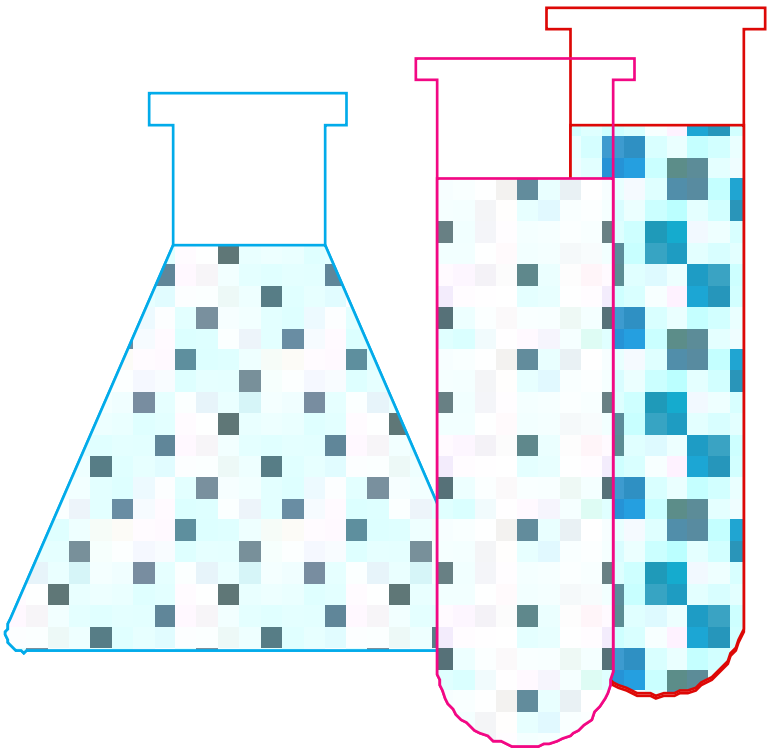
INFSCI 2040
ISSP 2250
Research Design

Syllabus



Marek J. Druzdzel

**School of Information Sciences
and Intelligent Systems Program,
University of Pittsburgh**

marek@sis.pitt.edu
http://www.pitt.edu/~druzdzel

Spring 2015 (15–2)

OVERVIEW:

> *"Science is not science fiction. It accepts the tests of observation and experiment, acknowledges the supremacy of fact over wish or hope. The smallest experiment can crash to earth the most attractive theory."* — Herbert A. Simon

INFSCI 2040 is an introductory course in research design for students in information science and related disciplines, such as computer science and intelligent systems. It is not designed to teach you the formal details of statistical procedures or to make you an expert practitioner of the specific design tools. That you can and should get in other specialized courses. The point of this class is to develop broad critical abilities and its emphasis is on the basic process of scientific inquiry. It aims at improving your ability to think about and frame research questions and to choose the ways to address them by empirical studies.

The approach taken in this course is somewhat unorthodox compared to what you can find in existing textbooks and courses on experimental design that I am aware of. We will start with the concept of causality and causal graphs and how they represent statistical independence. Causal graphs are close to directed probabilistic models, such as Bayesian networks, increasingly used in decision support systems. You will learn to reason about relevance and dependence. I believe this will help you in gaining insight into the structure of scientific experiments and in understanding what experimentation is about. Like any course on experimental design, this course will cover the basics of the design of experiments and topics that are directly related to it, such as identifying and articulating research problems, formulating testable hypotheses, measurement and data collection, subject-experimenter artifacts and their control, describing and displaying data, interpreting and drawing conclusions from data analysis, and reporting research findings and their implications. The course will also cover less orthodox topics such as: problems with the classical hypothesis testing, elements of Bayesian approaches to research design, and research methods used in information science and artificial intelligence, notably simulation and computer discovery.

The intended participants of this course are students who are interested in research and in pursuing academic careers. It is certainly a useful course for the doctoral students in the School of Information Sciences and the Intelligent Systems Program. I believe that it is also useful for those M.Sc. students who are working towards their Master's theses. It will quite likely be less useful and less interesting for those students who plan to pursue application-oriented careers. In the past, several students who were not interested in research found themselves struggling in this course.

As you might have already experienced as aspiring scientists, being a researcher requires intelligence, independent, creative thinking, and most of all commitment to hard working. This course reinforces this. There will a higher than usual amount of readings. I have selected them in such a way that they are fun to read and I expect that you will do them all with pleasure. You will be expected to prepare an abstract of one paper from among those on the list of readings and to present it in class. There will also be a term project that will involve writing a research proposal. This should be a real proposal. If you have not written your Ph.D. or M.Sc. proposal yet or if you would like to apply for funds for your studies, this is an excellent opportunity to do so. Students in this class have had an excellent track record in this respect (the total of well over $600,000 in research funding since fall 1994). Another experience that this course will provide you is peer review, subjecting your work to anonymous judgment of your colleagues and judging the work of others. The workload in this class will be heavy, but I believe that you will find it interesting and important. I require your commitment, doing the readings, coming to classes, and being their active participant. In return, I promise that you will have fun and you will learn many useful skills.

YOUR RESOURCES:

The course:

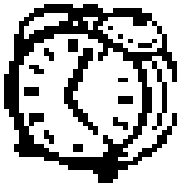
Name : INFSCI 2040/ISSP 2250: Research Design
CRN : 26111
Credits : 3.0

The instructor:

Marek J. Druzdzel
associate professor, School of Information Sciences and Intelligent Systems Program, University of Pittsburgh

Office : B-212 IS Building (Decision Systems Laboratory)
Email : *marek@sis.pitt.edu*
Phone : (412) 624-9432 (office, voice mail)
FAX : (412) 624-5231
WWW : http://www.pitt.edu/~druzdzel
FTP : ftp://ftp.pitt.edu/users/d/r/druzdzel
(host *ftp.pitt.edu*, directory *./users/d/r/druzdzel*)

Meeting times and locations:



Classes (502 IS Building):
Wednesdays, 6:00-8:50pm (break 7:20pm-7:35pm)
Office hours (B-212 IS Building):
Wednesdays, 5:00-5:50pm
Other times by appointment

Teaching Assistant:

There will be no teaching assistant for this class.

Your colleagues:

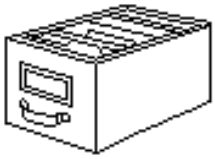
Name: _____
Phone: _____
Email: _____
Name: _____
Phone: _____
Email: _____
Name: _____
Phone: _____
Email: _____

THE COURSE:

Objectives:

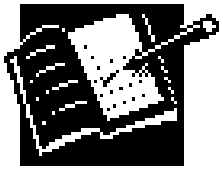
The primary objective of this course is to improve your ability to think about and frame research questions and to choose the ways to address them by empirical studies. It should increase your knowledge, confidence, and creativity in thinking about research problems. Central to that is the ability to evaluate critically the assumptions, strengths, and limitations of alternative analytical strategies and tools. This should make you a better-informed consumer of research results. Written assignments, including your term project, should increase your skills in scientific writing. Your oral in-class presentation will give you an opportunity to improve your presentation skills. Judging the work of your colleagues will expose you to the mechanism of peer review. The project that you will be performing for this course is an excellent opportunity to write a proposal for your Ph.D. dissertation, your M.Sc. thesis, or a proposal for funding your studies. Finally, being successful in the course should contribute to the development of your academic self-esteem.

Prerequisites:



All students in the course should have taken a statistics course (this is a prerequisite to the Information Sciences program!). Most important concepts will be briefly reviewed in class as needed, but this should not be seen as a substitute for formal preparation. The most important prerequisite, however, is your interest in scientific research, motivation, and commitment to learning.

Required reading:



Readings for this course will be taken from several sources. A copy of each of the readings will be put on reserve in the library. A substantial portion of the readings will come from the following textbook:

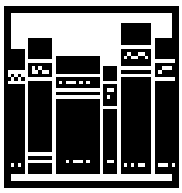
> Paul R. Cohen, "Empirical Methods for Artificial Intelligence", The MIT Press, Cambridge, MA, 1995, ISBN: 0-262-03225-2. (current price around $60.-)

While most issues covered by the book are universal and relevant for research design in general, Cohen's book deals with several computer-intensive experimental methods that are absent in most research design textbooks. I believe that this choice may benefit IS, ISP, and telecommunications students. Another book that we will be using is:

> Donald T. Campbell, Julian C. Stanley, "Experimental and Quasi-Experimental Designs for Research", Boston: Houghton Mifflin Co., 1966, ISBN 0-395-30787-2 (current price between $30 and $100)

It is a short but useful survey of various experimental and quasi-experimental designs.

Supplementary readings:



The required readings should be sufficient for the course. In case you want to read more or if you want to learn more about behavioral research, you can use any research design textbook for supplementary reading. The following three have been used before in teaching this course:

Robert Rosenthal, Ralph L. Rosnow, "Essentials of Behavioral Research: Methods and Data Analysis", New York: McGraw-Hill, 1991, ISBN 0-07-053929-4

Fred N. Kerlinger, "Foundations of Behavioral Research," Hartcourt Brace Jovanovich College Publishers, 1986, ISBN 0-03-041761-9

Geoffrey Keppel, "Design and Analysis: A Researcher's Handbook," third edition, Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1991, ISBN 0-13-200775-4

I used Rosenthal and Rosnow in this course some time in the past and I like the book very much. Its focus, however, is on behavioral sciences and I decided to use a textbook that would be focused more on information science. Kerlinger is somewhat more elementary than the other two in case you feel lost in the material and are looking for an accessible exposition.

WORK REQUIREMENTS:

Assignments:

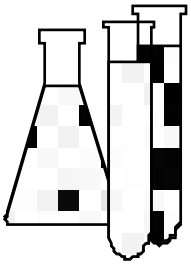
There will be five assignments that will help you to practice the material covered in class and will help me to identify those parts of the material that you have difficulties with. Assignments will usually be done in groups of two or three students, formed during the class meetings. The assignments have to be turned in on time, and all members of the group are responsible for meeting the deadline.

Class presentations:

You will be expected to give a 20 minutes long presentation of a paper selected from the list of readings (we will divide the readings during the first class). You will be also expected to prepare your presentation in PowerPoint (we will have a computer with a projection panel in the classroom). Sometimes I may be able to give you an old version of the summary and the slides, prepared by students in previous years. In that case, you will be expected to improve on them substantially. The presentation and your slides will be graded and you will also receive anonymous feedback on your presentation from the audience.

I would like to stress that the presentation is of a paper from the list of readings and not of your research work. If you would like to talk about your research, please volunteer for a seminar in the School of Information Sciences or the Intelligent Systems Program.

Term project:



A major part of the training that you will receive as part of this course will result from performing a project. The project will give you an opportunity to get experience with writing a real research proposal. The description of the project is attached to this syllabus. The due date (inflexible, again, due to the peer review process) is marked on the course schedule. I advise you to start working on the project as soon as possible, as it involves a considerable amount of work, especially if you are not yet focused in your research interests or are not too experienced in writing research proposals. Perhaps this is a good time to have a serious conversation with your advisor and to focus on a research problem, in case you have not already. I require each of you to provide a signed commitment from your advisor that he/she will be guiding you in your work over the course of the semester and that he/she will provide a written review of your work by a specific deadline.

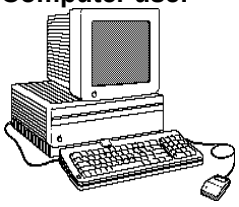
To help you in planning your semester, I have included two additional deadlines: pre-proposal and mid-semester progress report.

Aim at excelling in your project. It is very synergistic with your studies and your overall success in graduate school. Past students in this class have successfully used the project as a springboard to a dissertation proposal or to obtain funding from outside funding agencies. It could be a good start to winning a best paper award prize in one of the school's competitions. Excelling as a student and showing a great potential can open you the doors to top research environments after graduation.

Peer review:

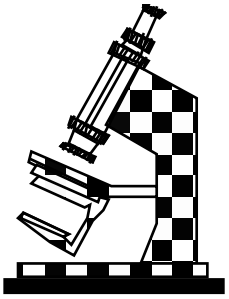
One of the basic mechanisms of the scientific enterprise is peer review: exposing your work to the judgment of other researchers possessing sufficient expertise to judge the merits of your work fairly. You will be expected to participate in this process as soon as you start your scientific career: your work will be reviewed and you will be asked to review the work of others. This class will provide you an opportunity to get acquainted with both sides of this mechanism by exposing your work to the judgment of your colleagues in the class and the instructor and by judging the work of others. The review process will be anonymous (except to the instructor). You can be harsh as long as your criticism is constructive. I will have my own opinion (and the resulting grade) of the work, so do not be afraid of harming the grade of your colleague by being frank. Praising poor work is a clear indication of a mediocre review. Of course, so is criticizing too harshly good work. If done well, a review is usually very useful for the author. When reviewing research proposals, be very detailed and make also corrections pertaining to spelling and language use. As many of the proposals from this class will end up being submitted somewhere, your comments will be very valuable for the author.

Computer use:



You will be required to use computers for typesetting your documents and for working with statistical packages. You will sometimes share your documents with your reviewers electronically. They have to be produced in an easily readable/printable format, such as PDF (recommended) or raw ASCII. The reviews should be submitted in raw ASCII. There will be no limitations on which system you use (except for natural limitations, such as functionality of the package). The peer review process and our communication will be electronic. You will be expected to use electronic mail on a daily basis.

Exam:

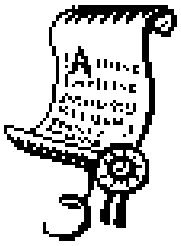


There will be an oral examination taken individually by appointment during the final week of the semester. Students who have missed at most one class, did well on the assignments (at least 80% of the possible score), and were active class participants will be exempted from the examination at my discretion. I reserve the right to change this examination into a written test if the class size is large and the number of exemptions is small.

Time load:

To help you with planning your semester, I would like to give you an idea of the minimal workload in this course. Expect to spend around ten hours quality time outside of class for every class meeting. I estimate that you will need about three hours to do the readings and two hours (on the average) to do the assignments. I assume here that you will be doing assignments in groups. In case you work on them individually, your time load may increase. The term project will demand as much as you decide to put into it, quite likely more than 50 hours of quality time. You are advised to put as much as you can, especially if it is your dissertation proposal or proposal for funding. It is explicitly allowed and encouraged to combine this with other coursework, a research seminar, or an independent study. The catch is that if you combine your project with another piece of work, I expect you to put an appropriately higher amount of effort in both (please do check with the other teacher if it is OK with him or her and do tell me that you are doing this!).

Grading:



Your final grade for the course will be the average of two grades, one for your term project (80% of which will be for the proposal and 20% for reviewing two proposals of others) and one for all other work, i.e., assignments, oral presentation, and the final examination (combined in the proportion 40%, 20%, and 40% respectively; on the top of this all, you can obtain up to 10% extra for in-class participation). Exemption from the final examination amounts to receiving the maximum score for the exam.

COURSE SCHEDULE

Those readings that are available for class presentations are underlined. You will be expected to choose a paper during one of the first classes. If there are more students in the class than there are underlined readings, readings that are not underlined can be chosen (only after the underlined readings have been claimed).

INTRODUCTION

The first class will be devoted to organizational matters and overview of the course. There are many readings that I would like you to make yourself familiar with but most of them are enjoyable and really useful. You will be glad you have read them (if you have not read them yet). We will spend some time on the issue of developing a research topic, as you will need to choose a research topic for the term project.

January 7 *[Readings: Bundy et al.; Zanna&Darley; desJardins; Agre; Cohen, Chapter 1; Peter&Olson; Baron; Shriver; Thomson; Psaraftis; Smith; Campanario]*
Getting to know each other. Organization of the course.
Characteristics of a good research problem.
How to develop a research topic.

UNCERTAINTY, STATISTICAL DECISION MAKING

This one-class block with help you in reviewing what you have learned in your statistics classes from the point of view of experimental design. In addition to the basic skills that we will just run through (you have learned most of these elsewhere), we will have a crash review of decision making under uncertainty and decision theory. This will be very useful for you in appreciating various components of statistical decision making and how they relate to experimentation.

January 14 ***** Signed commitment from the external reviewer due *****
[Readings: Your favorite statistics textbook; Cohen, Chapters 2&4]
Uncertainty, probability, variance, sampling, randomness,
elements of data analysis (a review).
Describing and displaying data, correlation.
Comparing means.
Decision making under uncertainty; elements of decision theory.

CLASSICAL DESIGN OF EXPERIMENTS

The three classes in this block form the core of classical experimental design. We will look at the classical experimental design from the modern perspective of graphical probabilistic models and decision theory. Causal graphs, a relative of Bayesian networks, will take a prominent place in the first class. It is crucial for your understanding of what experimentation is and, effectively, for understanding the rest of this course, that you know the basics of causal graphs and are able to "see" when variables are independent and when they are dependent purely from the structure of the graph. A bit of warning for the two classes that will follow: we will go through a quantity of statistical designs and this may be a little boring. Still, try to grasp the essentials: problems that a particular design is solving and problems that it introduces.

January 21 ***** Pre-proposal due *****
***** Homework assignment 1 due (describing and displaying data) *****
[Readings: C&S pages 1-26; Fisher; Druzdzel&Glymour 1995]
Causality, causal graphs, qualitative reasoning in causal graphs.

January 28 ***** Homework assignment 2 due (causal graphs) *****
[Readings: Cohen, Chapter 3; C&S pages 27-34]
Experimental designs.

February 4 Experimental designs.
Factorial designs.

February 11 *[Readings: C&S pages 34-64]*
Quasi-experimental designs.

FUNDAMENTAL DESIGN AND IMPLEMENTATION ISSUES

It is useful to realize that statistical designs are just an idealized setup and a lot in experimentation depends on the human side of it. In this one-class part of the course we will discuss dangers of experimentation, various biases that can enter a study, and resulting problems with the validity of the results. In addition, we will talk about scientific creativity, a topic that is too often neglected in teaching how to do science.

February 18 ***** Homework assignment 3 due (experimental design 1) *****
[Readings: Hanson; Freud; Thornton; Grünbaum; Kaplan]
Dangers of experimentation, experimenter's biases.
Problems in the laboratory. Scientific creativity.

COMPUTER-INTENSIVE STATISTICAL METHODS

The four classes in this block cover computer-intensive approaches to experimentation. Computers enter the domain of research design in two ways: as subjects — we often want to study how a new piece of hardware or software behaves, and as tools of experimentation. In both roles they are incredibly convenient and easy to use for an experimenter. We will pretty much follow Cohen's book.

February 25 ***** Homework assignment 4 due (experimental design 2) *****
[Readings: Cohen, Chapter 5]
Monte Carlo simulation, bootstrap methods, randomization procedures.

March 4 ***** Proposal progress report due *****
***** Homework assignment 5 due (simulation) *****
[Readings: Cohen, Chapters 6&7]
Performance assessment.

March 9–15 *Spring Recess for students (no classes)*

March 18 *[Readings: Cohen, Chapters 8&9]*
Modeling.

NON-CLASSICAL APPROACHES

The approaches to experimentation that we have discussed so far were classical in the sense of following the original orthodox design where manipulation and classical statistical hypothesis testing played crucial roles. In this block we will discuss what we can do when manipulation is not possible and how computers can help us in experimentation. In addition, we will talk in much detail about problems with the classical hypothesis testing and how Bayesian approach to statistics can be of help.

March 25 *[Readings: Epstein&Axtell]*
Simulation, artificial societies.

April 1 *[Readings: Glymour et al.; Langley et al.; Druzdzel&Glymour 1994]*
Computer discovery, causal discovery.

April 8 ***** Proposal abstract due *****
***** Homework assignment 6 due (causal discovery) *****
[Readings: Falk; Simon; Edwards et al.; Gregg&Simon; Jacob Cohen; Kadane]
Problems with the classical hypothesis testing.
Bayesian approaches to research design and statistical hypothesis testing.

CONCLUSION

Our last classroom meeting will be a grand conclusion of the course. Your proposals will have been completed and reviewed. The final grades for the class will be prepared and distributed along with proposals and their reviews (historically, all students in this class were exempted from the final examination). The winner of the Marek's Best Proposal Award and other interesting proposals will be announced and asked to tell something about his/her proposal. We will talk about the course and possible ways of improving it in the future.

April 10 (Friday) ***** Term proposal due *****

April 13 (Monday) ***** Proposal reviews due *****)

April 15 Presentation of proposals.
Conclusion of the course. General discussion.

April 20-24 Final exam (by appointment)

Sources of readings:

[Agre] Phil Agre. *"Networking on the Network: A Guide to Professional Skills for Ph.D. Students."* Unpublished paper, available electronically through WWW at http://dlis.gseis.ucla.edu/people/pagre/network.html

[Baron] Robert A. Baron. *"Research Grants: A Practical Guide."* In: Mark P. Zanna & John M. Darley (editors), *"The Compleat Academic: A Practical Guide for the Beginning Social Scientist,"* Chapter 7, pages 151-169, New York: Random House, 1987

[Bundy et al.] Alan Bundy, Ben du Boulay, Jim Howe & Gordon Plotkin. *"How to Get a Ph.D. in AI."* In Tim O'Shea & Marc Eisenstadt (editors) *"Artificial Intelligence: Tools, Techniques, and Applications,"* Chapter 5, pages 139-154, Cambridge: Harper & Row Publishers, 1984

[C&S] Donald T. Campbell & Julian C. Stanley. *"Experimental and Quasi-Experimental Designs for Research."* Boston, MA: Houghton Mifflin Co., 1966

[Campanario] Juan Miguel Campanario. *"Have Referees Rejected Some of the Most-Cited Articles of All Times?"* Journal of the American Society for Information Science, 47(4):302-310, 1996

[Cohen] Paul R. Cohen. *"Empirical Methods for Artificial Intelligence"*, The MIT Press, Cambridge, MA, 1995

[Jacob Cohen] Jacob Cohen. *"The Earth Is Round (p<.5)."* American Psychologist, 49(12), 997-1003, 1994

[desJardins] Marie desJardins. *"How to Be a Good Graduate Student/Advisor."* Unpublished paper, available electronically through WWW at http://www.ai.sri.com/~marie/papers/advice-summary

[Druzdzel&Glymour 1994] Marek J. Druzdzel & Clark Glymour. *"Application of the TETRAD II Program to the Study of Student Retention in U.S. Colleges."* In Proceedings of the AAAI-94 Workshop on Knowledge Discovery in Databases (KDD-94), pages 419-430, Seattle, WA, 1994, available at: http://www.pitt.edu/~druzdzel/abstracts/kdd94.html

[Druzdzel&Glymour 1995] Marek J. Druzdzel & Clark Glymour. *"Having the Right Tool: Causal Graphs in Teaching Research Design."* In *What Works in University Teaching: University of Pittsburgh Teaching Excellence Conference*, Pittsburgh, PA, 1995, available at: http://www.pitt.edu/~druzdzel/papers/cgraphs.html

[Edwards et al.] Ward Edwards, Harold Lindman & Leonard J. Savage. *"Bayesian Statistical Inference for Psychological Research."* Psychological Review, 70(3), 1963

[Epstein&Axtell] Joshua M. Epstein and Robert Axtell. *"Growing Artificial Societies: Social Science from the Bottom Up."* A Monograph of the 2050 Project, To be published by MIT Press, 1996. Chapters I & II.

[Falk] Ruma Falk. *"Misconceptions of Statistical Significance."* Journal of Structured Learning, 9, 83-96, 1986

[Fisher] Sir Ronald A. Fisher. *"The Design of Experiments."* Chapter 2, *"The Principles of Experimentation, Illustrated by a Psycho-Physical Experiment,"* pages 11-26, Oxford: Oxford University Press, 1990

[Freud] Sigmund Freud. *"Beitrag zur Kenntniss der Cocawirkung (Contribution to the Knowledge of the Effect of Cacaine)."* Wiener Medizinische Wochenschrift, 35(5), 130-133, 1885

[Glymour et al.] Clark Glymour, Richard Scheines, Peter Spirtes & Kevin Kelly. *"Discovering Causal Structure: Artificial Intelligence, Philosophy of Science, and Statistical Modeling."* Chapters 1-3, pages 3-59, San Diego, CA: Academic Press, Inc, 1987

[Gregg&Simon] Lee W. Gregg & Herbert A. Simon. *"Process Models and Stochastic Theories of Simple Concept Formation."* Journal of Mathematical Psychology, 4, 246-276, 1967

[Grünbaum] Adolf Grünbaum. *"Validation in the Clinical Theory of Psychoanalysis: A Study in the Philosophy of Psychoanalysis."* Chapter 10 on Freud's theory of dreams, Madison, CT: International Universities Press, 1993

[Hanson] Norwood R. Hanson. *"Patterns of Discovery."* Chapter 16, *"Observation,"* pages 164-178, Cambridge University Press, 1965

[Kadane] Joseph B. Kadane. *"Progress Towards a More Ethical Method for Clinical Trials."* The Journal of Medicine and Philosophy, 11, 385-404, 1986

[Kaplan] Craig Kaplan. *"Hatching a Theory of Incubation: Does Putting a Problem Aside Really Help? If So, Why?"* Ph.D. Dissertation, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, 1990 (selected fragments)

[Langley et al.] Pat Langley, Herbert A. Simon, Gary L. Bradshaw & Jan M. Zytkow. *"Scientific Discovery: Computational Explorations of the Creative Processes."* Chapters 1-2, pages 3-62, Cambridge, MA: The MIT Press, 1987

[Peter&Olson] J. Paul Peter & Jerry C. Olson. *"Is Science Marketing?"* Journal of Marketing, 47, 111-125, 1983

[Psaraftis] Harilaos N. Psaraftis. *"Review Standards for OR/MS Papers: How To Make an Imperfect Process Less Imperfect."* OR/MS Today, pages 54-57, June 1994

[Shriver] Bruce D. Shriver. *"The Benefits of Quality Refereeing."* IEEE Computer, pages 10-16, April 1990

[Simon] Herbert A. Simon. *"On Judging the Plausibility of Theories."* In Van Roostelaar & Staal (editors), *"Logic, Methodology and Philosophy of Science,"* pages 25-45, Amsterdam: North Holland Publishing Company, 1968

[Smith] Alan Jay Smith. *"The Task of the Referee."* IEEE Computer, pages 65-71, April 1990

[Thomson] Keith Steward Thomson. *"Scientific Publishing: An Embarrassment of Riches."* American Scientist, 82, 508-511, 1994

[Thornton] E.M. Thornton. *"The Freudian Fallacy: An Alternative View of Freudian Theory."* Chapter 11, *"First Encounter with Cocaine,"* pages 19-28, Garden City, NY: The Dial Press, 1984

[Zanna&Darley] Mark P. Zanna & John M. Darley. *"Everything You Always Wanted to Know About Research but Were Afraid to Ask (Your Advisor)."* In: Mark P. Zanna & John M. Darley (editors), *"The Compleat Academic: A Practical Guide for the Beginning Social Scientist,"* Chapter 5, pages 115-137, New York: Random House, 1987

TERM PROJECT

As the term project you will be expected to write a complete research proposal. This means a formal report describing in detail the proposed research, with all the necessary supporting materials and argumentation expected from a proposal. This includes the objective of your research (i.e., what problem you are planning to solve), its importance (i.e., why it is useful), prior work (i.e., how your problem fits into the existing body of knowledge), research strategy (i.e., how you plan to address the problem), expected results, time frame and, if applicable, the expected costs of the proposed work.

Choice of the problem:

The proposal that you write should be real. If you are a doctoral student, you might decide to write a Ph.D. dissertation proposal or, if you are not ready for this, a proposal for your qualifying examination paper. If you are a M.Sc. student interested in research (if you are not interested in research, do not take this course!), you might write a proposal for your Master's thesis. If you have already defended your dissertation proposal, you could refine it and turn it into a proposal for funding that you would then submit to a funding agency, such as the National Science Foundation. Whatever you decide to do, it should be a useful piece of work. Elaborating on your previous piece of work and/or combining your work on the proposal with another class are explicitly permitted and encouraged (please check with the other teacher to make sure that this is OK with him/her). If you do this, please state so explicitly in your pre-proposal and indicate ways in which you will extend on your work during the semester. Submission of an old piece of work without significant enhancement during this semester is not good for you and is explicitly forbidden.

You will be responsible for choosing a research problem yourself. There are no rules that would help you in choosing a research topic, but there are some heuristics that we will talk about during the first class meeting. Also, unless you are well on the way in your research, the first few weeks of the semester might be the best time to meet with your advisor and ask his or her advice about the possible directions that your research work should go. Ask your advisor for an example of a research proposal (ideally, but not necessarily, a successful one), and ask him/her what it takes to write a successful proposal, from the initial idea to submission. If you do not have an advisor with whom you work or are otherwise undecided, talk to me and I will be glad to help you in choosing a research topic, suggesting an advisor, or advising you. I have always a few interesting problems to work on myself and I will be glad to suggest a problem, give you initial pointers, and work it out into a proposal with you over the course of the semester. Feel free to get in touch with me if you are interested, but please do it early.

External reviewer:

You are responsible for finding at least one external reviewer for your proposal. This will be usually, but not necessarily, your advisor or another scientist at the University of Pittsburgh or elsewhere who agrees to support you in your work. I will act as an external reviewer for my research advisees. You are explicitly allowed and encouraged to work with the external reviewer on your proposal throughout the semester. I strongly advise that you do this, as having a good mentor will normally be an important factor in your success as a beginning scientist. The external reviewer should be knowledgeable in the area of your research, as his/her role is to ensure that your proposal is original, interesting, and feasible. The name and contact address (Email, if possible) of your external reviewer should be included in your pre-proposal. Your external reviewer will review and judge the final version of your proposal along with your peers and me. It is your responsibility to submit a copy of your proposal and the review form to the external reviewer and to make sure that he/she submits a review of your proposal to me by the deadline. Please treat this last requirement seriously, as I look very critically at proposals that the external reviewer did not even bother to evaluate and I usually reflect this in my final grade for your project.
In order to help you with making it clear to your external reviewer that the matter is serious, I require you to submit a signed commitment from the reviewer that he/she will review your proposal by the deadline (the form is attached to the syllabus).

Pre-proposal:

Your choice of topic will be a major factor in the success of your research endeavors. In order to prevent you from wasting your time and to make sure that your topic is realistic but challenging enough, I want you to submit a written pre-proposal for my approval. The pre-proposal should clearly state the recipient of your proposal (e.g., NSF, DARPA, a foundation, etc.), the name and Email address of your external reviewer, what problem you are planning to solve, why it is useful, and include a rough sketch of how you think you can address this problem. The pre-proposal can be fairly short — a couple of pages usually suffice. Research pre-proposals exist in real life, albeit they have a slightly more formal character than what you are expected to write. I will let you know whether your topic is acceptable as it stands or what you can do to improve it. If your pre-proposal is not approved, you will have to submit a revised version within a week. You should preserve the pre-proposal with my comments and submit it along with your mid-semester progress report and the final proposal.

Mid-semester progress report:

You will be expected to submit a mid-semester progress report containing at least an introduction to the final proposal (motivation, detailed description of the problem, and literature review, all within the bounds of the recipient's format), a sketch of the proposed approach, description of the work that you have completed so far, and a detailed plan of action for the remainder of the semester. The submitted work should conform to the format specified by the proposal recipient. The main purpose of this report is to help you in planning your work and spreading it over the course of the semester. The deadline for submitting the mid-semester progress report is marked on the syllabus. You are required to submit your pre-proposal with my comments along with your mid-semester progress report.

Empirical component of the proposal:

Although it would be nice to apply some of the techniques that you will learn in this class in your proposal, do not feel pressed to apply particular methods in your research design. Your proposal should most of all be real. Experiments and statistical design will in the fields of information science, library and information science, artificial intelligence, or telecommunications normally have a supportive character rather than being the focus of your proposal. A good study usually involves some empirical work of either exploratory or verifying character, so experimentation will naturally enter your work. If you believe that no empirical work should be involved, you had better provide a convincing argument that will support your belief and convince the reviewers, including your external reviewer and me.

The format:

The format of your proposal should conform to the format required by its recipients. Most funding agencies, for example, have a strict format, putting an upper limit on the number of pages, specifying the font size, sectioning units, etc. It is in your interest to follow the required format, as this will make your final product ready to go.
To make your choice of a format easier, I propose the National Science Foundation (NSF) format as described in their Grant Proposal Guide, available electronically through World Wide Web at
http://www.nsf.gov/publications/pub_summ.jsp?ods_key=gpg.

The default directorate to which you will be submitting is the Directorate for Computer and Information Sciences and Engineering (CISE), Division of Information and Intelligent Systems. Unless you attach a description of the format required by your recipient (web link is sufficient if the format is described on the web), you will be judged on your compliance to this default NSF format. In case you are working on a Ph.D. or a M.Sc. proposal, you should also follow the NSF format. It will be straightforward for you to convert an NSF proposal into the informal format of a Ph.D. proposal at the conclusion of this course.

Length of the proposal:

There are no limitations on the length of your proposal other than those imposed by the funding agency. NSF usually imposes a limit of 15 pages on the technical part of the proposal. Some fellowship competitions require a short proposal, such as four or even two pages. Whatever funding agency you choose, please adhere to its formats. A short proposal, if this is the format required by its recipient, is fine. In this case, however, I expect that your proposal is of a winning quality and I will not settle for less. I will be very strict in checking it, and that includes grammar and spelling in addition to the contents. In other words, if you put a semester's worth of work into four

pages, I expect these four pages to violate the laws of gravity rather than being some miserable brain dump performed the night before the deadline. My firmness usually benefits students – at least three fellowship applications were successful in obtaining funding in the past.

Submission requirements, review, and evaluation criteria:

Submit your proposal by the deadline in the syllabus. You are expected to submit a PDF version and, in addition to that, one hard copy version for me. (As far as your external reviewer goes, please follow his/her preferences.) A few days before the deadline you will be expected to submit the title and the abstract (up to 200 words) of your proposal by Email in raw ASCII format. This will aid in selecting in-class reviewers for your proposal. The deadlines are inflexible because of the review process that will follow. Your proposal will be judged by your external reviewer, two of your peers, and me.
Your proposal is going to be made accessible in electronic format to all your classmates, so please make sure that your topic and your writings are not "top secret." If they are, they are not suitable for a proposal for this class.
The criteria for grading your proposal are soundness, clarity of your writing and expressing your ideas, creativity, doability (the proposed work within the proposed amount of time), and in addition, local review criteria (set out, for example, by the funding agency) including your compliance with the required format. If this is different than the NSF format, please submit its description along with your proposal. Also, you should submit your pre-proposal and your mid-term progress report with my comments along with your final version of the proposal. Make sure that you save these.

What makes a proposal successful?

You need to realize that resources, such as money for your research, are scarce and there is a lot of competition for them. A winning proposal (if you decide to spend a considerable effort in writing a proposal, why wouldn't you want to write one that wins – the effort is going to be not much more) has to convince the reviewers that the proposed research is worth funding. While there are many ways of convincing a reader that your proposed work is really good, most winning proposals will have a strong, enthusiastic motivation section explaining what the problem is and why it is important. Another important component of a good proposal is a thorough literature review. Very few researchers are starting a completely new line of research and typically, you will be building on the shoulders of others who were before you. Literature review should acknowledge the work and findings of your predecessors. At the same time it should explain in what way your proposal is novel and show how the prior approaches to solving the problem stopped short of solving it. Please note that there is a fine line between acknowledging the works of other and arguing that that work has shortcomings. A useful mental exercise is imagining that the authors of the papers that you are citing are reviewing your proposal. What would you like to read in a proposal or a paper written by others about your own work? This exercise is very realistic – there is a very good chance that this indeed will happen. Being both frank and fair at the same time will help you in receiving favorable reviews. The next important task of a proposal is to convince the reader that you will be successful. Ask yourself the hard question why you think that you will be successful and then tell it in a not too complex way, injecting necessary enthusiasm. The remainder of your proposal, 40-80% of its volume, should contain an outline of your work, often a list of tasks that you are planning to perform, a rough schedule of your work, and a list of references. It is important that you follow the format specified by the funding agency because if you are violating it, your proposal may be rejected without even being reviewed.

Marek's Best Proposal Award:

Selected proposals will be discussed during the last class meeting. The best proposal in class, as judged by the founder (i.e., me ☺), will be awarded a doughnut and a cup of coffee (you can also opt for a comparable drink of your choice) along with an accompanying certificate. In addition, you will be invited to talk about your proposal at one of the Brown Bag seminars in the School of Information Sciences.

LETTER OF COMMITMENT FROM THE EXTERNAL ADVISOR / REVIEWER

(due 14 January 2015)

Name : _____
(advisor's name)

Title : _____

Phone number : _____

Email address : _____

I hereby confirm that I am familiar with the requirements of the term project in the course *INFSCI 2040/ISSP 2250, Research Design* and declare that I shall:

(1) guide the student _____ in his/her pursuits over the course of the semester
(student name)

(2) provide a written review of the final version of the term project by Monday, 13 April 2015.

I understand that if I miss the review submission deadline, I will seriously endanger the student's grade for the term project and, ultimately, the student's final grade for the course.

_____ _____
(advisor's signature) *(date)*