability to "see" what's going on

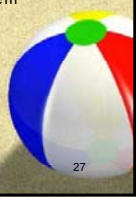


Effect of Eating Veggies on Body Weight

Manipulation

Veggies [lots, little] → ? → Body Weight Change [lbs]

- Imagine that for some people, eating vegetables makes them eat a lot more Ranch Dressing
- Imagine that for others, eating veggies makes them feel less overstuffed, so they move around more
  - To zero in on the effects of veggies, is it fair to try to equalize the use of Ranch Dressing between sample groups?
  - What about level of activity? Is it "getting in the way," or part of the mechanism by which eating vegetables is supposed to affect body weight in the first place?
  - Need to be thoughtful about these things.



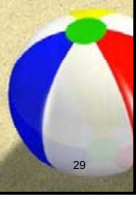
Clicker Question

In the previous "Vegetable Eating/Change in Body Weight" experiment, I manipulate subjects' intake of vegetables, and look for an effect on body weight after some interval of time. In both groups, though, some subjects gained weight, and some subjects lost weight, although on average the "lots of vegetables" group lost a little more. What can I tell from this information?

1. Based on the Method of Difference, it doesn't look like eating lots of vegetables is the common denominator among subjects who lost weight, so it's not likely to be a cause of weight loss.
2. It depends. If there's a big difference on average between groups and not so much variability within groups, then maybe eating lots of vegetables is a cause of weight loss.
3. It depends. If there's a lot of variability within groups and not so much difference on average between groups, then maybe eating lots of vegetable is a cause of weight loss.
4. You can conclude pretty confidently that vegetables cause weight loss, because of we still have lots of prior expectations that they probably would do so.

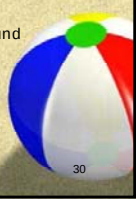
Example of confounding procedural variables

- The president of the AGL corporation wanted to get her workers to be more productive. She found that when each employee was presented with a jar of jellybeans, productivity increased.
- Was it the jellybeans that caused the increased productivity? Or was it:
  - Novelty of the situation
  - Attention from the president
  - Desire to reciprocate



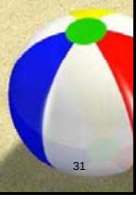
More sources of confounding

- **Stimulus confounding** Differences between the variable we *want* to manipulate and the concrete situation we *actually* manipulate
  - The physical situation of a manipulation always involves at least *some* other variables besides the intended independent variable, some of which may influence what you observe
  - Possible examples?
- **Response confounding** A result may be a compound of two or more response processes.
  - Ambiguity in a question, effects of being exposed to a new situation
  - Overly simplistic measure used to represent a very complex process



More sources of confounding

- **Error variability.** Sloppy procedures and measurements.
- **Conceptual confounding** Mixing up ideas that should be kept distinct.
  - For instance, in your own thinking you could start to confuse a theoretical construct with the specific operational definition
- **Fraud.** It happens.
  - In certain areas of research, regulators attempt to control this by requiring extensive records retention, audit trail, and cross-monitoring and data validation and verification procedures



Clicker Question

Study of strength of molded plastic. Hot plastic was injected into a mold, pressed for 10 seconds, then removed. Repeated using 20 seconds, 30 seconds, etc. Data showed nice curve of increasing strength as a function of pressing time. (Wilson 1952). Were there any obvious confounds in this study?

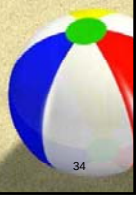
1) No. Looks good to me.
2) There was a subject confound. They probably used different kinds of plastic in the different conditions.
3) There might be a procedural confound because the order of trials was not randomized

Controlling for Confounding Variables

"…You gotta do the best you can…"

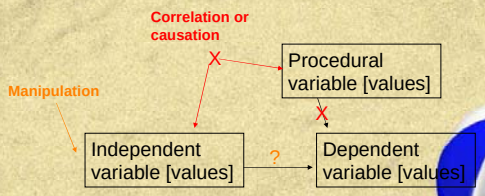
In general, how do you deal with confounds?

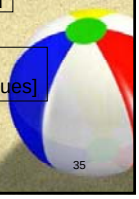
- Balance, hold constant, eliminate, or otherwise rule out the influence of extraneous variables, especially confounds.
- Ultimately, it will basically boil down to
  - using our best judgment,
  - actively seeking to learn from the past.
  - using as much background knowledge about the phenomena as we can, and
  - getting a "knack" for asking and assessing tough questions
- Much depends on details specific to each research area, the main goals of a particular study.
- Easy to be surprised by the unexpected.



Controlling confounding procedural variables

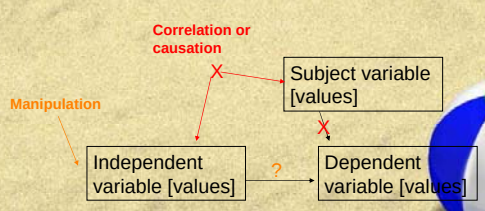
- Strategy: break the correlation—thereby breaking the effect of the confounding variable

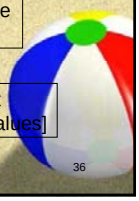




Controlling confounding subject variables

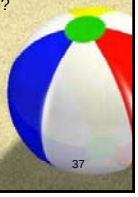
- Strategy: break the correlation—thereby breaking the effect of the confounding variable (e.g., by randomization)





Some strategies for controlling confounding variables - 1

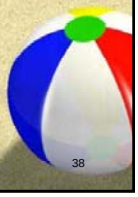
- Randomization. Aims to "balance out" groups before manipulation.
  - Use a random procedure to *assign* potential study participants to different experimental groups
  - How might this control for subject confounds?



2/28/08 37

Some strategies for controlling confounding variables - 2

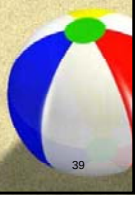
- Matching subjects/balancing sample groups on confounding variables
  - E.G. If you're aware of a few especially important subject variables, you might **ensure** ahead of time that similar proportions are enrolled in each group.
- Locking. Most commonly used to control confounding procedural variables
  - Try to make sure that as many extraneous variables as possible are held constant, and do not differ between groups.
  - It can be tricky to decide which variables you want to lock.



2/28/08 38

Some strategies for controlling confounding variables - 3

- Controlling for a confound by eliminating it.
  - Screening out undesirable subjects
  - Allowing enough time between conditions in a within-subjects design
  - Using background noise to mask the sound of an apparatus
- Careful and consistent planning and execution
  - Precise procedures and instructions
  - Validation procedures, audits, supervision
  - Operational definitions and measurements as precise and standardized as possible



2/28/08 39

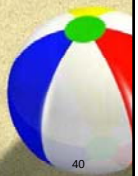
Effect of Eating Veggies on Body Weight

Manipulation

Veggies [lots, little] → ? → Body Weight Change [lbs]

- For some people, eating vegetables makes them eat a lot more Ranch Dressing
- For others, eating veggies makes them feel less overstuffed, so they move around more
  - Should we hold constant the use of ranch dressing? Activity level?
  - Should we try to balance each study group?
  - Should we have each subject try both?
  - Should we *measure* the intake of ranch dressing and activity level and factor them into our analysis?

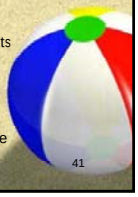
How do you decide?



2/28/08 40

Strategies for controlling the placebo effect

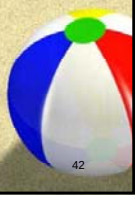
- Placebo effect: influence of *mere expectation* of effect
  - Adding a placebo "control group," and/or lead-in control periods before taking baseline measurements
  - Single, double, or triple blinding the study
    - Single blind: subjects don't know which group
    - Double blind: subjects AND whoever collects data in relatively close proximity to the subjects don't know
      - but people involved in safety monitoring, for instance, might know
    - Triple blind: Nobody directly or indirectly involved in study the knows, until the database is "frozen."



2/28/08 41

Control Groups

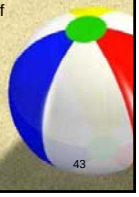
- Sometimes the "control group" is not a placebo, but a competing treatment or manipulation
  - It may be unsafe to leave people untreated
  - It may be *unimpressive* to produce a statistically significant result against a placebo.
    - You don't get to measure the "placebo effect."
    - Decision depends on the context, prior knowledge and research about the independent variable, your goals, etc.
    - This can be a sticky subject.



2/28/08 42

Between vs. Within Subjects Design

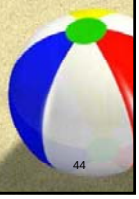
- Between subjects randomization -
  - assign subjects randomly to different groups,
  - manipulate the independent variable between groups
- Within subjects design -
  - subjects serving as their own control - each subject experiencing all the manipulations of the independent variable
  - the order of the conditions may be randomized to control for order effects.



Counterbalancing

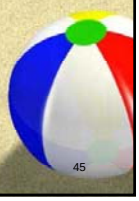
- Within subject counterbalancing
  - Reversing order: ABBA
- Across subject counterbalancing
  - Helps control for procedural variables such as order of exposure to different conditions
  - Complete: every possible sequence—if there are a lot of conditions, requires a lot of subjects!
  - Partial: Random, or Latin Square
    - Latin Square: each condition appears once and only once in a given ordinal position, no two conditions are juxtaposed in the same order more than once. E.G.:

      Order 1: A B D C
      Order 2: B C A D
      Order 3: C D B A
      Order 4: D A C B



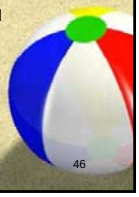
Between-subjects design

- **GOOD NEWS**:
  - Subjects are not "contaminated" in one condition as a result of having participated in the other
- **BAD NEWS**:
  - Requires a larger number of participants
  - Runs the risk of non-equivalence of subject groups



Within-subjects design

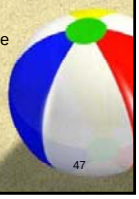
- **GOOD NEWS**:
  - Requires a smaller number of participants
  - Rules out any differences between subjects
- **BAD NEWS**:
  - Potential "contamination" of participants' behavior from previous trial: *carryover effect*
  - Subjects might learn from one condition and that could alter their behavior in the second condition
    - *Practice effect*
    - *Fatigue effect*



2/28/08 46

Example: alcohol and running speed

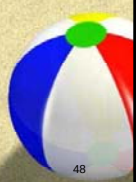
- Does alcohol affect speed of running a 100 yard dash?
- Between-subjects design
  - Different subjects would be used for the no-alcohol and alcohol condition, and each would be tested only once
- Within-subjects design
  - Each subject would be tested both under the no-alcohol and alcohol condition



2/28/08 47

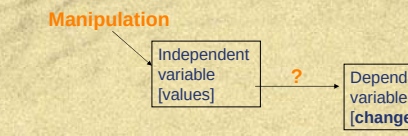
Pretest-posttest design

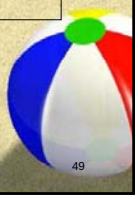
- There is always a danger in an experiment that the members of the two (or more) groups being studied already differ on the dependent variable
- Best control is to focus on change, not raw value of the dependent variable
  - Pretest: measure the dependent variable before the intervention
  - Posttest: measure the dependent variable after the intervention
- What's evaluated is the Change: Posttest – Pretest



2/28/08 48

Pretest-posttest design

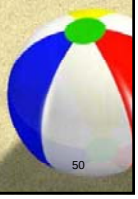




2/28/08 49

Limitations of pretest-posttest design

- Just measuring change in one group using a pretest and a posttest allows for confounds
  - Time has elapsed and subjects have gotten older (maturation)
  - Events occurring between the pretest and posttest could affect the dependent variable (history)
  - Experience with previous test may change performance
  - Pretest and posttest may vary in difficulty
- **Use of pretest-posttest does not obviate the need for a control group**



2/28/08 50

Clicker Question

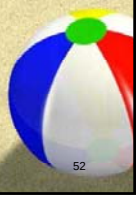
Which among the following is the best way to control for the confound of experimenter bias (placebo effect)?

1) Using a double or triple blind study design.
2) Using a within subjects rather than a between subjects design.
3) Balancing to ensure that each experimenter sees a similar number of subjects in each experimental condition.
4) Measuring the effect of the experimenter's opinion on the results, and subtracting that from the observed effect.

Sometimes it comes down to further research

- Sometimes you just have to wait for further research to be performed, and see what emerges over time.
  - But you may not get the opportunity to try again of your study "fails!"
    - Worth noting: Sometimes experimental design is constrained by safety, financial, or other practical considerations.



2/28/08 52

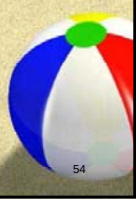
Internal and External Validity!


2/28/08 53

Internal and External Validity

- Not to be confused with validity in reference to arguments and sentential logic!
  - But in an intuitive sense, there's a connection.
  - We're sort of thinking of the following kind of question: Are some of the "conclusions" a researcher may be trying to draw *really* all that well-supported by his or her "premises" - - i. e., the data and the design details of an experiment?

(This mother and baby are skeptical that your conclusions are well supported…)

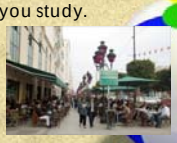



2/28/08 54

Internal and External Validity

- An important aspect of validity is whether your measures are a veridical/accurate/true/informative representation of the phenomenon or process you *seek* to study.
- It's also important to consider the extent to which your results extend to, or give you information about, situations outside of or other than the exact type of situation you study.





More "intuitively"…

- Internal Validity asks: *Within* the scope of this specific experiment with this subject population and this experimental manipulation, does your data really support the conclusions you made?
- External Validity asks: With respect to the "real world" outside of your little experimental bubble, does your experiment support any interesting claims, should we expect similar effects?

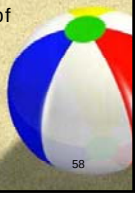
(Still skeptical…)

Internal validity

- An experiment is internally valid if it was in fact the manipulation of the independent variable that produced the change in the dependent variable
  - Was there an important confounding variable that did not get controlled?
  - Are the effects on the dependent variable due solely to the manipulation of the independent variable?
  - Was the study designed and powered well enough to avoid false-negative results?

External validity

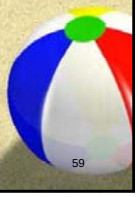
- To what extent can you ***generalize*** the results of your study?
  - Are they specific to a particular sample?
  - College sophomores or the general population
- Do they only apply in a particular (laboratory) setting?
- Do they generalize beyond the details of the manipulation?
- Sometimes called **ecological validity**



2/28/08 58

External Validity: Population generalization

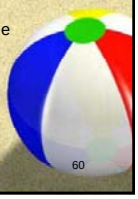
- Will a study using one population generalize to another population? For example:
  - Will a study of college sophomores generalize to middle-aged adults?
  - Will a study of chronically depressed patients generalize to patients who are acutely depressed?
  - Will a study of captive raised dolphins generalize to wild dolphins?
  - Will a study on mice generalize to humans?



2/28/08 59

External Validity: Setting Generalization

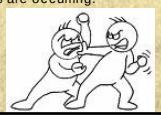
- Will a study conducted in one laboratory or clinical setting generalize to the setting of interest?
  - Will results obtained in a flight simulator generalize to an actual cockpit?
  - Will results obtained in an outpatient setting generalize to a psychiatric hospital?
  - Will results obtained in a laboratory generalize to customers in a store?

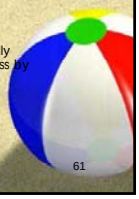


2/28/08 60

Competing Goals

- Sometimes, getting information about the *mechanism of a process*, and about "real-world" or "bottom-line" *outcomes*, are *competing* goals.
- Sometimes, researchers choose to focus on just one of these goals
  - For the goal of internal "process" validity, a primary consideration is to simplify and "lock" the situation in order to eliminate confounding from processes other than the process you mean to study.
  - External "outcome" validity is best obtained when the situation studied is more "realistic." These are often really complex, more confounds, harder to identify the process by which the effects are occurring.





2/28/08 61

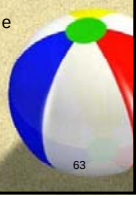
Clicker Question

Which of the following primarily expresses a concern with external validity?

1) I am worried that my study will not achieve statistical significance.
2) I am worried that I've limited the sorts of subjects who are eligible for my study so carefully, that I can't say what sorts of effects the independent variable has on other kinds of people.
3) I am worried that the subjects are responding to the novelty of the experiment more than the specific effect of the independent variable that I manipulated.
4) I am worried that some of my subjects figured out which treatment they were on.

Example: Rats and Saccharine

- 1977 Canadian study which fed pregnant rats up to 20% of their body weight per day in saccharine showed an increase in bladder tumors
- Saccharine was banned in Canada and the FDA was about to ban its use in the US when Congress intervened
- Assessing external validity. These would be the sorts of questions you might want to ask:
  - Are rats **relevantly** like humans in this context?
  - Is eating in the laboratory like eating at home, etc.?
  - Is feeding up to 20% of body weight like eating as part of diet?



2/28/08 63