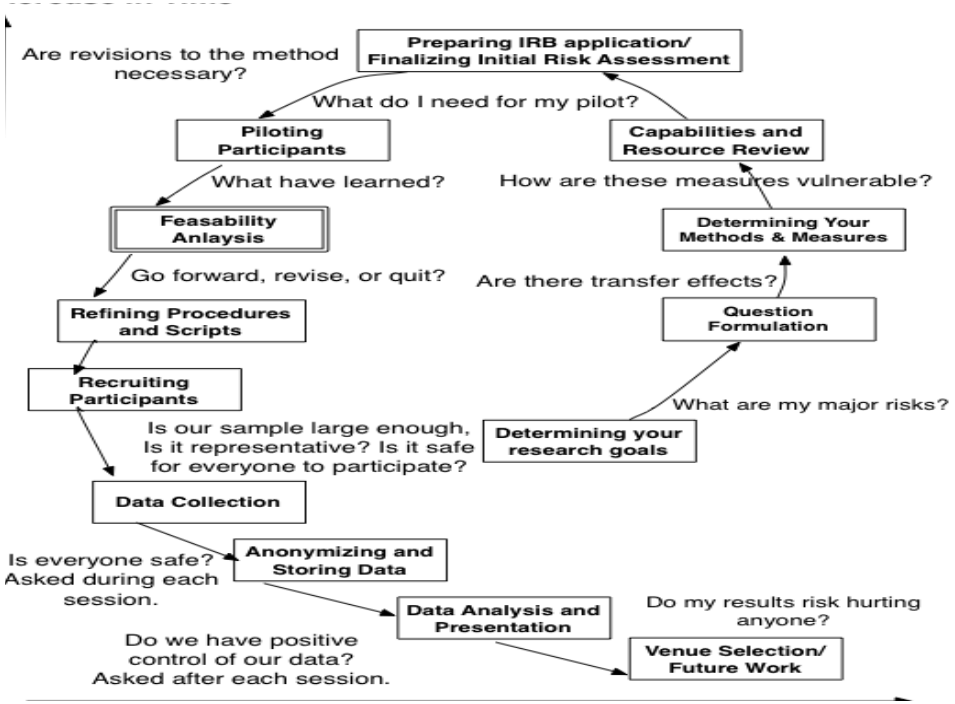
A Risk Driven Approach to Experimental Design and Practice

Frank E. Ritter and Jonathan H. Morgan (slides)

The College of IST
Penn State
&

Jong W. Kim and Richard Carlson (book)

Psychology, U. of Central Florida, and Psychology, Penn State



Supported by ONR, Contracts N00014-11-1-0275, N00014-10-1-0401, & W911QY-07-01-0004

Overview

📖 acs.ist.psu.edu/papers
acs.ist.psu.edu/reports/ritterKM09.pdf
www.frankritter.com/rbs/ rbs-handout-cogsci.pdf (TB, p. 3)

0900-0915 (0) Orientation

0915-0945 (1) An overview of risk-driven experimental design

0945-1015 (2) Preparation for running an experiment

1015-1040 break

1040-1115 (3) Ethical challenges in the experimental process

1115-1145 (4) Risks to validity, with class participation

1145-1200 Slack

1200-1215 (5) Conducting an experiment

1215-1230 (6) Concluding a study and reporting results, Summary

Who are you?

1) Name, organization, background, number of studies, what you want to get from this
2) Please form into pairs for later exercise

Summary 1 of tutorial:

(Re)Looking at failure: What constitutes a failure/risk?

- **Someone got hurt**
- **After committing significant resources, the study was never completed**
- **We have learned nothing new because our data is not repeatable or generalizable**
- **We have failed to communicate our results or their significance to anyone else**

Sources of Failure?

- **Why did someone get hurt?**
  - ***We failed to do a risk assessment***
  - ***Being prepared for unanticipated problems***
  - ***We failed to screen participants properly***
  - ***We failed to either develop or follow procedures, either experimental procedures or data management procedures***
  - ***We did not anticipate or mitigate situational risks either in our experimental setting or outside of it that hurt our participants***
  - ***We ignored additional insights we could have learned from the participants through observation or debriefing***
  - **Others?**

Sources of Failure?

- **Why we were unable to complete the study?**
  - ***We were overly ambitious***, **perhaps because we failed to fit the research question or methods to the problem at hand**
  - ***We ran out of time***
  - ***We ran out of resources or lacked them in the first place***
  - ***We lacked the people, either participants or staff, or trained staff***

  (experiments appear to have less risk than modeling)

Sources of Failure?

- **Why we were unable to reproduce our results or generalize them?**
  - ***We failed to use the same experimental procedures or test under the same conditions for each S***
  - ***We failed to achieve an adequate sample size or sufficient degree of representativeness in our sample***
  - ***Our task fidelity was poor. We failed to construct an experimental task that was analogous with respect to its key points.***

Sources of Failure?

- **Why have we been unable to report our results or communicate their significance?**
  - ***We failed to properly catalog or backup our data***
  - ***We failed to write as we went. We no longer remember some of the critical, early details.***
  - ***We made poor data analysis or display choices***
  - ***We failed to identify a venue early, or understand who we should consider our audience***

How do we avoid failure?

- We recognize that running a study is an incremental risk-driven process, similar in some respects to spiral development of systems
  (Boehm & Hansen, 2001; 🕮Pew & Mavor, 2007)
- To be successful, we need to:
  - Formulate a research question that meets our research goals
  - Have a theory of transfer effects that minimizes risks associated with confounding variables, and enables us to conserve time and resources
  - Pilot studies and study components
  - Be candid in our risk assessments and be willing to adapt and refine

What to get out of this Tutorial

1) Some feeling for how to run a study

- Cognitive science may be modeling + data
  So, to use data you have to know how it was gathered
- Modeling is slow, so data publication helps modelers
- If you are a computer scientist, you won't have taste in this area
  => Help you develop a green thumb
- Not how to *design* a study, but related

2) Some tools to help you set up a study

3) Materials

Book and report on this topic (please let me know if you use it for a class)

Handout (available online)

Example problems

4) A break at ~1015 am

5) A greater appreciation for mistakes to avoid and a theory of how to avoid them

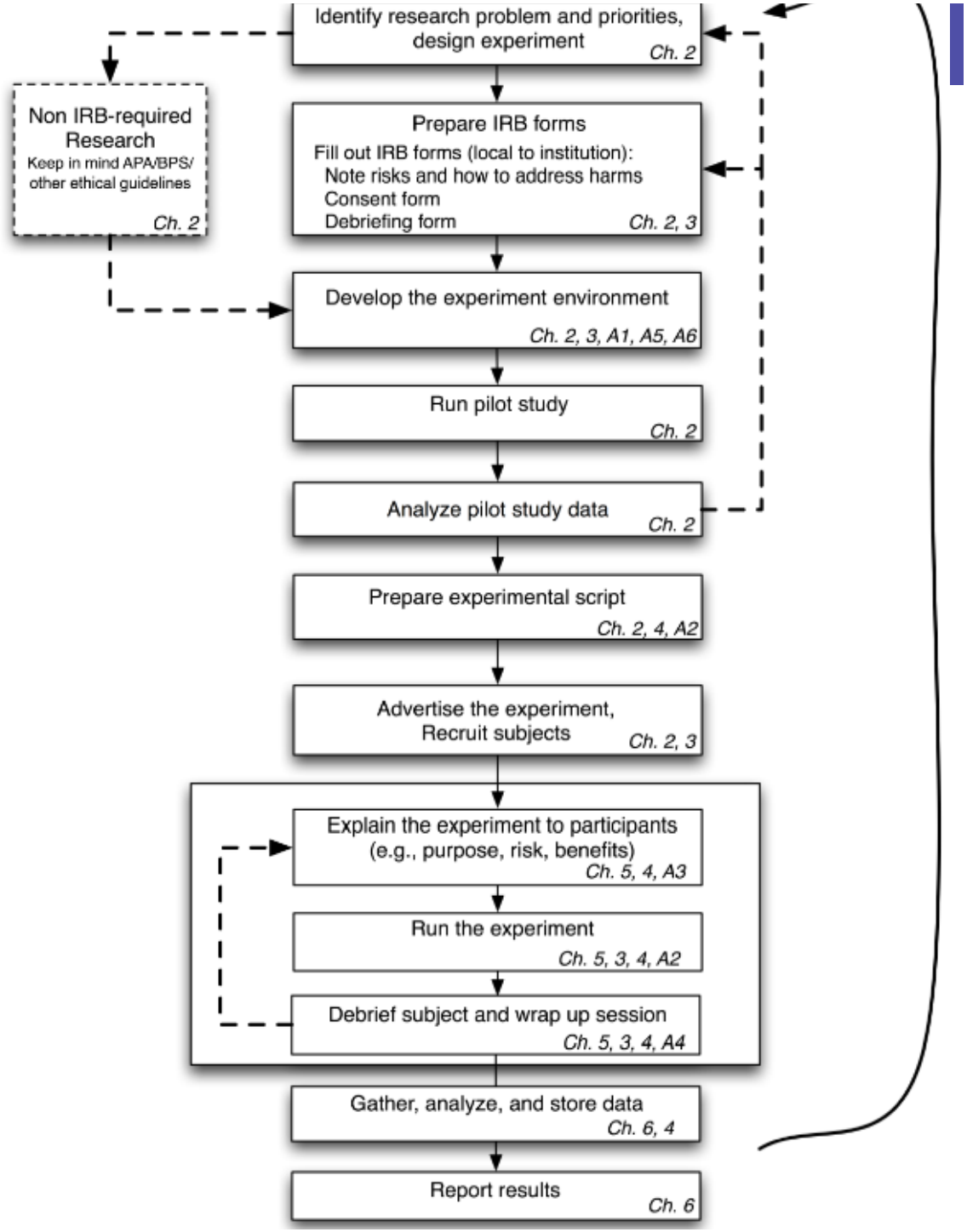
Ch 1. Overview
Some Terms used

- A study, varying an Independent variable (IV, e.g., amount of practice), to see the effect on a dependent variable (DV)
  - Worth reading a methods book(s)
- Subjects (Ss) or Participants (Ps), Users, learners, students, Experimenters (Es)
  - See *APA manual* and also Roediger (2006) for arguments for S and P and U/L/S/S
- Example studies
  - Multi-lingual fonts
  - Partially sighted and blind users
  - HRI

Experimental Process Overview, linear

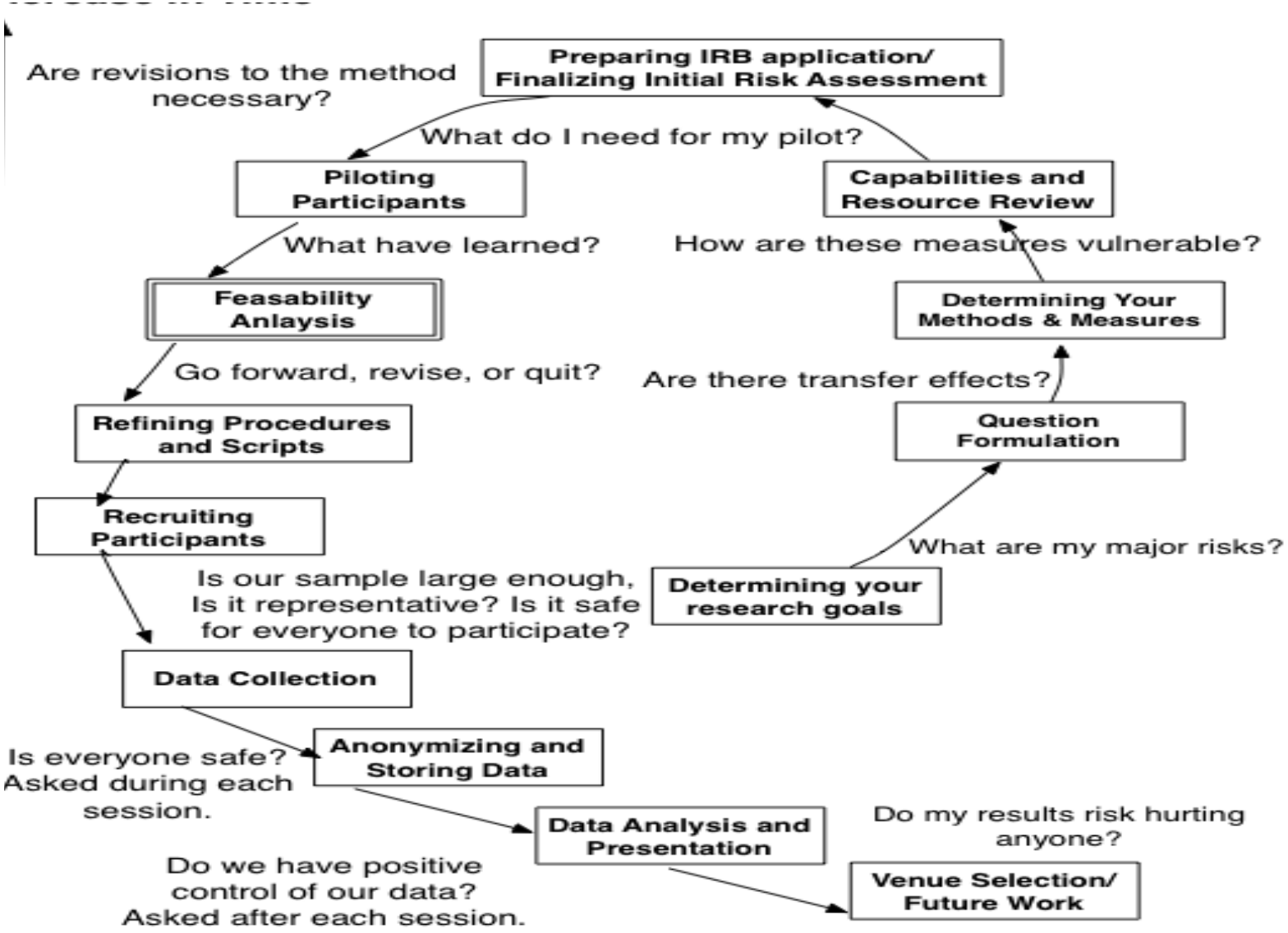
(TB, p. 11)

An iterative, and often over-lapping process



Experimental Process Overview

Risk Driven, more spiral (TB, p. 4)



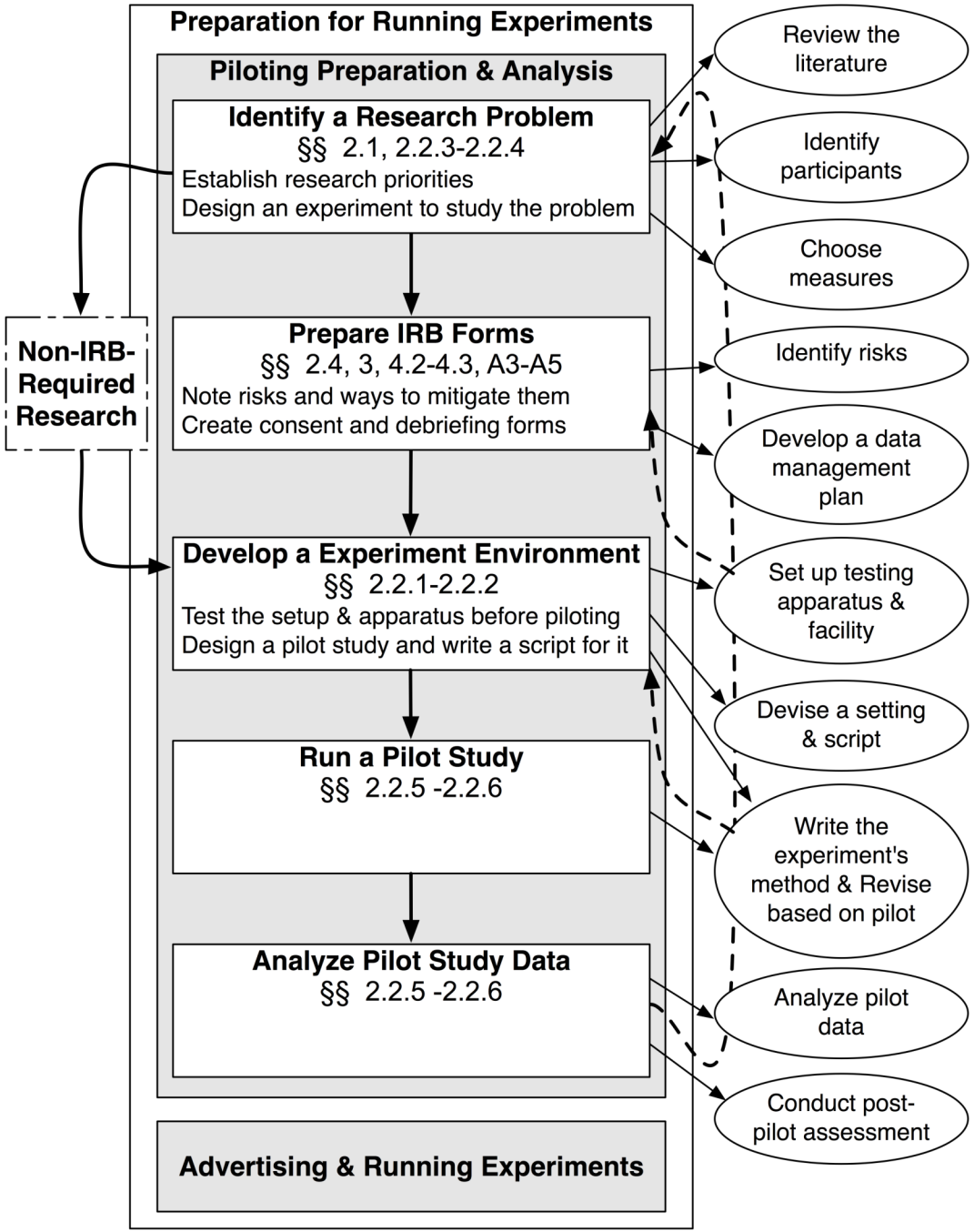
Summary: Lessons so Far

- More steps than I thought
- Iterative and risk-driven (if you pay attention)
- A process but not a set process
- Studies will overlap each other and inspire each other
- It is useful to have the RAs/Es pay attention
  - Ss suddenly 'get it'
  - Ss don't get some aspect
  - Ss comments
  - Ss 'cheat' somehow

Preparation for an experiment

(TB, p. 14)

- Experiments are driven by their questions and shaped by the methods available to explore those questions and existing results/lessons in that area
- This contributes to doing multi-disciplinary work



What studies need IRB?

- In the US
  - if not publishing no IRB (but, be careful), includes class projects
  - If only authors are Ss, no IRB!
  - If only published / publicly available data, no IRB but IRB has to ok this (!)
  - Else, IRB
  - Blood, sexual history, etc. are high-risk, => full IRB
- Outside US
  - Depends, UK used to do IRB only on high-risk studies
  - Can you tell me?
- In all cases, worth having someone check your work

IRB Forms

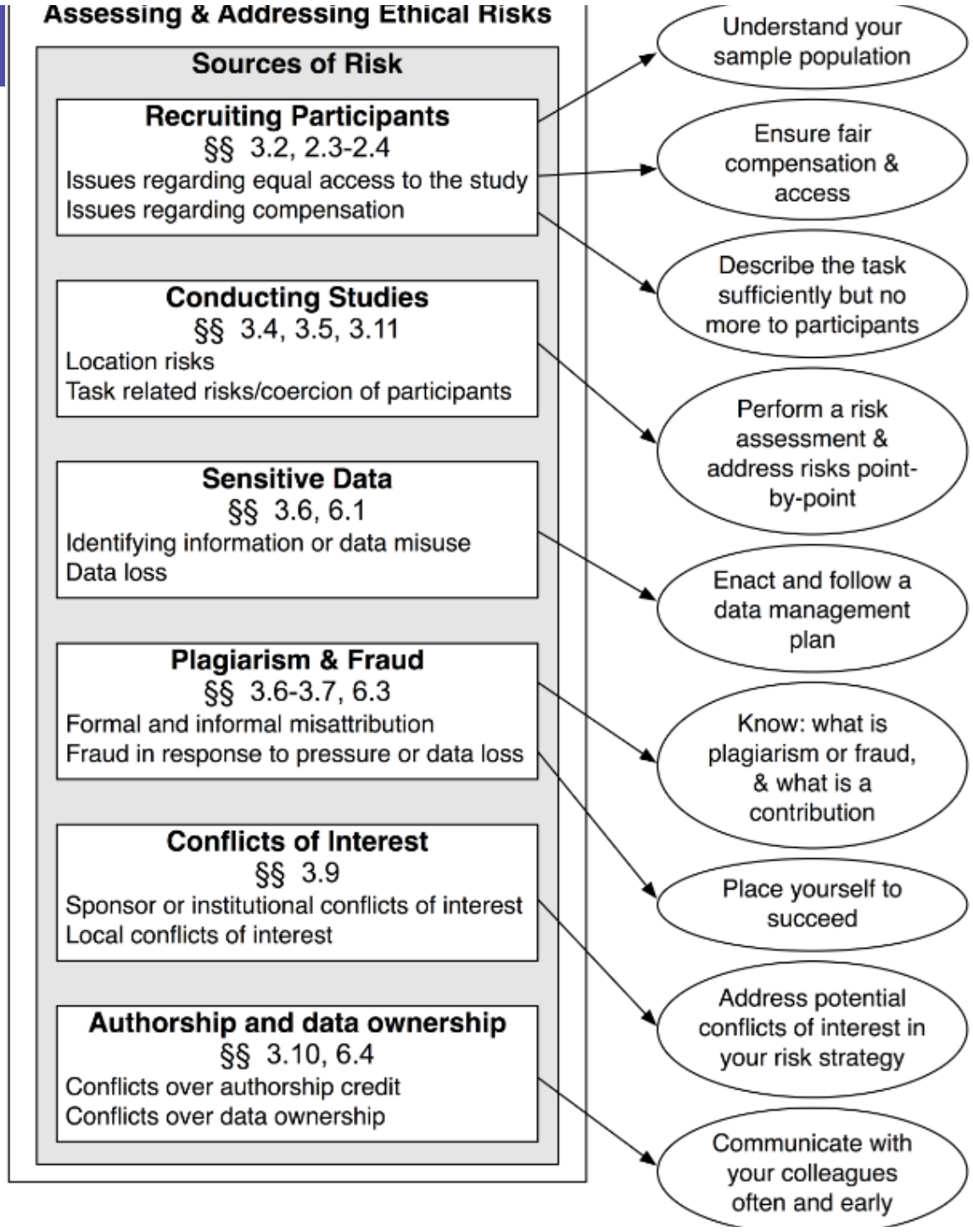
- Used to check your work
- May be worth being clear and concise
- Also check with example forms for language
- Draft for the Principle investigator (PI)

Summary: Piloting

- Write out method
- Used to check your work
- Use a script,
  Step 1, start program, Step 2 "Welcome to…"
- Start local, e.g., YOU, and then officemate, and then move further and further away
- Mount a scratch monkey
- Check your apparatus and data gathering and use of data
- Consider/reconsider, number of Ss to run
  - Previous studies
  - Power analyses (📖Cohen for Ss; 📖Ritter et al. for models)
  - Why not prefer large effects?

Ethical Challenges Associated with the Experimental Process

- Ethical problems can be decreased by deliberate proactive action.
- A couple of bad examples and then a general view



The Monster Study:
Wendell Johnson's Stuttering Study (1939)

- Evaluated the effect of external valuations on stuttering
  - interupting vs. non-interupting conditions
- Studied 22 orphans ranging in age from 5-15 years old, grouping them into 5 fluency categories
- Resulted in long-term developmental and psychological harm, with $925,000 awarded to six of the participants in 2007
- Avoid manipulations that can harm people

Jesse Gelsinger (1981-1999)

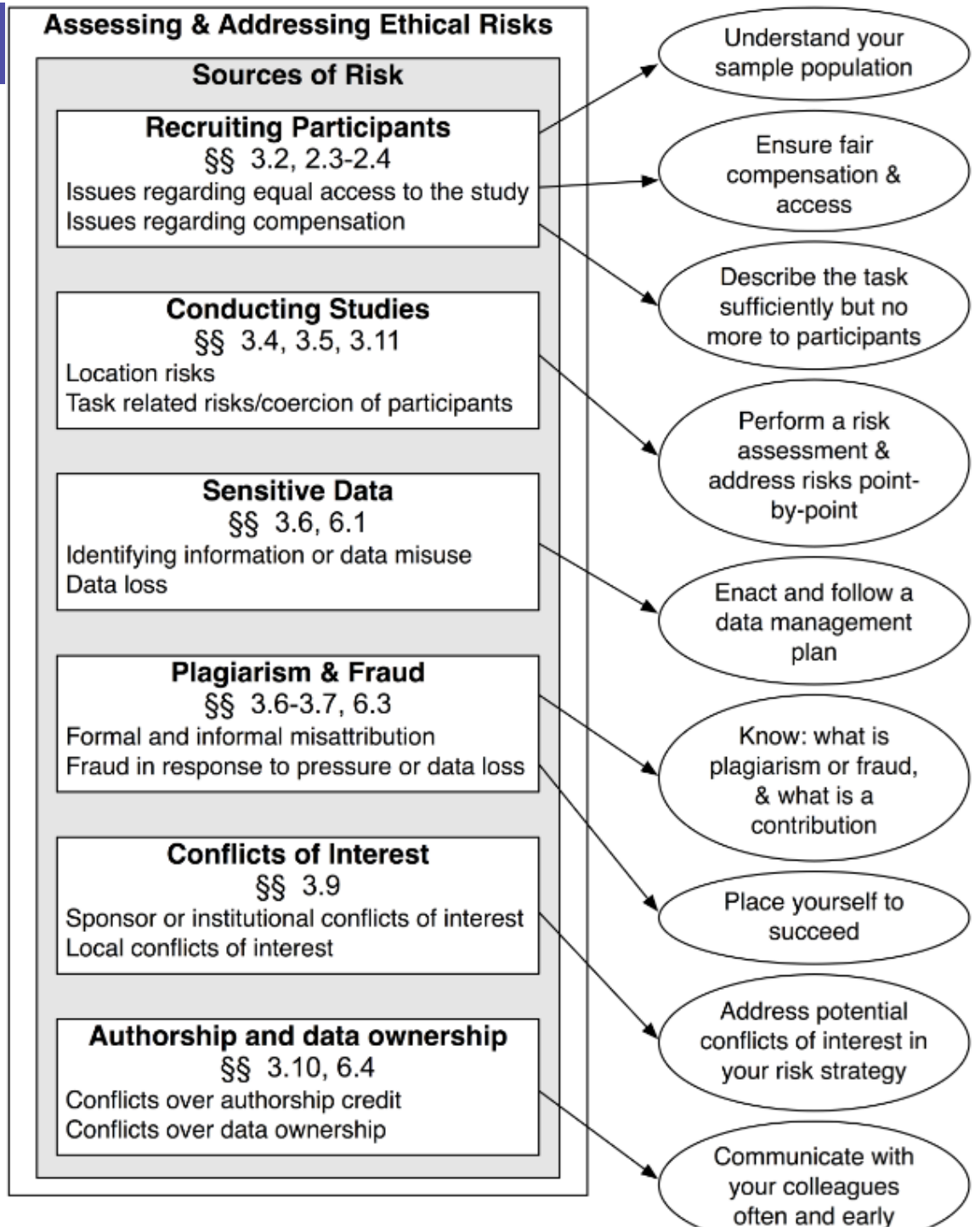
- Included in a bio-medical intervention study to replace a missing participant despite testing positive for high ammonia levels
- The informed consent agreement failed to disclose either known adverse drug effects or the death of two monkeys in animal trials.
- A profound conflict-of-interest existed
  - ↳ Avoid conflict of interests
- Cases like this give rise to the need for IRBs

A HCI Study Gone Wrong (circa 2008)

- No informed consent
- No privacy grantees or data management plan
- "You have no friends." Yes, a student researcher felt compelled to inform a participant and the S's teachers and Dean of this fact.
  - ↪ Even "HCI" studies can hurt people
  - ➡ Know your methods, protect Ss

Ethical Challenges Associated with the Experimental Process

Ethical problems can be decreased by deliberate proactive action.



Exercise: Two ethical dilemmas
[iff time]

B. In screening candidates for a stress study, you discover one of your P's heart rate suggests a medical condition. (or, in any study situation, a subject arrives in an altered state.) Do you have an ethical obligation to report this to them?

A. In collaboration with Dept. of Veterans Affairs, you & your team are evaluating long term a learning theory and a tutor based on that theory where some learners have PTSD. As the study progresses, many of the learners experience significant personal hardship and prolonged unemployment.
Does this change in status present an ethical challenge with regards to the participants' freedom of consent? If so, does the veterans' right to participate and their self-felt obligation to help, and their increasing interest in the payments, outweigh this potential threat to consent? Also, what if the nature of the content knowledge (e.g., battlefield first-aid) interacts badly with their PTSD?

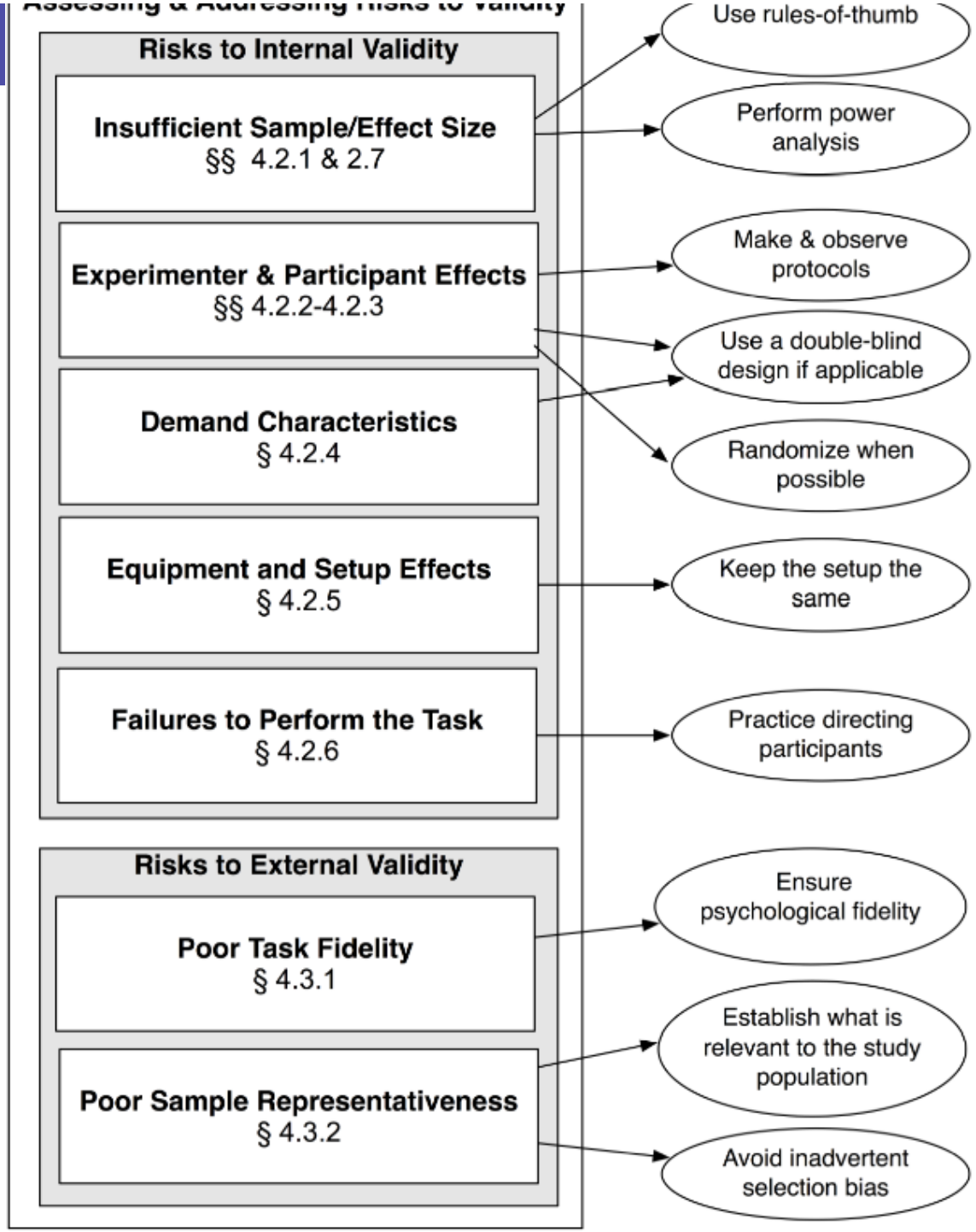
Summary: How to avoid ethical problems

- Recruit fairly
- Look out for your Ss
- Anonymise data at the beginning of each session by using subject IDs, not names
- Have a plan for surprising data (e.g., high BP)
- Communicate early and relatively often about publication plans and data ownership 🕮(Diguisto, 1994)
- Some argue that you have an obligation to use the data you gather

Challenges to Validity:
Constraints on your study

Or: alternative hypothesis for results (TB, p. 21)

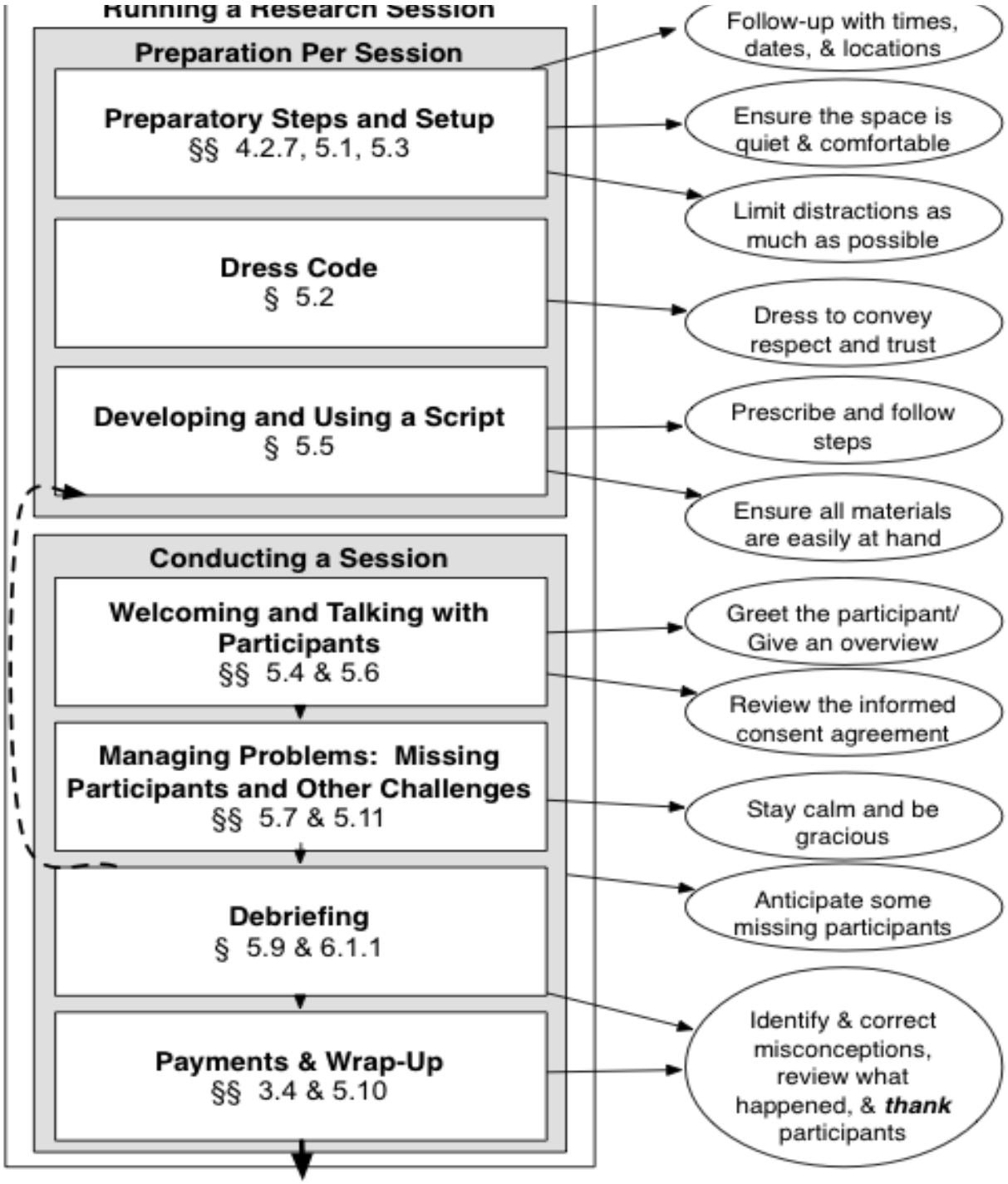
Challenges to validity can be anticipated and mitigated.



Conducting an Experiment

(TB, p. 24-25)

uccess in execution is directly correlated to careful preparation



Exercise: Two running problems
[iff time]

A. *In a developmental cognition study, you are working with 10 parents & their infants. You have found in your piloting that many of the parents are late b/c the building is confusing. In addition, some mothers have inquired whether there might be a play space for their older children. But you don't have one.*

- *How will instruct your RAs to deal with late parents and older children, particularly children alarmed at being separated from their parents?*

B. *In a study examining language acquisition in multilingual families (or, indeed any study), you find that some of the participants are concerned about signing the informed consent agreement. While you have provided translations of the agreement, there is still some obvious tension regarding the agreement.*

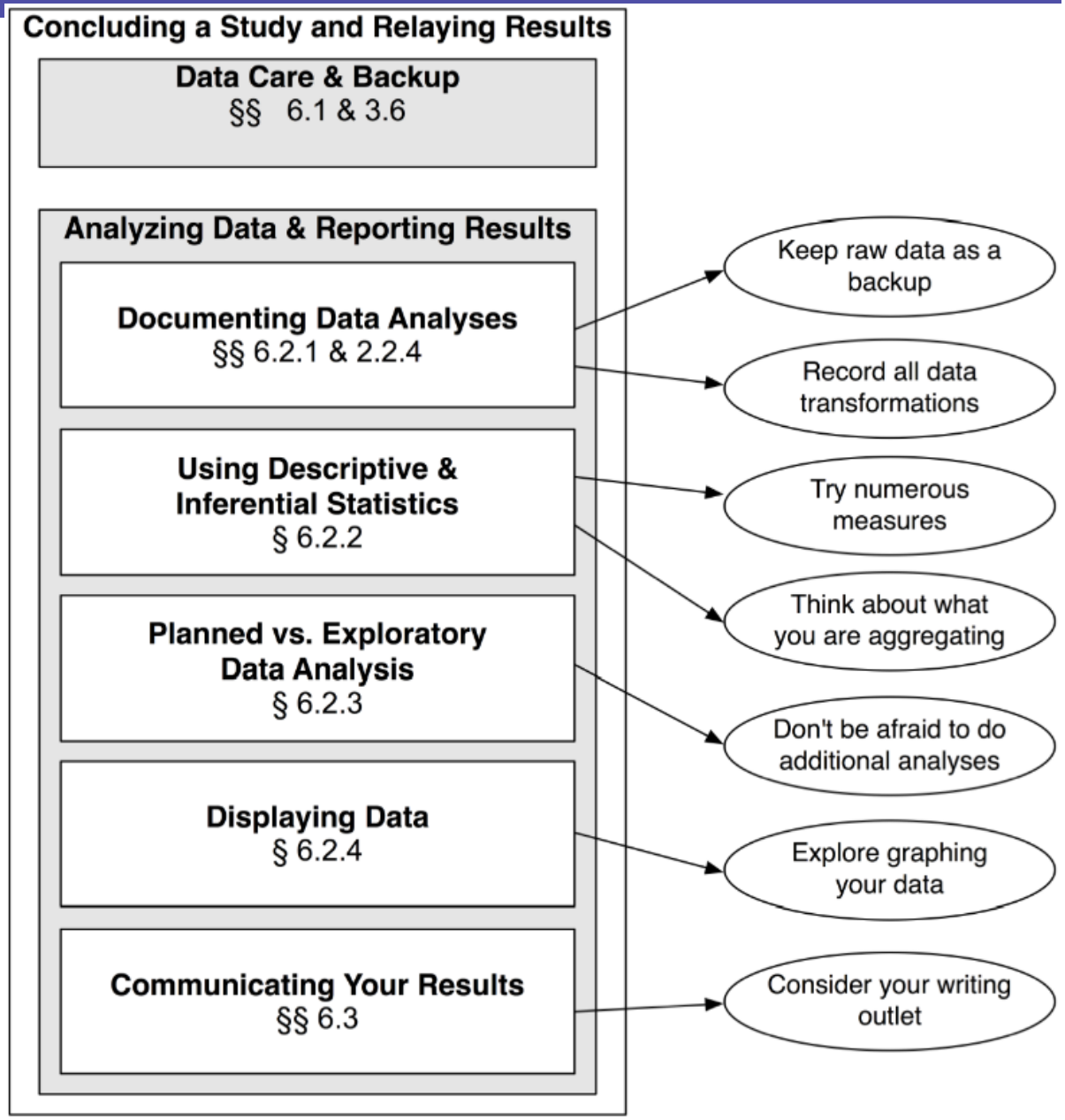
- *How would resolve this tension?*
- *Also, do you have to exclude participants who are unwilling to sign the informed consent agreement?*

Summary: Running a session

- Use of piloting means no surprises (except for the data!)
- Script keeps treatment the same, it includes session set up
- Keep eyes open while running for further insights
- Anonymise data as soon as possible

Concluding an Experiment and Reporting Your Results

(TB, p.27)



Debrief, debrief, debrief!

Summary: Concluding an Experiment and Reporting Your Results

- Concluding a session
  - Finish with the subject (thank, debrief, check paperwork)
  - Check the data was collected and saved
  - Comment on the data if anomolies
- Data care, security and privacy
  - Anonymizing removes nearly all ills
- Back up data (daily, weekly)
- Data analysis
  - Not how, but note how (document and keep track of)
  - Know your data if you are the RA that analyses
  - Save the analyses, time is not important, space is not important, the insights and results are important
  - Aside: we prefer regression
  - Aside: we prefer individual analyse

Ch 6.5 Communicating your results

- Start with a target in mind
  (if you can)
- Work to larger publications
  (workshop, conf, journal, book)
- Rewrite, rewrite, rewrite
  (the book was draft #53 turned in,
  revised twice in pageproofs)

Exercise: setting up space [iff time]

- (a) Describe your space with your partner for your next study
- (b) Are there any ethical risks or risks to validity?
- (c) How could you improve it?
- (d) Should you improve it?

Ch. 7 Afterward

- Appropriate behavior with subjects
- Insights
- Repeatability
- Reportability

Summary 2 of Tutorial

- There are steps to running a study separate from design and analysis
- These are practical, hands-on, implicit knowledge
- They are informed by previous studies
- To be successful, we need to:
  - Formulate a research question that meets our research goals
  - Pilot studies and study components
  - Be candid in our risk assessments and be willing to adapt and refine
  - Be aware of alternative hypotheses, and avoid what we can and control what we cannot avoid
  - Plan for reporting results early

If you will teach this….

- Full book available from Sage & Sage online
- Slides available as ppt or pdf
- Workbook available as pdf

References

📖 acs.ist.psu.edu/papers
acs.ist.psu.edu/reports/ritterKM09.pdf
www.frankritter.com/rbs/ rbs-handout-cogsci.pdf (TB, p. 3)

Boehm, B., & Hansen, W. (2001). The Spiral Model as a tool for evolutionary acquisition. *Crosstalk: The Journal of Defense Software Engineering, 14*(5), 4-11.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

📖 Delaney, P. F., Reder, L. M., Staszewski, J. J., & Ritter, F. E. (1998). The strategy specific nature of improvement: The power law applies by strategy within task. *Psychological Science, 9*(1), 1-8.

Digiusto, E. (1994). Equity in authorship: A strategy for assigning credit when publishing. *Social Science & Medicine, 38*(I), 55-58.

📖 Kim, J. W., Koubek, R. J., & Ritter, F. E. (2007). Investigation of procedural skills degradation from different modalities. In *Proceedings of the 8th International Conference on Cognitive Modeling*, 255-260. Taylor & Francis/Psychology Press: Oxford, UK.

Pew, R. W., & Mavor, A. S. (Eds.). (2007). *Human-system integration in the system development process: A new look*. Washington, DC: National Academy Press. http://books.nap.edu/catalog.php?record_id=11893, checked March 2012.

📖 Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not the answer. *Journal of Experimental Psychology : Learning, Memory & Cognition, 18*(3), 435-451.

Ritter, F. E., Kim, J. W., Morgan, J. H., & Carlson, R. A. (in press, 2012). *How to run experiments: A practical guide to research with human participants*. Currently 150 pages. Thousand Oaks, CA: Sage.

📖 Ritter, F. E., Schoelles, M. J., Quigley, K. S., & Klein, L. C. (2011). Determining the number of model runs: Treating cognitive models as theories by not sampling their behavior. In L. Rothrock & S. Narayanan (Eds.), *Human-in-the-loop simulations: Methods and practice* (pp. 97-116). London: Springer-Verlag.

Roediger, R. (2004). What should they be called? *APS Observer, 17*(4), 5, 46-48. Online: www.psychologicalscience.org/observer/getArticle.cfm?id=1549.

Winston, A. S. (1990). Robert Sessions Woodworth and the "Columbia Bible": How the psychological experiment was redefined. *The American Journal of Psychology, 103*(3), 391-401.