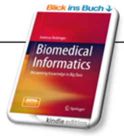


Andreas Holzinger

VO 709.049 Medical Informatics

25.11.2015 11:15-12:45

Lecture 07

Knowledge, Decision, Uncertainty, Bayesian Statistics, Probabilistic Modelling

a.holzinger@tugraz.at

Tutor: markus.plass@student.tugraz.at

http://hci-kdd.org/biomedical-informatics-big-data

Status as of 24.11.2015 13:30

Dear Students, welcome to the 7th lecture of our course. Please remember from the last lecture: the differences between data mining, knowledge discovery and the importance of machine learning for those fields. You should now be aware of some methods and their usage in biomedical informatics.

In this lecture we learn some basics which will we then use in lecture 08 on decision support – the most difficult but at the same time very important topic.

Please always be aware of the definition of biomedical informatics (Medizinische Informatik):
Biomedical Informatics is the inter-disciplinary field that studies and pursues the effective use of biomedical data, information, and knowledge for scientific inquiry, problem solving, and decision making, motivated by efforts to improve human health (and well-being).

Schedule

- 1. Intro: Computer Science meets Life Sciences, challenges, future directions
- 2. Back to the future: Fundamentals of Data, Information and Knowledge
- 3. Structured Data: Coding, Classification (ICD, SNOMED, MeSH, UMLS)
- 4. Biomedical Databases: Acquisition, Storage, Information Retrieval and Use
- 5. Semi structured and weakly structured data (structural homologies)
- 6. Multimedia Data Mining and Knowledge Discovery
- **7. Knowledge, Decision, Cognition, Probability, Uncertainty, Bayes & Co**
- 8. Biomedical Decision Making: Reasoning and Decision Support
- 9. Intelligent Information Visualization and Visual Analytics
- 10. Biomedical Information Systems and Medical Knowledge Management
- 11. Biomedical Data: Privacy, Safety and Security
- 12. Methodology for Info Systems: System Design, Usability & Evaluation

Today we tackle some important basics for decision making and deal with probability, uncertainty, Bayesian statistics and probabilistic modelling, Which represents the fundamentals for machine learning

Keywords of 7th Lecture

- Bayes theorem
- Case based reasoning
- Differential diagnosis
- Human decision making
- Hypothetico-deductive method
- Incomplete data
- Model of human information processing
- Modeling patient health
- PDCA-Deming wheel
- Receiver operating characteristics
- Rough set theory
- Selected attention
- Signal detection theory
- Triage

Advance Organizer (1/2)

- **Brute Force** = a trivial very general problem-solving technique that consists of systematically enumerating all possible candidates for the solution and checking whether each candidate satisfies the problem's statement;
- **Cognition** = mental processes of gaining knowledge, comprehension, including thinking, attention, remembering, language understanding, decision making and problem-solving;
- **Cognitive load** = According to Sweller (1996) a measure of complexity and difficulty of a task, related to the executive control of the short-term memory, correlating with factors including (human) performance; based on the chunk-theory of Miller (1956);
- **Cognitive Science** = interdisciplinary study of human information processing, including perception, language, memory, reasoning, and emotion;
- **Confounding Variable** = an unforeseen, unwanted variable that jeopardizes reliability and validity of a study outcome.
- **Correlation coefficient** = measures the relationship between pairs of interval variables in a sample, from r = -1.00 to 0 (no correlation) to r = +1.00
- **Decision Making** = a central cognitive process in every medical activity, resulting in the selection of a final choice of action out of alternatives; according to Shortliffe (2011) DM is still the key topic in medical informatics;
- **Diagnosis** = classification of a patient's condition into separate and distinct categories that allow medical decisions about treatment and prognostic;
- **Differential Diagnosis (DDx)** = a systematic method to identify the presence of an entity where multiple alternatives are possible, and the process of elimination, or interpretation of the probabilities of conditions to negligible levels;
- **Evidence-based medicine (EBM)** = aiming at the best available evidence gained from the scientific method to clinical decision making. It seeks to assess the strength of evidence of the risks and benefits of treatments (including lack of treatment) and diagnostic tests. Evidence quality can range from meta-analyses and systematic reviews of double-blind, placebo-controlled clinical trials at the top end, down to conventional wisdom at the bottom;

Advance Organizer (2/2)

- **External Validity** = the extent to which the results of a study are generalizable or transferable;
- **Hypothetico-Deductive Model (HDM)** = formulating a hypothesis in a form that could conceivably be falsified by a test on observable data, e.g. a test which shows results contrary to the prediction of the hypothesis is the falsification, a test that could but is not contrary to the hypothesis corroborates the theory – then you need to compare the explanatory value of competing hypotheses by testing how strong they are supported by their predictions;
- **Internal Validity** = the rigor with which a study was conducted (e.g., the design, the care taken to conduct measurements, and decisions concerning what was and was not measured);
- **PDCA** = Plan-Do-Check-Act, The so called PDCA-cycle or Deming-wheel can be used to coordinate a systematic and continuous improvement. Every improvement starts with a goal and with a plan on how to achieve that goal, followed by action, measurement and comparison of the gained output.
- **Perception** = sensory experience of the world, involving the recognition of environmental stimuli and actions in response to these stimuli;
- **Qualitative Research** = empirical research exploring relationships using textual, rather than quantitative data, e.g. case study, observation, ethnography; Results are not considered generalizable, but sometimes at least transferable.
- **Quantitative Research** = empirical research exploring relationships using numeric data, e.g. surveys, quasi-experiments, experiments. Results should be generalized, although it is not always possible.
- **Reasoning** = cognitive (thought) processes involved in making medical decisions (clinical reasoning, medical problem solving, diagnostic reasoning, behind every action;
- **Receiver-operating characteristic (ROC)** = in signal detection theory this is a graphical plot of the sensitivity, or true positive rate, vs. false positive rate (1 − specificity or 1 − true negative rate), for a binary classifier system as its discrimination threshold is varied;
- **Symbolic reasoning** = logical deduction
- **Triage** = process of judging the priority of patients' treatments based on the severity of their condition;

Glossary

- CES = Central Executive System
- DDx = Differential Diagnosis
- DM = Decision Making
- DSS = Decision Support System
- EBM = Evidence-based medicine
- fMRI = functional Magnetic Resonance Image
- HDM = Hypothetico-Deductive Model
- IOM = Institute of Medicine
- LTS = Long Term Storage
- ME = Medical Error
- PDCA = Plan-Do-Check-Act
- QM = Quality Management
- ROC = Receiver Operating Characteristic
- ROC = Receiver-operating characteristic
- RST = Rough Set Theory
- STS = Short Term Storage
- USTS = Ultra Short Term Storage

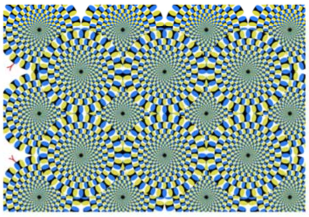
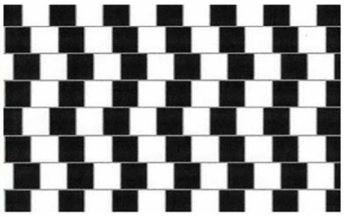
Learning Goals: At the end of this 7th lecture you …

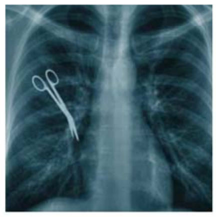
- … are familiar with some principles and elements of human information processing;
- … can discriminate between perception, cognition, thinking, reasoning & problem solving;
- … have got insight into some basics of human decision making processes;
- … got an overview of the Hypothetico-Deductive Method HDM versus PCDA Deming approach;
- … have acquired some basics on modeling patient health, differential diagnosis, case-based reasoning and medical errors;

Slide 7-1 Key Challenges

- Time to make a decision = “5 Minutes” [1], [2]

- Limited perceptual, attentive and cognitive human resources [3], and Human error
- Noisy, missing, probabilistic, uncertain data



[1] Gigerenzer, G. 2008. *Gut Feelings: Short Cuts to Better Decision Making* London, Penguin.
[2] Gigerenzer, G. & Gaissmaier, W. 2011. Heuristic Decision Making. In: Fiske, S. T., Schacter, D. L. & Taylor, S. E. (eds.) Annual Review of Psychology, Vol 62. pp. 451-482.
[3] Bialek, W. 1987. Physical Limits to Sensation and Perception. Annual Review of Biophysics and Biophysical Chemistry, 16, 455-478.

A. Holzinger 709.049 8/79 Med Informatics L07

- The short time to make a decision “5 Minutes”
- Limited perceptual, attentive and cognitive human resources
- Human error

According to Gerd Gigerenzer from the Max Planck Insitute in Berlin the average time spend for a patient, hence to make a decision in a public hospital is 5 minutes. This is the famous “5 Minutes” approach (Gigerenzer, 2008); however, in essence time is fundamentally critical in medicine and health care.

Medical Diagnosis - Decision Making

Slide 7-2: Decision Making is central in Biomedical Informatics

Source: Cisco (2008). Cisco Health Presence Trial at Aberdeen Royal Infirmary in Scotland

A. Holzinger 709.049 10/79 Med Informatics L07

De-cision (Ent-Scheidung) is a cognitive process resulting in the selection of one of several alternatives.
Every decision making process produces a final choice and the output can be an action or an opinion.
This slide shows a typical clinical decision making process supported by a “second opinion” during a project of the Cognitive Engineering Research Group and the Scottish Centre for Telehealth at Aberdeen Royal Infirmary.
The project aims to investigate human errors in carrying out diagnosis of dermatological conditions using remote imaging. A particular focus is on the role of color images, i.e. the effect of image quality and characteristics as a function of human color perception. The research will develop a model of human information processing that incorporates the user in a “human-in-the-loop-system”. Such a model can form the basis for predicting and preventing human errors relating to color in remote imaging of dermatological problems.
Remark: If it were not for human error there would be little justification for the area of Human Factors Engineering as an applied discipline. Research during the past 20 years has made progress in understanding and controlling human errors in human-computer interaction but technology develops fast which poses new challenges for researchers. All users of IT systems make errors and this is accepted.
http://www.comp.rgu.ac.uk/staff/ph/html_docs/Human_error_telehealth.htm

Slide 7-3: Reasoning Foundations of Medical Diagnosis

3 July 1959, Volume 130, Number 3366 SCIENCE

Reasoning Foundations of Medical Diagnosis

Symbolic logic, probability, and value theory aid our understanding of how physicians reason.

Robert S. Ledley and Lee B. Lusted

The purpose of this article is to analyze the complicated reasoning processes inherent in medical diagnosis. The importance of this problem has received recent emphasis by the increasing interest in the use of electronic computers as an aid to medical diagnostic processes

fitted into a definite disease category, or that it may be one of several possible diseases, or else that its exact nature cannot be determined." This, obviously, is a greatly simplified explanation of the process of diagnosis, for the physician might also comment that after seeing a

ance are the ones who do remember and consider the most possibilities."

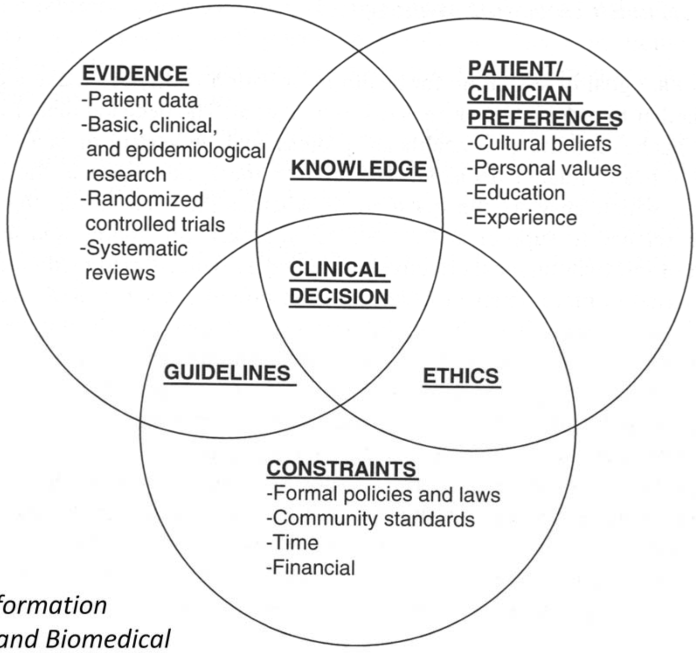
Computers are especially suited to help the physician collect and process clinical information and remind him of diagnoses which he may have overlooked. In many cases computers may be as simple as a set of hand-sorted cards, whereas in other cases the use of a large-scale digital electronic computer may be indicated. There are other ways in which computers may serve the physician, and some of these are suggested in this paper. For example, medical students might find the computer an important aid in learning the methods of differential diagnosis. But to use the computer thus we must understand how the physician makes a medical diagnosis. This, then, brings us to the subject of our investigation: the reasoning foundations of medical diagnosis and treatment.

Medical diagnosis involves processes that can be systematically analyzed, as well as those characterized as "intangible." For instance, the reasoning foundations of medical diagnostic procedures

If a physician is asked, "How do you make a medical diagnosis?" his explanation of the process might be as follows. "First, I obtain the case facts from the patient's history, physical examination, and laboratory tests. Second, I evaluate the relative importance of the different signs and symptoms. Some of the data may be of first-order importance and other data of less importance. Third, to make a differential diagnosis I list all the diseases which the specific case can reasonably resemble. Then I exclude one disease after another from the list until it becomes apparent that the case can be fitted into a definite disease category, or that it may be one of several possible diseases, or else that its exact nature cannot be determined.“

This, obviously, is a greatly simplified explanation of the process of diagnosis, for the physician might also comment that after seeing a patient he often has a "feeling about the case." This "feeling," although hard to explain, may be a summation of his impressions concerning the way the data seem to fit together, the patient's reliability, general appearance, facial expression, and so forth; and the physician might add that such thoughts do influence the considered diagnoses. No one can doubt that complex reasoning processes are involved in making a medical diagnosis. The diagnosis is important because it helps the physician to choose an optimum therapy, a decision which in itself demands another complex reasoning process (Ledley & Lusted, 1959)

Slide 7-4 Decision Making is central in Medicine!

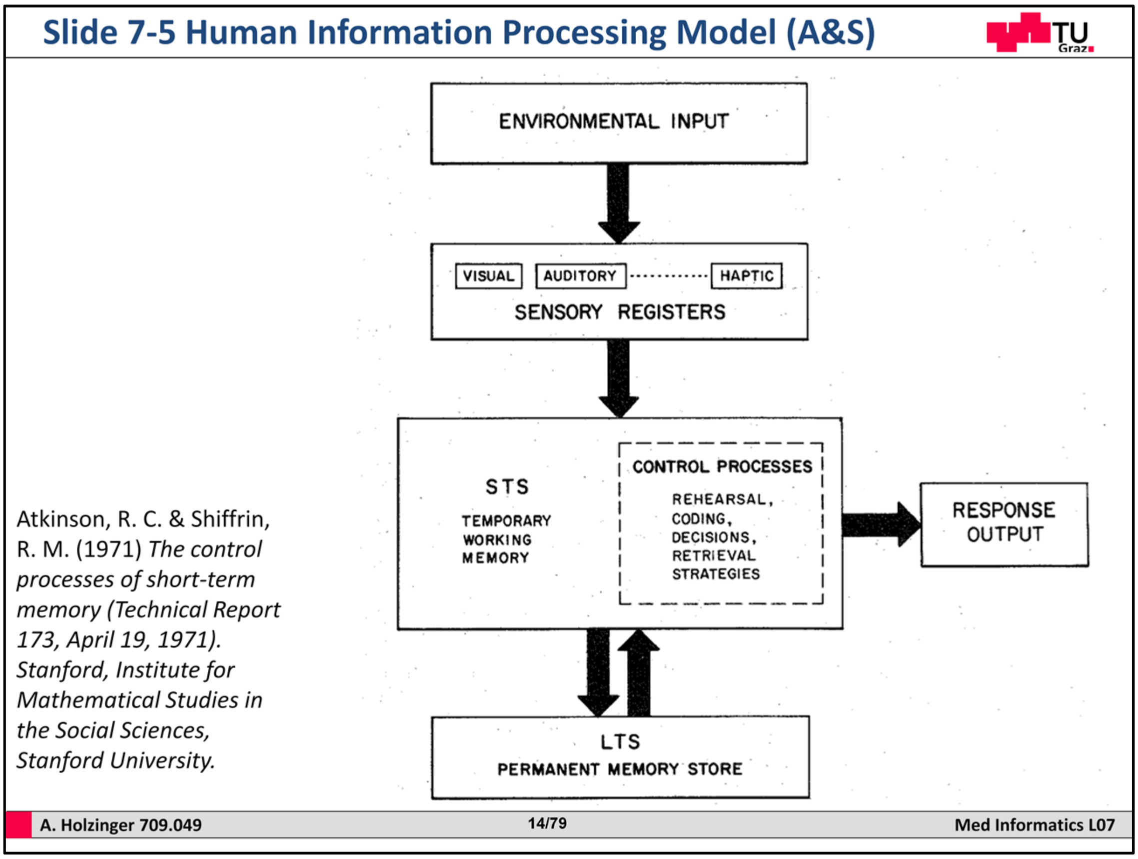


Hersh, W. (2010) *Information Retrieval: A Health and Biomedical Perspective. New York, Springer.*

A. Holzinger 709.049 12/79 Med Informatics L07

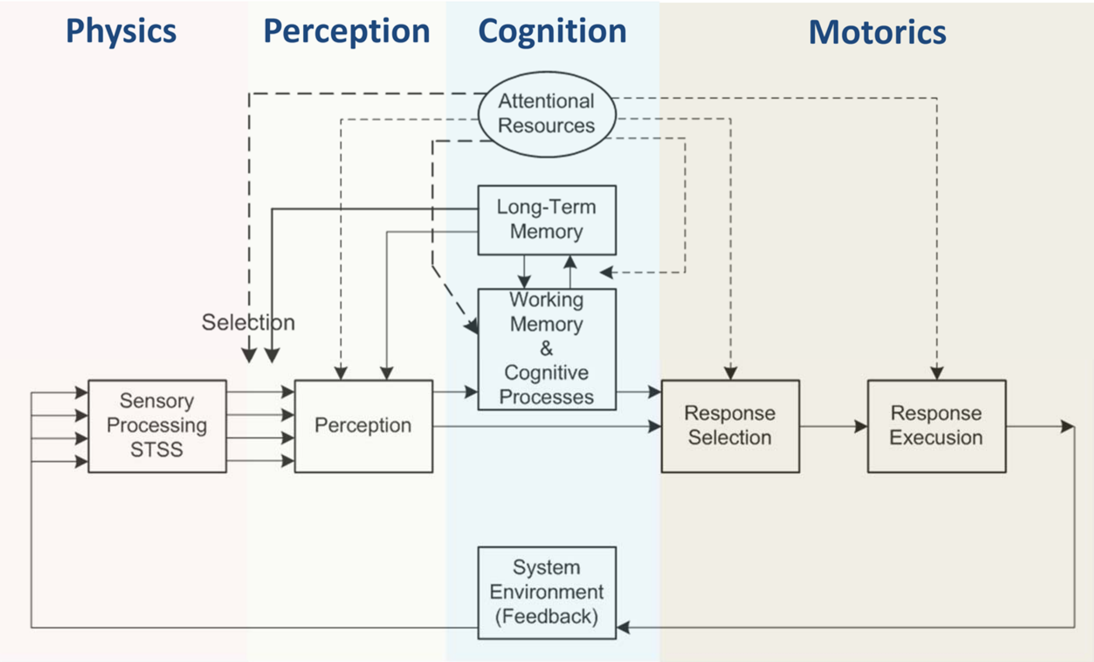
In practice the clinicians use Decision-Making Frameworks to guide patient management, communicate with other health care providers and educate patients and their families. A number of frameworks have been applied to guide clinical practice. (Schenkman, Deutsch & Gill-Body, 2006) have proposed a unifying framework for application to decision making in the management of individuals on the example of neurologic dysfunction. The framework integrates both enablement and disablement perspectives (Schenkman, Deutsch & Gill-Body, 2006). A good primer of the theory of medical decision making is (McNeil, Keeler & Adelstein, 1975). Clinical decisions are dependent on evidence (e.g. patient daa, epidemiological research, randomized controlled trials, etc.), guidelines, ethics, knowledge, patient/clinicians subjective preferences (cultural beliefs, see e.g. on the influence of organizational culture and communication (Xie et al., 2013), personal values, education and experience; and last but not least: constraints (e.g. formal policies, restrictions, laws, community standards, time, and financial issues) (Hersh, 2010). Before we go into more detail let us review how humans process information.

Human Information Processing



Information processing is the core of research in decision support and cognitive performance and we must mention that there are dozens of different models available in this scientific community. Here we see the information flow in the famous three-storage memory system developed by (Atkinson & Shiffrin, 1971): The environmental information is processed by sensory registers in the various physical modalities (visual, auditory, haptic, gustatoric, olfactoric) and entered into the short-term store (STS). The information remains temporarily in STS, the length of stay depending on control processes. Whilst the information remains in STS it may be copied into the long-term store (LTS). While information remains in STS, information in LTS associated with it may also be activated and entered in STS. Background Information: Lev Vygotsky (1896 - 1934) and Jean Piaget (1896 - 1980) were the first who indicated the importance of the end user's knowledge base for interpreting information. New knowledge should not be redundant but should have sufficient reference to what the user already knows. Moreover, sufficient overlap fosters the understandability of information, in particular deep understanding (Kintsch, 1994), is an important determinant of interest (Schraw & Lehman, 2001), and, through both understandability and interest, the knowledge base has an indirect effect on motivation. Concluding, we can say that it is of pivotal importance to have a **model of the user's knowledge** allowing to measure the distance between a topic and the user's Knowledge Model (KM) (van der Sluis & van den Broek, 2010).

Slide 7-6 General Model of Human Information Processing

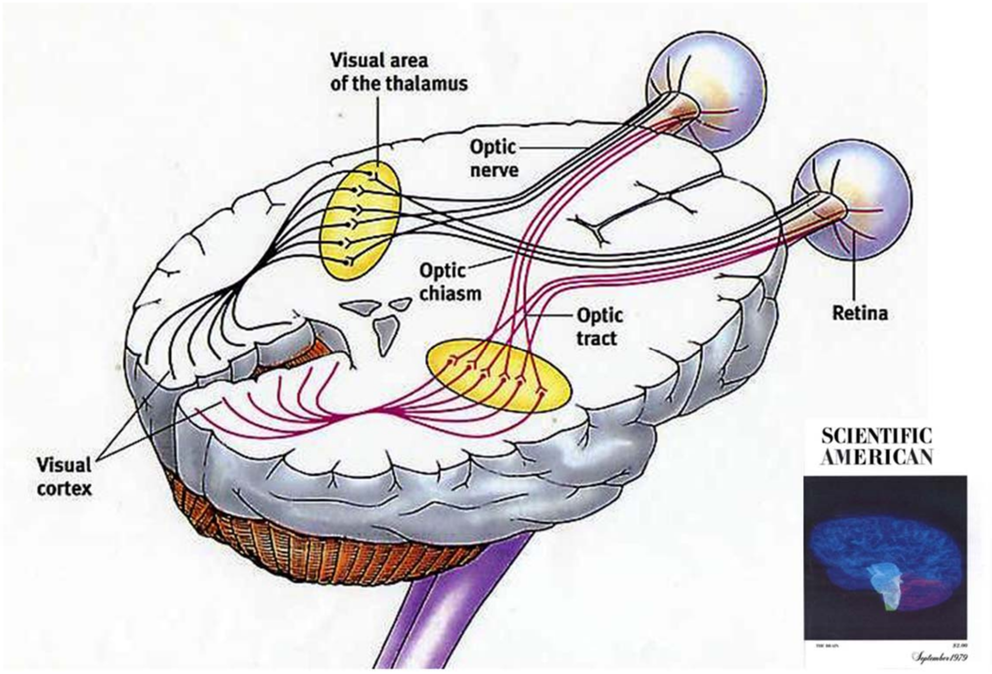


Wickens, C., Lee, J., Liu, Y. & Gordon-Becker, S. (2004) *Introduction to Human Factors Engineering: Second Edition. Upper Saddle River (NJ), Prentice-Hall.*

A. Holzinger 709.049 15/79 Med Informatics L07

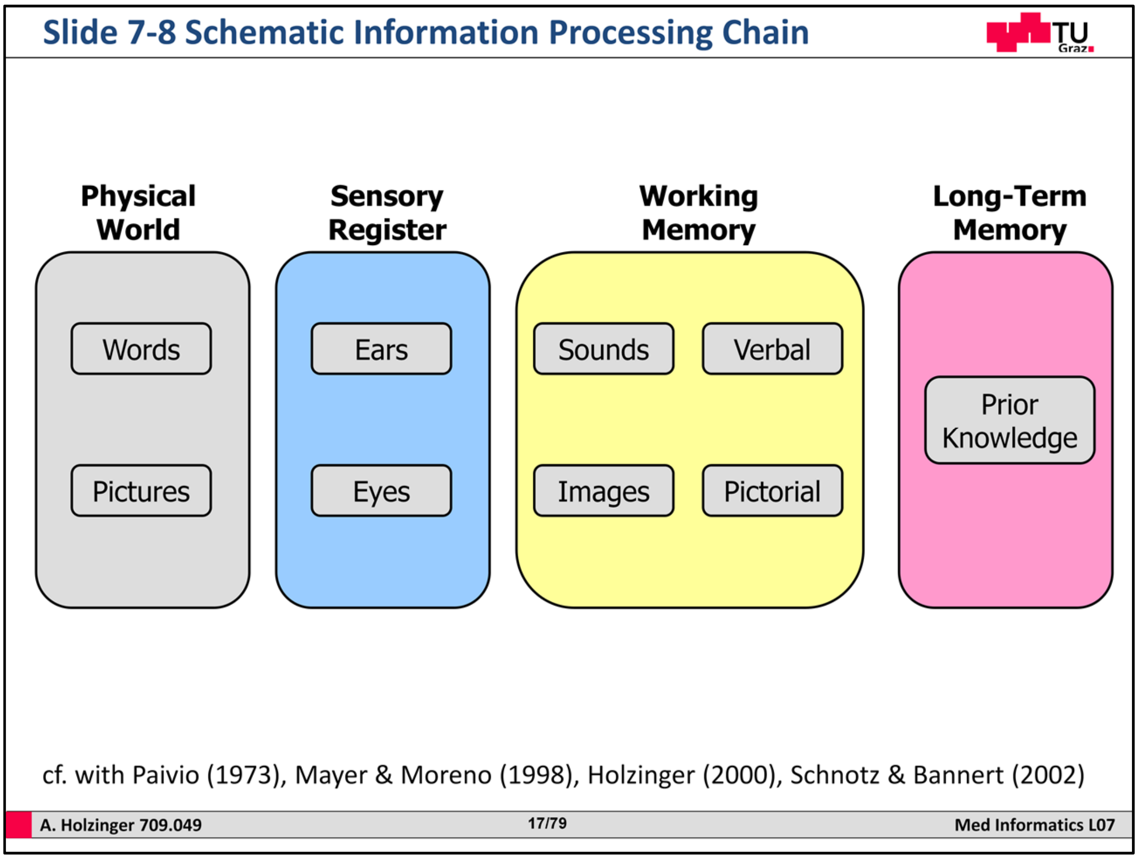
The three-storage model of Atkinson & Shiffrin is up to the present the standard information-processing model, in which each sensory system has its own store.
A current view on the three-stage human memory model (Wickens et al., 2004) makes this obvious: We can determine the four different areas: Physics (of information), perception, cognition and action (motorics) – the system having a feedback loop (e.g. the outcome of a decision is seen and accordingly will be perceived and cognitively processed).
The sensory processing unit is equivalent to the Ultra-short term storage (USTS) and holds the incoming sensory information long enough so that unconscious processes can determine whether the input should be let into the working memory, or should immediately be discarded (controlled by attentional resources). Selected stimuli which passes through this USTS will be perceived and hold within the Short-term Storage (STM, working memory), which is regarded as the center of consciousness, analog to the central processing unit (CPU) of a Von-Neumann machine (see Lecture 1 and lecture 2).
The capacity of the STM is small, limiting the input to 7 +/- 2 chunks (Miller, 1956).
Note: A chunk can be any meaningful information – for a little child this might be one single letter of the alphabet – for a medical professional this might be half the page of the medical report – it depends on many factors. Miller (1956) found that the capacity of WM is limited to this certain span of so called chunks: "With binary items this span is about nine and although it decreases to five with monosyllabic English words, the difference is far less than the hypothesis of constant information would require".
Always remember that a chunk is a flexible amount of information, dependent on the previous knowledge (Simon, 1974).
Finally we have the Long-term Storage (LTM, permanent memory) is often compared as the external storage of a Von-Neumann machine.

Slide 7-7 Example: Visual Information Processing



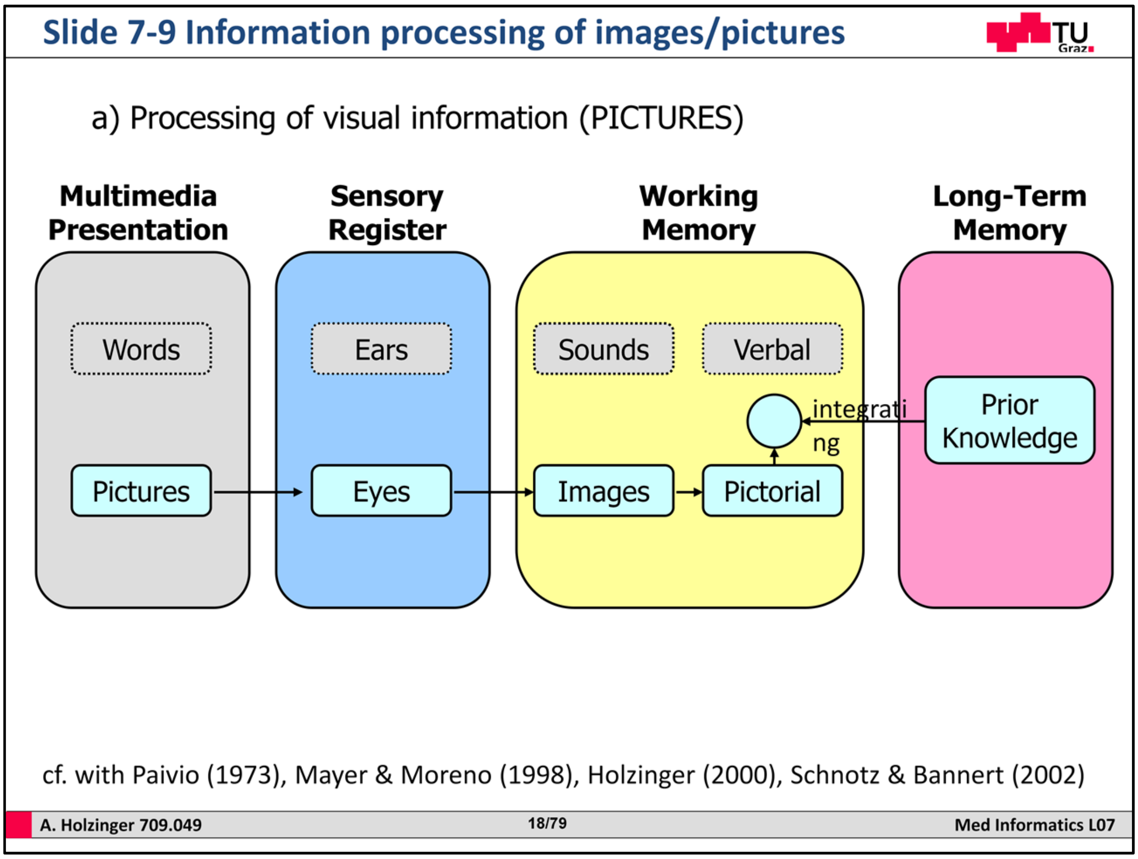
Source: Department of Neuroscience, The Mount Sinai School of Medicine (2004)

We recognize physical objects (colour, shapes, etc.), we touch a surface of this object, we listen to a speech. For the human brain, this is only an information process. It is amazing that literally our physical world is consisting of information. The perceived neural information can be measured (Yu et al., 2010) and although all brain functions require coordinated activity of many neurons, it has been difficult to estimate the amount of information carried by a population of spiking neurons. Let us look for the example of visual information processing.
The visual system is the most complex neural circuitry of all the sensory systems. The auditory nerve contains about 30,000 fibers, but the optic nerve contains over one million. Most of what we know about the functional organization of the visual system is derived from experiments similar to those used to investigate the somatic sensory system. The similarities of these systems allow us to identify general principles governing the transformation of sensory information in the brain as well as the organization and functioning of the cerebral cortex - processing of visual information is still poorly understood (state of 2013). In the slide you can see a contemporary principal sketch of the visual system in the brain. A good primer to understand these processes is still the special issue of Scientific American from September 1979 – still valid!



Slide 7-8 Schematic Information Processing Chain

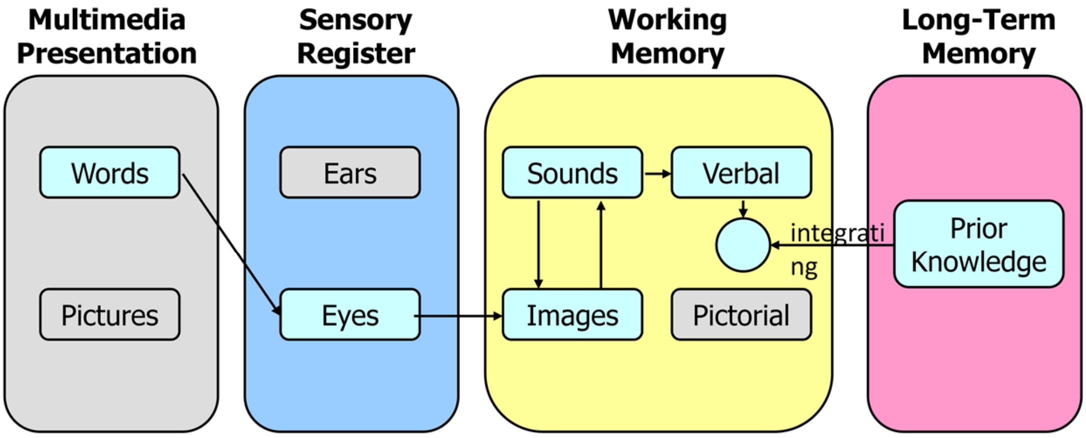
A simplified schematic process chain consists of the physical world (grey area), the sensory registers (ears, eyes, blue area), the working memory (verbal and pictorial, yellow area) and the long term memory (rose area). This will help us to quickly understand the basic processes of visual and verbal information processing (see next slides) (Holzinger, 2002).



In this slide we see the visual information processing chain. The perception of images is similar to the perception of printed words, although within the working memory we have to determine between pictorial/visual information processing and verbal/visual information processing (see next slide).

Slide 7-10 Information processing of words/pictures

b) Processing of visual information (PRINTED WORDS)



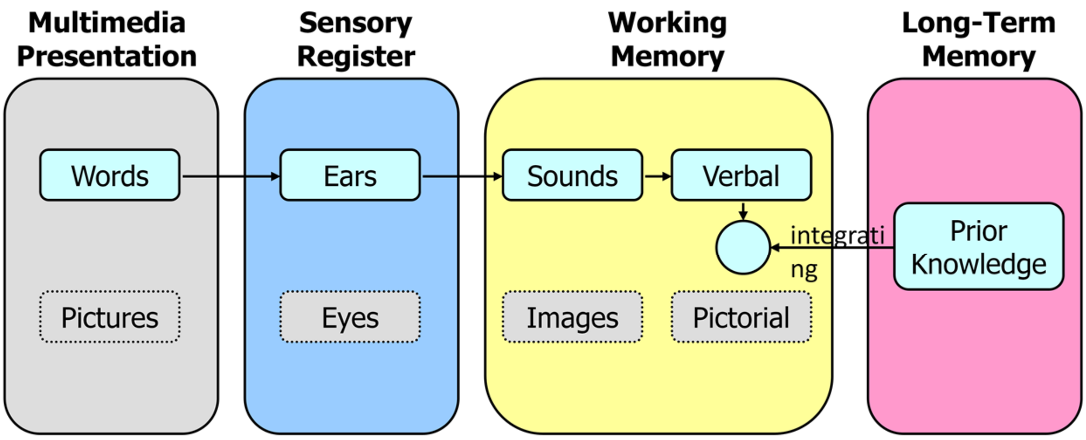
cf. with Paivio (1973), Mayer & Moreno (1998), Holzinger (2000), Schnotz & Bannert (2002)

A. Holzinger 709.049 19/79 Med Informatics L07

Here the processing of printed words, perceived as images but interpreted, processed as verbal information.

Slide 7-11 Information processing of words/sounds

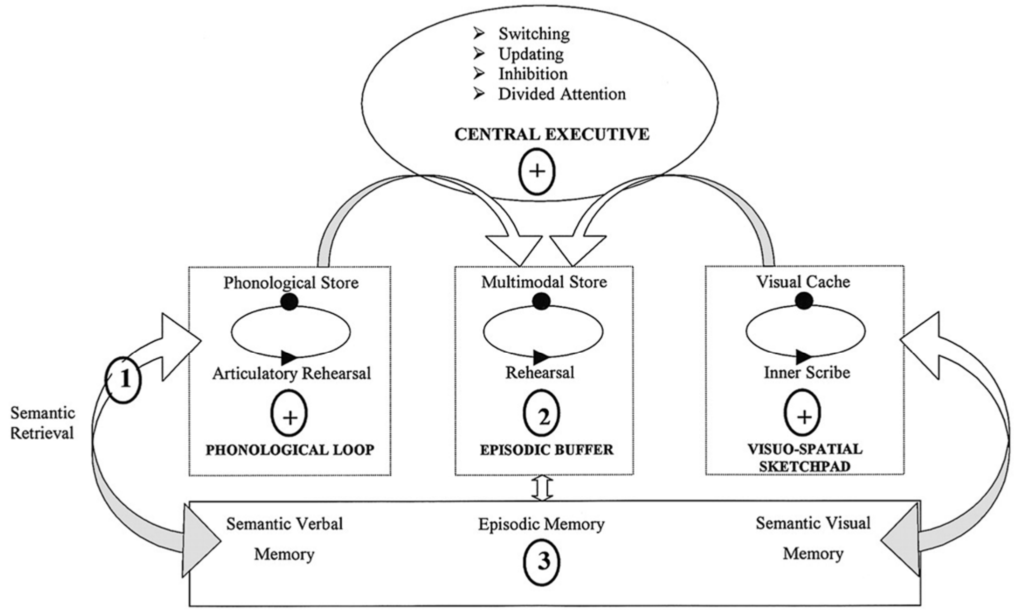
c) Processing of audio information (SPOKEN WORDS)



cf. with Paivio (1973), Mayer & Moreno (1998), Holzinger (2000), Schnotz & Bannert (2002)

Finally, the processing of spoken words, perceived as sounds via the ears and the auditory sensor register and processed as verbal information.

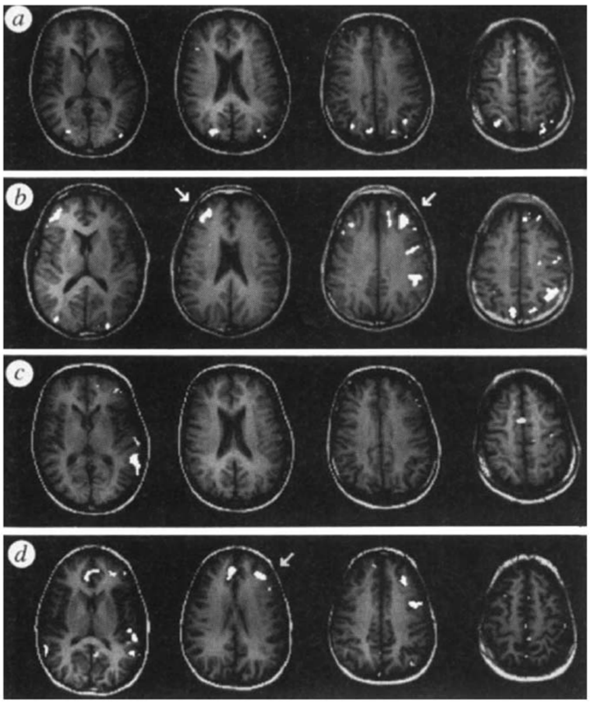
Slide 7-12 Alternative Model: Baddeley - Central Executive



Quinette, P., Guillery, B., Desgranges, B., de la Sayette, V., Viader, F. & Eustache, F. (2003) Working memory and executive functions in transient global amnesia. *Brain, 126, 9, 1917-1934.*

The model by Atkinson & Shiffrin which we have seen in Slide 7-5 and Slide 7-6 can explain much, but not everything, some aspects remain unexplained. For example, working memory (the STS) refers to a whole system for the temporary storage and manipulation of information. This is function critical for a wide range of cognitive operations. Consequently, it has been proposed that the working memory includes a central executive system (CES) to control attention and information flow to and from verbal and spatial short-term memory buffers – as can be seen in this slide – a image from (Quinette et al., 2003):
This is a schema originally going back to the concept of "central executive" by (Baddeley, 1981): A central executive system (CES) has direct influence on all information processes.

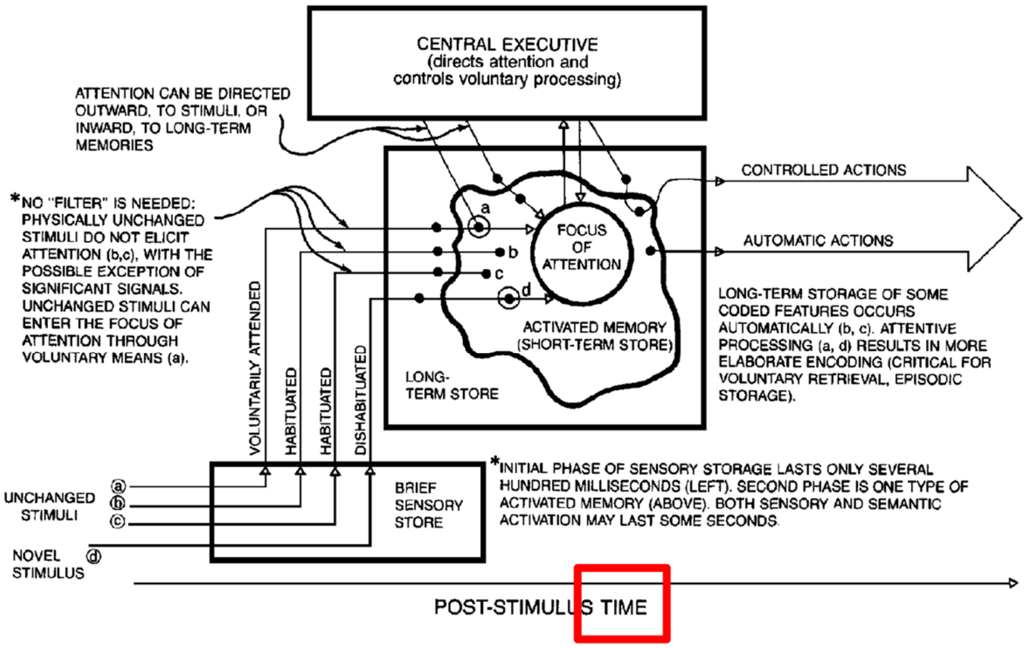
Slide 7-13 Neural Basis for the "Central Executive System"



D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S. & Grossman, M. (1995) The neural basis of the central executive system of working memory. *Nature, 378, 6554, 279-281.*

D'Esposito et al. (1995) have used functional magnetic resonance imaging (fMRI) to examine such brain activations during the concurrent performance of two tasks, which is expected to engage the CES. Activation of the prefrontal cortex was observed when both tasks are performed together, but not when they are performed separately. These results support the view that the prefrontal cortex is involved in human working memory and such a central executive is involved.
In this slide regions significantly activated in the single- and dual-task conditions in individual subjects are shown in a-d. Single-task activations are from the same subject; dual-task activations are from two different subjects. Comparison of the single-task versus baseline conditions show activation in the bilateral parieto-occipital regions for the dot-location task (a) and the left temporal region for the semantic-judgement task (c). Regions significantly activated during the dual-task versus singletask conditions are shown in band d. In the first comparison (dual task minus semantic taSk), activation is seen in prefrontal cortex, anterior cingulate and premotor cortex (b). In addition, activation is seen in bilateral parieto-occipital regions, which represents activation during the dot-location task. In the second comparison (dual task minus spatial task), activation again is seen in prefrontal cortex and anterior cingulate (d) (D'Esposito et al., 1995).

Slide 7-14 Central Executive – Selected Attention



Cowan, N. (1988) Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104, 2, 163.*

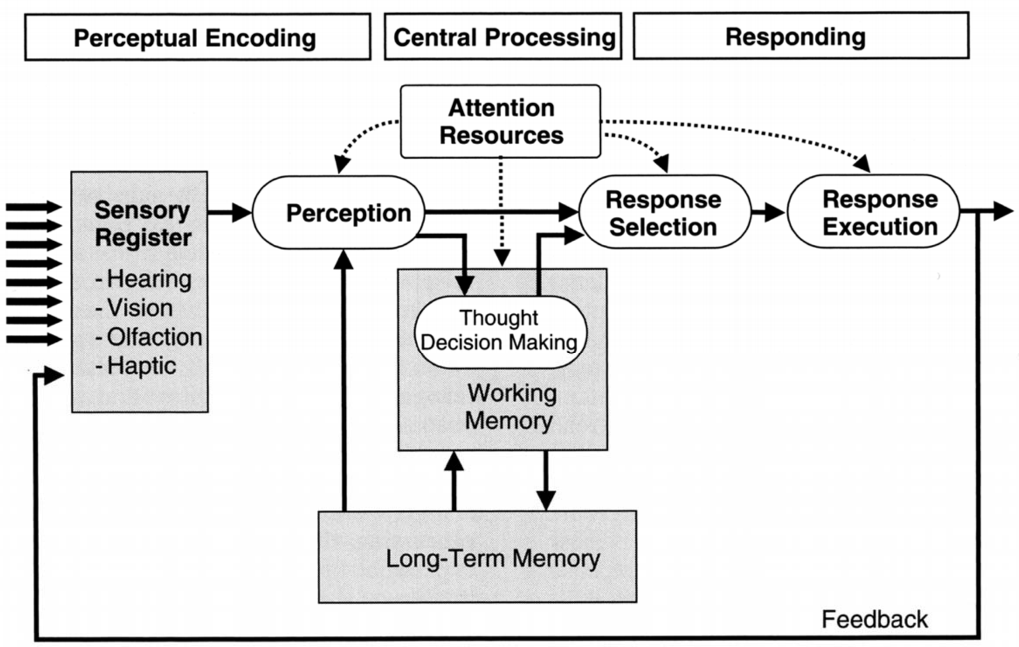
Selected Attention – going back to the Central Executive System - is a term originally coined by Broadbent already in the 1950ies where he proposed a revised model of the human information-processing system: In the Slide we see a classic sketch by (Cowan, 1988): on the x-axis we have the time since a stimulus reception is represented. The components are arranged in real time and the stimulus information can be present in more than one component at the same time. The STS is represented as an activated subset of LTS and the focus of attention is represented as a subset of the STS. The timing of involvement of the central executive in processing is flexible. The arrows represent the transfer of information from one form to another; these are discrete approximations to continuous processes that can occur in parallel or cascade. Pathways leading to awareness can come from three sources:

1) changed stimuli for which there is dishabituation;
2) items selected through effortful processing (whether of sensory origin or not), and
3) the spontaneous activation of long-term memory information based on associations (not shown).

Simons, D. J. & Chabris, C. F. 1999. Gorillas in our midst: sustained inattentional blindness for dynamic events. Perception, 28, (9), 1059-1074.

Look at Youtube for “Gorilla experiment”

Slide 7-15 Human Attention is central for problem solving

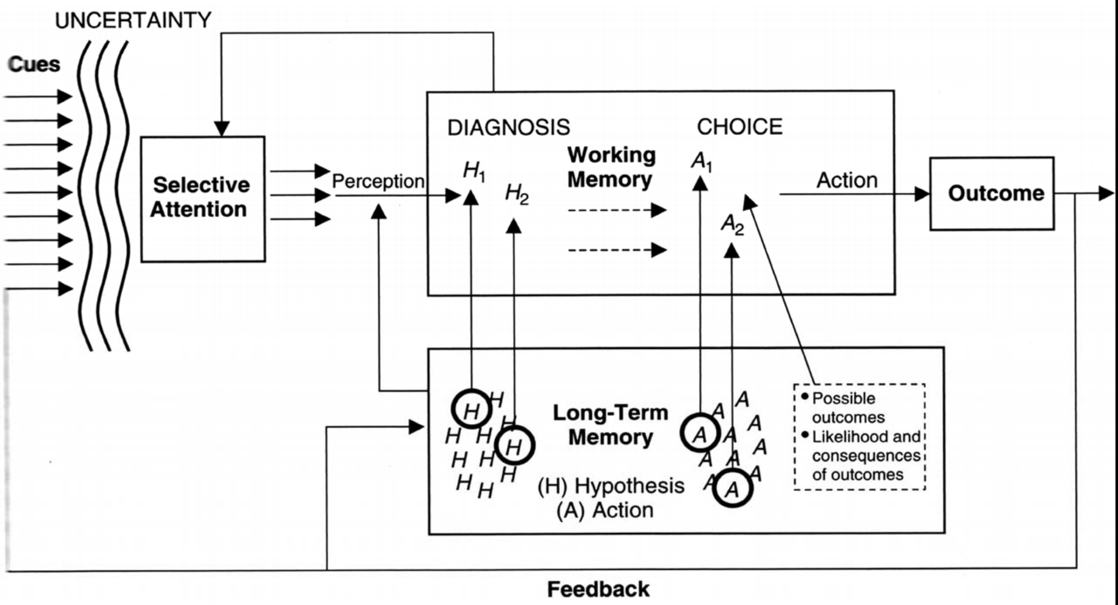


Wickens, C. D. (1984) *Engineering psychology and human performance. Columbus (OH), Charles Merrill.*

A. Holzinger 709.049 25/79 Med Informatics L07

As we can see in this slide, the information received by the receptor cells gets stored in a system of sensory registers (vision, hearing, olfaction, haptic, gustatory). Although the storage capacity would be theoretically enormous, the information is only available a very short time for further processing. Through a process of selective attention (controlled by the central executive system), only selected subsets of the vast collection of incoming information become designated for further processing (we call exactly this: perception). Here information can also become meaningful trough comparison with information from the LTS (previous knowledge, hypotheses, models of the world etc.). A large portion of our conscious effort (especially in making decisions) is dedicated to these processes in the STS - you now understand why the STS is called working memory (WM): Rehearsal within the WM enables it finally to become the information encoded into LTS – where it remains a lifetime; otherwise it decays rapidly. Moreover, the WM has severe capacity constraints, governing the amount of information kept active (Wickens, Lee, Liu & Gordon-Becker, 2004) p.713.

Slide 7-16 Human Decision Making



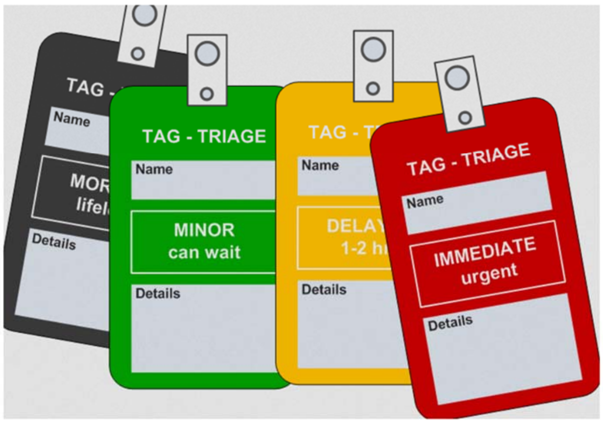
Wickens, C. D. (1984) *Engineering psychology and human performance. Columbus (OH), Charles Merrill.*

Human decision making is perception and response execution (outcome) and is an activity tight to fallibility, especially in complex and dynamic situations. If the perceived information is incomplete or fuzzy, intensive interpretation and integration of this information is required. Any hypothesis which the decision maker generates regarding the incoming information is highly dependent on information available in the LTS (=previous knowledge). Constraints often lead to shortcuts in decision making, whereby people opt for choices that are good enough for their purposes and adopt strategies for sampling information that they perceive to be most relevant to solve their current problem. Clinical reasoning (problem solving, decision making, judgement) is a central component of the physicians competence (Norman, 2005).

To help us to understand the decision making process we can refer to (Wickens, 1984): The physical stimuli (cues) are selected by the attentional resources and the perceived information builds working hypotheses H1, H2, … . Those are compared and judged against available Hypotheses, already present in the long-term memory. On this basis the best alternative will be chosen and actions A1, A2, … performed according to likelihoods and consequences of outcomes – which can be perceived again via the feedback loop.

Note: The “nebula of uncertainty” emphasizes what we have already learned in our course: we deal always with probable information! Each information chunk, item or whatever you call it, has always a certain probability aspect (thanks to Bayes ;-). The model of Wickens is still the best model to explain clinical decision making.

Start with the most simplest decision support system



Iserson, K. V. & Moskop, J. C. 2007. Triage in Medicine, Part I: Concept, History, and Types. Annals of Emergency Medicine, 49, (3), 275-281.

Slide 7-17 Example: Triage Tags - International Triage Tags

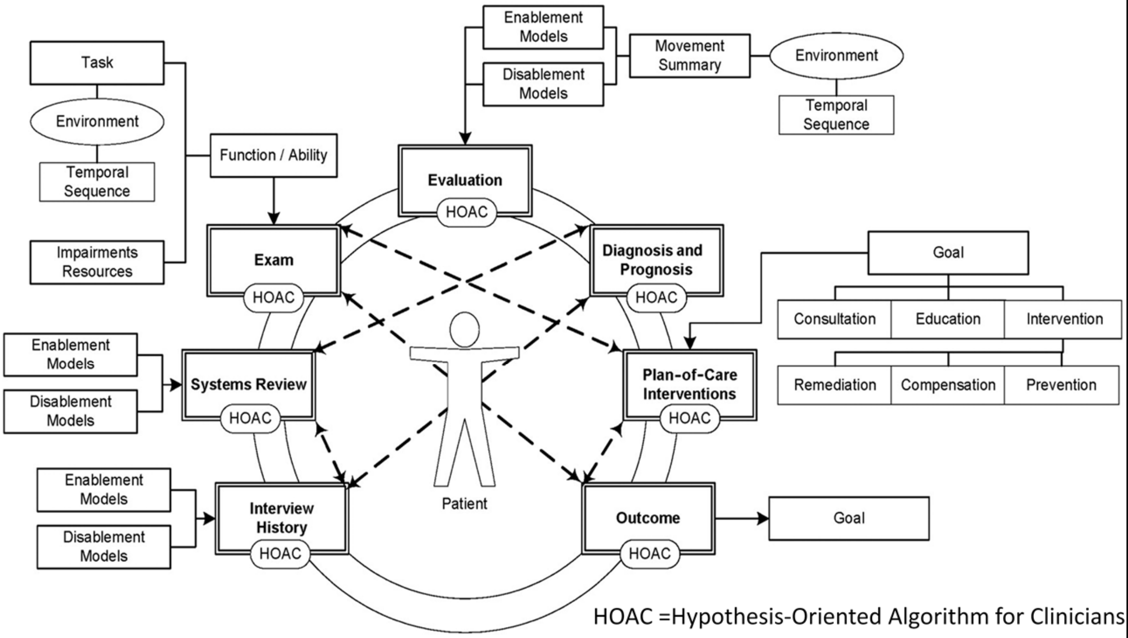
Source: http://store.gomed-tech.com

International Triage Tags
(Symbol Only Triage Tags)

The MT-137 provides first responders and rescue personnel with the standard triaging features needed to enable them to perform a prompt and accurate assessment of a Mass Casualty Incident (MCI) victim's injuries and to easily record the data.

The simplest decision support process is triage, which is the process of determining the priority of patients' treatments based on the severity of their conditions. The idea comes from times where resources are insufficient for everybody to be treated immediately – for example in mass incidents. Triage is also used in hospitals for patients arriving at the emergency department.

Slide 7-18 Clinical DM: Hypothesis-Oriented Algorithm



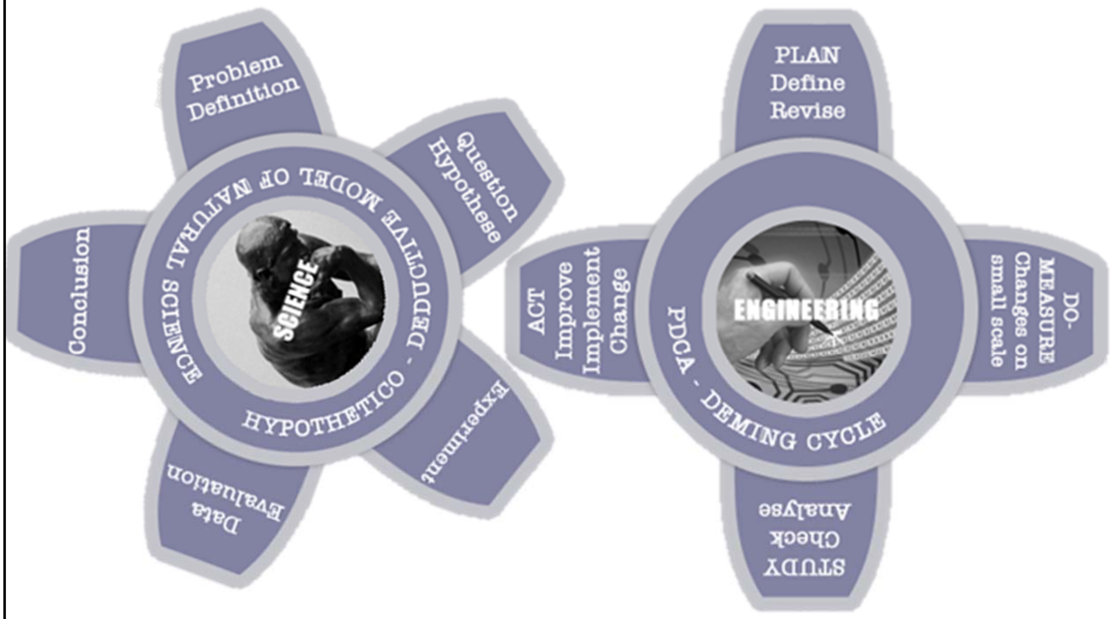
HOAC =Hypothesis-Oriented Algorithm for Clinicians

Schenkman, M., Deutsch, J. E. & Gill-Body, K. M. (2006) An Integrated Framework for Decision Making in Neurologic Physical Therapist Practice. *Physical Therapy, 86, 12, 1681-1702.*

Overview of the unifying framework demonstrating the steps in the clinical decision-making process. This framework is patient centered, integrates
both enablement and disablement perspectives, and incorporates a variety of existing conceptual models and analyses at various points in the patient
management process. HOAC =Hypothesis-Oriented Algorithm for Clinicians.

In practice the clinicians use Decision-Making Frameworks to guide patient management, communicate with other health care providers and educate patients and their families. A number of frameworks have been applied to guide clinical practice. (Schenkman, Deutsch & Gill-Body, 2006) have proposed a unifying framework for application to decision making in the management of individuals on the example of neurologic dysfunction. The framework integrates both enablement and disablement perspectives (Schenkman, Deutsch & Gill-Body, 2006):
The process in Slide 7-18 is patient-centered, which is in contrast to a pathology-driven approach in which the process is disease-centered and culminates in disablement. The patient-centered approach emphasizes roles and functions of which the individual is capable and at the same time identifies limitations in the individual's abilities, with the goal of minimizing barriers to full participation within society or the environment. The purpose of the history and interview is to gain an understanding of the patient as an individual, determine why the individual seeks physical therapy, identify what he or she hopes to achieve through physical therapy, and begin to formulate the examination strategy. The purpose of the systems review is to rule out those body systems with which the physical therapist need not be concerned, identify systems that are resources for the patient, guide choices regarding which aspects of the remaining systems to examine in detail, and determine whether the physical therapist should proceed or should refer the patient to other health care providers. The specific purposes of the examination may vary and depend on the reason for which it is carried out. The level of examination is adjusted to reflect patient-identified problems and goals and can be drawn from elements of the continuum of the ICF model. The evaluation consists of an interpretation of findings in order to develop a realistic plan of care. The plan of care is based on a synthesis of the information from all of the previous steps, including the patient's goals and expectations, task performance, the patient's resources and impairments, and the medical diagnosis and prognosis for the condition. The plan of care is organized around the patient's goals. Goal-directed therapy has been identified as improving motor function and promoting cortical reorganization. HOAC = Hypothesis Oriented Algorithm for Clinicians.

Slide 7-19 Hypothetico-Deductive vs PDCA Deming wheel

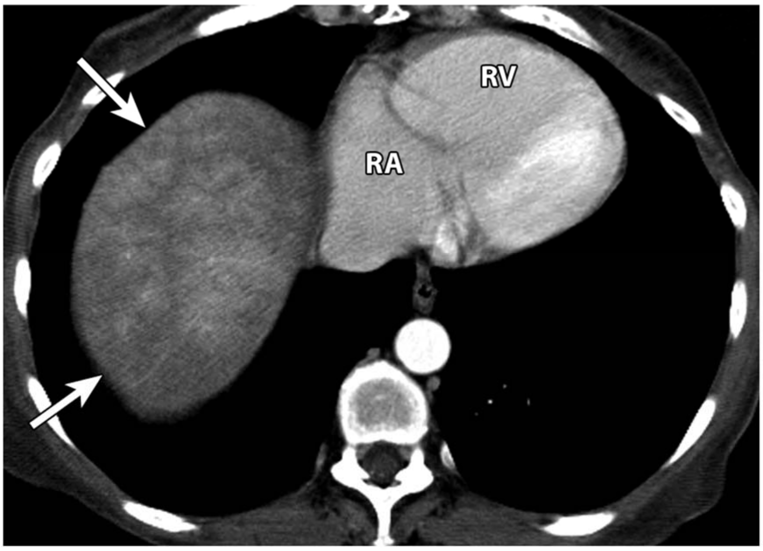


Holzinger, A. (2010) *Process Guide for Students for Interdisciplinary Work in Computer Science/Informatics. Second Edition. Norderstedt, BoD.* Online: http://www.hci4all.at

The HDM (Newton – Scientific method, in the slide on the right) assumes that properly formed theories arise as generalizations from observable data that they are intended to explain. These hypotheses, however, cannot be conclusively established until the consequences, which logically follow from them, are verified through additional observations and experiments. In conformity with the rationalism of Descartes, the HDM treats theory as a deductive system, in which particular empirical phenomena are explained by relating them back to general principles and definitions. This method abandons the Cartesian claim that those principles and definitions are self-evident and valid; it assumes that their validity is determined only by their consequences on previously unexplained phenomena or on actual scientific problems. Scientific methods include techniques for investigating phenomena with the aim of acquiring new knowledge, or correcting and integrating previous knowledge, i.e. going from state of the art to beyond state of the art (Holzinger, 2010).
Hypothesis. A working hypothesis (Arbeitshypothese) develops from the formulation of a question, e.g.: "The learning success is greater with the use of the simulation X than with the traditional method Y". Again, using exact definitions of the terms used. According to the circumstances, the working hypothesis must be continually modified (see section 9), until it represents, exactly what one wants to examine, e.g.: "by the employment of the simulation X, learning time for the topic Y can be significantly shortened".
Here, a very careful clarification is necessary: Which variables are to be examined and which expectations exist (again with reference to the theory and/or to preceding work).
Formal scientific hypotheses (Popper, 2005) can be brought formally into the form: When X then Y; X implies Y; X Y
It is important to understand that hypotheses are only conditional statements: The effect occurs, not under all possible circumstances, but only under completely pre-determined circumstances. The hypothesis must be evaluated experimentally. Whereby there are exactly three possibilities:
1) TRUE. The hypothesis is significantly proved (verified = is true)
2) FALSE. The hypothesis is significantly disproved (falsified = is wrong)
3) UNDECIDED. The hypothesis can (on the basis of the existing data) neither be verified nor falsified (no statement can be made).
Note: Whether a hypothesis is true (thus ALWAYS applies), can not be confirmed due to the universality of the statement. Scientific hypotheses can therefore only be falsified. Nevertheless hypotheses can be differentiated with respect to the degree of their confirmation, whereby certain criteria is taken into account, such as how frequently and in which critical places the hypothesis has already been confirmed (Popper, 2005).
PDCA: This concept was developed by (Shewhart, 1958) as PDSA cycle The roots can be tracked back to Aristotle (384–322 BC) and Francis Bacon (1561–1626). The PDSA cycle consists of four steps:
1) PLAN: Study the process;
2) DO: Make changes on a small scale;
3) STUDY: Observe the effects and
4) ACT: Identify what you can learn from your observation.

Example 1/4: carcinoid heart disease (chd)

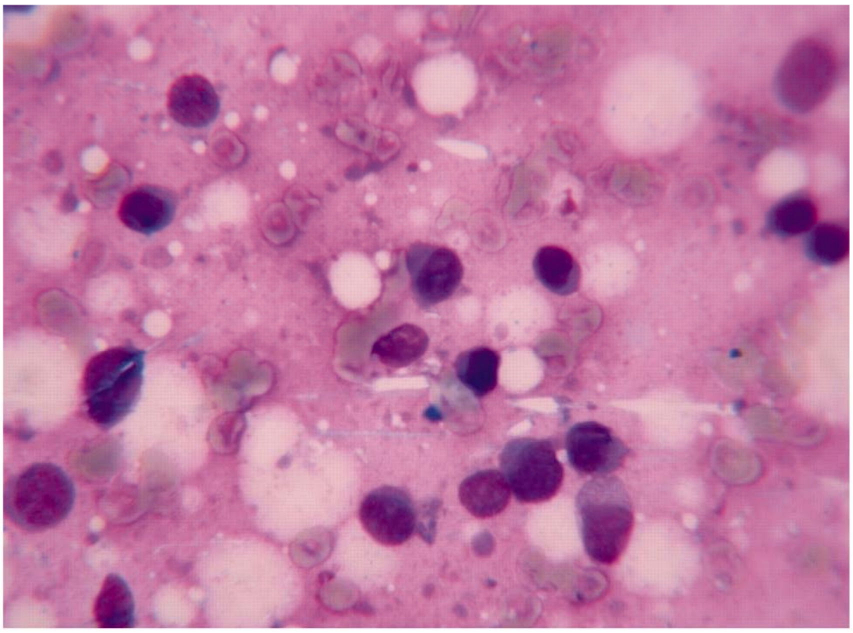
Hepatic venous congestion and carcinoid heart disease secondary to an ovarian carcinoid tumor in a 56-year-old woman with elevated liver enzyme levels and right upper quadrant pain.



Shanbhogue, A. K. P., Shanbhogue, D. K. P., Prasad, S. R., Surabhi, V. R., Fasih, N. & Menias, C. O. (2010) Clinical Syndromes Associated with Ovarian Neoplasms: A Review. *Radiographics, 30, 4, 903-919.*

Hepatic venous congestion and carcinoid heart disease secondary to an ovarian carcinoid tumor in a 56-year-old woman with elevated liver enzyme levels and right upper quadrant pain. Axial contrast-enhanced CT images of the upper abdomen (b at a lower level than a) show diffuse heterogeneous enhancement ("mosaic" attenuation) of the liver (arrows in a), a finding suggestive of hepatic venous congestion. Here the dilatation of the right atrium (RA) and right ventricle (RV) is seen.

Example 2/4: bone-marrow depression (bmd)



Prasad, M., Maitra, A., Sethiya, N., Bharadwaj, V. K., Chowdhury, V., Valecha, J. & Biswas, R. (2009) Acute renal failure followed by low back ache. *BMJ Case Reports, 2009.*

The CT of the spine taken for low back ache and of the cranium taken for altered sensorium revealed multiple punched out osteolytic lesions (evident on figs 1 and 2). This was immediately suggestive of myeloma-like plasma cell dyscrasia.

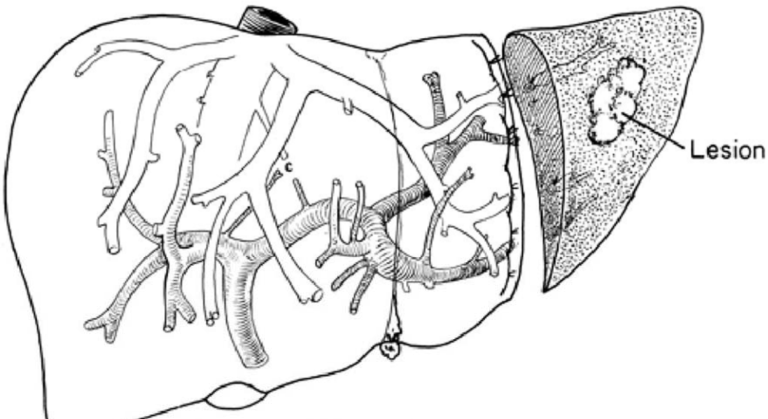
The bone marrow showed a severely depressed erythroid element, and the myeloid elements replaced with lymphoid cells showed reactive lymphocytes with moderate basophilic cytoplasm suggestive of a lymphoplasmacytic monoclonal proliferation.

These bone marrow findings are suggestive of Waldenstrom's macroglobulinaemia (WM). A serum protein agarose gel electrophoresis revealed an M spike in the γ globulin region.

WM is less common than multiple myeloma, and must be differentiated on predominantly clinical grounds from other entities presenting with monoclonal spikes of IgM. A diagnosis of WM can be made irrespective of IgM concentration if there is evidence of bone marrow infiltration by lymphoplasmacytoid lymphoma.1 Lytic bone disease is very uncommon in WM but well described in the literature.2 Although renal complications of plasma cell dyscrasias with IgM proliferations such as WM are rare, a wide spectrum of kidney lesions has been reported in patients, many of whom have presented with acute renal failure.3 The second hospital admission of our patient was with altered sensorium which was thought to be due to uraemic encephalopathy (along with twitching of the limbs that could again be a part of the uraemic encephalopathy).

Our patient received one cycle of chemotherapy with standard alkylating agents and steroids, following which he developed a severe bout of diarrhoea and fever. He was treated for the gastrointestinal sepsis but he could not recover and succumbed to the infection 4 weeks after.

Example 3/4: partial liver resection (plr)



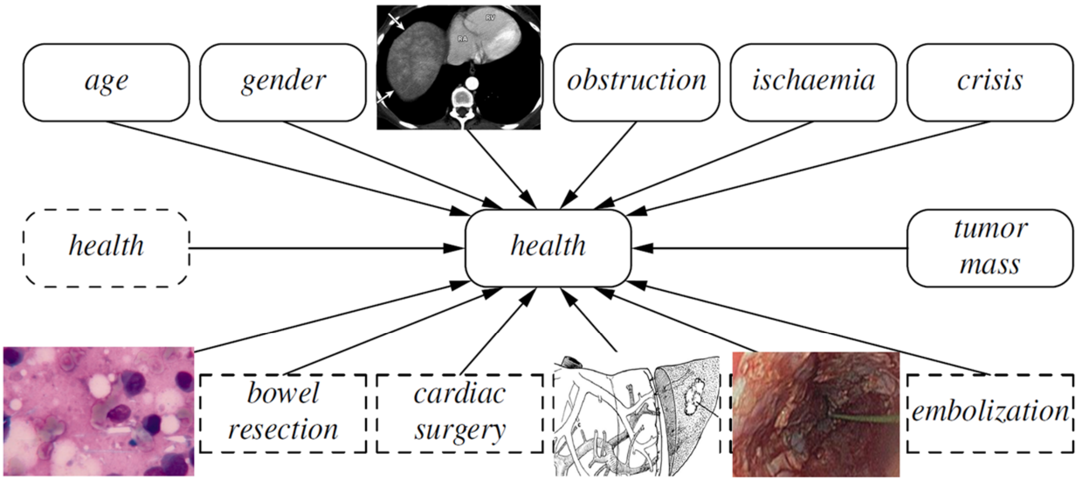
Zollinger, R. M. & Ellison, C. (2010) *Zollinger's Atlas of Surgical Operations (9th Edition). New York, McGraw Hill.*

Example 4/4: radiofrequency ablation (rfa)

van Vilsteren, F. G. I. et al. (2011) Stepwise radical endoscopic resection versus radiofrequency ablation for Barrett's oesophagus with high-grade dysplasia or early cancer: a multicentre randomised trial. *GUT.*

Ablation = Abtragung, Amputation, Entfernung – surgical removal

Slide 7-20 Modeling Patient health (1/2)



chd = carcinoid heart disease; bmd = bone-marrow depression; plr = partial liver resection;
rfa = radiofrequency ablation; dashed … past states; square objects … treatments

van Gerven, M. A. J., Taal, B. G. & Lucas, P. J. F. (2008) Dynamic Bayesian networks as prognostic models for clinical patient management. *Journal of Biomedical Informatics, 41, 4, 515-529.*

We have discussed modeling already in →Lecture 3. Now we see how we can apply this knowledge for decision support: For example, in oncology, one way to represent the patient's health status is in terms of the performance status, which is distinguished into normal, mild complaints, ambulatory, nursing care, intensive care, and death. In this Slide we see how we can model the various influences on carcinoid patient health and they can be categorized into the described factors: The variables age, gender, and (past) health determine the patient health independent of the disease. The variable tumor mass represents the direct effect of the tumor on patient health. The dashed objects denote past states and square objects denote treatments (van Gerven, Taal & Lucas, 2008).
chd = carcinoid heart disease; bmd = bone-marrow depression; plr = partial liver resection; rfa = radiofrequency ablation;

Slide 7-21 Modeling Patient Health (2/2)

Let $\mathbf{U} \subseteq \mathbf{X}$ denote this risk factors and
Let $\mathbf{V} = \mathbf{X} \setminus \mathbf{U}$ denote the complement.

The risk of immediate death $p(health(t) = death|X)$ can be expressed by calculation of the following product:

$$\prod_{U \in \mathbf{U}} p(U)$$

Further, we obtain

$$p(health(t) = h|\boldsymbol{X}) = p(h|\boldsymbol{V}) \prod_{U \in \boldsymbol{U}} p(health(t) \neq death|U, health(t-1))$$

for $h \neq death$

van Gerven, M. A. J., Taal, B. G. & Lucas, P. J. F. (2008) Dynamic Bayesian networks as prognostic models for clinical patient management. *Journal of Biomedical Informatics, 41, 4, 515-529.*

U is a subset of or equal to X
V not X or not U denote the complement
Probability p = death is the product of all risk factors
and the complement is the probability health status
P(A|B) means the probability of the event A occurring - given that B occurs.
The vertical dash | means "such that"

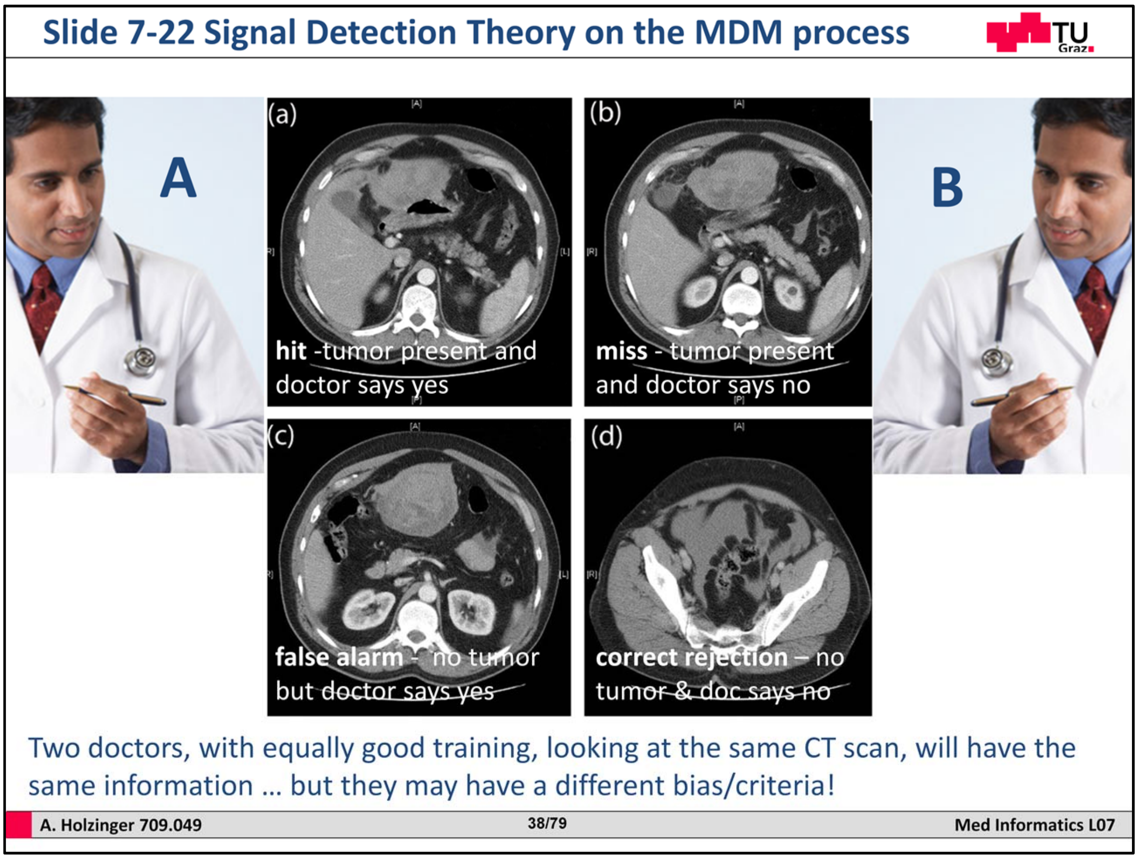
The large number of conditioning variables that (partially) determine patient health, makes estimation of conditional probabilities for this variable very difficult. However, a large subset of these variables are risk factors that influence health only due to the fact that they may cause immediate patient death, thereby simplifying the specification as follows.
Let $\mathbf{U} \subseteq \mathbf{X}$ denote this risk factors and
Let $\mathbf{V} = \mathbf{X} \setminus \mathbf{U}$ denote the complement.
The risk of immediate death $p(health(t) = death|X)$ can be expressed as:

$$1 - p(health(t) \neq death|\boldsymbol{V}) \prod_{U \in \boldsymbol{U}} p(health(t) \neq death|U, health(t-1))$$

Further, we obtain:

$$p(health(t) = h|\boldsymbol{X}) = p(h|\boldsymbol{V}) \prod_{U \in \boldsymbol{U}} p(health(t) \neq death|U, health(t-1))$$

for $h \neq death$
These simplifications greatly reduce the number of parameters that need to be estimated (van Gerven, Taal & Lucas, 2008).
To understand these processes we must have a look on some basics of human decision making, and we start with one of the oldest approaches: signal detection theory.



Look at Slide 7-22: You see two doctors (doctor A and doctor B), with equally good training, looking at the same CT scan data, so both will have the same information … but they may gain different knowledge out of the data due to many intervening factors, bias, criteria.

Interpreting CT data is difficult and it takes a lot of training and there is always some uncertainty as to what is correct or not. In principal we have four possibilities: Either there is a tumor (signal present) or there is not (signal absent). Either the doctor sees a tumor (they respond "yes'') or does not (they respond "no'').

There are four possible outcomes:

(a) hit - tumor is present and doctor says "yes''

(b) miss - tumor is present and doctor says "no''

(c) false alarm - tumor is absent and doctor says "yes"

and

(d) correct rejection – tumor is absent and doctor says "no".

Let us look more closely on the information process.

Slide 7-23 Information Acquisition and criteria - bias

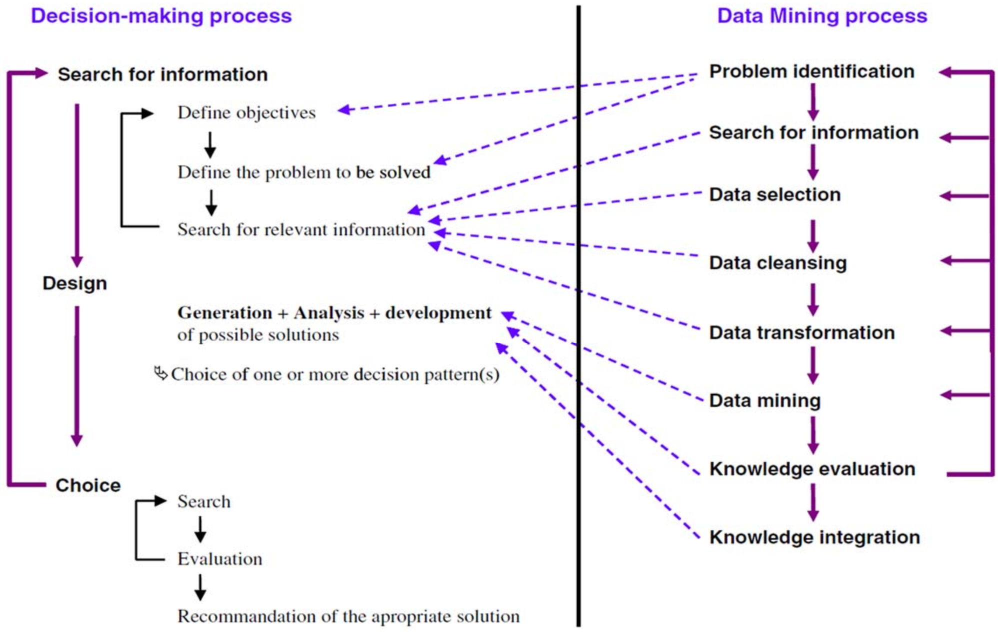
- **Information acquisition:** in the CT data, e.g. healthy lungs have a characteristic shape; the presence of a tumor might distort that shape (= anomaly).
- Tumors have different image characteristics: brighter or darker, different texture, etc.
- With proper training a doctor learns what kinds of things to look for, so with more practice/training they will be able to acquire more (and more reliable) information.
- Running another test (e.g., MRI) can be used to acquire more (relevant!) information.
- The effect of information is to increase the likelihood of getting either a hit or a correct rejection, while reducing the likelihood of an outcome in the two error boxes (slide 33).
- **Criterion:** Additionally to relying on technology/testing, the medical profession allows doctors to use their own judgment.
- Different doctors may feel that the different types of errors are not equal.
- For example, a doctor may feel that missing an opportunity for early diagnosis may mean the difference between life and death.
- A false alarm, on the other hand, may result only in a routine biopsy operation. They may chose to err toward ``yes'' (tumor present) decisions.
- Other doctors, however, may feel that unnecessary surgeries (even routine ones) are very bad (expensive, stress, etc.).
- They may chose to be more conservative and say ``no'' (no turmor) more often. They will miss more tumors, but they will be doing their part to reduce unnecessary surgeries. And they may feel that a tumor, if there really is one, will be picked up at the next check-up.

Mohamed, A. et al. (2010) Traumatic rupture of a gastrointestinal stromal tumour with intraperitoneal bleeding and haematoma formation. *BMJ Case Reports, 2010.*

Information acquisition: First, there is information in the CT data, e.g. healthy lungs have a characteristic shape; the presence of a tumor might distort that shape. Tumors may have different image characteristics: brighter or darker, different texture, etc. With proper training a doctor learns what kinds of things to look for, so with more practice/training they will be able to acquire more (and more reliable) information. Running another test (e.g., MRI) can also be used to acquire more information. Regardless, acquiring more information is good. The effect of information is to increase the likelihood of getting either a hit or a correct rejection, while reducing the likelihood of an outcome in the two error boxes. Criterion: The second component of the decision process is different: Additionally to technical information acquisition, medical doctors use their own judgment. Different doctors may feel that the different types of errors are not equal. One doctor may feel that missing an opportunity for early diagnosis may mean the difference between life and death; on the other hand a false alarm may result only in a routine biopsy operation. They may choose to err toward yes (tumor present) decisions. Other doctors may feel that unnecessary surgeries are very bad (expensive, stress, etc.). They may choose to be conservative and say no (no turmor) more often. They will miss more tumors, but they will be doing their part to reduce unnecessary surgeries. And they may feel that a tumor, if there really is one, will be picked up at the next check-up. Two doctors, with equal training, looking at the same CT, have the same information; but having a different bias/criteria.

For differences DSS vs. DM see the next slide.

Slide 7-24: Decision Making Process vs. Data Mining process

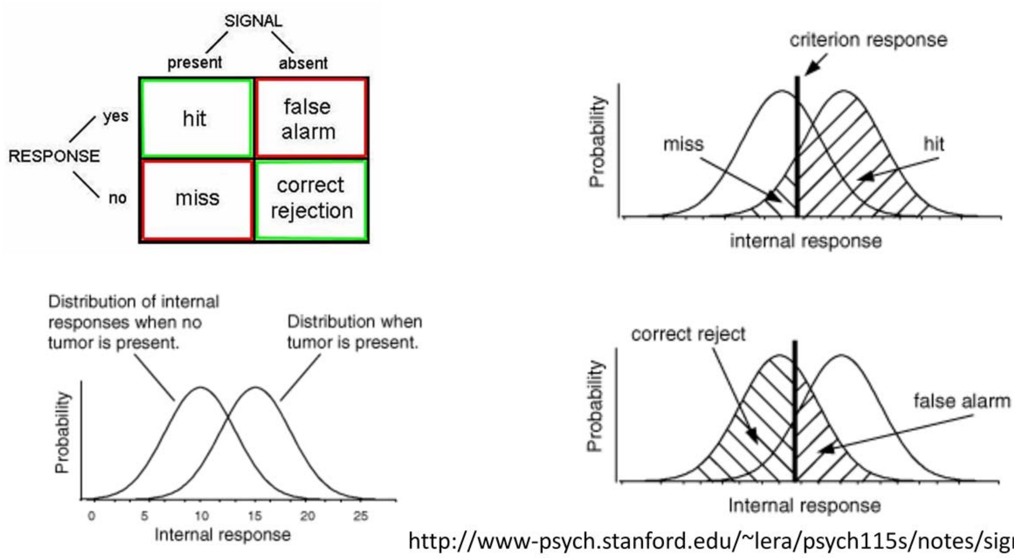


Ayed, B. M., Ltifi, H., Kolski, C. & Alimi, A. (2010) A user-centered approach for the design & implementation of KDD-based DSS: A case study in the healthcare domain. *Decision Support Systems, 50, 64-78.*

The decision making process described in Slide 7-23 and visualized in this slide can be compared to the data mining process (right in this Slide) (Ayed et al., 2010).

Slide 7-25 Decision Making Process - Signal Detection

Remember: Two doctors, with equally good training, looking at the same CT scan data, will have the same information … but they may gain different knowledge due to *bias/criteria*.

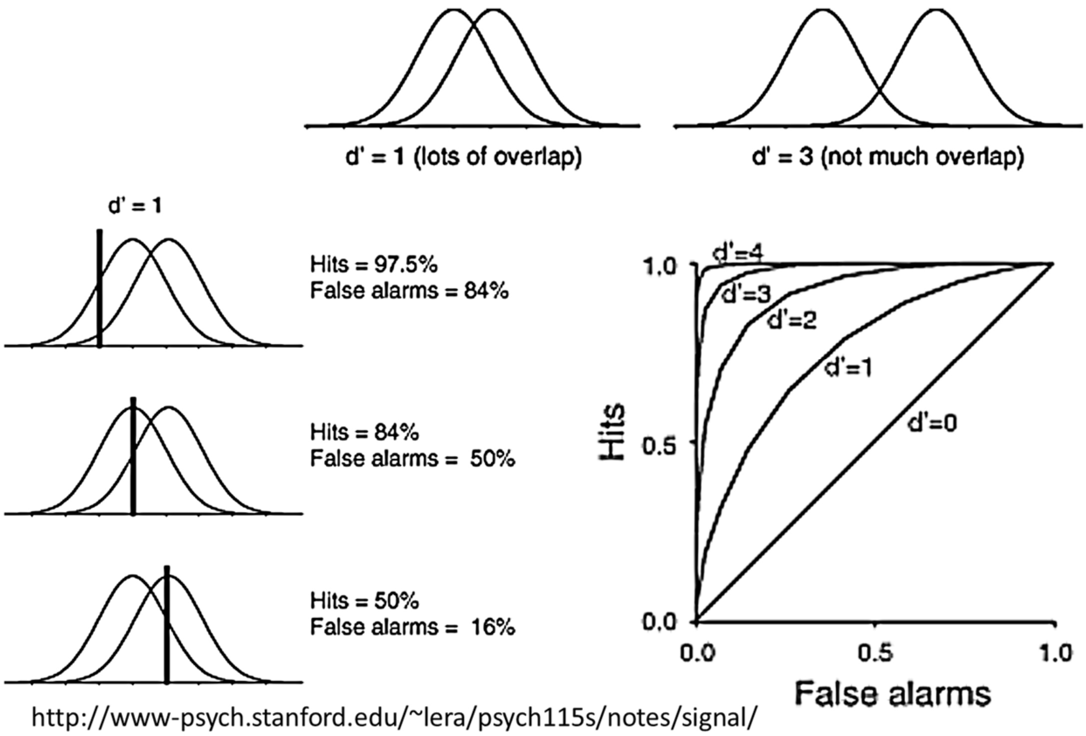


http://www-psych.stanford.edu/~lera/psych115s/notes/signal

For an example see: Braga & Oliveira (2003) Diagnostic analysis based on ROC curves: theory and applications in medicine. *Int. Journal of Health Care Quality Assurance, 16, 4, 191-198.*

Remember our example from slide 7-22: Two doctors, with equally good training, looking at the same CT scan data, will have the same information … but they may gain different knowledge due to bias/criteria.
Noise: Detecting a tumor is very difficult and there will always be some amount of uncertainty (as generally in the medical area, always remember that we speak of probable information). There are two kinds of noise factors that contribute to this uncertainty: internal noise and external noise.
There are many possible sources of external noise. There can be noise factors that are part of the imaging process, a smudge, or bad spots. While the doctor makes every effort possible to reduce the external noise, there is little or nothing that they can do to reduce internal noise. Internal noise refers to the fact that neural responses are “noisy” per se. Look at the sketches in this slide – based on the classical Signal Detection Theory (SDT):
Note: The starting point for signal detection theory was that all reasoning and decision making takes place in the presence of uncertainty. Signal detection theory provides a precise language and graphic notation for analyzing decision making under uncertainty.
SDT was developed by (Peterson, Birdsall & Fox, 1954) and applied to psychophysics by (Tanner Jr & Swets, 1954) and is a very generic and widely used method for drawing inferences from data. The data needed for an SDT analysis are just the counts of hits, false alarms, misses, and correct rejections.

Slide 7-26 Receiver Operating Characteristics (ROC curve)



http://www-psych.stanford.edu/~lera/psych115s/notes/signal/

D = 0 = random guessing
Within our course we have seen already examples for ROC curves (for example in →Slide 6-39, where we compared the performance of SVM with ANN).
In signal detection theory (SDT), a receiver operating characteristic (ROC) is the curve which illustrates the performance of a binary classifier system as its discrimination threshold is varied. It is created by plotting the fraction of true positives out of the total actual positives (TPR = true positive rate) vs. the fraction of false positives out of the total actual negatives (FPR = false positive rate), at various threshold settings. TPR is also known as sensitivity (also called recall in some fields), and FPR is one minus the specificity or true negative rate.
In classification tasks the more the curve is in the upper left corner, the better.

Slide 7-27b Bayes Law of Total Probability = data modelling

Bayes' Rule in words
d … data; h … hypothesis

$$P(h|d) = \frac{P(d|h)P(h)}{P(d)}$$

$P(h)$: prior belief (probability of hypothesis h before seeing any data)

$P(d \mid h)$: likelihood (probability of the data if the hypothesis h is true)

$P(d) = \sum_h P(d \mid h)P(h)$: data evidence (marginal probability of the data)

$P(h \mid d)$: posterior (probability of hypothesis h after having seen the data d)

$$posterior = \frac{likelyhood * prior}{evidence}$$

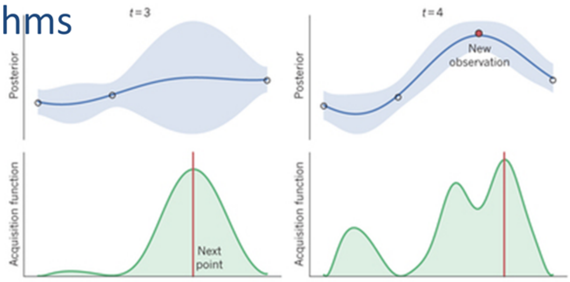
The inverse probability allows to infer unknowns, learn from data and make predictions = machine learning!

A. Holzinger 709.049 44/79 Med Informatics L07

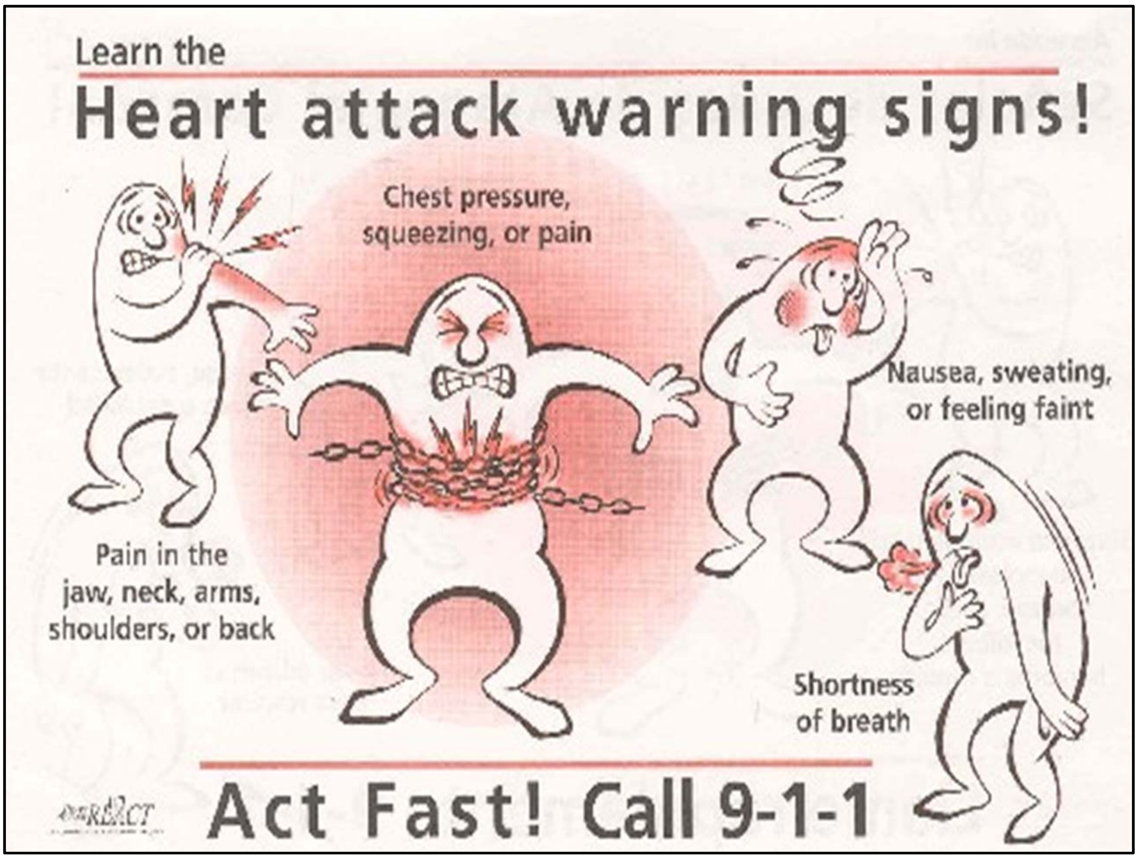
A model describes data that one observes – the inverse probability allows to infer unknowns, learn fron
Goal: To determine the most probable hypothesis, given the data D plus any initial knowledge about th
Prior probability of h, P(h): it reflects any background knowledge we have about the chance that h is a
Prior probability of D, P(D): it reflects the probability that training data D will be observed given no kr
Conditional Probability of observation D, P(D|h): it denotes the probability of observing data D given s

Slide 7-27c Representation of uncertainty

- Many aspects of intelligence and learning depend on **probabilistic representation of uncertainty:**
- Forecasting
- Decision support
- Learning from noisy, missing, uncertain data …
- Knowledge discovery
- Probabilistic programming (e.g. Stochastic Python, Julia)
- Universal inference algorithms
- Global optimization



Ghahramani, Z. 2015. Probabilistic machine learning and artificial intelligence. Nature, 521, (7553), 452-459.


Learn the
Heart attack warning signs!
Chest pressure, squeezing, or pain
Pain in the jaw, neck, arms, shoulders, or back
Nausea, sweating, or feeling faint
Shortness of breath
Act Fast! Call 9-1-1

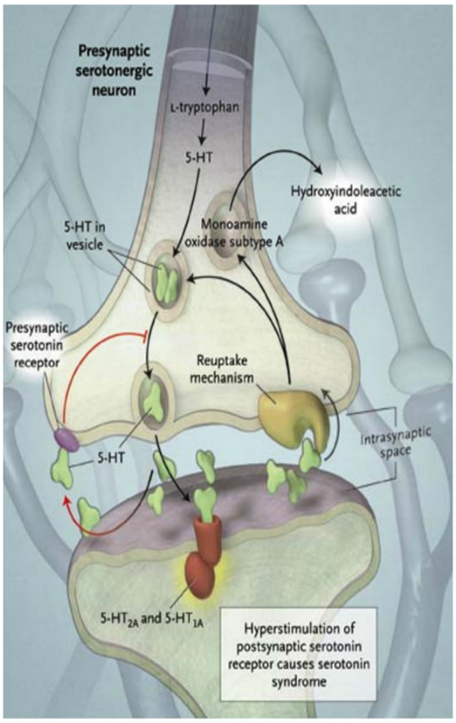
Slide 7-27d Bayes' in clinical practice

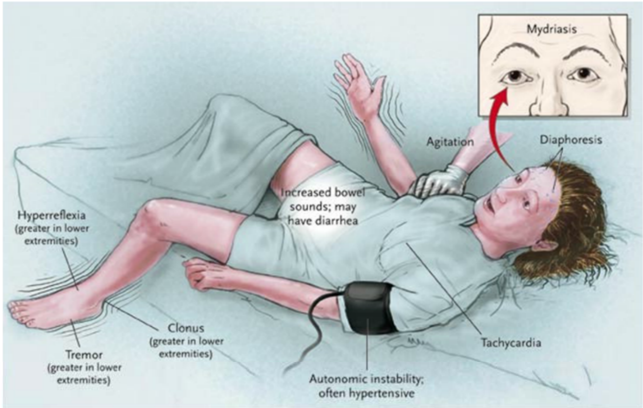
- Clinical Example:
- D ... acute heart attack
- U_+... instable chest pain
- $p(D)$ ... 37 of 1000 = 0,037 (heart attack)
- $p(\overline{D})$ ... 963 of 1000 = 0,963 (no heart attack)
- 40% of patients report on instable chest pain
- $p(U_+|D)$ = 0,4
- Unfortunately this symptoms also occur in 5 % of the healthy population
- $p(U_+|\overline{D})$ = 0,05
- We find the probability for a heart attack during this symptoms therefore by using Bayes' Rule:
- $p(D|U_+) = \frac{p(U_+|D)*p(D)}{p(U_+|D)*p(D)+p(U_+|\overline{D})*p(\overline{D})}$ = 0,235

A. Holzinger 709.049 47/79 Med Informatics L07

Let us take a classic clinical example:
D ... acute heart attack
U+... instable chest pain
p(D) ... 37 of 1000 = 0,037 (heart attack)
p(D̅) ... 963 of 1000 = 0,963 (no heart attack)
40% of patients report on instable chest pain
p(U_+ |D┤) = 0,4
Unfortunately this symptoms also occur in 5 % of the healthy population
p(U_+ |D̅)┤ = 0,05
We find the probability for a heart attack during this symptoms therefore by using Bayes' Rule:
p(D│U_+)= (p(U_+ ├|D┤)*p(D))/(p(U_+ ├|D┤)*p(D)+p(U_+ ├|D̅)┤ *p(D̅)) = 0,235
P(A|B) means the probability of the event A occurring - given that B occurs.
The vertical dash | means "such that"

Slide 7-28: Example Clinical Case: Serotonin Syndrome

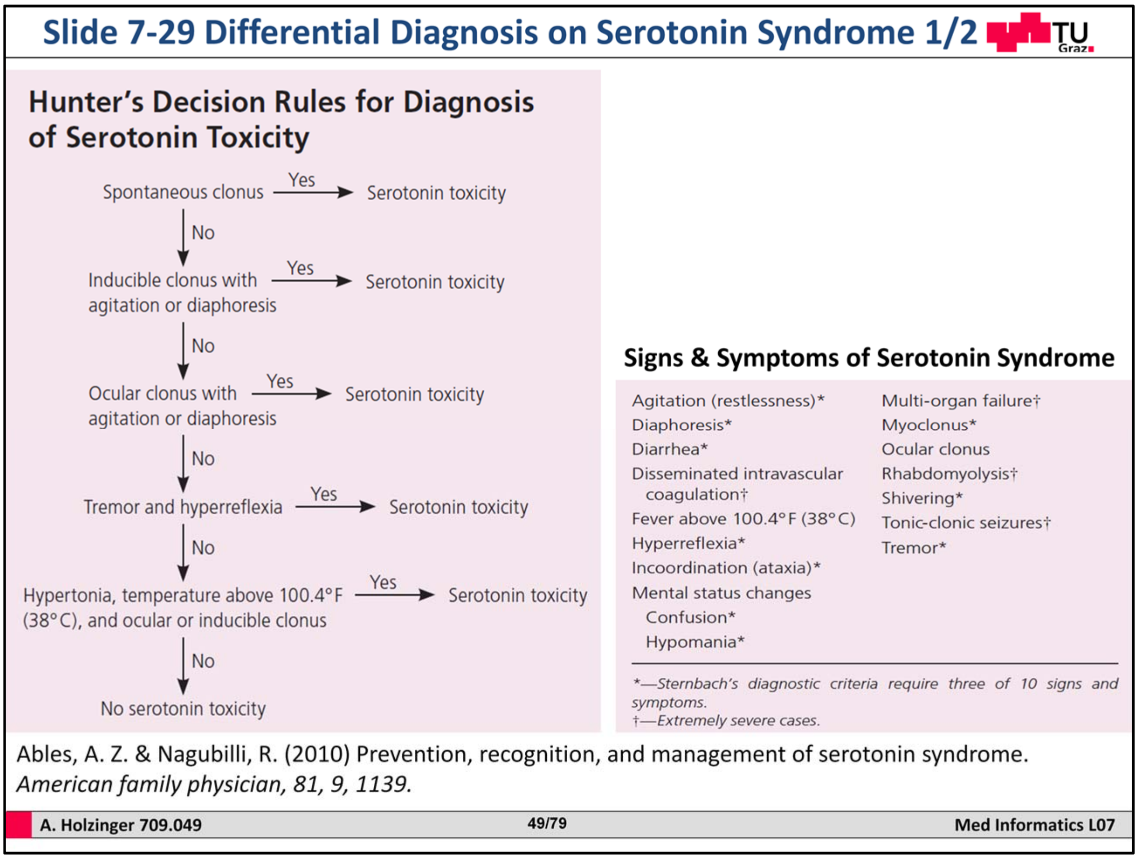




Boyer, E. W. & Shannon, M. (2005) The Serotonin Syndrome. *New England Journal of Medicine, 352, 11, 1112-1120.*

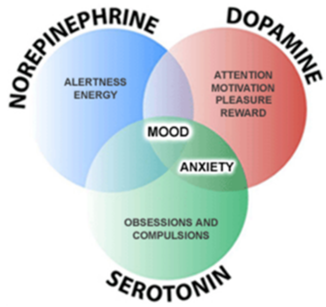
A. Holzinger 709.049 48/79 Med Informatics L07

Let us practice differential diagnosis on a clinical case: the serotonin syndrome is a potentially life-threatening adverse drug reaction that results from therapeutic drug use, intentional self-poisoning, or inadvertent interactions between drugs. Three features of the serotonin syndrome are critical to an understanding of the disorder. First, the serotonin syndrome is not an idiopathic drug reaction; rather it is a predictable consequence of excess serotonergic agonism of the central nervous system's receptors and peripheral serotonergic receptors. Second, excess serotonin produces a spectrum of clinical findings. Third, clinical manifestations of the serotonin syndrome range from barely perceptible to lethal. The death of an 18-year-old patient in New York City more than 20 years ago, which resulted from coadminstration of meperidine and phenelzine, remains the most widely recognized and dramatic example of this preventable condition (Boyer & Shannon, 2005).
In this slide we see that the serotonin syndrome encompasses a range of clinical findings. Patients with mild cases may be afebrile but have tachycardia, with a physical examination that is notable for autonomic findings such as shivering, diaphoresis (sweating), or mydriasis (blown pupil).



Differential Diagnosis (DD) is principally a systematic diagnostic method used to identify the presence of an entity, where multiple alternatives are possible and may also refer to any of the candidate alternatives. It is essentially a process of elimination, or at least, rendering of the probabilities of candidate conditions to negligible levels. The correct medical definition: Differential diagnosis is the process of weighing the p(x) of one disease versus the p(x) of other diseases possibly accounting for a patient's illness. The differential diagnosis of rhinitis (a runny nose) includes allergic rhinitis (hayfever), the abuse of nasal decongestants and the common cold.
In this slide we see Hunter's Decision Rules for Diagnosis of Serotonin Toxicity (left) from the example in Slide 7-28 and the signs (gamuts, patterns) and symptoms related (Ables & Nagubilli, 2010).

Slide 7-30 Differential Diagnosis on Serotonin Syndrome 2/2



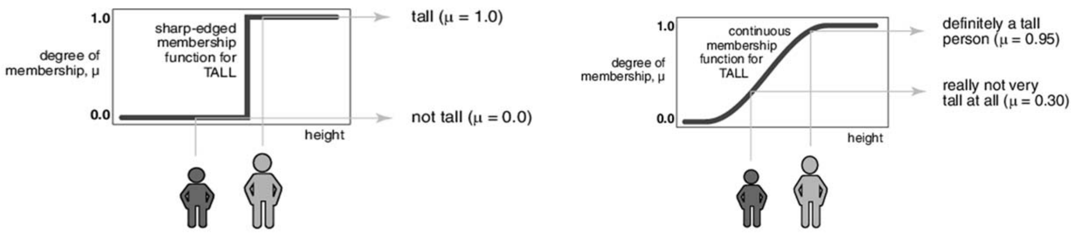
Clinical condition	*History*	*Vital signs*	*Clinical features*
Anticholinergic syndrome	Use of tricyclic antidepressants or other anticholinergic drugs	Tachycardia, tachypnea, hyperthermia (usually 102.2°F [39°C] or below)	Dry mouth, blurred vision, mydriasis, flushed skin, agitation/delirium, decreased bowel sounds
Malignant hyperthermia	Administration of halogenated inhalational anesthetics or depolarizing muscle relaxants	Hypertension, tachycardia, tachypnea, hyperthermia (up to 114.8°F [46°C])	Diaphoresis, mottled skin, agitation, decreased bowel sounds, muscular rigidity, hyporeflexia
Neuroleptic malignant syndrome	Ingestion of antipsychotic medications	Hypertension, tachycardia, tachypnea, hyperthermia (above 105.8°F [41°C])	Sialorrhea, diaphoresis, pallor, stupor, mutism, coma, normal or decreased bowel sounds, lead-pipe rigidity, bradyreflexia

Ables, A. Z. & Nagubilli, R. (2010) Prevention, recognition, and management of serotonin syndrome. *American family physician, 81, **9, 1139.***

The primary differential diagnosis of serotonin syndrome includes anticholinergic syndrome, malignant hyperthermia, and neuroleptic malignant syndrome as shown in this slide, along with the history, vital signs and clinical features (Ables & Nagubilli, 2010).

What can we do if we have not only probabilistic, but also I c pl e data …

Rough Set Theory



https://www.calvin.edu/~pribeiro/othrlnks/Fuzzy/fuzzyeng.htm

Rough estimate = über den Daumen gepeilt

Slide 7-31 Rough Set Theory for dealing with incomplete data

- … is an extension of the Classical Set Theory, for use when representing incomplete knowledge.
- RS are sets *with fuzzy boundaries* – sets that cannot be precisely characterized using the available set of attributes, exactly like it is in medical decision making; based on 2 ideas:
- 1) a given concept can be approximated by partition-based knowledge as upper and lower approximation – which corresponds to the focusing mechanism of differential medical diagnosis: upper approximation as selection of candidates and lower approximation as concluding a final diagnosis.
- 2) a concept or observation can be represented as partitions in a given data set, where rough sets provides a rule induction method from given data. Thus, this model can be used to extract rule-based knowledge from medical databases.

A. Holzinger 709.049 53/79 Med Informatics L07

RST is an extension of the Classical Set Theory, for use when representing incomplete knowledge. Rough sets are sets with fuzzy boundaries – sets that cannot be precisely characterized using the available set of attributes, exactly like it is in medical decision making; they are based on two ideas:
1) a given concept can be approximated by partition-based knowledge as upper and lower approximation – which corresponds exactly to the focusing mechanism of differential medical diagnosis: upper approximation as selection of candidates and lower approximation as concluding a final diagnosis.
2) a concept or observations can be represented as partitions in a given data set, where rough sets provides a rule induction method from given data. Consequently, this model can be used to extract rule-based knowledge from medical databases.
Note: Medical reasoning is a focusing mechanism, which is used to select the final diagnosis from many candidates. For example, in the differential diagnosis of headache, more than 60 diseases will be checked (present history, physical examinations and laboratory examinations). In diagnostic procedures, a candidate is excluded if a symptom necessary to diagnose is not observed.
This style of reasoning consists of the following two kinds of reasoning processes: exclusive reasoning and inclusive reasoning. The diagnostic procedure is shown in the next slide on the example of symptom “headache”.

Slide 7-32 Diagnostic Procedure (Differential Diagnostic)

Example Symptom: Headache

Focusing Mechanism (Selection of Candidates)	Characterization (Negative Rules)
Differential Diagnosis	Discrimination (Positive Rules)
Detection of Complications	Complications

Tsumoto, S. (2006) Pawlak Rough Set Model, Medical Reasoning and Rule Mining. In: Greco, S., Hata, Y., Hirano, S., Inuiguchi, M., Miyamoto, S., Nguyen, H. & Slowinski, R. (Eds.) *Rough Sets and Current Trends in Computing. Berlin, Heidelberg, Springer, 53-70.*

Exclusive reasoning excludes a disease from candidates when a patient does not have a symptom which is necessary to diagnose that disease.
Inclusive reasoning suspects a disease in the output of the exclusive process when a patient has symptoms specific to a disease. These two steps are modelled as usage of two kinds of rules, negative rules (or exclusive rules) and positive rules, the former of which corresponds to exclusive reasoning and the latter of which corresponds to inclusive reasoning. In the next two subsections, these two rules are represented as special kinds of probabilistic rules.
In the following slide, we use the notations introduced by (Skowron & Grzymala-Busse, 1994), (refer also to (Pawlak et al., 1995)), which are based on rough set theory (Pawlak, 1982). These notations are illustrated by a small dataset shown in an example, which includes symptoms exhibited by six patients who complained of headache .

Slide 7-33 Rough Set Theory Example Symptom: Headache 1

Let U denote a non-empty, finite set called the universe and A denote a non-empty, finite set of attributes:

- $a : U \to V_a$ for $a \in A$
- where V_a is called the domain of a
- Then, the decision table is defined as an information system:
- $A = (U, A \cup \{d\})$.
- The table shows an example of an information system with
- $U = \{1, 2, 3, 4, 5, 6\}$ and
- A = {age, location, nature, prodrome, nausea,M1} and
- d = class.
- For *location* $\in A$, $V_{location}$ is defined as {occular, lateral,whole}

No.	age	location	nature	prodrome	nausea	M1	class
1	50-59	occular	persistent	no	no	yes	m.c.h.
2	40-49	whole	persistent	no	no	yes	m.c.h.
3	40-49	lateral	throbbing	no	yes	no	migra
4	40-49	whole	throbbing	yes	yes	no	migra
5	40-49	whole	radiating	no	no	yes	m.c.h.
6	50-59	whole	persistent	no	yes	yes	psycho

DEFINITIONS. M1: tenderness of M1, m.c.h.: muscle contraction headache, migra: migraine, psycho: psychological pain.

Tsumoto, S. (2006) Pawlak Rough Set Model, Medical Reasoning and Rule Mining. In: Greco, S., Hata, Y., Hirano, S., Inuiguchi, M., Miyamoto, S., Nguyen, H. & Slowinski, R. (Eds.) *Rough Sets and Current Trends in Computing. Berlin, Heidelberg, Springer, 53-70.*

Slide 7-33 Rough Set Theory Example Symptom: Headache 1/3
Let U denote a non-empty, finite set called the universe and A denote a non-empty, finite set of attributes:
$a : U \to V_a$ for $a \in A$
where Va is called the domain of a
Then, the decision table is defined as an information system:
$A=(U,A \cup\{d\})$
Let us have an information system (see table right in the slide) with
$U=\{1,2,3,4,5,6\}$
A={age,location,nature,prodrome,nausea,M1} and
d = class.
For location $\in$ A, Vlocation is defined as {occular,lateral,whole}.
The table right shows an example data set from (Tsumoto, 2006);
Note: prodrome = in medicine, a prodrome is an early symptom (or set of symptoms) that might indicate the start of a disease before specific symptoms occur.

Is there another possibility?

1) Logic
2) Statistics/Probability
3) Heuristics

Slide 7-41: Science Vol. 185, pp.1124-1131, Sept. 1974

TU Graz

Judgment under Uncertainty: Heuristics and Biases

Biases in judgments reveal some heuristics of thinking under uncertainty.

Amos Tversky and Daniel Kahneman

Many decisions are based on beliefs concerning the likelihood of uncertain events such as the outcome of an election, the guilt of a defendant, or the future value of the dollar. These beliefs are usually expressed in statements such as "I think that . . . ," "chances are . . . ," "it is unlikely that . . . ," and so forth. Occasionally, beliefs concerning uncertain events are expressed in numerical form as odds or subjective probabilities. What determines such beliefs? How do people assess the prob-

mated when visibility is good because the objects are seen sharply. Thus, the reliance on clarity as an indication of distance leads to common biases. Such biases are also found in the intuitive judgment of probability. This article describes three heuristics that are employed to assess probabilities and to predict values. Biases to which these heuristics lead are enumerated, and the applied and theoretical implications of these observations are discussed.

occupation from a list of possibilities (for example, farmer, salesman, airline pilot, librarian, or physician)? How do people order these occupations from most to least likely? In the representativeness heuristic, the probability that Steve is a librarian, for example, is assessed by the degree to which he is representative of, or similar to, the stereotype of a librarian. Indeed, research with problems of this type has shown that people order the occupations by probability and by similarity in exactly the same way (*1*). This approach to the judgment of probability leads to serious errors, because similarity, or representativeness, is not influenced by several factors that should affect judgments of probability.

Insensitivity to prior probability of outcomes. One of the factors that have no effect on representativeness but should have a major effect on probability is the prior probability, or base-rate frequency, of the outcomes. In the case of Steve, for example, the fact that there are many more farmers than librarians in the population should enter into any reasonable estimate of the probability that Steve is a librarian rather than a farmer. Considerations of base-rate frequency, however, do not

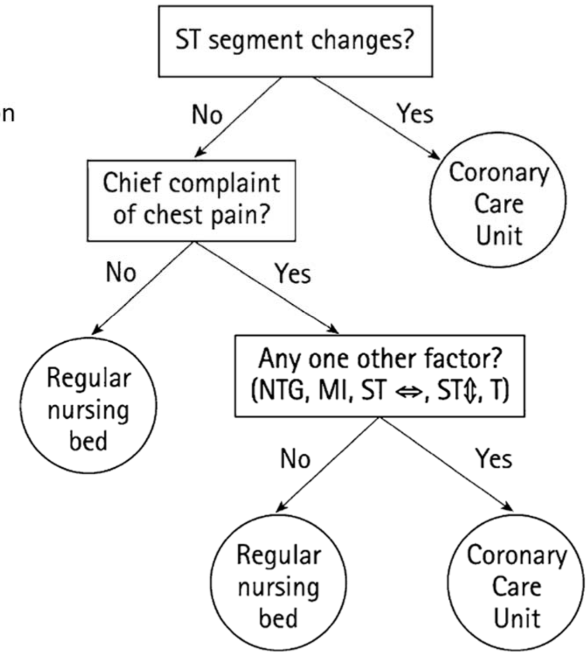
A. Holzinger 709.049 57/79 Med Informatics L07

Slide 7-41: Science Vol. 185, pp.1124-1131, Sept. 1974
In 1974 Amos Tversky and Daniel Kahneman wrote an article (meanwhile it received 20k citations) about Judgement under Uncertainty, and they describe the approach of Heuristic Decision Making (Tversky & Kahneman, 1974).
Remember: There are three major decision making strategies:
1) logic,
2) statistics, or
3) heuristics.

Each strategy is suited to a particular kind of problem. Rules of logic and statistics have been linked to rational reasoning. Heuristics have been linked to error-prone intuitions or irrationality. As Simon (1979) emphasized, the classical model of rationality requires knowledge of all relevant alternatives, their consequences and probabilities, and a predictable world without surprises. These conditions, however, are rarely met for the problems that individuals and organizations face. Savage (1954), known as the founder of modern Bayesian decision theory, called such perfect knowledge small worlds, to be distinguished from large worlds. Heuristics are strategies that ignore certain information – to make decisions faster.

Slide 7-42 Heuristic Decision Making

MI = myocardial infarction
N.A. = not applicable
NTG = nitroglycerin
T= T-waves with peaking or inversion

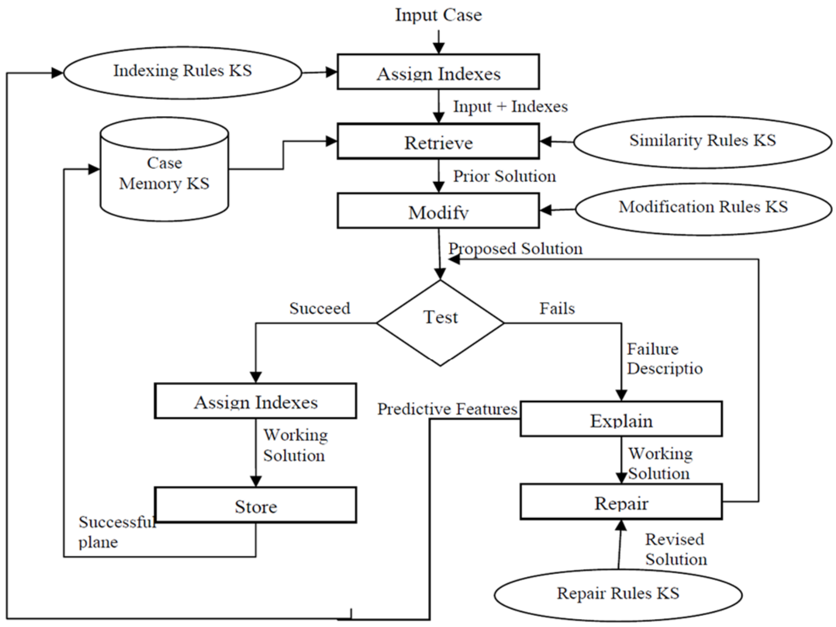


Gigerenzer, G. & Gaissmaier, W. (2011) Heuristic Decision Making. In: Fiske, S. T., Schacter, D. L. & Taylor, S. E. (Eds.) *Annual Review of Psychology, Vol 62. Palo Alto, Annual Reviews, 451-482.*

Slide 7-42: Heuristic Decision Making
1. Heuristics can be more accurate than more complex strategies even though they process less information (less-is-more effects).
2. A heuristic is not good or bad, rational or irrational; its accuracy depends on the structure of the environment (ecological rationality).
3. Heuristics are embodied and situated in the sense that they exploit core capacities of the brain and their success depends on the structure of the environment. They provide an alternative to stable traits, attitudes, preferences, and other internal explanations of behavior.
4. With sufficient experience, people learn to select proper heuristics from their adaptive toolbox.
5. Usually, the same heuristic can be used both consciously and unconsciously, for inferences and preferences, and underlies social as well as nonsocial intelligence.
6. Decision making in organizations typically involves heuristics because the conditions for rational models rarely hold in an uncertain world (Gigerenzer & Gaissmaier, 2011).
In this slide we see a tree which prescribes how emergency physicians can detect acute ischemic heart disease. It only asks up to three yes/no questions, namely whether the patient’s electrocardiogram shows a certain anomaly (“ST segment changes”), whether chest pain is the patient’s primary complaint, and whether there is any other factor (Gigerenzer & Gaissmaier, 2011); MI = myocardial infarction; N.A. = not applicable; NTG = nitroglycerin; T= T-waves with peaking or inversion.

Slide 7-43 Case Based Reasoning (CBR)

Boxes = processes; ovals = knowledge structures (KS)

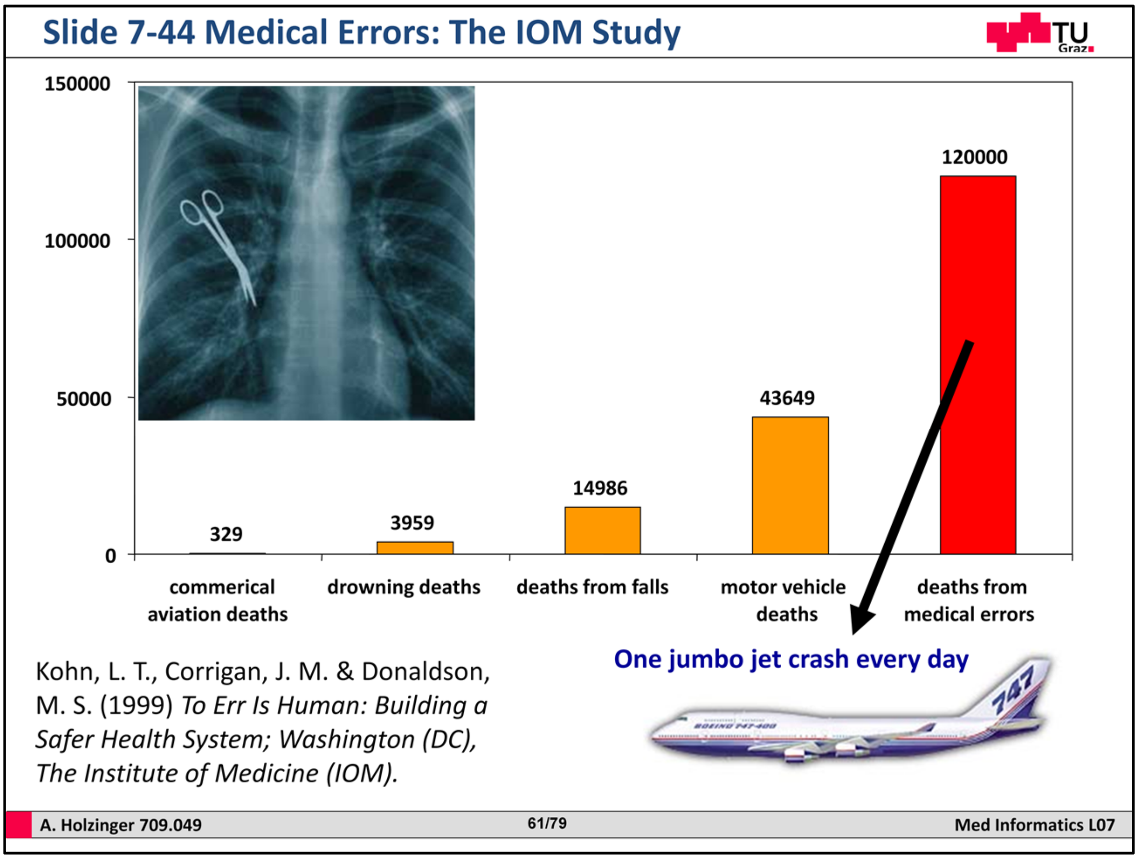


Salem, A. B. M. (2007) Case based reasoning technology for medical diagnosis. *Proc. World Academy of Science, Engineering and Technology, 25, 9-13.*

A. Holzinger 709.049 59/79 Med Informatics L07

CBR is an analogical reasoning method providing both a methodology for problem solving and a cognitive model. CBR means reasoning from experience or "old cases" in an effort to solve problems, critique solutions, and explaining anomalous situations. Some historical interesting medical CBR-Systems include: CASEY that gives a diagnosis for heart disorders, NIMON is a renal function monitoring system, COSYL that gives a consultation for a liver transplanted patient, or ICONS that presents suitable calculated antibiotics therapy advised for intensive care patients (Salem, 2007). In this slide we see a typical CBR Methodology. Boxes represent processes and ovals represent KS (Salem, 2007).

Human Error



Kohn L.T., Corrigan, J.M., Donaldson, M.S. (1999): To Err is Human: Building a Safer Health System, National Academy Press, Washington (DC)

Let us finalize this lecture with a short look on human errors. Preventable medical mistakes and infections are responsible for about 200,000 deaths in the U.S. each year, according to an investigation by the Hearst media corporation. The report comes 10 years after the Institute of Medicine's "To Err Is Human" analysis, which found that 44,000 to 98,000 people were dying annually due to these errors and called for the medical community and government to cut that number in half by 2004 (Kohn, Corrigan & Donaldson, 1999).

Slide 7-45 Definitions of medical errors

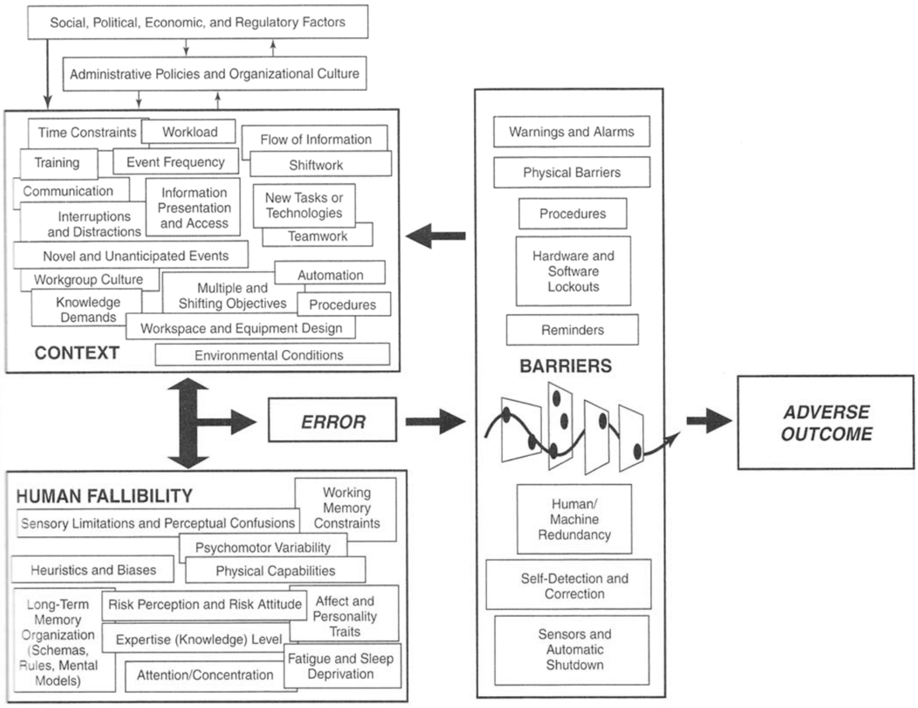
- Medical error = any failure of a planned action;
- Serious ME = causes harm; includes preventable adverse events, intercepted serious errors, and non-intercepted serious errors. Does not include trivial errors with little or no potential for harm or non-preventable adverse events;
- Intercepted serious error = is caught before reaching patients;
- Non-intercepted serious error = reaches the patient but of good fortune or sufficient reserves to buffer the error, it did not cause harm;
- Adverse event = any injury (e.g. a rash caused by an antibiotic, deep vein thrombosis following omission to continue prophylactic subcutaneous heparin orders on transfer to the critical care unit, ventricular tachycardia due to placement of a central venous catheter tip in the right ventricle etc.);
- Non-preventable adverse event = Unavoidable injury due to appropriate medical care.
- Preventable adverse event = Injury due to a non- intercepted serious error in medical care.

Rothschild et al. (2005) The Critical Care Safety Study: The incidence and nature of adverse events and serious medical errors in intensive care. *Critical Care Medicine, 33, 8, 1694.*

Rothschild et al. (2005) provides some definitions of medical errors:
Medical error (ME) = any failure of a planned action;
Serious ME = causes harm; includes preventable adverse events, intercepted serious errors, and non-intercepted serious errors. Does not include trivial errors with little or no potential for harm or non-preventable adverse events;
Intercepted serious error = is caught before reaching patients;
Non-intercepted serious error = reaches the patient but of good fortune or sufficient reserves to buffer the error, it did not cause harm;
Adverse event = any injury (e.g. a rash caused by an antibiotic, deep vein thrombosis following omission to continue prophylactic subcutaneous heparin orders on transfer to the critical care unit, ventricular tachycardia due to placement of a central venous catheter tip in the right ventricle etc.);
Non-preventable adverse event = Unavoidable injury due to appropriate medical care.
Preventable adverse event = Injury due to a non- intercepted serious error in medical care.
Finally in the next slide we see the Swiss-Cheese Model – we will discuss more details in Lecture 11.

Slide 7-46 Framework for understanding human error

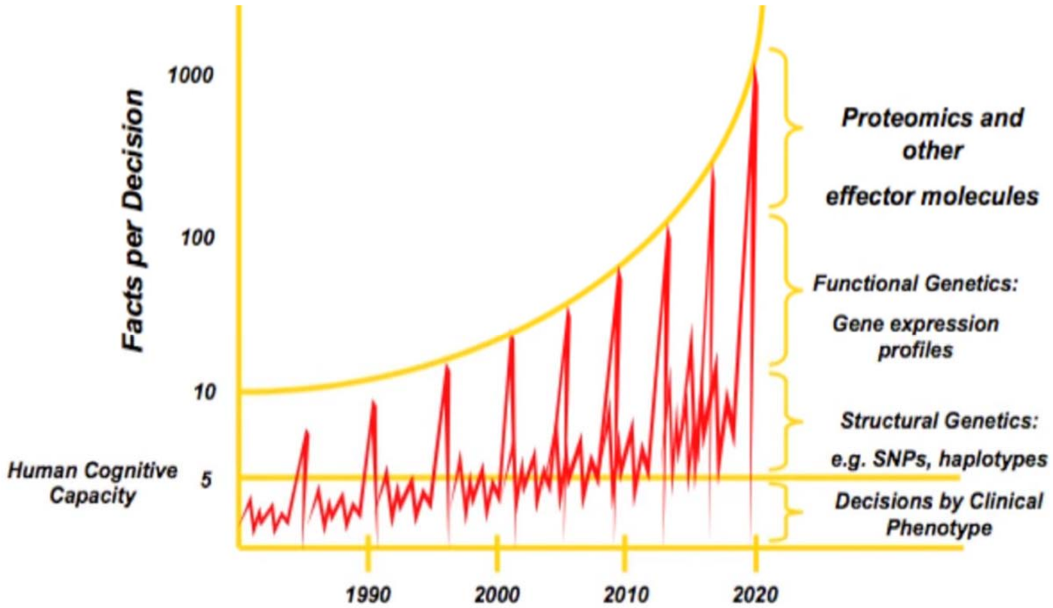
Sharit, J. (2006) Human Error. In: Salvendy, G. (Ed.) *Handbook of Human Factors and Ergonomics, Third Edition. Hoboken (NJ), Wiley, 708-760.*



This is a standard framework for understanding human error, it describes how human error arises and can result in adverse outcomes. There are three major components: 1. Human fallibility, 2. Context and 3. Barriers (Sharit, 2006).

My DEDICATION is to make data valuable … Thank you!

Slide 7-47 Future Outlook



William Stead, IOM Meeting, 8 October 2007. Growth in facts affecting provider decisions versus human cognitive capacity.

Sample Questions (1)

- What is still considered the main and central topic in medical informatics?
- Please explain the information flow within the memory system according to Atkinson & Shiffrin!
- Explain the general model of human information processing following the model of Wickens!
- Explain the processing of visual (image, pictorial) information!
- What is so different in the alternative memory model according to Baddeley (1986)?
- Why is Attention of importance for medical informatics?
- Please explain the process of human decision making according to the model of Wickens (1984)!
- What is Triage?
- Please explain the hypothesis-oriented algorithm for Clinicians!
- What is the big difference between the Hypothetico-Deductive Method and the Plan-Do-Check-Act Deming Model?
- How can we model patient health – please provide an example!

Sample Questions (2)

- Please contrast the decision making process with the data mining process!
- Why is Signal Detection Theory important for us?
- Please provide an Example for the application of Bayes' Theorem!
- How does Differential Diagnosis work?
- How can we apply Rough Set Theory for differential diagnostics?
- What is Heuristic Decision Making?
- What is problematic when dealing with heuristic decision making from an informatics viewpoint?
- What is Case Based Reasoning (CBR)?
- How are medical errors defined?
- How does the framework for understanding human error work?

Some Useful Links

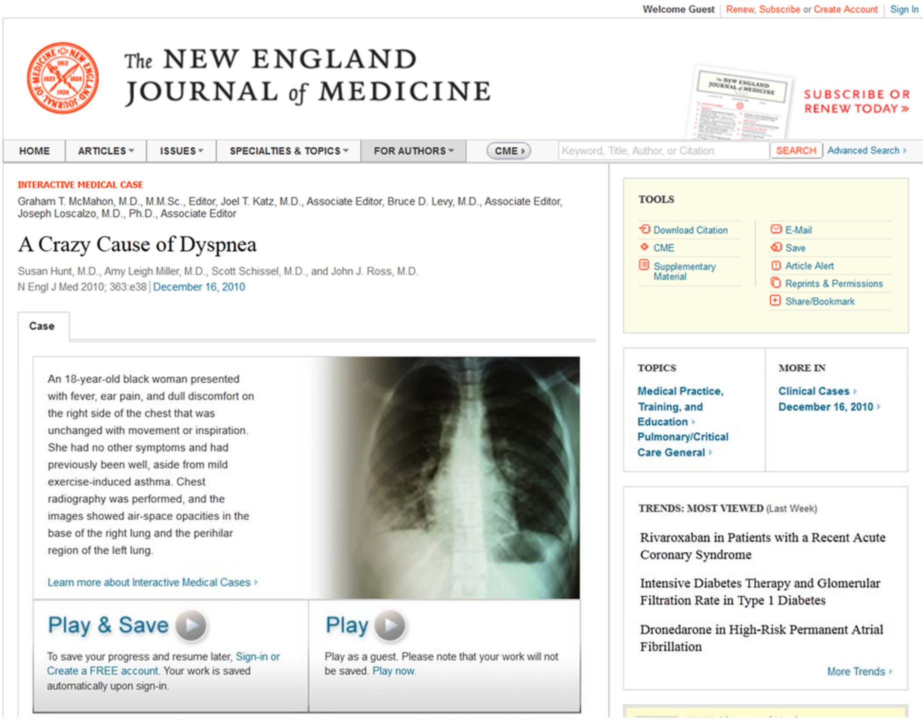
- http://www.anaesthetist.com/mnm/stats/roc
- http://sbml.org
- http://www.hci4all.at
- http://www.lcb.uu.se/tools/rosetta
- http://wise.cgu.edu/sdtmod/overview.asp (excellent Tutorial on SDT)
- http://www.iom.edu (Institute of Medicine)
- http://www.ahrq.gov/qual/patientsafetyix.htm (Agency for Health Care Research and Quality)
- http://www.fda.gov/drugs/drugsafety/medicationerrors/default.htm (Food and Drug Administration, FDA, medication errors)

(1) Playing with ROCs – very good

http://psychology.wikia.com/wiki/Information_retrieval

Appendix: NEJM Interactive Multimedia Cases

Hunt, S., Miller, A. L., Schissel, S. & Ross, J. J. (2010) A Crazy Cause of Dyspnea. Interactive Multimedia Case *New England Journal of Medicine, 363, 25, e38.*

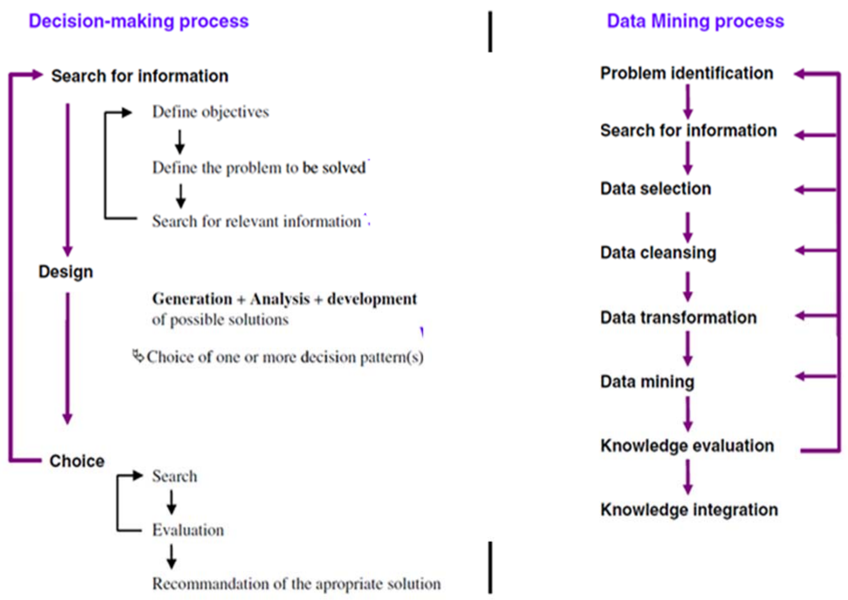


http://www.nejm.org/doi/full/10.1056/NEJMimc1008281

Schedule

- 1. Intro: Computer Science meets Life Sciences, challenges, future directions
- 2. Back to the future: Fundamentals of Data, Information and Knowledge
- **3. Structured Data: Coding, Classification (ICD, SNOMED, MeSH, UMLS)**
- 4. Biomedical Databases: Acquisition, Storage, Information Retrieval and Use
- 5. Semi structured and weakly structured data (structural homologies)
- 6. Multimedia Data Mining and Knowledge Discovery
- 7. Knowledge and Decision: Cognitive Science & Human-Computer Interaction
- 8. Biomedical Decision Making: Reasoning and Decision Support
- 9. Intelligent Information Visualization and Visual Analytics
- 10. Biomedical Information Systems and Medical Knowledge Management
- 11. Biomedical Data: Privacy, Safety and Security
- 12. Methodology for Info Systems: System Design, Usability & Evaluation

test



A. Holzinger 709.049 71/79 Med Informatics L07

Slide 7-34 Rough Set Theory Example Symptom: Headache 2

- The atomic formula over
- B ⊆ A ∪ {d} and V are expressions of the form [a = v]
- called descriptors over B, where a ∈ B and v ∈ Va.
- The set F(B, V) of formulas over B is the least set containing all atomic formulas over B and closed with respect to disjunction, conjunction and negation. For example, [location = occular] is a descriptor of B.
- For each f ∈ F(B, V), fA denote the meaning of f in A, i.e., the set of all objects in U with property f, defined inductively as follows.
- 1. If f is of the form [a = v] then, fA = {s ∈ U|a(s) = v}
- 2. (f ∧ g)A = fA ∩ gA; (f ∨ g)A = fA ∨ gA; (¬f)A = U − fa
- For example, f = [location = whole] and fA = {2, 4, 5, 6}. As an example of a conjunctive formula, g = [location = whole] ∧ [nausea = no] is a descriptor of U and fA is equal to glocation, nausea = {2, 5}.

A. Holzinger 709.049 72/79 Med Informatics L07

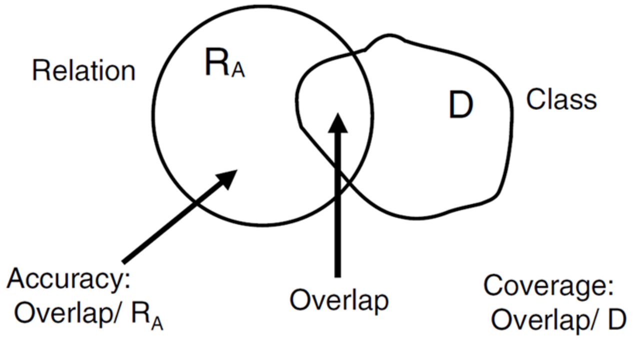
Slide 7-34 Rough Set Theory Example Symptom: Headache 2/3
The atomic formula (Tsumoto, 2006) over
B⊆A ∪ {d}
and V are expressions of the form [a = v]
called descriptors over B, where a ∈ B and v ∈ Va.
The set F(B, V) of formulas over B is the least set containing all atomic formulas over B and closed with respect to disjunction, conjunction and negation. For example, [location = occular] is a descriptor of B.
For each f ∈ F(B, V), fA denote the meaning of f in A, i.e., the set of all objects in U with property f, defined inductively as follows.
1. If f is of the form [a = v] then, fA = {s∈U ├|a(s)=v┤ }
2. (f ∧ g)A = fA ∩ gA; (f ∨ g)A = fA ∨ gA; (¬f)A = U − fa
For example, f = [location = whole] and fA = {2,4,5,6}. As an example of a conjunctive formula, g = [location = whole] ∧ [nausea = no] is a descriptor of U and fA is equal to glocation, nausea = {2,5}.

Slide 7-35 Classification Accuracy and Coverage

Definition 1. *Let R and D denote a formula in $F(B, V)$ and a set of objects which belong to a decision d. Classification accuracy and coverage(true positive rate) for $R \rightarrow d$ is defined as:*

$$\alpha_R(D) = \frac{|R_A \cap D|}{|R_A|} (= P(D|R)), \text{ and}$$
$$\kappa_R(D) = \frac{|R_A \cap D|}{|D|} (= P(R|D)),$$

where $|S|$, $\alpha_R(D)$, $\kappa_R(D)$ and $P(S)$ denote the cardinality of a set S, a classification accuracy of R as to classification of D and coverage (a true positive rate of R to D), and probability of S, respectively.



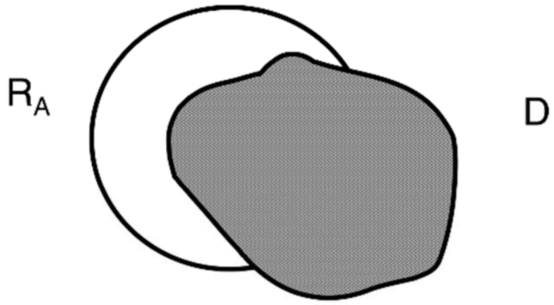
Tsumoto (2006)

Slide 7-35 Classification Accuracy and Coverage
Definition 1. Let R and D denote a formula in F(B, V) and a set of objects which belong to a decision d. Classification accuracy and coverage(true positive rate) for R → d is defined as:
αR(D) = |RA ∩ D|/ |RA| (= P(D|R)), and
κR(D) = |RA ∩ D|/|D| (= P(R|D)),
where |S|, αR(D), κR(D) and P(S) denote the cardinality of a set S, a classification accuracy of R as to classification of D and coverage (a true positive rate of R to D), and probability of S, respectively, which can be seen in this slide (Tsumoto, 2006).

Slide 7-36 Probabilistic Rules – modus ponens

By the use of accuracy and coverage, a probabilistic rule is defined as:

$$R \overset{\alpha,\kappa}{\rightarrow} d \quad s.t.\ R = \wedge_j [a_j = v_k], \alpha_R(D) \geq \delta_\alpha$$
$$\text{and } \kappa_R(D) \geq \delta_\kappa,$$



$$\boxed{R \rightarrow D \quad s.t. \quad \alpha_R(D) > \delta_\alpha, \ \kappa_R(D) > \delta_\kappa}$$

Tsumoto (2006)

Slide 7-36 Probabilistic Rules – modus ponens

In the above example, when R and D are set to [nau = yes] and [class =migraine], $\alpha R(D) = 2/3 = 0.67$ and $\kappa R(D) = 2/2 = 1.0$.

It is notable that $\alpha R(D)$ measures the degree of the sufficiency of a proposition, $R \rightarrow D$, and that $\kappa R(D)$ measures the degree of its necessity. For example, if $\alpha R(D)$ is equal to 1.0, then $R \rightarrow D$ is true. On the other hand, if $\kappa R(D)$ is equal to 1.0, then $D \rightarrow R$ is true. Thus, if both measures are 1.0, then $R \leftrightarrow D$.

By the use of accuracy and coverage, a probabilistic rule is defined as:

R

$\alpha\rightarrow,\kappa d$ s.t. $R= \wedge j\ [aj = vk]$, $\alpha R(D) \geq \delta\alpha$

and $\kappa R(D) \geq \delta\kappa$,

If the thresholds for accuracy and coverage are set to high values, the meaning of the conditional part of probabilistic rules corresponds the highly overlapped region.

This slide shows the Venn diagram of probabilistic rules with highly overlapped regions. This rule is a kind of probabilistic proposition with two statistical measures, which is an extension of the Ziarko's variable precision model (VPRS) (Tsumoto, 2006).

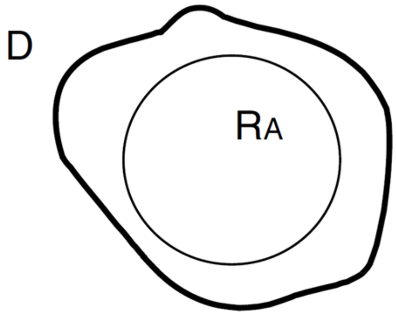
Slide 7-37 Positive Rules

A positive rule is defined as a rule supported by only positive examples, the classification accuracy of which is equal to 1.0. It is notable that the set supporting this rule corresponds to a subset of the lower approximation of a target concept, which is introduced in rough sets [1]. Thus, a positive rule is represented as:

$$R \to d \quad s.t. \qquad R = \wedge_j [a_j = v_k], \quad \alpha_R(D) = 1.0$$

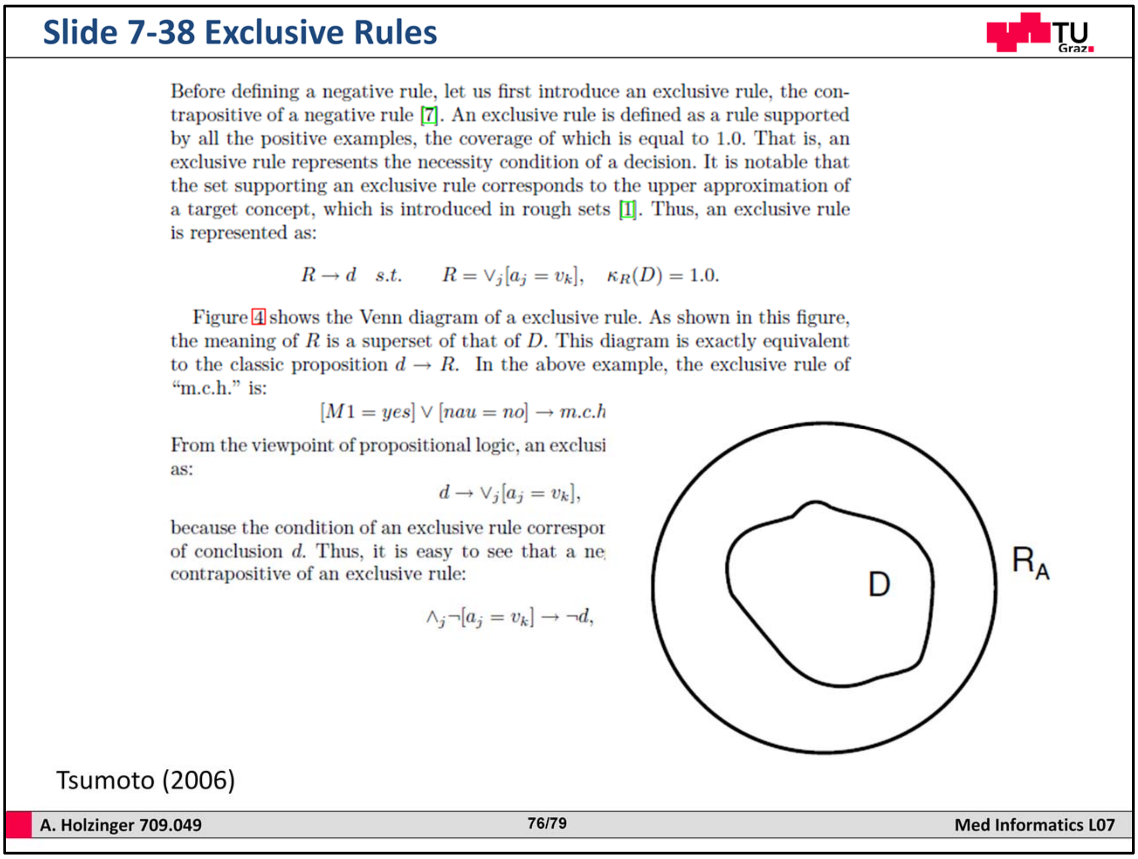
Figure 4 shows the Venn diagram of a positive rule. As shown in this figure, the meaning of R is a subset of that of D. This diagram is exactly equivalent to the classic proposition $R \to d$. In the above example, one positive rule of "m.c.h." (muscle contraction headache) is:

$$[nausea = no] \to m.c.h. \quad \alpha = 3/3 = 1.0.$$



Tsumoto (2006)

Slide 7-37: Positive Rules
Positive Rules. A positive rule is defined as a rule supported by only positive examples, the classification accuracy of which is equal to 1.0. It is notable that the set supporting this rule corresponds to a subset of the lower approximation of a target concept, which is introduced in rough sets.
Thus, a positive rule is represented as:
R →d s.t. R= ∧j [aj = vk], αR(D) = 1.0
This slide shows the Venn diagram of such a positive rule. As shown in this Figure, the meaning of R is a subset of that of D. This diagram is exactly equivalent to the classic proposition R → d. In the above example, one positive rule of "m.c.h." (muscle contraction headache) is:
[nausea = no] → m.c.h. α = 3/3 = 1.0.
This positive rule is often called a deterministic rule (Tsumoto, 2006).



Slide 7-38: Exclusive Rules

Exclusive Rules. It is also called contrapositive of a negative rule and is defined as a rule supported by all the positive examples, the coverage of which is equal to 1.0. That is, an exclusive rule represents the necessity condition of a decision. It is notable that the set supporting an exclusive rule corresponds to the upper approximation of a target concept, which is introduced in rough sets. Thus, an exclusive rule is represented as:

$R \to d$ s.t. $R = \vee_j [a_j = v_k]$, $\kappa_R(D) = 1.0$.

As shown this slide, the meaning of R is a superset of that of D. This diagram is exactly equivalent to the classic proposition $d \to R$. In the above example, the exclusive rule of

"m.c.h." is:

$[M1 = yes] \vee [nau = no] \to m.c.h.\ \kappa = 1.0$

From the viewpoint of propositional logic, an exclusive rule should be represented as:

$d \to \vee_j [a_j = v_k]$,

because the condition of an exclusive rule corresponds to the necessity condition of conclusion d. Thus, it is easy to see that a negative rule is defined as the contrapositive of an exclusive rule:

$\wedge_j \neg [a_j = v_k] \to \neg d$

Slide 7-39 Negative Rule

of conclusion d. Thus, it is easy to see that a negative rule is defined as the contrapositive of an exclusive rule:

$$\wedge_j \neg [a_j = v_k] \rightarrow \neg d,$$

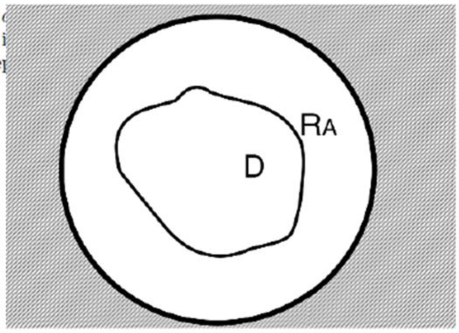
which means that if a case does not satisfy any attribute value pairs in the condition of a negative rules, then we can exclude a decision d from candidates. For example, the negative rule of m.c.h. is:

$$\neg [M1 = yes] \wedge \neg [nausea = no] \rightarrow \neg m.c.h.$$

In summary, a negative rule is defined as:

$$\wedge_j \neg [a_j = v_k] \rightarrow \neg d \quad s.t. \quad \forall [a_j = v_k] \; \kappa_{[a_j = v_k]}(D) = 1.0,$$

where D denotes a set of samples which belong to a class
Venn diagram of a negative rule. As shown in this figure,
negative region is the "positive region" of "negative conce



Tsumoto (2006)

Slide 7-39: Negative Rule
Negative Rule. It is now easy to see that a negative rule is defined as the contrapositive of an exclusive rule:
$\wedge j \neg [aj = vk] \rightarrow \neg d$,
which means that if a case does not satisfy any attribute value pairs in the condition of a negative rules, then we can exclude a decision d from candidates. For example, the negative rule of m.c.h. is:
$\neg [M1 = yes] \wedge \neg [nausea = no] \rightarrow \neg$m.c.h.
In summary, a negative rule is defined as:
$\wedge j \neg [aj = vk] \rightarrow \neg d$ s.t. $\forall [aj = vk]$ $\kappa[aj=vk](D) = 1.0$,
where D denotes a set of samples which belong to a class d.
This slide shows the Venn diagram of a negative rule. As shown in this slide, it is notable that this negative region is the "positive region" of "negative concept" (Tsumoto, 2006).

Slide 7-40 Example: Algorithms for Rule Induction

procedure *Exclusive and Negative Rules*;
  var
    L : *List*;
      /* A list of elementary attribute-value pairs */
  begin
    $L := P_0$;
    /* P_0: A list of elementary attribute-value pairs given in a database */
    while $(L \neq \{\})$ **do**
      begin
        Select one pair $[a_i = v_j]$ from L;
        if $([a_i = v_j]_A \cap D \neq \phi)$ **then do** /* D: positive examples of a target class d */
          begin
            $L_{ir} := L_{ir} + [a_i = v_j]$;   /* Candidates for Positive Rules */
            if $(\kappa_{[a_i=v_j]}(D) = 1.0)$
            then $R_{er} := R_{er} \wedge [a_i = v_j]$;
              /* Include $[a_i = v_j]$ into the formula of Exclusive Rule */
          end
        $L := L - [a_i = v_j]$;
      end
    Construct Negative Rules:
        Take the contrapositive of R_{er}.
  end {*Exclusive and Negative Rules*};

A. Holzinger 709.049 78/79 Med Informatics L07

Slide 7-40 Example: Algorithms for Rule Induction
The contrapositive of a negative rule, an exclusive rule, is induced as an exclusive rule by the modification of the algorithm introduced in PRIMEROSE-REX (seen in this slide). This algorithm works as follows. (1) First, it selects a descriptor [ai = vj] from the list of attribute-value pairs, denoted by L. (2) Then, it checks whether this descriptor overlaps with a set of positive examples, denoted by D. (3) If so, this descriptor is included into a list of candidates for positive rules and the algorithm checks whether its coverage is equal to 1.0 or not. If the coverage is equal to 1.0, then this descriptor is added to Rer, the formula for the conditional part of the exclusive rule of D. (4) Then, [ai = vj] is deleted from the list L. This procedure, from (1) to (4) will continue unless L is empty. (5) Finally, when L is empty, this algorithm generates negative rules by taking the contrapositive of induced exclusive rules. On the other hand, positive rules are induced as inclusive rules by the algorithm introduced in PRIMEROSE-REX. For induction of positive rules, the threshold of accuracy and coverage is set to 1.0 and 0.0, respectively.

The Naïve Bayes Classifier

- What can we do if our data *d* has several attributes?
- Naïve Bayes assumption: Attributes that describe data instances are conditionally independent given the classification hypothesis

$$P(\mathbf{d} \mid h) = P(a_1, \dots, a_T \mid h) = \prod_t P(a_t \mid h)$$

  - it is a simplifying assumption, obviously it may be violated in reality
  - in spite of that, it works well in practice
- The Bayesian classifier that uses the Naïve Bayes assumption and computes the MAP hypothesis is called Naïve Bayes classifier
- One of the most practical learning methods
- Successful applications:
  - Medical Diagnosis