Introduction to Scientific Research

Master 2 Info – Polytech

November 2019

Lecturers: Marco Winckler (UCA, I3S)
winckler@unice.fr

UNIVERSITÉ CÔTE D'AZUR

Overview

- Introduction to Scientific Research methods
- Lifecycle of scientific knowledge acquisition, validation and dissemination
- Methods for reading and writing scientific articles
- The carrier of researcher
- Scientific Research in the industry

Planning

- Two classes
  - November 19th 2019 (today) 14h30-17h30
  - November 24th 2019 (Monday) 14h-17h
- A panel with industrialists
  - December 18th 2019 – 17h30-19h30

Assessment

- Two assessments
  - QCM (50%)
  - State-of-Art (50%)
    - Synthesis of a bunch of scientific articles
    - The same for the PFE
    - Anyone not doing PFE ?

Agenda today

- Science and Methods for conducting research
- Bibliographical research
- Selecting bibliographic sources
  – Difference between types of publications (journals, conference, workshops, etc)
- Methods for reading and writing scientific articles
  – How to read, classify and analyze papers

Scientific research is …….

- Knowledge acquisition gained
  - through reasoning
  - through intuition
  - but most importantly through the use of appropriate methods

The Scientific Method

Basic Elements of the Scientific Method

- **Empiricism**: the notion that enquiry is conducted through observation and knowledge verified through evidence
- **Determinism**: the notion that events occur according to regular laws and causes. The goal of research is to discover these
- **Scepticism**: the notion that any proposition is open to analysis and critique

Components of Science

- Evidence
- Logic
- Hypotheses
- Theories
- Induction
- Deduction
- Validity
- Reliability

Evidence

- Definition: empirically observed experiences, i.e. from the five senses
- Words that signal the use of evidence:
  - Data
  - Observations
  - Empirical
  - Empiricism
- Kinds of evidence
  - Quantitative vs. Qualitative

Logic

- Definition: the study of necessary connection or outcome.

- Science must avoid logical mistakes that make incorrect connections and result in faulty information outcomes.

Some Common Errors of Logic

- **Ad hominem** – attacking the person making the argument rather than producing evidence to contradict the argument itself
- **False dilemma** – presenting the issue as having only two opposing explanations
- **Questionable cause** – assuming cause and effect when one or more of the three requirements for cause are not present
- **Reductionism** – attempting to explain all questions using a single type of explanation
- **Ethnocentrism** – applying your culture's explanations to situations derived from other cultures.

Hypotheses

- Definition: a trial explanation that can be empirically tested.
- Once tested and supported, an hypothesis can be used as an assumption on which other hypotheses are constructed.
- If refuted, an hypothesis must be discarded or changed and retested.

NEVER USE THE WORDS PROVE AND HYPOTHESIS IN THE SAME SENTENCE!

Theory

- Definition: a set of related hypotheses that act together to provide a better explanation than any single hypothesis.
- Theories are only scientific if the hypotheses that make them up can be empirically tested.

NEVER USE THE WORDS PROVE AND
THEORY IN THE SAME SENTENCE!

Induction

- Definition: the generation of hypotheses and/or theories from empirical evidence.
- Induction takes specific data and makes generalizations from them.
- These generalizations take the form of hypotheses and theories.

Deduction

- Definition: the testing of hypotheses and/or theories with empirical evidence.

- Deduction takes generalizations and collects evidence to see if they are supported under a variety of circumstances.

Validity

- Definition: You actually measure what you say you are measuring.

- One would not measure self esteem by collecting data on shoe size. Shoe size is an invalid measure of self esteem. There are a number of self esteem measurement instruments developed to measure the specific aspects of self esteem as the scientific community defines the concept. These instruments have been tested and accepted as valid instruments.

Reliability

- Definition: The results of a research project have been repeated and the same findings result each time.

- The knowledge that boys are, on average, more active and aggressive as children than girls is a reliable finding because dozens of studies have independently come to that same conclusion.

Quantitative Data

- Definition: data collected in the form of numbers or categories that can be labeled with numbers.

- Quantitative data is analyzed using descriptive and inferential statistical methods. The results are frequently presented in tables.

- The collection of quantitative data lends itself to certain kinds of research designs.

Quantitative Research

- Focuses on variables that can be measured or labeled with numbers
- Studies many cases or participants.
- Produces nomothetic (generalizations) results.
- Attempts to stay independent of the context and setting.
- Detaches the researcher from the study.
- Analyzes numeric results using descriptive, and inferential statistics.
- Usually collects data through surveys, experiments, content analyses or structured interviews.

Qualitative Data

- Definition: data collected in the form of descriptions.
- Qualitative data is summarized in written descriptions of the contents of interviews, content analyses, observations or participant observation. The results are presented in essay format sometimes including quoted examples.
- The collection of qualitative data lends itself to certain kinds of research designs.

Qualitative Research

- Focuses on interactive processes, events, ideas and emotions.
- It involves the researcher intimately in the study being conducted.
- Is used to produce case studies of societies, groups or individuals.
- Provides the only way to collect truly emic data.
- Produces idiographic (descriptions of individual things or people) results.
- Usually collects evidence through unstructured interviewing, observation or participant observation or content analysis.

Steps in the Scientific Research Process

1. Choose a question to investigate
2. Identify a hypothesis related to the question
3. Make testable predictions in the hypothesis
4. Design an experiment to answer hypothesis question
5. Collect data in experiment
6. Determine results and assess their validity
7. Determine if results support or refute your hypothesis

The Scientific Method

1. "Suspicion that a factor (exposure) may influence occurrence of disease or a noted health outcome"
   - Observations in clinical practice
   - Examination of disease/outcome patterns
     - Do subpopulations have higher or lower rates?
     - Are disease rates increased in the presence of certain factors?
   - Observations in laboratory research
   - Theoretical speculation

The Scientific Method

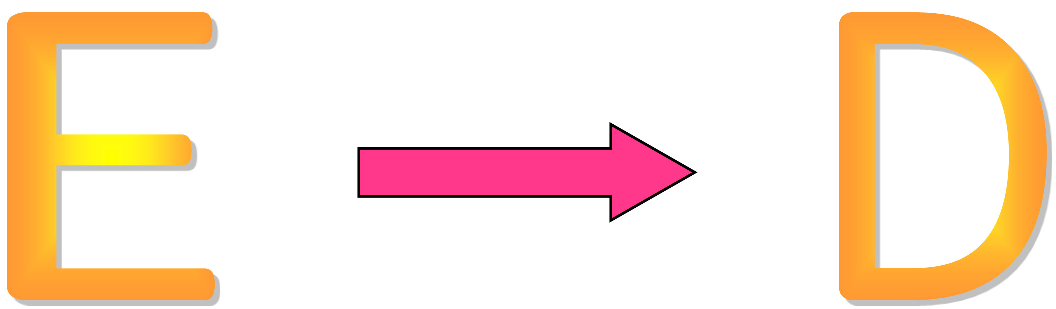
2. Identify variables you are interested in:
   - Exposure - (risk factor, protective factor, predictor variable, treatment)
   - Outcome - (disease, event)

3. Formulate a specific hypothesis
   - Frame a hypothesis which seeks to answer a specific question about the relationship between an exposure and an outcome

Basic Question in Research

Are exposure and disease/outcome linked?

Is there an association between them?



Exposure

Disease / Health Outcome

Next Step: Design Study

- Study Designs …(not exhaustive)
  - Case series
  - Cross-sectional
  - Case-control
  - Cohort
  - Randomized controlled clinical trial

Statistical Association

The degree to which the rate of disease or outcome in persons with a specific exposure is either higher or lower than the rate of disease or outcome among those without that exposure.

The Scientific Method

- Assess validity of association
  - Does the observed association really exist?
    - Is the association valid?
    - Are there alternative explanations for the association?
      - Chance
      - Bias
      - Confounding

Hypotheses

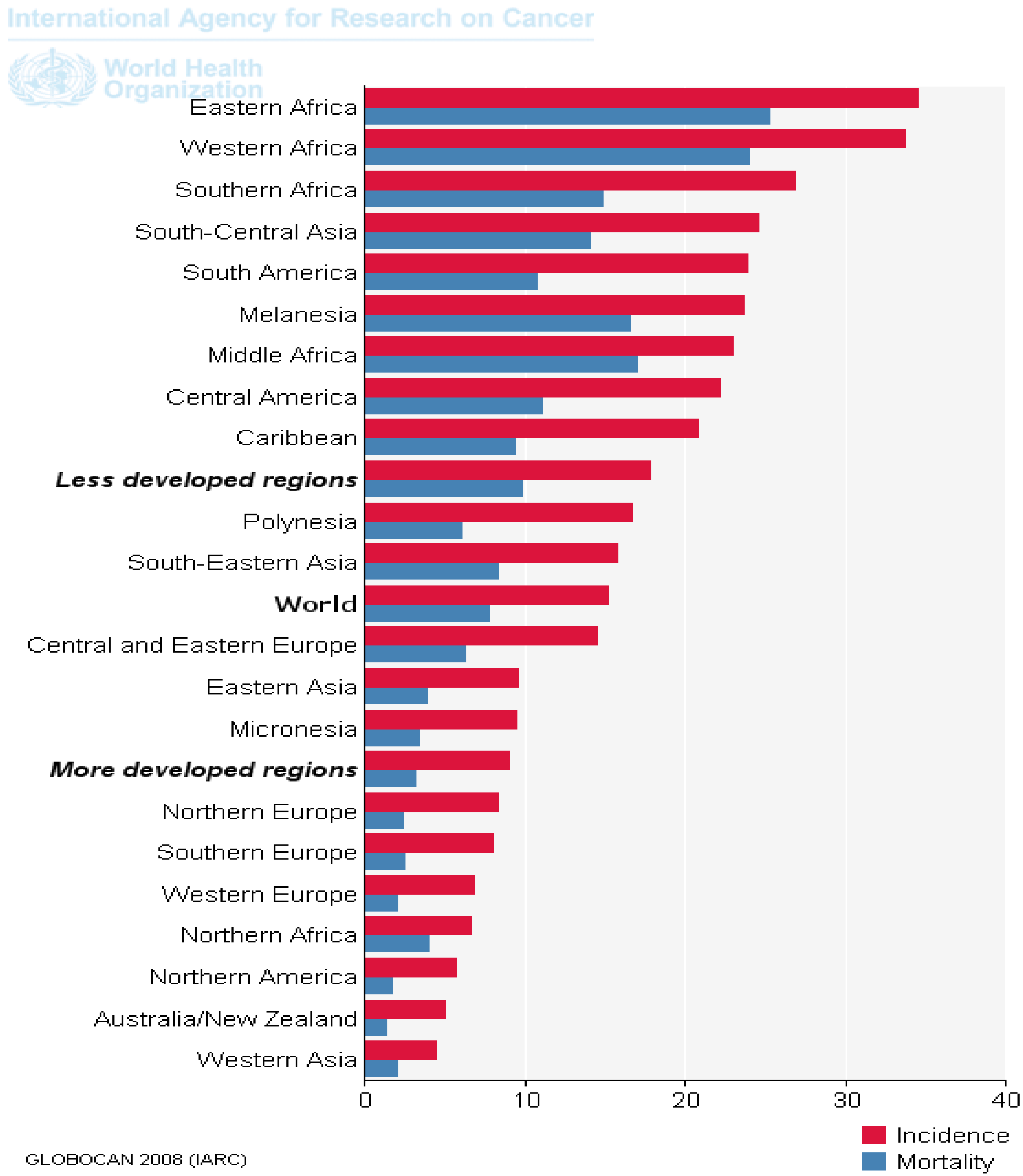
- Shape and guide a research study in terms of:

- Identification of study sample size
- What issues should be involved in
  - data collection
  - the proper analysis of the data
  - data interpretation

Hypothesis Formulation

- Formulate a hypothesis
- Frame the hypothesis in a format that is testable
- Test the hypothesis

Cervical Cancer



- Infectious and chronic diseases show great variation from one country to another.
- Some differences may be attributed to:
  - Climate
  - Cultural factors
  - Diet
  - Genetics

Hypothesis Framing

H_0: There is <u>no association</u> between the exposure and disease of interest

H_1: There <u>is</u> an association between the exposure and disease of interest (beyond what might be expected from random error alone)

Hypothesis Framing

Another Type of Framing:

What is the best estimate of the risk of disease in those who are exposed compared to those who are unexposed (i.e. exposed are at XX times higher risk of disease).

This moves away from the simple dichotomy of yes or no for an exposure/disease association – to the estimated magnitude of effect irrespective of whether it differs from the null hypothesis.

Hypothesis Framing

- Guidelines for Framing Hypotheses:
- State the exposure to be measured as specifically as possible.
- State the health outcome as
  - Specifically as possible.
  - Strive to explain the smallest amount of ignorance

Hypothesis Framing

- Example Hypotheses:
- POOR
  - Eating junk food is associated with the development of cancer.
- GOOD
  - The human papilloma virus (HPV) subtype 16 is associated with the development of cervical cancer.

The Next Step

- Formally test the identified hypotheses in a research study
  - The study should follow a specific plan or protocol (the study design)
  - Study designs direct how the investigation is conducted and allows for the translation of a conceptual hypothesis into an operational one

Why Conduct Research?

- To develop knowledge for professions.
- To develop effective policies.
- To solve practical problems.
- To make informed decisions.
- To increase the knowledge base of larger society.

- Huge amounts of daily life and experience in our society are based on what we have learned using the logic and evidence involved in scientific research.

BRETAM model of the development of science technology

- Gaines, B.: Modeling and forecasting the information sciences. Inf Sci 57/58: (1999) 13-22

Modeling and Forecasting the Information Sciences

Brian R. Gaines
Knowledge Science Institute
University of Calgary
Alberta, Canada T2N 1N4

Abstract

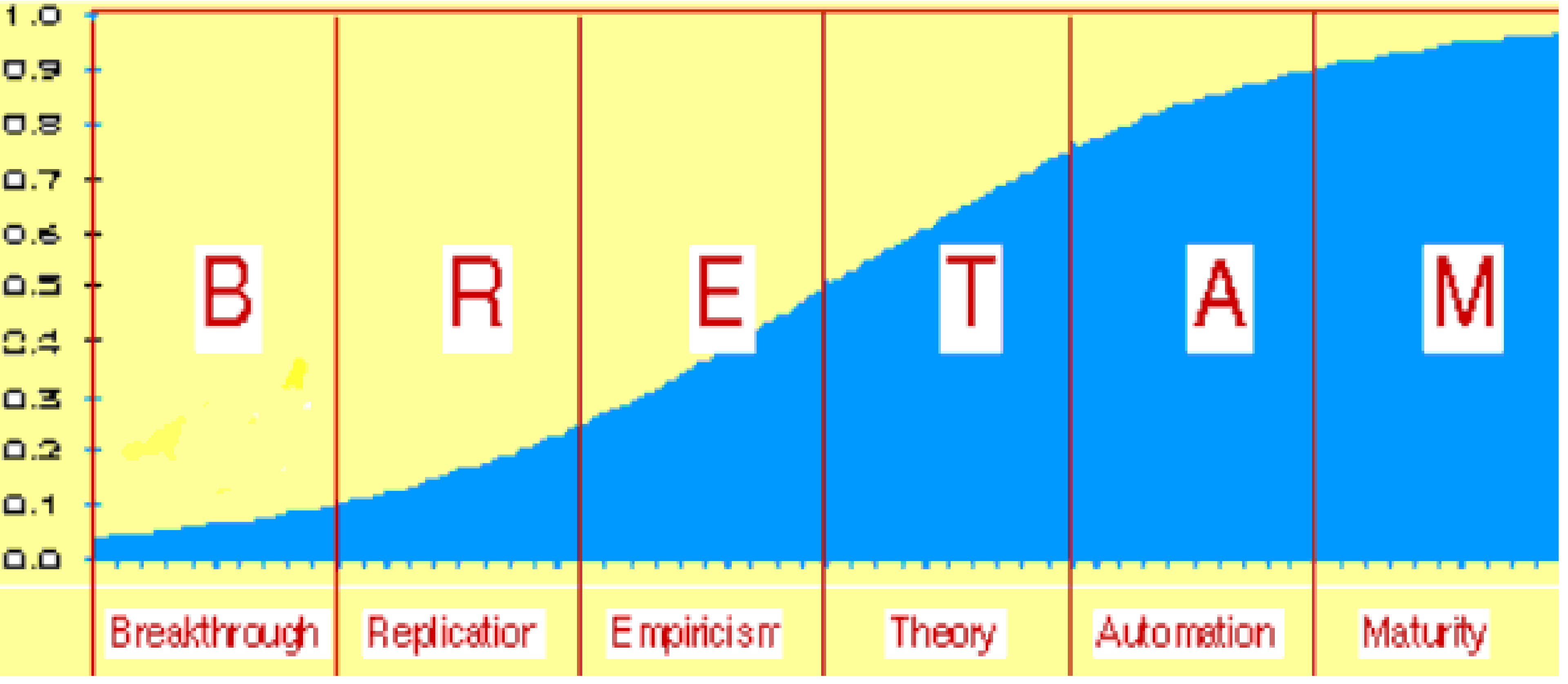
A model of the development of the information sciences is described and used to account for past events and predict future trends, particularly fifth and sixth generation priorities. The information sciences came into prominence as electronic device technology enabled the social need to cope with an increasingly complex world to be satisfied. Underlying all developments in computing is a tiered succession of learning curves which make up the infrastructure of the computing industry. The paper provides a framework for the information sciences based on this logical progression of developments. It links this empirically to key events in the development of computing. It links it theoretically to a model of economic, social, scientific and individual development as related learning processes with a simple phenomenological model. The fifth generation development program with its emphasis on human-computer interaction and artificial intelligence, and the sixth generation research program with its emphasis on knowledge science are natural developments in the foci of attention indicated by the model.

1 Introduction

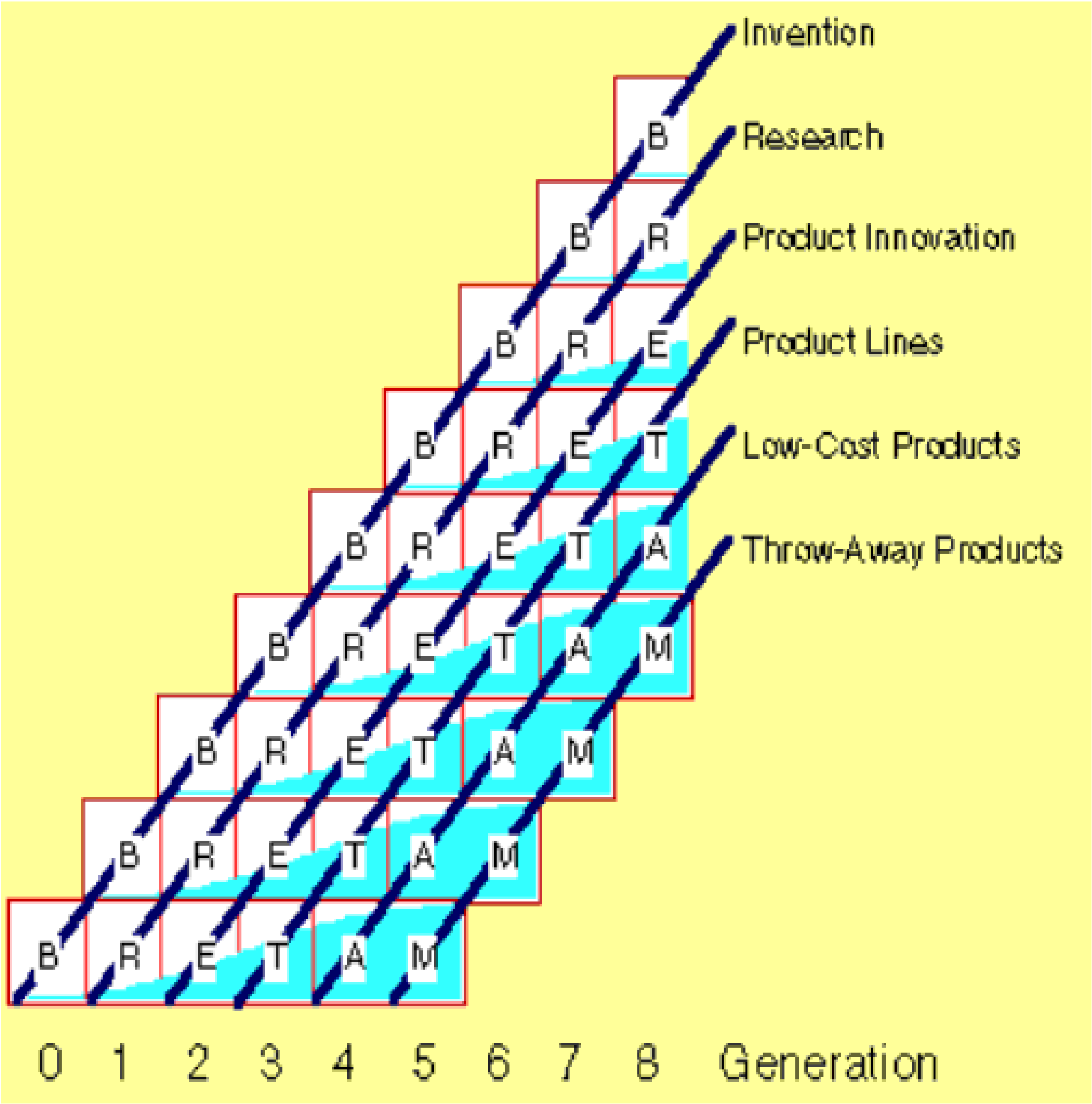
Forecasting advances in technology and their impact has a track record of making fools of the forecasters (Schnaars, 1989). However, the game of life is one of anticipating the future. We model the past that we may learn the lessons of experience, and we extrapolate our models into the future. That the future reserves the right not be anticipated is a meta-lesson that Hume taught many years ago. Nevertheless, our individual behaviors and our civilizations are founded on the assumption that anticipation is possible and, at some level of modeling, history does repeat itself. Our technological civilizations go one step further and change the universe to reify our anticipations. The information sciences, in particular, structure a wholly artificial reality composed from the ultimate abstractions of the human mind. Paradoxically, they should be readily modeled because they are artifacts of our own mentation, but they may also be beyond modeling because to do so fully may involve ultimate understanding of ourselves.

This paper presents an integrative model of the information sciences that shows them as a tightly coupled, mutually supportive system. It first presents the underlying electronic device technology which provides both the basic support and physical constraints on information technology. It then presents a model of the learning curves of scientific and technological knowledge acquisition that underlies the development of the information sciences. The historic opportunities triggering successive advances in the information sciences are then analyzed and their learning curves superimposed to provide a model of the infrastructure of the information sciences. This model is extrapolated to provide forecasts of future directions, and fitted to various scientific and technological developments. The interactive synergies between levels are analyzed to show the basis of the positive feedback phenomena which continue to support the exponential growth of the information sciences and technologies.

Evolution of domain



The process ladder



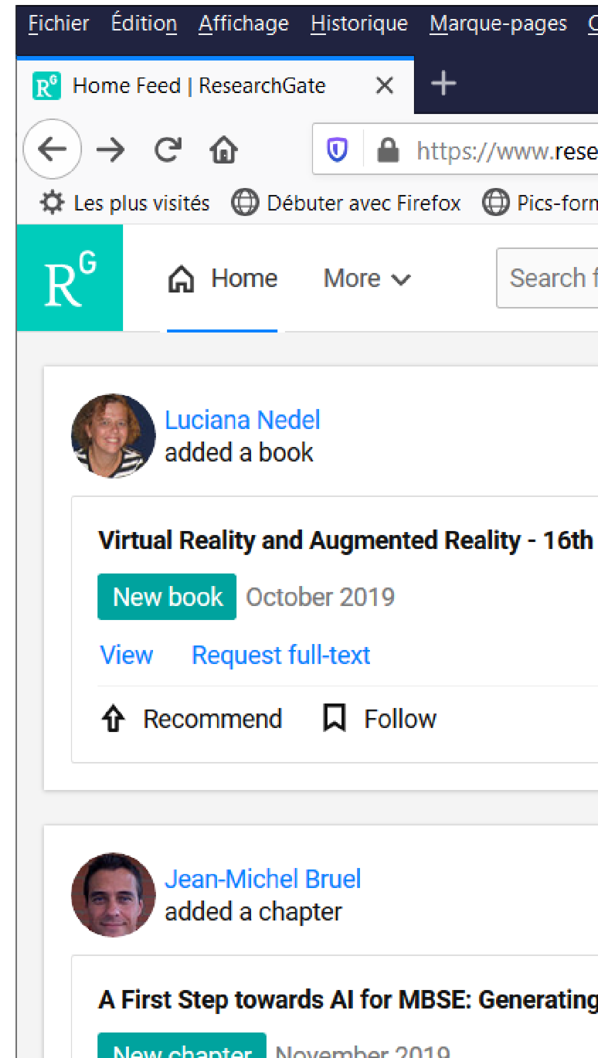
Keep updated

Sources of information about

- Digital libraries and Collaborative Platforms
- Scientific Associations
- Attending Scientific events
  - conferences, symposiums, workshops, …

Digital Library and Collaborative Platforms

- Digital Libraries
  - PubMed: https://www.ncbi.nlm.nih.gov/pubmed/
  - ACM Digital Library: http://dl.acm.org/
  - Springer Link: https://link.springer.com/
  - IEEExplorer: https://ieeexplore.ieee.org/
  - HAL Archives Ouvertes (France): https://hal.archives-ouvertes.fr
- Collaborative Platforms
  - Interaction Design Fondation: https://www.interaction-design.org/
  - Research Gate https://www.researchgate.net/
  - Mendely: https://www.mendeley.com/
- Also indexes
  - DBLP: https://dblp.uni-trier.de/
  - Google Scholars: https://scholar.google.com/



Scientific Associations

Association for Computing Machinery

http://ifip.org/

- Association for Computing Machinery (ACM)
  - 37 Special Interest Groups (SIGs)

ifip

http://ifip.org/

- International Federation for Information Society (IFIP)
  - Countries representation

IEEE
Advancing Technology for Humanity

https://www.ieee.org/

- Institute of Electrical and Electronics Engineers (IEEE)

Scientific events

- Forums for meeting other researchers and industrialists
- Place to…
  – Report our scientific findings
  – Getting critics about our work
  – Learn from others
  – Assess overall knowledge about a topic/domain
  – Create scientific networking
    - Make project proposals, writing papers, jury, etc…

Conference, Workshop, Journal?

- National OR international
- **Workshop:** work in progress and preliminary results, might be subject of a publication
- **Conference:** mature work, scientific publication in the form of **proceedings**…
  - **Many venues/tracks:** short/full papers, PhD Symposium, Keynotes (notable invitees)
- **Journal:** extended work, comprehensive material, do not require a live presentation, often many rounds of review

Reviewing process


Insert Your Title Here*
Insert Subtitle Here

- Submissions
  - Follow specific format/template
  - Often anonymous
  - Attention to deadlines…
- Quality check
  - Anonymity, scope, etc.
- Peer reviewing process
  - 3 or more anonymous reviewers (blind review)
  - Might have a meta-reviewer (coordinator)
- Editor/Committee decision
  - Accept as it is, minor review, major review, reject
- Preparation for publication
  - Revision, publication consent, etc.

How to read a scientific paper ?

(usual) Anatomy of a scientific paper

- Title and authors
- Abstract/summary
- Introduction
- Materials and Methods
- Results
- Discussion
- Acknowledgements
- References
- Figures/Tables

Title and authors

- Title is very descriptive (often states the main finding) and is not about being creative and "catchy"!
- Order of authors is important. What can you tell from it?

Example:

VEGF, a prosurvival factor, acts in concert with TGF-beta1 to induce endothelial cell apoptosis.
Ferrari G, Pintucci G, Seghezzi G, Hyman K, Galloway AC, Mignatti P.

Abstract/Summary

- Brief background of subject
- Purpose for the study
- Major findings of the study
- Relationship between these findings and the field

This is what you see when you do a pubmed search. You can decide if the paper is worth reading based upon this.

Introduction

- Presents the background information for a fellow scientist (possibly in another field) to understand why the findings of this paper are significant.
- Structure is usually:
  - Accepted state of knowledge in the field
  - Focus on a particular aspect of the field, often the set(s) of data that led directly to the work of this paper
  - Hypothesis being tested
  - Conclusions (scientists don't really like surprise endings!)

How to approach the introduction

- **Grab a blank piece of paper:**
  - Take notes
  - Draw mini figures
  - Define vocabulary
    (wikipedia is a quick reference)

- **Answer these questions:**
  - What is the accepted state of knowledge?
  - What data led directly to the work of this paper?
  - What is the hypothesis being tested?
  - What are the basic conclusions? (Scientists don’t really like surprise endings and this is usually stated in the last paragraph.)

Notes allow you to take a break (hours to days) and come back to your thoughts…you won't have to re-read the parts you completed.

Materials and Methods

- Should be detailed enough for another scientist to replicate the work (volumes, times, company material was purchased from etc.)
- In reality, often compressed and you may need to look up another paper that is referenced for more detail.

Should you read the materials and methods?

- Often you can skim over them before the results.
- However, when you get to the results, you will need to flip back to them often to clarify how experiment was done.
  - Sample number? (Did they do this more than once?)
  - Conditions? (Am I looking at a reduced or non-reduced protein gel?)

Results

- While the introduction poses the questions being asked, the results describes the outcome of the experiments that were done to answer the questions.
- Results are often simply stated with *interpretation* of them coming later in the discussion.
- Figures and tables allow the reader to see the outcomes of the experiments for themselves!

How to read the results:

- Read the text straight through, but as a figure is referred to, examine the figure.
- Take notes, giving yourself a place to refer to about each figure.
- With each experiment/figure you should be able to explain :

1) the basic procedure
2) the question it sought to answer
3) the results
4) the conclusion
5) criticisms

Discussion

- Data is analyzed to show what the authors believe the data show. (You don’t have to agree with their interpretations!)
- Findings are related to other findings in the field (contribute to knowledge, correct errors, etc.)– How is this work significant?

How to read a discussion

- Take notes and answer these questions:
  - What conclusions do the authors draw? Be sure to separate fact from their opinion/interpretation?
  - Describe for yourself why these data significant. (Does it contribute to knowledge or correct errors?)

By now, you may be tired of this paper…
but don't relax yet.

Save energy for the overall reflection and criticism.

Acknowledgements

- Thank people who contributed materials.
- Thank people who contributed technically but maybe not intellectually (would not be authors).

References

- Papers cited in the text
- What parts of the paper cite other papers?
  – Introduction
  – Materials and Methods
  – Discussion
  – (Maybe a few in Results)

Question: How should we READ a scientific paper?

Answer: not necessarily in order!

A four-step method based on: **Ann McNeal, School of Natural Science, Hampshire College, Amherst MA**

http://hampshire.edu/~apmNS/design/RESOURCES/HOW_READ.html

Step 1: Skim the entire paper

- Look at the major headings (do they follow the “anatomy” we just described?)
- How many figures are there, what kinds of figures are they (gels, graphs, microscopic images?)
- What is the conclusion of the paper?
  - (It may not make sense to you at the moment, but note what it is.)

Step 2:Vocabulary

- Go through the paper as a whole simply underlining words and phrases you do not understand.
- You are not reading the paper for comprehension of the whole paper yet, just making sure you have understanding of the words to then comprehend it.

Step 2: Vocabulary continued

- Look up simple words and phrases, where?
  - Biology textbooks
  - Online at biology dictionaries or encyclopedias (www.wikepedia.com seems to be a good resource for basic definitions and procedures)
  - Look up methods that you are not familiar with

  (i.e. what is an immunoprecipitation or a transformation?)

- Note important phrases that are part of a major concept and are bigger than just vocabulary (i.e. "risk reduction"). You will come back to them in context while reading for comprehension.

Step 3: Read for comprehension, section by section

- Introduction
  - What is the accepted state of knowledge in the field (take notes and even draw your own figures)?
  - What data led directly to this work?
  - What question are they answering? (Is there a clear hypothesis?)
  - What are their conclusions?

Step 3: Read for comprehension, section by section

- Materials and Methods and Results:
  - Read the methods first or read them as you read the results. (I prefer the latter)
  - With each experiment/figure you should be able to explain
    - 1) the basic procedure
    - 2) the question it sought to answer
    - 3) the results
    - 4) the conclusion

- You should be able to explain all of these (1-4) to another classmate clearly!

Step 3: Read for comprehension, section by section

- Discussion
  - What conclusions do the authors draw? Be sure to separate fact from their opinion/interpretation?
  - Describe for yourself why these data significant. (Does it contribute to knowledge or correct errors?)

Step 4: Reflection and criticism

- Do you agree with the authors' rationale for setting up the experiments as they did?
- Did they perform the experiments appropriately? (Repeated a number of times, used correct control groups, used appropriate measurements etc)
- Were there enough experiments to support the one major finding they are claiming?
- Do you see patterns/trends in their data that are problems that were not mentioned?
- Do you agree with the authors' conclusions from these data? Are they over-generalized or too grand? Or are there other factors that they neglect that could have accounted for their data?
- What further questions do you have? What might you suggest they do next?

Tips for success:

- Spend a lot of time on each paper NOW look up every detail that you are unsure of. (Time you invest now will payoff in the long run). Discovering the answers for yourself is one of the best ways to learn and have the information be retained.
- Imagine yourself teaching the paper or figures to classmates—teaching something to others is also another great way to learn.

Tips for success:

- Start a database of procedures that you take the time to look up and teach to yourself. What are some of the common procedures that are used in various papers? (e.g. western, immunoblots, RT-PCR, apoptosis assays, yeast two hybrids, etc.)
- Watch others in your lab experiences and find out what they are doing…you may never get the opportunity to do RT-PCR, but the more you understand the procedure, the more critical you can be of data you need to interpret.

Tips for success:

- Read papers when you are awake and interested in reading. If you are going to break up a paper and read it over several days be sure to summarize before continuing each day.
- If you are already in the field you plan to stay in, consider starting a database on papers that relate to your lab/project. You will want to be able to impress your P.I. with your quick analysis and summary of a monumental paper from another lab!

Homework

- Search relevant literature for your project
- Create a bibliography dataset
  - Ex. Excel file
  - Add info citation paper and category (ex; conf./workshop/journal, etc.)
- Create a grid for comparing the work
  - Start with a initial list of attributes
    - Select obvious attributes for the domain. Ex. List of topic
    - Define range of values for attributes
  - For each paper
    - Classify each paper according to attributes
    - Add keywords, personal notes, etc.…
  - If new topics or attributes are cited, revise the grid and check papers again
- Bring you classification for the next class