Pitfalls in Research

Things that can go wrong in the process of doing science

Pitfalls in Research

(adapted from T.X. Barber)

- Changing nature of science
  - Classic Image: Little Science
    - Scientist in white lab coat working by self
  - Reality: Medium to Big Science
    - Research Teams
    - Research Labs
    - Cross university-cross national research groups
- Investigators and Experimenters
- Sources of error due to Investigators and Experimenters

Investigator Paradigm Effects

- Paradigm: a logical or conceptual structure serving us as a form of thought within a given area of experience
- Kuhn and the philosophy of science: each period of normal science in the development of a scientific discipline corresponds to one and only one methodological framework or *paradigm*. In a nut-shell, paradigms are `universally recognized scientific achievements that for a time provide model problems and solutions to a community of practitioners'

(OED)

Investigator Paradigm Effects

- Paradigms and paradigm shifts (Kuhn)
- Paradigms as a shared collection of beliefs, methods, and problems to be addressed
- Paradigms help organize research and data
- Paradigmatic research facilitates communications with a research community
- Shared methods and shared analytical techniques

Paradigmatic thinking as a potential source of error

- Tenacity of paradigms and resistance to new discoveries
  - theories aren't disproved, old theorists die
- Failing to see events that do not fit within paradigm
    - Continental drift and plate tectonics
      - Alfred Wegner and theory of continental drift (1912)
      - Early suggestions by Snider-Pellegrini (1858) and Seuss (1885)
      - But no mechanism to explain it
      - Harry Hess (1962) and sea floor spreading to account for oceanographic findings of ridges, trenches, magnetic striping
    - Glacial dams and the Grand Coulee
      - Harlan Bretz (1922) described the channeled scablands of Montana and Eastern Washington
      - His catastrophic theory was rejected until 1965 by geologists holding to uniformitarianism

Paradigmatic thinking as a potential source of error

- Seeing non "events" that fit within a paradigm
    - Blondlot and N-Rays
        - Original claims of N-rays
        - Robert Wood and the critical experiments
    - Cold Fusion
        - Fleischman and Pons
    - Lysenko and the transmission of environmental effects
- "They laughed at Galileo, they laughed at Columbus, they also laughed at Bozo the clown." (Sagan, 1980)

Paradigms in Psychology

- Behaviorism
  - Forces operationalizations, ignores constructs
- Cognitivism (cognitive psychology, cognitive social, cognitive learning, …)
- Tabula Rasa paradigm of development
  - Ignores evidence for genetic causes
- Neuroscience reductionism
  - Ignores evidence for social causes

Paradigms in Personality

- Freudian dynamics
  - Importance of childhood in adult behavior
  - Unconscious processes driving thought
- Trait theories of consistency
  - Assumes that people reflect underlying trait differences
- Social learning theories of variability
  - Everything is learned, no trans-situational consistency
- Biological bases of personality

Investigator Design Effects

- Confirmatory designs
  - If A then B
  - Frequently test by doing A and observing B
  - But what about observing Not B?
- Demonstrations versus tests
  - Experiments that are consistent with theoretical predictions, but do not test the theory
- Failure to pit theory against theory
  - Strong inference (Platt)

Reasoning in Research

- Karl Popper and the testability of theory
  - The hallmark of science is the testability of theory
  - Non-testable theories are not science
  - "it must be possible for all empirical scientific system to be refuted by experience"
  - Theories are not shown to be correct, they are shown to be incorrect
- Science is the process of asking questions that have answers (Rep. Rush Holt)

J. Platt and Strong Inference

(Science, 1964)

- 4 signs of strong science
  - Devising alternative hypotheses;
  - Devising a crucial experiment (or several of them), with alternative possible outcomes, each of which will, as nearly is possible, exclude one or more of the hypotheses;
  - Carrying out the experiment so as to get a clean result;
  - Recycling the procedure, making subhypotheses or sequential hypotheses to refine the possibilities that remain, and so on.

Strong inference

- A theory which cannot be mortally endangered cannot be alive. (Rushton, as cited by Platt)
- "The problems of how enzymes are induced, of how proteins are synthesized, of how antibodies are formed, are closer to solution than is generally believed. If you do stupid experiments, and finish one a year, it can take 50 years. But if you stop doing experiments for a little while and *think* how proteins can possibly be synthesized, there are only about 5 different ways, not 50! And it will take only a few experiments to distinguish these." (Szilzard, as cited by Platt)

Platt and Strong Inference

"I will mention one severe but useful private test - a touchstone of strong inference - that removes the necessity for third-person criticism, because it is a test that anyone can learn to carry with him for use as needed. It is our old friend the Baconian "exclusion," but I call it "The Question." Obviously it should be applied as much to one's own thinking as to others'. It consists of asking in your own mind, on hearing any scientific explanation or theory put forward, "But sir, what experiment could *dis* prove your hypothesis?"; or, on hearing a scientific experiment described, "But sir, what hypothesis does your experiment *dis* prove?" Platt, *Science*, 1964

Confirmatory designs and weak inference

1. If Introverts are more aroused than extraverts
2. If arousal varies throughout the day
3. If arousal is increased by caffeine
4. If cognitive performance is curvilinearly related to arousal
5. Then introverts should be hurt (helped less) by caffeine than extraverts (in the morning) and extraverts should be hurt (helped less) by caffeine than introverts in the evening.

- We observe 5, can we infer 1-4?

Confirmatory designs and problems of inference

- If A implies B and we observe B, does this imply A?
  - No. Not B implies not A.
- If A and B and C and D imply E, and we observe E, does that imply A, B, C, & D?
  - No.
  - But Not E implies one of A, B, C, & D is not true, but which one?

Correlation and inverse probabilities

- Does observing that B almost always happens when we do A imply that doing A almost always leads to B?
- Examples:
  - B A
  - Auto accidents and drinking
  - Lung cancer and smoking
  - Pregnancy and intercourse

Correlation and inverse probabilities

- If one has disease B, then one tests A+ with p=.99
- If one tests A-, then one has disease B with probability .01
- 99% of people do not have the disease
- If one is Test A+, what the probability that they have disease B?

Correlation and inverse probabilities

		Then Test A+	Then Test A-	
If	Disease	0.99	0.01	
	Not disease	0.01	0.99	

Correlation and inverse probabilities

	Test A+	Test A-	frequency
Disease B+	0.99	0.01	100
Not disease B-	0.01	0.99	9900

Correlation and inverse probabilities

	Test A+	Test A-	frequency
Disease	99	1	100
Not disease	99	9801	9900
	198	9802	10,000

Probability of Disease|Test A = .5!

Investigator Loose Procedure Effects

- Poor specification of how to conduct experiment
  - “make the subject relaxed”
  - Record memories of childhood
- Badly defined specification of manipulation
  - Threaten with fear of public speaking

Investigator Data Analysis Effects

- “Pilot Studies’ and the filing cabinet syndrome
  - How many studies are done before “the experiment”
  - The meaning of significance tests when many studies are done but only one is reported
- Stopping rules, data “purification”, deletion of “outliers”
- Hypothesis development and “confirmation”
- Over interpretation of findings
- Meaning of significance tests versus experiment-wise error rates

Investigator Data Analysis Effects

- Significance tests and experiment wide significance
- Interpretability of randomness
- Meta analytic technique for averaging results across studies
- emphasis upon effect sizes and confidence estimates

Investigator "Fudging" Effects and Scientific Fraud

- Trimming, Cooking, idealizing
  - Galileo, Newton, Dalton, Mendel, Millikan
- Generation of new "confirming cases"
  - Burt?
- Outright Fraud
  - Piltdown Man
  - Summerlin, Schon, Darsee, Hwang
  - Unknown prevalence assumed (hoped) to be small
  - Nonexistent data, nonexistent subjects

Examples of fraud in science

(from Judson, The Great Betrayal cited by swartz http://www.aaronsw.com/weblog/001616

- William T. Summerlin (chief of transplantation immunology at Sloan-Kettering) claimed he could transplant onto animals corneas, glands, and skin that would normally be rejected — sometimes even across species. He was discovered only after three years of this when a lab assistant noticed that the black "skin graphs" were drawn on with a marker (all the rest of his work turned out to be fake as well).

- John Long (a resident) studied Hodgkins's cell lines at Mass General in collaboration with MIT. A year later, a junior colleague charge fraud and it was discovered that the cell lines were from monkeys and healthy people.

- Elias A. K. Alsabti (a researcher at Boston University) had published sixty papers by his mid-twenties, when it turned out that most of them were papers published in obscure foreign journals with only slight changes (like a new title).

- Vijay Soman, an assistant professor at Yale, was asked to peer review a paper by Helena Wachslicht-Rodbard. He sent back a negative review, delaying publication, then turned around and submitted the same paper to another journal. He was found out when, in an amazing twist of fate, Helena Wachslicht-Rodbard was asked to peer review Soman's paper and recognized it as her own.

- John Darsee had published dozens of papers with completely made up data — and done an incredibly bad job making up the data. (One paper claimed a father had four children — conceived when he was 8, 9, 11, and 12 years old, respectively.) To cover up this fact, Darsee had practiced "gift authorship" — adding people as co-authors even when they didn't do any work. Darsee had been at Harvard for three years before he was discovered by some postdocs, even then it took the university five months to admit the fraud.

The Darsee case

Dr. John Darsee was regarded a brilliant student and medical researcher at the University of Notre Dame (1966-70), Indiana University (1970-74), Emory University (1974-9), and Harvard University (1979-1981). He was regarded by faculty at all four institutions as a potential "all-star" with a great research future ahead of him. At Harvard he reportedly often worked more than 90 hours a week as a Research Fellow in the Cardiac Research Laboratory headed by Dr. Eugene Braunwald. In less than two years at Harvard he was first author of seven publications in very good scientific journals. His special area of research concerned the testing of heart drugs on dogs.

In May 1981, three colleagues in the Cardiac Research Laboratory observed Darsee labeling data recordings "24 seconds," "72 hours," "one week," and "two weeks." In reality, only minutes had transpired. Confronted by his mentor Braunwald, Darsee admitted the fabrication; but he insisted that this was the only time he had done this, and that he had been under intense pressure to complete the study quickly. Shocked, Braunwald and Darsee's immediate supervisor, Dr. Robert Kroner, spent the next several months checking other research conducted by Darsee in their lab. Darsee's research fellowships were terminated, and an offer of a faculty position was withdrawn. However, he was allowed to continue his research projects at Harvard for the next several months (during which time Braunwald and Kroner observed his work very closely).

Hopeful that this was an isolated incident, Braunwald and Kroner were shocked again in October. A comparison of results from four different laboratories in a National Heart, Lung and Blood Institute (NHLBI) Models Study revealed an implausibly low degree of invariability in data provided by Darsee. In short, his data looked "too good." Since these data had been submitted in April, there was strong suspicion that Darsee had been fabricating or falsifying data for some time. Subsequent investigations seemed to indicate questionable research practices dating back as far as his undergraduate days.

http://www.wmich.edu/ethics/ESC/cs1.html

The Bruening case

In December 1983, Dr. Robert Sprague wrote an eight page letter, with 44 pages of appendices, to the National Institute of Mental Health (NIMH) documenting the fraudulent research of Dr. Stephen Breuning.(53) Breuning fabricated data concerning the effects psychotropic medication have on mentally retarded patients. Despite Breuning's admission of fabricating data only three months after Sprague sent his letter, the case was not finally resolved until July 1989. (Sprague credits media attention with speeding things along!) ... Most painful of all, Sprague's wife died in 1986 after a lengthy bout with diabetes. In fact, his wife's serious illness was one of the major factors prompting his "whistleblowing" to NIH. Realizing how dependent his diabetic wife was on reliable research and medication, Sprague was particularly sensitive to the dependency the mentally retarded, clearly a vulnerable population, have on the trustworthiness of not only their care givers, but also those who use them in experimental drug research.

Writing nine years after the closing of the Bruening case, Sprague obviously has vivid memories of the painful experiences he endured and of the potential harms to participants in Bruening's studies. However, he closes the account of his own experiences by reminding us of other victims of Bruening's misconduct--namely, psychologists and other researchers who collaborated with Bruening, but without being aware that he had fabricated data.

Dr. Alan Poling, one of those psychologists, writes about the consequences of Bruening's misconduct for his collaborators in research. Strikingly, Poling points out that between 1979 and 1983, Bruening was a contributor to 34% of all published research on the psychopharmacology of mentally retarded people. For those not involved in the research, initial doubts may, however unfairly, be cast on all these publications. For those involved in the research, efforts need to be made in each case to determine to what extent, if any, the validity of the research was affected by Bruening's role in the study. Even though Bruening was the only researcher to fabricate data, his role could contaminate an entire study. In fact, however, not all of Bruening's research did involve fabrication. Yet, convincing others of this is a time-consuming, demanding task. Finally, those who cited Bruening's publications in their own work may also suffer "guilt by association." As Poling points out, this is especially unfair in those instances where Bruening collaborations with others involved no fraud at all.

http://www.wmich.edu/ethics/ESC/cs1.html

Reasons for Fraud

- Pressure for success
  - Grant funding
  - Tenure review
- Hubris/Psychopathology
  - I know I am right, they are so stupid not to recognize it
- Poor supervision -- too large a lab
  - Schon’s data not analyzed by supervisors
- Lack of replication, pressure to be new

Experimenter Effects

- Experimenter personal attributes
  - Stable
    - Sex, age, ethnicity, height, weight, age
  - Modifiable
    - Prestige, anxiety, friendliness, warmth
- Possible interactions of experimenter characteristics and subject characteristics
  - Sex of experimenter by sex of subject
  - Prestige of experimenter by anxiety
  - Etc.
- Partial control for these by making every experimenter run all conditions (avoids experimenter by condition effects)

Experimenter failing to follow procedures effects

- Impossible, difficult and inconvenient procedures
  - Have they been pilot tested
- Poorly specified procedures
  - What actually is supposed to be said or done
- Poor training, lack of practice, lack of supervision
  - Does the experimenter know how to do it?

Experimenter Misrecording Effect

- Mistakes of data collection
  - What condition was the subject supposed to be
  - What manipulation was actually done
- Mistakes at data entry
  - Data entry and data checking
  - What protection against mistakes in entry
    - Double entry (two independent sources)
    - Automatic entry (but what about programming errors)

Experimenter Fudging Effects

- Lack of interest in the outcome
  - Who cares, why bother
- Too much interest in the outcome
  - Grade depends upon significance! (no!)

Experimenter Expectancy Effects

- Pygmalion in the classroom
  - An supposed example of expectancy effects
  - An actual example of sloppy procedures and misanalyzed data
  - (see Thorndike, R.S. "Review of Pygmalion in the Classroom." AMERICAN EDUCATIONAL RESEARCH JOURNAL 5 (1968): 708-711)
  - But a real effect when done properly

    Younger sibs do better with same teacher if older sib did better, no difference with different teacher

Recommendations-1

- Be aware of one's own and alternative paradigms
  - Alternative procedures
  - Alternative explanations
- Try to do strong inference
- Tighten protocols
- Replicate
  - If it is worth doing and worth reporting, it is worth replicating
  - If you can't replicate your findings, who can?

Recommendations-2

- Delicate balance between theory testing and theory development
  - Methodological rigor versus theoretical speculation
  - Significance tests based upon prior hypotheses
  - Speculation based upon fortuitous findings
    - Issues of type I and Type II error
  - Examine the data, create hypotheses, design new studies to test those hypotheses.

Recommendations - 3

- Things to consider
  - Statistical Significance versus real significance
    - How likely is result to happen by chance (alpha level)
    - How important is this effect in real world
  - Point estimates and confidence intervals
    - What is the value of the effect and what is the range
  - Effect sizes:
    - differences in terms of within cell error
    - Size of correlation coefficient
  - Power
    - How likely can you find a result if it is there?
    - Expressed in probability of significance given effect size of X.