DEPARTMENT OF PSYCHOLOGY, COLLEGE OF LIBERAL ARTS
ROCHESTER INSTITUTE OF TECHNOLOGY

PSYC 719—HUMAN FACTORS IN AI

HANDOUT: INTRODUCTION TO HUMAN FACTORS

PROF. RANTANEN

August 23, 2022

1 A Little History

Human factors evolved as a distinct discipline during and after World War II. Three forces may be identified behind this [1]:

1. *Practical needs* that arose from the accelerating advancement of technology, which was a direct result of the war effort.
2. *Technological advancements* were particularly pronounced in the aviation domain, where aircraft speed, capabilities, and complexity increased at an unprecedented rate. This resulted in unacceptable accident rates and loss of life before the pilots ever saw combat. It was no longer possible to *fit the human to the machine* through selection and training, but human capabilities and limitations had to be considered in the *design of the machines*.
3. Linguistic developments was brought about by the combined effort of both engineers and psychologists to address the novel human-machine interface problems. The role of the human operator was acknowledged as that of a system component and human behavior was described in similar terms as the systems they were interacting with. Thus terminology and concepts common in electrical and systems engineering (e.g., channel capacity, feedback, optimal control, etc.) replaced the stimulus-feedback language of behavioral psychology and further facilitated the integration of engineering and psychology for the design and evaluation of human-machine systems.

2 Terms and Definitions

2.1 Human Factors/Ergonomics

Human Factors/Ergonomics (HF/E) is the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data, and other methods to design in order to optimize human well-being and overall system performance.

In short, Human Factors/Ergonomics studies human capabilities and limitations as they apply to the design of systems and products.

Ergonomics (from gk. ergon "work" + nomics) is the application of scientific information concerning humans to the design of objects, systems and environment for human use.

For many more definitions, please see
`https://www.hfes.org/About-HFES/What-is-Human-Factors-and-Ergonomics#professional_societies`

2.2 Engineering Psychology

To define engineering psychology as a discipline it is important to distinguish it from both psychology and engineering as well as from several other, closely related, disciplines. Although there is a relationship between applied psychology and engineering psychology, there is also an important difference: Where applied psychology seeks to control and influence people, the goal of engineering psychology is the design of a better machine. Although this at the time was a very non-traditional objective from the psychologists' point of view, psychology made substantial contribution to the design of machines: Knowledge of human variability and methods of dealing with it, factors engineers were not used to accounting for. Thus the role of the engineering psychologist in machine design is thus both that of a scientist, seeking knowledge of human behavior, capabilities, and limitations for engineers to use, and that of a technologist, actively participating in the design of human-machine systems [2].

Although engineering psychology shares the practical orientation with applied psychology, the methods employed in research of human-machine systems are primarily those of experimental psychology [3]. To differentiate engineering psychology from applied experimental psychology, then, one must again consider the specific domain of applications of engineering psychology, the human-machine systems. To further underscore the unique nature of engineering psychology, the discipline was accorded a divisional status (Division 21) by the American Psychological Association. The mission for Engineering Psychologists is defined by APA as "to promote research, development, application and evaluation of psychological principles relating human behavior to the characteristics, design and use of environments and systems within which people work and live" (American Psychological Association)

Engineering psychology differs from the closely related discipline of *human factors*, or *human factors engineering*, in two important aspects. On one hand, human factors is a much broader discipline which encompasses such diverse sub-disciplines as anthropometry and biomechanics [4]. The same is true for *ergonomics*, a term commonly used in Europe and essentially synonymous with human factors, as the discipline is referred to in the United States. Engineering psychology, true to its roots in psychology, is concerned predominantly with the information processing aspects of human performance. On the other hand, human factors can be seen as a purely applied discipline, while engineering psychology, albeit motivated by applications in human-machine systems design, is also concerned with more basic research [1]. The ultimate goal of human factors is to improve system design, not to seek understanding of human behavior, whereas "the aim of engineering psychology is not simply to compare two possible designs for a piece of equipment, but to specify the capacities and limitations of the human, from which the choice of the better design should be deducible directly" [5, p. 178].

The cognitive focus of engineering psychology has recently become increasingly pronounced. This reflects the shift of interest towards cognition in psychology in general but also the new demands increasingly complex systems place on the operators. The emphasis on cognition is today the main scientific force driving application efforts [6]. This fact is underlined by the emergence of such disciplines as *cognitive engineering* or *knowledge engineering*. More about cognitive engineering below.

Because the main area of application for engineering psychology is systems design and evaluation, a quantitative approach to the description of human behavior is imperative. These efforts have benefited substantially from the influence of the traditional engineering disciplines [2, 1]. In addition to the methods of experimental psychology, mathematical modeling is an essential tool used by engineering psychologists [7]. Extensive reviews of the various modeling approaches are provided by [8, 9, 10, 11].

2.3 Cognitive Engineering

We should also distinguish between **cognitive engineering** and **engineering psychology**. The distinction is in the emphasis, that is, the second word in the names, either on engineering or psychology. Although the goal of engineering *psychology* is the design of a better machine, the role of the engineering psychologist in machine design is both that of a scientist (seeking knowledge of human behavior, capabilities, and limitations for engineers to use) and that of a technologist (actively participating in the design of human-machine systems), but with emphasis on the former. Engineering psychology shares the practical orientation with engineering disciplines but the methods employed in research of human-machine systems are primarily those of experimental psychology. Engineering psychology is also concerned with basic research and understanding of human behavior. Cognitive *engineering*, as the name implies, is an engineering discipline with focus on systems *design* rather than basic research on human capabilities and limitations. The following definitions hopefully highlight these distinctions.

- "**Cognitive engineering** is an applied cognitive science that draws on the knowledge and techniques of cognitive psychology and related disciplines to provide the foundation for principle-driven design of person-machine systems." [12]

- **Cognitive engineering** is comprised of "...observation, modeling, analysis and design of complex work domains in which human expertise is paramount and multiple aspects of the work environment may drive performance." Research in cognitive engineering "...seeks to understand how people engage in cognitive work in real-world settings and the development of systems that support that work." Research "...on human cognition and the application of this knowledge to the design and development of system interfaces, automation, aids and other support systems, training programs, personnel selection devices, and coordination environments for people who work in teams or groups." (Description of the scope of the *Journal of Cognitive Engineering and Decision Making*, Sage Publications.)

- "**Cognitive Engineering**, a term invented to reflect the enterprise I find myself engaged in: neither Cognitive Psychology, nor Cognitive Science, nor Human Factors. It is a type of applied Cognitive Science, trying to apply what is known from science to the design and construction of machines. It is a surprising business. On the one hand, there actually is quite a lot known in Cognitive Science that can be applied. But on the other hand, our lack of knowledge is appalling. On the one hand, computers are ridiculously difficult to use. On the other hand, many devices are difficult to use—the problem is not restricted to computers, there are fundamental difficulties in understanding and using most complex devices. So the goal of Cognitive Engineering is to come to understand the issues, to show how to make better choices when they exist, and to show what the tradeoffs are when, as is the usual case, an improvement in one domain leads to deficits in another." [13]

- "A **cognitive system** performs the cognitive work of knowing, understanding, planning, deciding, problem solving, analyzing, synthesizing, assessing, and judging as they are fully integrated with perceiving and acting. A cognitive system is a distributed system in which people with diverse roles and capabilities, and with the assistance of technological functions, collaborate in the planning and performance of cognitive work. Cognitive work is accomplished by people, either by themselves or in collaboration with others and with the support of technological functions. A cognitive system comprises one or more individuals (necessarily, at least one) and will typically (although not essentially) comprise diverse technological functions." (Lintern, `http://www.cognitivesystemsdesign.net/home.html`)

- Please see also the very good definitions in [14] and [15].

Two figures may illustrate the relationships between the many different disciplines that are related to human factors. Figure 1 seeks to organize these disciplines along two continua, or axes. Figure 2 illustrates the key focus areas of cognitive engineering and methods associated with them.

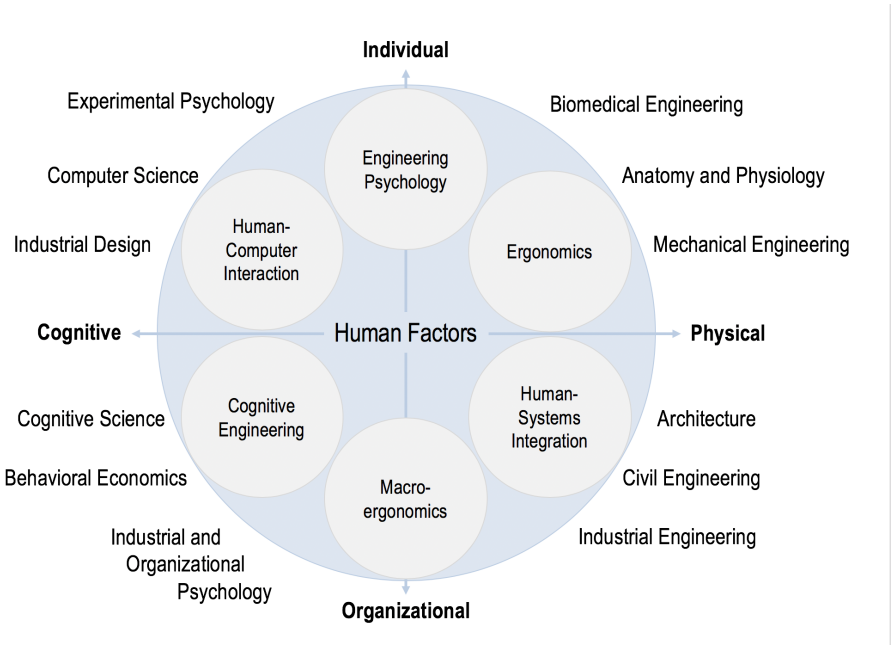


Figure 1. The domains of human factors organized along the continua from individual to organizational (macro) and from cognitive to physical [17]. The disciplines placed in the periphery of this figure represent myriad engineering and psychological disciplines that *ought* to consider human factors within themselves.

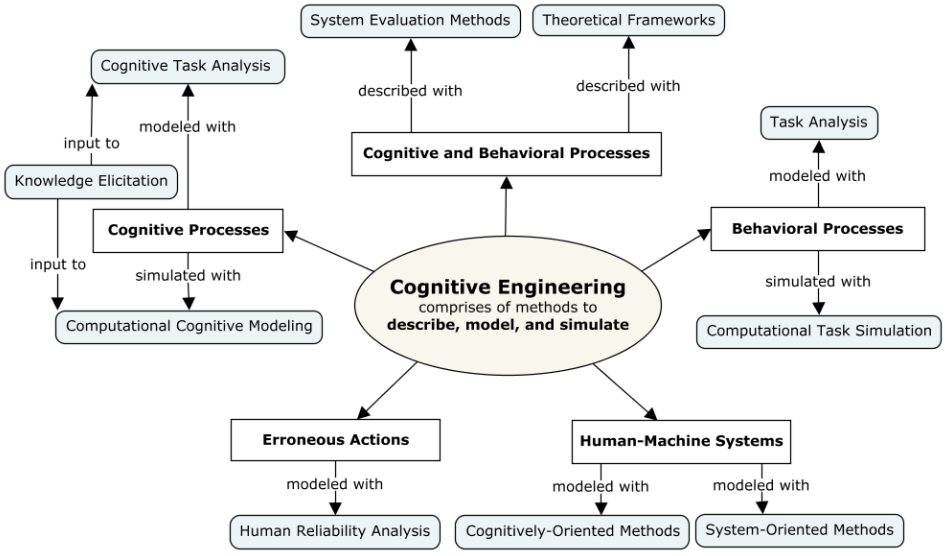


Figure 2. The cognitive engineering discipline with key focus areas and methods. Adapted from [16].

3 Survey of the Field

A useful method to survey the field of engineering psychology and related disciplines is to look at the journals that publish papers on relevant topics. The following contain descriptions of the field as cited in a few of the major journals in engineering psychology:

- *Human Factors: The Journal of the Human Factors and Ergonomics Society* "publishes peer-reviewed scientific studies in human factors/ergonomics that present theoretical and practical advances concerning the relationship between people and technologies, tools, environments, and systems. Papers published in Human Factors leverage fundamental knowledge of human capabilities and limitations—and the basic understanding of cognitive, physical, behavioral, physiological, social, developmental, affective, and motivational aspects of human performance—to yield design principles; enhance training, selection, and communication; and ultimately improve human-system interfaces and sociotechnical systems that lead to safer and more effective outcomes."

- *Journal of Cognitive Engineering and Decision Making* (JCEDM), focuses "on research that seeks to understand how people engage in cognitive work in real-world settings and the development of systems that support that work."

- *Ergonomics in Design: The Quarterly of Human Factors Applications* (EID), is intended to serve the needs of practicing human factors/ergonomics professionals who are concerned with the usability of products, systems, tools, and environments.

- *Ergonomics*, the official journal of the Institute for Ergonomics and Human Factors: "Ergonomics, also known as human factors, is the scientific discipline that seeks to understand and improve human interactions with products, equipment, environments and systems. Drawing upon human biology, psychology, engineering and design, ergonomics aims to develop and apply knowledge and techniques to optimise system performance, whilst protecting the health, safety and well-being of individuals involved. The attention of ergonomics extends across work, leisure and other aspects of our daily lives."

- *The International Journal of Aviation Psychology*, journal of the Association for Aviation Psychology (AAP) publishes "scholarly papers developed within this increasingly important field of study–the development and management of safe, effective aviation systems from the standpoint of the human operators. Four divergent academic disciplines contribute heavily to its contents, making it truly interdisciplinary in nature and scope. These fields are engineering and computer science, psychology, education, and physiology."

- *The Institute of Electrical and Electronics Engineers* (IEEE; pronounced "Eye-triple-E") *Systems, Man, and Cybernetics Society* publishes papers in several areas:

  - Systems engineering including efforts that involve issue formulation, issue analysis and modeling, and decision making and issue interpretation at any of the lifecycle phases associated with the definition, development, and implementation of large systems. It also include efforts that relate to systems management, systems engineering processes and a variety of systems engineering methods such as optimization, decision making, modeling and simulation.
  - Human system and human organizational interactions, cognitive ergonomics, system test and evaluation, and human information processing and decision concerns in systems and organizations.

- Cybernetics including communication and control across humans, machines and organizations at the structural or neural level, as well as at functional and purposeful levels; design and development of biologically and linguistically motivated computational paradigms emphasizing vision, neural networks, genetic algorithms, fuzzy systems, automated planning, computational intelligence, and robotics.
- Applications of these concepts, in terms of hardware and software, to the design, quality assurance, risk assessment and management, development, implementation, systems management, quality assessment and management, and reengineering and systems integration of realistic systems in any of several contemporary application areas.

4 Human-Machine Systems

A **system** is a regularly interacting or interdependent group of items forming an integrated whole. A **human-machine system** is a system in which the functions of a human operator (or a group of operators) and a machine are integrated. This term can also be used to emphasize the view of such a system as a single entity that interacts with external environment. This kind of systems view is common for both engineering psychology and cognitive engineering.

Note that to interact with any system, the human operator needs an *interface*. While cognitive engineers may design the actual system, too, the most important design challenge is usually the interface. An interface must necessarily have two parts: (1) displays that provide information about the system status to the human operator, and (2) controls that allow the operator change the system status.

5 Links

- APA Division 21: `http://www.apa.org/about/division/div21.aspx`
- Division 21 website: `http://www.apadivisions.org/division-21/`
- Human Factors and Ergonomics Society (HFES): `https://www.hfes.org/home`
- International Ergonomics Association (IEA): `http://www.iea.cc`

References

[1] C. D. Wickens. *Engineering Psychology and Human Performance*. HarperCollins Publishers, 2nd edition, 1992.

[2] F. V. Taylor. Psychology and the design of machines. *American Psychologist*, 12(5):249–258, 1957.

[3] P. M. Fitts. Engineering psychology and equipment design. In S. S. Stevens, editor, *Handbook of Experimental Psychology*, pages 1287–1340. John Wiley & Sons, Inc., New York, 1951.

[4] C. D. Wickens and A Kramer. Engineering psychology. *Annual Review of Psychology*, 36(1):307–348, 1985.

[5] E. C. Poulton. Engineering psychology. *Annual Review of Psychology*, 17(1):177–200, 1966.

[6] D. Gopher and R. Kimchi. Engineering psychology. *Annual Review of Psychology*, 40:431–455, 1989.

[7] B. M. Huey, S. Baron, and D. S. Kruser. *Quantitative Modeling of Human Performance in Complex, Dynamic Systems*. National Academies Press, 1990.

[8] T. B. Sheridan and W. R. Ferrell. *Man-machine systems; Information, control, and decision models of human performance.* The MIT Press, 1974.

[9] W. B. Rouse. *Systems engineering models of human-machine interaction*. North Holland, New York, 1980.

[10] R. W. Pew and S. Baron. Perspectives on human performance modelling. *Automatica*, 19(6):663–676, 1983.

[11] W. B. Rouse. Models of human problem solving: Detection, diagnosis, and compensation for system failures. *Automatica*, 19(6):613–625, 1983.

[12] D. D. Woods and E. M. Roth. Cognitive engineering: Human problem solving with tools. *Human Factors*, 30(4):415–430, 1988.

[13] D. A. Norman. Cognitive engineering. In D. A. Norman and S. W. Draper, editors, *User centered system design*, pages 31–61. Erlbaum, Hillsdale, NJ, 1986.

[14] K. M. Wilson, W. S. Helton, and M. W. Wiggins. Cognitive engineering. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(1):17–31, 2013.

[15] J. R. Gersh, J. A. McKneely, and R. W. Remington. Cognitive engineering: Understanding human interaction with complex systems. *Johns Hopkins APL Technical Digest*, 26(4):377–382, 2005.

[16] C. Bonaceto and K. Burns. Using cognitive engineering to improve systems engineering. In *Manuscript submitted for presentation at the 2006 International Council on Systems Engineering Conference*, 2006.

[17] J. D. Lee, C. D. Wickens, Y. Liu, and L. N. Boyle. *Designing for people: An introduction to human factors engineering*. CreateSpace, Charleston, SC, 2017.

DEPARTMENT OF PSYCHOLOGY, COLLEGE OF LIBERAL ARTS
ROCHESTER INSTITUTE OF TECHNOLOGY

PSYC 719—HUMAN FACTORS IN AI

HANDOUT: WHITHER HUMAN FACTORS IN THE AGE OF AI

PROF. RANTANEN

August 25, 2022

1 HF/E: Where do we come from?

The history of human factors as a distinct discipline can be traced to *applied* psychology and production efficiency in the late 1800's and early 1900's. Walter Dill Scott (1869–1955) was the first to apply psychology to advertising, employee selection, and management. Hugo Münsterberg (1863-1916), one of the pioneers in applied psychology, published a book titled "The Psychology of Industrial Efficiency" (1913) with a goal of improving worker efficiency. Scientific management, also known as *Taylorism* after Frederick Winslow Taylor (1856-1915), attempted to improve economic efficiency through labor productivity. Another milestone in the history of applied psychology was the Hawthorne studies, begun in 1924, at the Hawthorne, Illinois, Western Electric Company plant, to study the effects of work environment on employee efficiency.

World War I marked the emergence of human factors as an important discipline. The Army tested for recruits who could read and write as well as recruits who could not read or speak English. Tests were also used to detect neurotic tendencies and in officer and pilot training. However, it was really in the aviation front during WW II where human factors became of age. Consider a WW I fighter plane, Sopwith Camel, with a top speed 115 mph. A mere 30 years later, the P-51 "Mustang" fighter plane used in WW II had a top speed 437 mph. That kind of technological advancement upended the old paradigm of fitting the man to the machine (through selection and training), requiring a new paradigm: fitting the machine to the man, through design and engineering.

The subsequent decades produced a wealth of knowledge in both theory and applications of human factors engineering. The number of these textbooks is too big to be listed here (but ask me; you would do good to collect many of them in your personal libraries). Aviation continued to lead human factors science and practice. The "jet age" and multiplication of aircraft speeds and changes in their handling characteristics continued to challenge the pilots flying them and the engineers designing displays and controls for them. The 1980s saw development of the so-called "glass cocklpits" where all-electronic instruments and fly-by-wire controls replaced previous electromechanical instruments in cockpits and direct, mechanical, linkages between cockpit controls and the control surfaces in the aircraft. At the same time, autopilots became increasingly capable, and complex.

The Airbus family of jetliners led the way in cockpit automation. There was nothing inherently wrong with their airplanes, but the sophistication of automation in them went beyond of what pilots, airlines, and aviation regulators for that matter, were prepared for. There were a string of accidents involving Airbus aircraft caused by pilots' not understanding what the autopilot was doing and putting their aircraft into positions they could not recover from. Although human factors eventually caught up and accidents involving Airbus aircraft are (mostly) history, two very recent aircraft accidents unfortunately repeat the story. On October 29, 2018, Lion Air Flight 610 crashed in Jakarta, Indonesia, and on March, 10, 2019, Ethiopian

Airlines Flight 302 crashed in Nairobi, Kenya. Both involved the Boeing 737 MAX 8 aircraft, which was a recent redesign of the B737 aircraft family, with bigger engines mounted further forward in front of the wings. This required installation of the Maneuvering Characteristics Augmentation System (MCAS), a flight stabilizing program, to the aircraft to counter a tendency of the engines to push aircraft nose up during certain maneuvers. However, Boeing did not think of requiring training for the pilots on the MCAS system, and the regulator (FAA) was not up its job, either.

2 AI-Driven Automation

The above examples illustrate the problems in human-automation interactions (HAI), or the unintended consequences from overreliance on automation. Already in the 1990s terms such as "Clumsy Automation" and "Automation Surprises" were coined. Back then, autopilots were (and still are) straightforward, deterministic systems, yet they can cause operator confusion (as recently illustrated with the Boeing 737 MAX 8 flight accidents). There are two fundamental problems with the next generation, AI-driven, automation:

1. Automation based on AI is often completely opaque and its actions inscrutable by humans. If engineers cannot describe how a system based on machine learning (ML) or AI works, they also cannot provide training material to the people operating the system. Moreover, systems based on ML and AI often evolve on their own in time, making past experience with them moot for training new operators.

2. Human factors has traditionally dealt with relatively small and very specialized populations, mostly well-trained and highly experienced operators of complex technological systems (such as airline pilots, nuclear power operators, medical professionals). Widespread applications of AI in myriad domains and across all societal elements means that unintended and unforeseeable consequences will be felt on a larger, societal, scale than has been a case with earlier technologies. Current and future AI-based automation will penetrate every aspect of people's lives (e.g., IoT, collection of vast amounts of personal data from social media, &c.) requiring that past approaches to human-system integration and cognitive systems engineering must be correspondingly "scaled up". Moreover, increasing human variability across heterogeneous user groups presents additional challenges.

We should also not forget about the "Dark Side" of ML/AI-driven automation. "Built-in" bias in ML is a widely acknowledged problem. In healthcare applications there is seen a rise of false positives due to new and "better" machines and diagnostic aids [1]. Review of Facebook patents reveal how algorithms they used created a "filter bubble", an "an echo-chamber factory" [2]. The so-called "attention economy" or monetizing attention in the form of clicks and websites visited is also a result of AI driven automation in digital media.

The term "artificial intelligence" was coined in the late 1950s with aspiration of human-level intelligence in software and hardware. However, it is very important to separate the reality of AI technology from the hype surrounding it. Most of the hype concerns human-imitative AI, for example humanoid robots and machines that can pass the Turing test. A critical question for interesting discussions is *why*, or for what *purpose*, do we want to imitate human intelligence? Moreover, we are very far from realizing human-imitative AI aspirations [1]. Yet, the thrill (and fear) of making even limited progress on human-imitative AI gives rise to levels of over-exuberance and media attention that is not present in other areas of engineering.

More realistic questions concern the "Intelligent Infrastructure" (II) (e.g., so-called "smart" devices). Do ML (and AI) -driven automation applications work? Are self-driving cars a feasible form of transportation? What has been the impact of business applications and industrial relevance of ML from the early

1990s on (e.g., with Amazon, Google, Facebook/Meta). What are the contributions of AI in engineering applications (e.g., in space flight, document retrieval, text classification, fraud detection, recommendation systems, personalized search, social network analysis, planning, diagnostics)? Where are the human factors contributions to these applications?

3 HF/E: Where are we going to?

To plot the course forward for HF/E in the age of AI, we should ask more fundamental questions that go beyond usability and user experience about the relationships between people and technology. Such questions whould shift the focus in human factors to the *human*, away from *factors*. For example, ask:

- What is a human?
- What is the ultimate purpose of a human?
- Are we using technology for human flourishing or dulling of the human mind?
- What does "human flourishing" mean?

4 Claude Shannon

The life of Claude Elwood Shannon (April 30, 1916–February 24, 2001), an American mathematician, electrical engineer, and cryptographer, may offer an intriguing angle to human flourishing. Shannon is widely known as a "father of information theory", but he may also be considered a "father of a digital computer". In his master's thesis in 1937 at MIT [3] he applied Boolean algebra to the design of logic circuits using electromechanical relays, showing that machines could make decisions using states of "on" and "off" (or 0 and 1). Shannon is best known for his 1948 paper "A Mathematical Theory of Communication", or information theory [4].

Shannon has been quoted of saying "Human is a machine". Consider these two other, contradictory, quotes, by Shannon:

1. "I have great hopes that this direction for machines that will rival or even surpass the human brain. This area, known as artificial intelligence, has been developing for some thirty or forty years... It is difficult to predict the future, but it is my feeling that by 2001 AD we will have machines which can walk as well, see as well, and think as well as we do."
2. "The important people and events of history are the thinkers and innovators, the Darwins, Newtons, and Beethovens whose work continues to grow in influence in a positive fashion."

Given his view that a human is a machine, or that machines might surpass humans in "thinking", it is quite paradoxical to consider the environment where Shannon did his most important work: the Bell Labs in New Jersey. Bell Labs was known for giving people working there complete freedom to do whatever they liked. Shannon was known to ride his unicycle while juggling along the hallways of the lab and otherwise staying in his office behind a closed door, without anyone knowing what he might have been working on [5]. Of course, results of his work emerged at a steady pace. Perhaps Bell Labs may offer us an idea of what might be required for human flourishing. Certainly, Shannon did not lead a very machine-like life there.

References

[1] M. I. Jordan. Artificial intelligence–the revolution hasn't happened yet. *Harvard Data Science Review*, (1.1), 2019.

[2] Bob Garfield. The revolution will not be monetized. *IEEE Spectrum*, 48(6):34–39, 2011.

[3] C. E. Shannon. A symbolic analysis of relay and switching circuits. *Electrical Engineering*, 57(12):713–723, 1938.

[4] C. E. Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[5] J. Soni and R. Goodman. *A mind at play: How Claude Shannon invented the information age*. Simon and Schuster, 2017.

DEPARTMENT OF PSYCHOLOGY, COLLEGE OF LIBERAL ARTS
ROCHESTER INSTITUTE OF TECHNOLOGY

PSYC 719—HUMAN FACTORS IN AI

HANDOUT: DESIGN METHODS IN HUMAN FACTORS

PROF. RANTANEN

August 31, 2022

1 A Systems Approach

The study of human factors is the study of a moving target. Technology is created by humans for human use. However, as technology is designed (properly) to fit human needs, humans also adapt to technology in unforeseeable ways, changing their behaviors to match the often imperfect usability of their technological tools (e.g., workarounds) or misuse or abuse the technology in ways it was never intended to be used [1]. Such *interconnectedness* necessitates systems thinking in human factors engineering, and recognition that design requires tradeoffs with many competing objectives and conflicting guidelines. Moreover, how people interact with technology also depends on the environment and situational factors where the interaction takes place, and these present much variability, further challenging the designers [2].

The systems thinking may be exemplified in systems design by application of the "Five Whys" of Toyota Motor Corporation [3]. Ask "why" (at least) five times to define why the system is being built, or what the purpose of the system is. Also ask "what" (at least) five times to explore what could happen that might not be expected, that is to find out what could go wrong [2].

2 Human Role in a Technological System

What is the human role in a technological system is an all-important question. The answers to this question have ranged from "cogs in machines" of the early industrialization to "spam in a can" in the early US space program (project Mercury). Other suggestions for human role in technological systems are in the Fitts' List [4]:

Humans appear to surpass present-day machines in respect to the following:

1. Ability to detect a small amount of visual or acoustic energy
2. Ability to perceive patterns of light or sound
3. Ability to improvise and use flexible procedures
4. Ability to store very large amounts of information for long periods and to recall relevant facts at the appropriate time
5. Ability to reason inductively
6. Ability to exercise judgment

Present-day machines appear to surpass humans in respect to the following:

1. Ability to respond quickly to control signals and to apply great force smoothly and precisely
2. Ability to perform repetitive, routine tasks
3. Ability to store information briefly and then to erase it completely

4. Ability to reason deductively, including computational ability
5. Ability to handle highly complex operations, i.e. to do many different things at once

Another list relevant to this course is the 10 levels of automation [5]:

1. The computer offers no assistance; the human must do it all
2. The computer offer a complete set of alternatives,
3. ...and narrows the selection down to few,
4. ...or suggests one,
5. ...and executes the suggestion if the human approves,
6. ...or allows the human a restricted time to veto before automatic execution,
7. ...or executes automatically, then necessarily informs the human,
8. ...or informs him or her after execution only if he or she asks,
9. ...or informs him or her after execution if it, the computer, decides to do so.
10. The computer decides everything and acts autonomously, ignoring the human.

However, automation can never be legally, ethically, and socially responsible for its actions. Hence, regardless of systems' scale and sophistication, humans will always have the ultimate responsibility of their operation. Because people will inevitably be responsible for system operation, they must (1) perceive the nature of these responsibilities, and (2) have appropriate levels of authority to fulfill them. That is, people have to be "in charge."

Design objectives should be to support humans to achieve the operational objectives for which they are responsible. Consider these two examples and how they shift the design paradigms: The purpose of a pilot is *not* to fly the airplane that takes people from point A to point B; instead, the purpose of the airplane is to support the pilot, whose responsibility is to take people from A to B. The purpose of factory workers is *not* to staff a factory designed to achieve some productivity goals; instead, the purpose of the factory is to support the workers who are responsible for achieving productivity goals. These are the *human-centered design* objectives.

Contrast the human-centered design with technology-driven design. Technological "fixes" to poorly designed systems always increase complexity, resulting in an "arms race of technological warfare". Radar detectors detected by radar detector detectors, anyone?

Finally, we must not forget the *human contribution* in technological systems in terms of training, discipline, and leadership, and sheer unadulterated professionalism, as well as skill and luck, inspired improvisations, and heroic recoveries from seemingly hopeless situations. Examples of such human contributions include the Apollo 13 flight [6], the United Airlines flight 232 on July 19, 1989 [7], and the "The Miracle on Hudson", or ditching of the US Airways Flight 1549 on Hudson River on January 15, 2009 [8].

3 The Design Process

Human factors is a process. This process varies depending on different purposes. Here, we focus on the design process. There are three major stages in the design process: (1) Front-end analyses. (2) iterative design and testing, and (3) final test and evaluation [9]. It might be argued that the most important stage is the earliest one, for there is little even well-designed interfaces can do if the overall system design is flawed. Figures 1 and 2 illustrate the design process with an example of a medical device development.

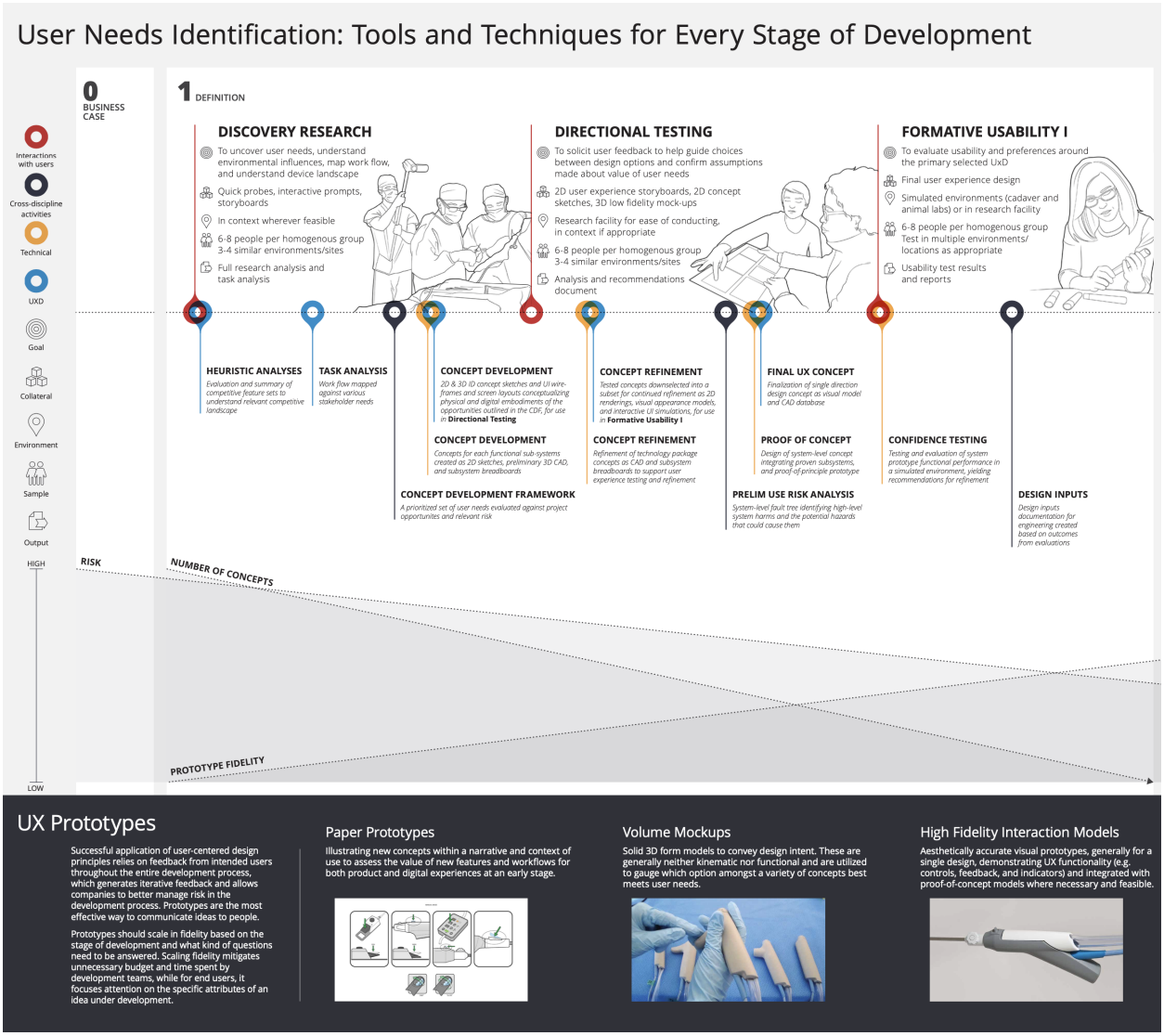


Figure 1. The design process of a medical device from discovery research to the first formative usability stage. Note the timeline and the many analyses and concept development, refinement, and testing activities. Graph courtesy of Alisa Rantanen, Nemera Insight Chicago LLC.

3.1 Front-End Analysis

Front-end analyses are needed to understand the users, their needs, and the demands of the work situation. There are formal methods for each of these analyses.

- **User analysis** seeks to identify and characterize potential system users for each stage of the system life cycle. Note that in case of most AI applications, these users are *everybody*, introducing tremendous diversity and variability to the analysis.
- **Environment analysis** is concerned with the use environments of the system, again through each stage of the system life cycle.
- **Function analysis** is done to identify the general transformations of information and system states to

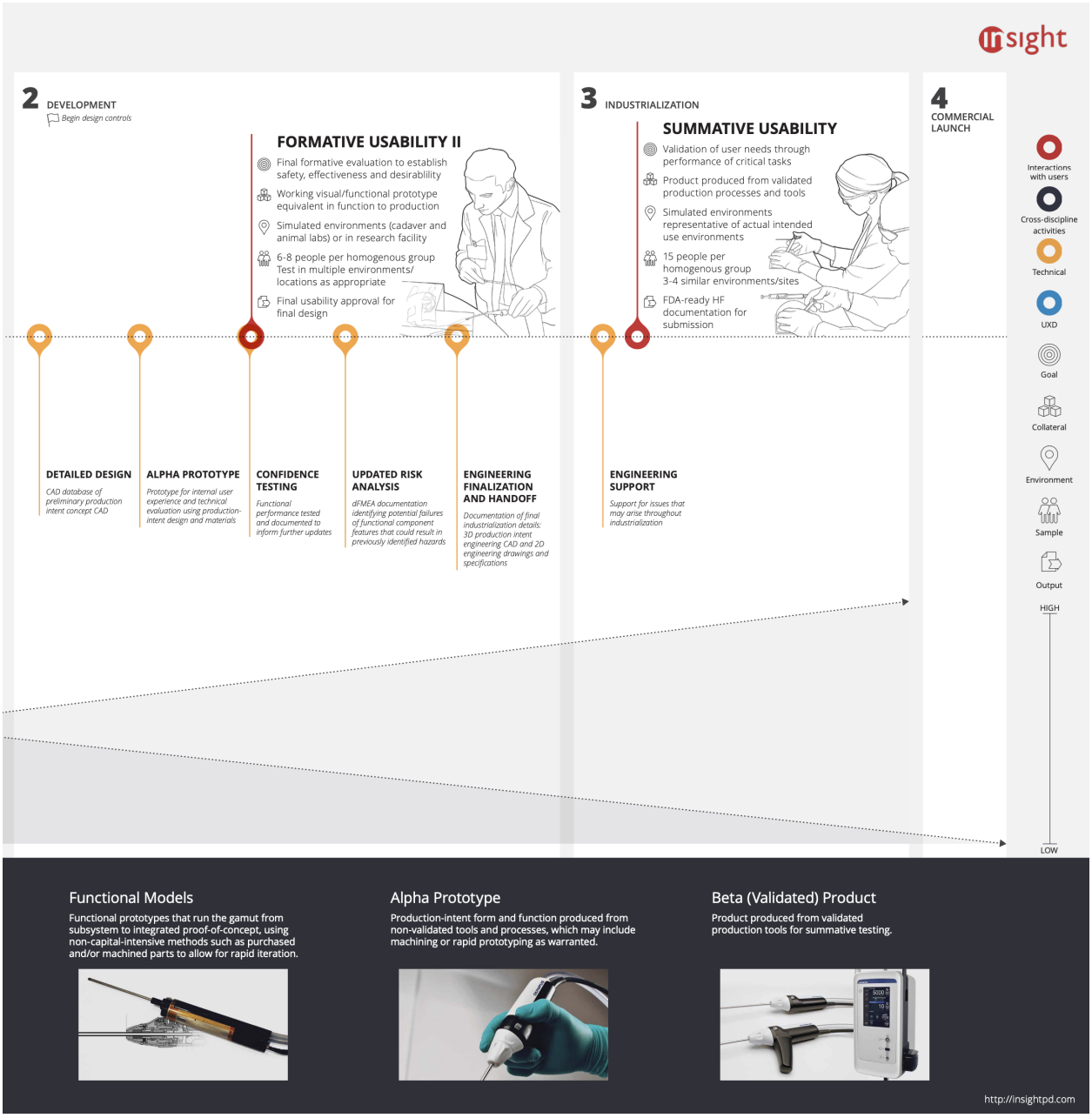


Figure 2. The design process of a medical device through the second formative usability and summative usability stages. Graph courtesy of Alisa Rantanen, Nemera Insight Chicago LLC, LLC.

achieve the system goals. Function allocation is part of this analysis (cf. the Fitts' list and the levels of automation).

- **Task analysis** is perhaps the most important part of front-end analysis and the entire design process, and warrants a bit lengthier discussion (a semester-long course in and of itself, really!).

Task analysis (TA) refers to a *collection* of techniques for describing how people interact with systems. In other words, TA is a way of *systematically* describing a system to better understand how to match the de-

mands of the system to human capabilities. Function analysis is similar to task analysis except that it focuses on basic functions that will be performed by the human-machine system. A function is generally composed of many tasks.

TA seeks to identify and document

- User goals and their associated activities
- Tasks and subtasks necessary to achieve the goal
- Conditions under which tasks are performed
- Outcomes of the tasks and subtasks
- Information or knowledge needed to perform the tasks and subtasks
- Communication with others for performing the tasks
- Equipment needed to perform the tasks

The above documentation are used for

- Training requirements
- Interface requirements
- Process redesign
- System reliability
- Staffing requirements
- Workload estimation

TA uses different methods for

- Data collection
- Data representation (lists, outlines, and matrices)
- Data analysis and modeling
- Observable (physical) and unobservable (cognitive) activity

The point of the above—admittedly rather tedious—lists is to emphasize the fact that TA is not a singular "thing", but that TA does *many* different things, for *many* different purposes, using *many* different methods. The general TA methods can be divided into two classes: (1) methods for data *collection* and (2) methods for data *representation*. Methods for data collection include:

- Document and equipment analysis
- Unstructured and structured interviews
- Group interviews and focus groups
- Sorting and rating tasks
- Direct observation
- Verbal protocol analysis
- Questionnaires and surveys

Methods for data representation include:

- Lists and outlines
- Matrices (cross-tabulation tables)
- Structural and hierarchical networks
- Flowcharts
- Time-line charts
- Concept maps

Link analysis is a simple but very powerful tool that identifies links between an individual and some part of the system. Such a link occurs when a person shifts his or her attention or physically moves between two parts of the system [10].

Flow diagrams are a very useful way to represent various aspects of different tasks. In particular, flow diagrams capture the temporal sequence of tasks better than, for example, hierarchical task analysis.

Hierarchical Task Analysis (HTA) is one of the most useful TA techniques. The HTA produces a hierarchy of operations (what people must *do* within a system) and plans (*conditions* necessary to undertake the operations. HTA may be organized in a graphical form, but as a comprehensive HTA is typically very large, a tabular form is more convenient.

Critical incident technique (CTI) studies both positive and negative incidents. Its premises are that critical incidents will be inherently memorable to those working in a system; therefore, they will be able to recall recentnd even earliervents that were deemed as critical. However, as a prerequisite, "critical" must be defined a priori. Also, although CTI is open-ended and flexible, it is difficult to define exactly what the technique sets out to do.

The above examples work when the performance of the task is *observable*. What about analysis of *unobservable* cognitive tasks? **Cognitive task analysis** (CTA) represents collection of techniques for such situations. CTA provides for description of the cognitive skills needed to perform a task proficiently. CTA, too, includes two parts, (1) knowledge *elicitation* and knowledge *representation*. CTA is valuable for tasks that depend on cognitive aspects of expertise, for example, decision making and problem solving.

The key attributes of CTA are *how* to look (interview, or self-reports, or observation, or automated capture) and *where* to look: Where in time (past, present, or future), where in realism (real-world or simulated scenarios), where in difficulty (routine or challenging tasks), and where in generality (abstract knowledge or specific events). Conducting a CTA is far from a trivial undertaking. The following are the main steps necessary:

1. Preparation: Due to the very involved nature of CTA, careful preparation is a prerequisite for capturing useful data.
2. Knowledge elicitation: Interview-, observation-, modeling- and experimental methods
3. Data analysis
4. Knowledge representation: Decision flow diagrams, expert/novice contrasts, and knowledge audit
5. Application of result: Development of instructional materials, design guidelines, and discovery of interface issues

Knowledge audit examines the nature of the expertise needed to perform work skillfully [11]. Knowledge audit is also useful as an interview technique. Knowledge audit contrasts expert and novice performance and elicits detailed and specific information about interesting incidents. Knowledge audit covers 8 aspects of expertise; below are also example questions, or knowledge audit probes, to be used in interviews:

1. Past and future: "Is there a time when you walked in the middle of a situation and knew exactly how things got there and where they were headed?"
2. The Big Picture: "Can you give me an example of what is important about the big picture for this task? What are the main elements you have to know and keep track of?"
3. Noticing: "Have you had experiences where part of the situation just 'popped' out at you? Where you noticed things others did not catch? What are the examples?"

4. Job smarts" “When you do this task, are there ways of working smart, or accomplishing more with less, that you have found especially useful?”
5. Opportunities, improvising: “What are some examples when you have improvised in this task, or noticed an opportunity to do something more quickly or better, and followed up on it?”
6. Anomalies: “Can you describe an instance when you spotted a deviation from the norm, or knew something was amiss?”
7. Equipment difficulties: “Have there been times when the equipment pointed in one direction, but your judgment told you to do something else? Or when you had to rely on experience to avoid being led astray by the equipment?”

Consider also a “scenario from hell”: “If you were to devise a scenario to show someone what this job is really about, what would you put in that scenario? Have you had experiences along the way that have made you humble about performing this task?” Other probes may include cognitive demands, difficult cognitive elements (why they are difficult?), common errors, and cues and strategies used. All questions serve to sharpen the contrast between novices and experts. The goal is to understand and describe expert performance.

Self-report data are affected by demand pressures, inaccuracies about participant judgment, and analyst variance. They generally have good reliability and validity, but be careful about asking for introspection of cognitive processes themselves.

Behavioral measures are affected by method variance and obtrusiveness. They are generally flexible, general, logical, and linked to training interventions. However, behavioral measures have emphasis on procedures, may oversimplify complex tasks, and—obviously—are less relevant to tasks with cognitive components.

Cognitive work analysis (CWA) is a framework for systematic and detailed analysis of cognitive work. As the name implies, its scope is broader that CTA, and one can see CTA as a component of CWA. CWA consists of 6 distinct stages [12].

1. **Functional Work Structure** or **Work Domain Analysis** (WDA) is a technique within CWA that creates a representation of a socio- technical systems work domain, known as the abstraction-decomposition space (ADS). The ADS identifies the important, activity-independent structure of the work domain, to aid researchers in understanding the necessary values and priorities, work functions, technical functions, and physical resources to fulfill the domain purpose of the complex socio-technical system.

   The purpose of an ADS is to identify aspects of a work domain that either support the achievement of the domain purpose or constrain against it. The typical ADS representation portrays the domain purpose as the final element composed of more detailed levels that follow in a hierarchical fashion. The domain purpose is listed at the top of the representation, followed by the domain values and priorities, the work function to obtain the values and priorities, the technical functions necessary to fulfill the work functions, and ending in the physical resources required to fulfill the technical functions (either people for socio-technical systems, or technological components for technical systems). Within each of the aforementioned levels of the ADS, functions of the work domain are placed as nodes. Links between nodes at different levels represent means-ends relationships between the linked nodes.

2. **Partitioning and Organization of Work** or **Work Organization Analysis** (WOA) focuses on domain functions, as identified in the ADS, work situations, which are the various situational contexts

in which work takes place, and work tasks, which are the distinctive outcomes to be achieved. The product of this stage of analysis is a Contextual Activity Matrix.

3. **Cognitive Transformations Analysis** examines cognitive states established during task execution, and cognitive processes used to effect the transitions between states. The product of this stage of analysis is a suite of decision ladders, originally developed by Rasmussen [13]. Decision ladders are an extremely powerful tool to investigate just how operators perform their tasks and what information they need to do so.

4. **Cognitive Strategies Analysis** and

5. **Cognitive Processing Analysis** focus on the reasons that a worker may select one strategy in preference to another or may transition between strategies during execution of a cognitive process and identifies the skills-, rules, or knowledge-based modes of cognition [14]. In ecological interface design, the designer may choose to encourage and induce one cognitive mode over another as dictated by the situation (e.g., in an emergency response, skill-based performance is preferred). The products of these stages of analysis are detailed description of potential strategies and of the factors that will prompt selection of one strategy over another, as well as of the activity elements associated with the different modes of cognitive processing.

6. **Social Transactions Analysis** results in a Social Transactions Matrix, which maps agents (either human or technological or some combination) to Work Tasks and maps Work Tasks to Transaction Demands and Transaction Modes. A second product is a Transaction Network in which the transactions between agents (either human or technological) are identified and characterized in terms of fundamental or generic properties relevant to design.

Figure 6 shows the 6 stages of CWA, what products follow from each process, and how each product may be used to inform design.

References

[1] R. Parasuraman and V. Riley. Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2):230–253, 1997.

[2] J. D. Lee, C. D. Wickens, Y. Liu, and L. N. Boyle. *Designing for people: An introduction to human factors engineering*. CreateSpace, Charleston, SC, 2017.

[3] T. Ohno and N. Bodek. *Toyota production system: beyond large-scale production*. Productivity press, 2019.

[4] P. M. Fitts, M. S. Viteles, N. L. Barr, D. R. Brimhall, G. Finch, E. Gardner, W. F. Grether, W. E. Kellum, and S. .S Stevens. Human engineering for an effective air-navigation and traffic-control system, and appendixes 1 thru 3. Technical report, Ohio State Univ Research Foundation, 1951.

[5] T. B. Sheridan and W. L. Verplank. Human and computer control of undersea teleoperators. Technical report, Massachusetts Inst of Tech, Man-Machine Systems Lab, 1978.

[6] J. Kluger and J. Lovell. *Lost Moon: The Perilous Voyage of Apollo 13*. Houghton Mifflin, 1994.

[7] National Transportation Safety Board (NTSB). Aircraft Accident Report, United Airlines Flight 232, McDonnell Douglas DC-10-10, Sioux Gateway Airport, Sioux City, Iowa, July 19, 1989. Technical Report NTSB/AAR-90/06, 1990.

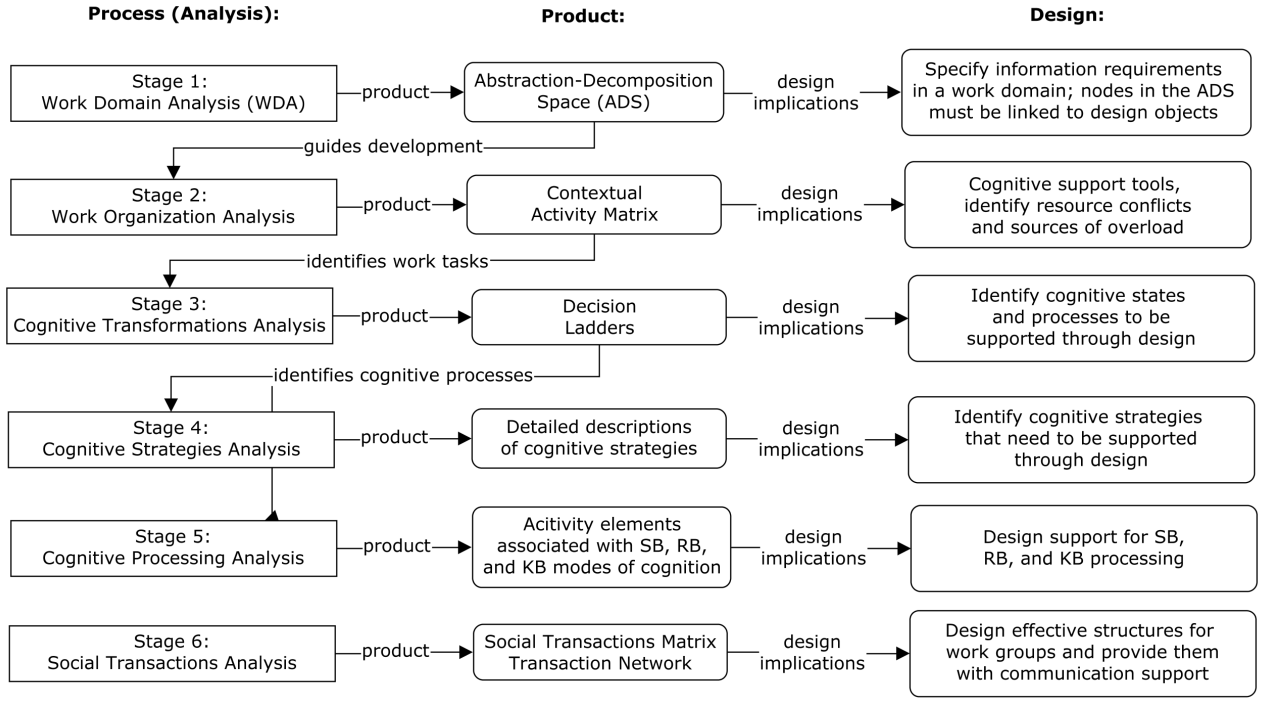


Figure 3. The 6 stages of CWA, showing the products that follow from each process, and how each product may be used to inform design.

[8] C. Sullenberger, J. Zaslow, and M. McConnohie. *Highest duty: My search for what really matters*. HarperAudio, 2009.

[9] C. .D Wickens, S. E. Gordon, Y. Liu, and J. D. Lee. *An introduction to human factors engineering*, volume 2. Pearson Prentice Hall, 2004.

[10] B. Kirwan and L. K Ainsworth. *A guide to task analysis*. CRC press, 1992.

[11] B. Crandall, G. A. Klein, and R. R. Hoffman. *Working minds: A practitioner's guide to cognitive task analysis*. MIT Press, 2006.

[12] G. Lintern. The foundations and pragmatics of cognitive work analysis: A systematic approach to design of large-scale information systems. http://www. cognitivesystemsdesign.net/home.html, 2009.

[13] J. Rasmussen, A. M. Pejtersen, and L. P. Goodstein. *Cognitive systems engineering*. Wiley, 1994.

[14] J. Rasmussen. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man and Cybernetics*, (3):257–266, 1983.

DEPARTMENT OF PSYCHOLOGY, COLLEGE OF LIBERAL ARTS
ROCHESTER INSTITUTE OF TECHNOLOGY

PSYC 719—HUMAN FACTORS IN AI

HANDOUT: EVALUATION METHODS IN HUMAN FACTORS

PROF. RANTANEN

September 7, 2022

1 Formal Evaluation Methods

There is much evaluation going on in the system or product design process, as I hope was clear in the previous handout. In this handout I will attempt to provide an introduction to the more formal evaluation methods, to discover how people actually interact with technological systems. Evaluation methods may be classified (roughly) into three main categories [1]:

1. **Understand** how to improve the system. These evaluations include open-ended surveys and task analysis methods, and collect both qualitative and quantitative data;
2. **Diagnose** problems with prototypes. The methods include heuristic evaluation, cognitive walk-throughs, and usability testing. The data collected are mostly quantitative;
3. **Verify** that the system meets its design criteria or performs better than benchmark systems. These methods yield mostly quantitative data.

1.1 Review of Relevant Literature

Any scientist and engineer stands on the shoulders of those who came before them. The work of those who allow the scientist to add to the body of knowledge on any subject, or an engineer to develop new and improved systems and products, must be rightly acknowledged. Furthermore, for any scientific research to make a contribution to the body of knowledge on any topic, it has to be presented in a *context* of past and current research on the same and relevant subjects. This requirement for context also necessitates several reviews of relevant literature throughout the research, each progressively more focused than the previous one.

1.2 Heuristic Evaluation

Heuristic evaluation involves application of heuristics, that is, rules-of-thumb, principles, guidelines, and standards to the system under development. Although this evaluation method may sound daunting, note that it relies on "ready-made" materials that have been developed and tested on myriad projects before. Equally important to system development is to also identify where it may violate established design principles. Heuristic evaluations are best suited for interface design [1]. A quick intro into heuristic evaluation is a free poster of Jakob Nielsen's *10 Usability Heuristics*, available for download here: `https://www.nngroup.com/articles/ten-usability-heuristics/#poster`.

1.3 Cognitive Walkthroughs

Cognitive walkthroughs are best suited for *interaction design*. This method involves a user of the system to try each task top be performed with the system under development answering questions of whether the user understands exactly what to do throughout the interaction, how to do the task, and notice the next task to be performed [1]. Cognitive walkthroughs may be conducted on prototypes of various functionality, from simple (non-functional) mock-ups to "black box" systems that lack the ultimate functionality but may be programmed to mimic the final system demands and responses.

1.4 Usability Testing

Usability belongs to a domain in and of itself, and you can take several courses on usability and usability testing right here at RIT. Usability refers to "ease of use" of a system or product and has the following common dimensions [2]:

1. **Learnability**; the system should be easy to learn so that training time may be minimized and productive work maximized;
2. **Efficiency** for maximum productivity;
3. **Memorability** to minimize retraining demands after periods of disuse;
4. **Errors**; the users should make few errors in the system use, but if they make an error, they shod be able to recover from it with minimal conseqeunces;
5. **Satisfaction**; the system should be pleasant to use so that users feel satisfied using the system.

2 Measurement

Evaluation methods can range from micro (e.g., neuroscience dealing with biological and biochemical phenomena) to macro (e.g., sociology focusing on societal systems). Measurement considerations must reflect these scales.

It is very useful (essential?) to think about evaluation in terms of *variables*. At least four categories of variables may be identified. It is critical that variables are classified correctly into these categories:

1. Independent (or treatment-, or stimulus-, or predictor-) variable(s): Levels of independent variables are established by the experimenter before the experiment and are thus independent of anything that happens in the experiment (i.e., they are *manipulated*)

   Note that the objective of research is to establish causal relationship between independent and dependent variables. That is a tall order, indeed, but it can be done if the study is designed well. More about that later.

2. Dependent (or or criterion-, or response) variable(s): If a relationship exists, the values of the dependent variable will depend on the level of the independent variable.

3. Subject variables: In psychological research human subjects bring with them innumerable variables that *may* also affect the dependent variable(s) in addition to—or instead of—the independent variable. These variables must be identified and *controlled* for the research to be valid.

4. Extraneous Variables: Strictly speaking, any variable that is not treated as an independent variable and manipulated accordingly would be classified as an extraneous variable. In this presentation the distinction between subject and extraneous variables is that the latter are present in the environment or experimental task and—just as uncontrolled subject variables—may *confound* independent variables and render the research (or rather, conclusions drawn from data) invalid.

2.1 Operational Definition

Operationalization of variables is defined as representation of ideas or concepts in terms of specific behaviors or concrete activities that anyone can witness or repeat. Operationalization of variables takes the critical task of defining your variables a step further: It is essential for manipulation of independent variables and *measurement* of dependent variables. In other words, operational definitions are necessary to creation of *testable* hypotheses.

Operational definition of a variable means describing the variable in terms of the operations (procedures, actions, or processes) by which it could be observed and measured. In other words, a construct of interest (say, mental workload) must be defined in terms of *observable* behaviors or physiological responses that independent observers may all observe and values of which they should agree (in case of mental workload, heart rate variability, or pupil dilation, or galvanic skin response). Note below how operationalization determines the construct validity in research.

Operational definitions also help answer two critical questions:

1. What can we measure? Much of behavioral sciences deal with theoretical *constructs*. By definition, a construct (n.) is a concept, model, or schematic idea. How do you measure an idea?

2. What do the measurements mean? By definition, measurement = assigning numbers to things. So, what do the numbers mean? (i.e., how to interpret them?)

Operational definitions of the dependent variables and development of appropriate measures are all-important for the construct validity of research. Does the measure really measure the desired variable?

2.2 Measurement Error

There are many other criteria for measures. Measures should be objective, quantitative, unobtrusive, easy to collect, inexpensive, reliable, valid, free from contamination, and sensitive (i.e., free from floor- or ceiling effects).

However, it is important to keep in mind that *all* measures contain *error*. In other words, the observed value = true value + random error. Random error is also known as *variable* error, and there is little one can do about it. Variable error may stem from myriad sources, ranging from an observer's inability to make accurate or consistent observations to inherent inaccuracies in instruments. Another type of error is *constant* error, or emphbias. Although bias may be just as big—or bigger—threat to validity of research, it is, if known, easier to deal with (to correct) than variable error. For example, if a scale consistently reads 2 lbs too high, simply subtract that 2 lbs from the weight of the people using the scale.

2.3 Reliability and Validity

A measure is *reliable* if it yields the same value when repeated on the same object (subject) under the same conditions. A measure is *valid* if it measures what it is supposed to measure. A measure that is not reliable

cannot be valid. However, reliability of a measure does not guarantee its validity. Hence, reliability is a necessary but *not sufficient* condition to validity.

Validity comes in multiple flavors. Three of the most important varieties of validity are:

1. **Construct validity** refers to the degree to which the researchers (a) manipulated the independent variable they wanted and (b) measured the dependent variable they wanted. Note how both of these depend on operational definitions of the variables.

2. **Internal validity** refers to the situation or experiment where the causal or independent variable, and no other extraneous variables, caused the change in the dependent variables being measured. It is often impossible to know about the existence of confounding variables and their influence on results.

3. **External validity** refers to the degree to which the experimental results can be generalized to other situations, tasks, settings, and people. Threats to external validity include unrealistically simple experimental task and subject sample that is unrepresentative of the target population (cf. "All we know about psychological theory we have learned from American undergraduate college students").

2.4 Measurement Scales

There are four scales of measurement. It is critical to know exactly what scale a measure belongs to to know what operations (e.g., statistical) are permissible on the data.

1. **Nominal** scale: Identity. For example, numbers on baseball playersniforms, employment status. Nominal measures only assign *labels* on things.

2. **Ordinal** scale: Identity and magnitude. For example college football polls are on ordinal scale and they tell us that one team may be better than another, but not by how much.

3. **Interval** scale: Identity, magnitude, and equal intervals. For example, temperature on the Fahrenheit and Celsius scales is measured on the interval scale. We may say that a daytime high of, say, 60 F is 30 degrees warmer than the low of 30 degrees, but it is nonsensical to say that the high was *twice* as warm as the low because the scales continue to negative degrees.

4. **Ratio** scale: Identity, magnitude, equal intervals and absolute zero. Examples include length, weight, force, response time, number of responses, &c.

3 Experimental Methods

The experimental method is the best way to establish a *causal* relationship between two variables. We may define *experimentation* broadly as "a systematic study design to examine the consequences of deliberately varying a potential causal agent" [3]

There are two noteworthy keywords in this definition: First, the study must be *systematic*. That means that the variables are chosen based on some *theoretical* understanding of the phenomenon under study, and the manipulation of the independent variables should be informed by their effect *predicted* by theory.

The second keyword is *deliberate*. This means that the experimenter deliberately "messes with" the system to bring about some effect.

Pause for a moment to think how powerful this is: A theoretical *causal* link between two variables may be *tested* by manipulating one of the variables (the independent variable) and observing the effect on the other (the dependent variable). If no effect is observed, the theory is *wrong* (provided that the experiment was run properly).

However, it is important to add a qualifier to the above statement: A causal link between variables exist if manipulation of the independent variable results in observable effect on the dependent variable, *all other things being equal*. Keeping all other things equal is one of the most difficult problems in experimental research.

3.1 Features of Experiments

Three features of true experiments are [4]:

1. They test a hypothesis that makes a causal statement about the relations among variables;
2. A comparison of the dependent measure is made at least for two levels of an independent variable;
3. High level of control.

Note how very restrictive these features are. There are many interesting research questions that simply cannot be examined experimentally. However, even in such cases the features of true experiments may be used to create quasi-expermental designs and to improve the validity of the research.

3.2 Steps in Conducting an Experiment

The following are generic steps that are necessary for any true experiment.

1. Define the research problem and formulate hypotheses.
2. Operationally define the independent (treatment) variables
3. Operationally define the dependent (response) variables. Operational definitions allow for unambiguous empirical testability of hypotheses and theories (cf. falsifictionism; also, working definitions).
4. Define the expected relationship between in dependent and dependent variables as supported by theory.
5. Specify the experimental plan.
6. Check for the exact definitions of "all of the above".
7. Choose an experimental design.
8. Determine the equipment, task(s), environment, participants.
9. Conduct the study (determine if a pilot study is necessary)
10. Analyze data, moving from exploratory data analysis (look at the data) to descriptive statistical analysis and to inferential statistical analysis.
11. Draw conclusions. Also ask why the results turned out the way they did.
12. Formulate new questions for further research.

3.3 A Graphical Method for Tracking Experimental Variables

This handout demonstrates a method for tracking different variables (independent-, dependent-, subject-, and extraneous), their *provenance*, their derivatives, and their uses in hypothesis testing. The method is graphical, and the examples provided here were created with the CmapTools concept mapping software (`http://cmap.ihmc.us/products/`).

The basic, generic, structure of a diagram describing the variables of interest and their relationships is depicted in Figure 1. Note that the leftmost "column"may be read down: experimental stimuli are presented to participants whose responses are recorded. At each stage, different variables are identified (reading to the right).

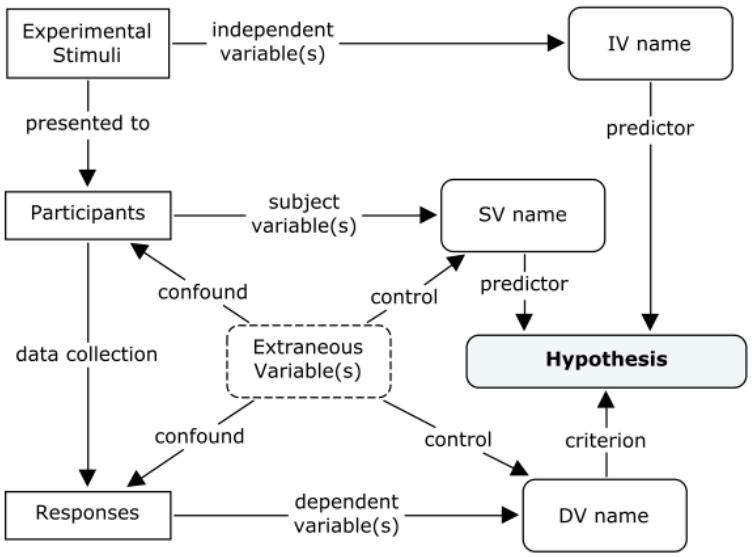


Figure 1. The generic structure of a graphical representation of variables and their relationships in research.

Manipulation of experimental stimuli represents the *independent* variable. For example, two different kinds of stimuli may be presented, and it may me hypothesized that one kind of stimuli result in different responses than the other kind. Participants may have characteristics that may also influence the results; these are *subject* variables (e.g., age, or sex). Responses are measured to yield *dependent* variables (e.g., response time). Finally, any number of *extraneous* or *nuisance* variables may be identified that may *confound* the results and that need to be *controlled*.

All of the above variables come together in hypothesis testing. Independent variables (and also subject variables, if they are of interest) are *predictors*. That is, they make a *prediction* about how their manipulation may impact the dependent variable. The prediction is tested against the dependent variable; did the data indeed differ depending on the condition they were recorded from? The dependent variable this becomes the *criterion* against which the hypothesis may be tested.

3.4 Control

Recall the definitions and varieties of validity. Control pertains to *internal* validity and is used to eliminate of control *threats* to internal validity. In an experiment, one manipulates an independent variable in order to measure its effect on a dependent variable. For the experiment to be valid, the only factor operating should be the independent variable. *Confounding* occurs when two variables that can influence the dependent variable are allowed to covary simultaneously, for example, when the experimenter accidentally manipulates the subjects in an unintended fashion, or the subjects are influenced by merely being in an experiment and this is mistaken for a treatment effect, or the groups are selected such that there is a bias between groups. Controls are used to fix all other variables than the one manipulated (i.e., to prevent covariation of external variables with the independent variable) or to measure potential covariates to account for their influence.

3.5 Threats to Internal Validity

This is an non-exhaustive list of some of the most common threats to internal validity:

1. **History**: Events external to the experiment that unfold in time might influence measurements taken in the course of the experiment. Examples include external factors affecting sensory processes (e.g., lighting conditions, ambient noise, clothing), factors influencing emotions (e.g., major news events such as terrorism, election results, sports, financial events and stock market), and factors affecting psychophysiological measures (e.g., fatigue, circadian rhythms, drugs broadly understood, incl. caffeine, nicotine, alcohol, as well as prescription drugs).

2. **Maturation**: Changes in participant characteristics during the experiment For example, the participants may become more tired, bored, experienced, &c. while participating in the experiment.

3. **Testing**: Problems with test-retest reliability, for examples, repeated blood pressure measurements, and learning and memory effects.

4. **Instrumentation**: Instruments may drift from calibration during the experiment. For example eye-trackers may lose the pupil during an experiment.

5. **Statistical regression** (also, regression to the mean): Extreme scores on any measure at one point in time will, for purely statistical reasons, probably return less extreme scores the next time they are measured, accounting for inherent variability apart from independent variable manipulation. Examples include selection of participants when they feel ill and pseudoscientific claims of "miracle cures".

6. **Selection**: Groups in between-subjects experiments should be homogenous

7. Mortality: Some participants sharing some characteristics drop out from the experiment. For example, say you want to study the effects of violent TV on mood for over a period of a week. You set up a protocol where your subjects have to watch certain programs. It turns out that women dislike watching violent TV and do not watch the shows. If you do not notice this, your results would be confounded by gender leading to differential mortality. Your results would only be valid for men.

8. Selection-maturation interaction.

9. Diffusion or imitation of treatments.

3.6 Experimental Designs

Design of Experiments (DOE) is a branch of applied statistics that deals with planning, conducting, analyzing, and interpreting controlled tests to evaluate the factors that control the value of a parameter or group of parameters. The primary purpose of DOE is control, and through control the validity and reliability of the data collected from the experiment. The experimental design must be such that the experimenter (as well as the audience of there research) may be convinced that *only* the experimental manipulation of treatment can be responsible for any observed effects on the dependent variable.

There are probably innumerable different, valid experimental designs. The actual design should reflect the research problem at hand. In other words, choose or develop a design that best allows for testing of your hypothesis and control of extraneous variables rather than forcing the study to a potentially inappropriate design. However, what might be called "standard" designs offer true and tested safeguards against threats to validity and are often closely associated with established statistical tests (e.g., factorial designs and Analyses of Variance). The point is that choosing or developing of an experimental design requires lots of thought! The following is just a *sample* (!) of few of the most common ones. Note how each of the designs achieve *control*.

3.7 Between-Subjects Design

These designs have two or more groups of participants who receive different treatment (e.g., a treatment group and a control group). The pros and cons of between-subjects designs include:

+ Sometimes necessary, for example, when differences between independent variables are beyond manipulation, or when carryover effects cannot be effectively controlled;

+ Include a control or comparison group that helps to control for many extraneous variables.

+ Can randomly assign participants to groups, which helps to balance extraneous variables across the groups.

- Large samples needed; for example, if we agree that the minimum N is 30, then a two-group study needs 60 participants, and a three-group study 90 participants. Think about the logistics of recruiting that many volunteers and running the experiment.

- One cannot always randomly assign participants to groups, which may introduce confounds.

- Large individual differences between participants may make it difficult to observe meaningful group differences.

3.8 Within-Subjects or Repeated Measures Design

Many of the cons with between-subjects designs can be addressed by within-subjects experiments. In within-subjects designs the same participant is exposed to all the experimental conditions, and each participant is only compared to him or herself under different conditions, not to other participants.

For example, say we are running an experiment testing two different interfaces and measuring the time-on-task, hypothesizing that a new and improved interface B would result in faster task completion times. Let us examine just two participants: S1 is very quick, and gets the task done in 4.8 s with the old interface A and in 4.1 s with the new interface B. Another participant, S2, produces the following data: 6.1 s for A and

5.5 s for B. This participant is much slower than the first, and the data shows large between-subjects variability. However, we do not need to worry about that because we we only compare the differences between the interfaces, which are A – B = 0.7 s for S1 and A – B = 0.6 s for S2 (both are positive, indicating faster performance with B). These data show very little variability and would be tested against the null hypothesis A – B = 0.

The pros and cons of within-subjects designs include:

+ Increased sensitivity, have more statistical power (e.g., matched t-test).
+ Require fewer participants.
- Carryover effects (e.g., practice and fatigue).
- Sometimes differences of interest cannot be measured within subjects (for example, variables such as age and gender).

A few more words about carryover effects in repeated measures designs. If all subjects performed the experimental task in all three conditions (say, B1, B2, B3) in the same order, then treatment may be confounded with practice if practice indeed helps performance. Practice effects can be mitigated by counterbalancing conditions or equalizing the amount of practice. For example, you could test one person in the B1, B2, B3 order, another with B1, B3, B2, until every order is tested. Repeating trials so that people are tested like B1, B2, B3, B3, B2, B1, which would also balance practice effects. Or, one could randomize the order of conditions; performing the tasks in different orders will minimize the effects of practice, fatigue, boredom, etc. carryover effects. However, counterbalancing and randomization of trials necessitate more subjects.

3.9 Mixed Designs

The experimental design does not have to be either between- or within-subjects. For example, only one variable may be repeated and the other is a between-subjects variable (can be applied to any number of IVs). Mixed Designs are useful when one might end up testing the subjects too much in a wholly within-subjects design (too fatiguing, too much time required from the subjects). Another example: An independent variable is a subject variable like gender and not manipulatable or the two levels would interact. If A1 and A1 were gender (Male vs. Female), one could not use a totally repeated design. Similarly, if one were testing two drugs (A) over the period of three days, B would be days (1,2,3) but A1 and A2 would have to be different groups in most cases.

3.10 Multilevel Designs

Although the minimum of 2 levels of an independent variable may be adequate for establishing a causal relationship between it and the dependent variable, 3 or more levels required to discover the type of relationship, that is. whether the relationship is linear or non-linear. Multilevel designs should be used when the independent variable is continuous rather than categorical. One should always perform pairwise (Post Hoc) comparisons between multiple levels, too.

3.11 Factors and Designs

Factors are the main independent variables (i.e., to be manipulated). Levels are subdivisions of factors (i.e., actual manipulations). We can distinguish several *types* of factors:

1. Treatment factors: Any subjects can be assigned to any one of the levels of the factor (randomly). The different levels of the factor consist of explicitly distinguishable stimuli or situations in the environment of the experimental unit.

2. Control factors: Systematization of possible effects of extraneous variables.

3. Trial factors: Possible change of scores over trials. There are as many levels of trial factors as there are trials, to distinguish between treatment and trial effects.

4. Blocking factors (see above).

5. Group factors (differences between groups).

6. Random factors: The results are intended to be generalized beyond the experimental design. The factor levels are determined by a random choice from a very large population of levels. For any replication, the levels of the factor should be determined by a new random selection. Examples include subjects and error.

7. Fixed factors: Results are not to be generalized, or any generalizations are subjective. Levels of the factor can be determined by any procedure. Replication of the experiment requires exactly the same levels of the factor as the original experiment. Examples include treatment and trial.

8. Crossing factors: To examine interactions between factors. For example: Factor A with 2 levels and Factor B with 3 levels: $2 \times 3 = 6$ combinations: AB11, AB12, AB13, AB21, AB22, AB23. Whenever 3 or more factors are crossed, all combinations at all levels must occur. For example, factors A, B, and C with 2 levels each:results in $2 \times 2 \times 2 = 8$ combinations: ABC111, ABC112, ABC121, ABC211, ABC212, ABC221, ABC122, ABC222.

9. Nested factors: Factor B is nested within Factor A if each meaningful level of factor B occurs with only one level of Factor A; B(A): B = nested factor, A = nest factor. For example, factor A with 2 levels and Factor B with 3 levels: B(A): 4 combinations: AB11, AB12, AB13, AB21

3.12 Factorial Designs

Factorial designs combine two or more independent variables each with two or more levels to examine the effect of each independent variable (main effects) and interactions between independent variables (interaction effects). Factorial designs allow for "packing" multiple factors into a single experiment, thus making the research more efficient. However, be careful about "mushrooming" designs; the simplest factorial design is a 2 (levels of factor 1) $\times$ 2 (levels of factor 2) = 4 experimental conditions. But say you think of a third factor of interest, also at 2 levels; then the design will have $2 \times 2 \times 2 = 8$ conditions to run. Or what if we suspect nonlinear relationships between the factors (independent variables) and dependent variables and want to manipulate each factor at 3 levels, which will yield a $3 \times 3 \times 3 = 27$ unique conditions to run. For a between-subjects design that is 27 *groups*, or with a within-subjects design *each* participant has to be run through 27 conditions!

In naturalistic settings there are always not only multiple factors that affect dependent variables of interests, but the factors also *interact* bringing about effects they separately would not. Factorial design therefore allow for a better mimicking realistic situations in experimental settings than single-factor designs. Note that interaction effects must be interpreted first, and if an interaction exists, the main effect must be discussed with a qualifier.

3.13 Some Other Considerations

As important as careful DOE is, it is not a panacea, necessarily—or sufficiently—guaranteeing validity of the research. There are several, higher-level, considerations to be always kept in mind to guide formulating research questions, theses, and hypotheses, and the DOE itself:

1. Common sense in the design of experiments;
2. Objectives of the research;
3. Hypothetical relationships between variables and their interactions, supported by theory;
4. Practical considerations, including interpretability of results from complex experimental designs.

References

[1] J. D. Lee, C. D. Wickens, Y. Liu, and L. N. Boyle. *Designing for people: An introduction to human factors engineering*. CreateSpace, Charleston, SC, 2017.

[2] J. Nielsen. *Usability Engineering*. Morgan Kaufmann, 1994.

[3] W. R. Shadish, T. D. Cook, and D. T. Campbell. *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin, Boston, MA, 2002.

[4] W. J. Ray. *Methods Toward a Science of Behavior and Experience*. Cengage Learning, 10th edition, 2011.

[5] C. D. Wickens, S. E. Gordon, and Y. Liu. *An Introduction to Human Factors Engineering*. Longman, New York, 1998.

DEPARTMENT OF PSYCHOLOGY, COLLEGE OF LIBERAL ARTS
ROCHESTER INSTITUTE OF TECHNOLOGY

PSYC 719—HUMAN FACTORS IN AI

HANDOUT: THE LOGIC OF EXPERIMENTATION

PROF. RANTANEN

September 12, 2022

1 A Primer on Logic

I would like to present an argument (see definition below) that all research should be based on a solid philosophical foundation and sound and explicit logic. The materials in this section are from [1]. Before elaborating, I must define the terms I use (note how each definition requires defining the terms it uses):

1.1 Definitions

1. *logic* (n.)
   1. A system, or mode, or reasoning
   2. A particular system or codification of the principles of proof and inference: Aristotelian logic.

2. *reasoning* (n.)
   1. "Movement of the mind from premises to a conclusion".
   2. Deductive reasoning: Reasoning from at least one general (or universal) premise to (usually) to a more particular conclusion.
   3. Inductive reasoning: Reasoning from particular premises to a more general (or universal) conclusion.
   Note: Inductive reasoning can yield only probability. Deductive reasoning (if correct) yields certainty.

3. *argument* (n.) [2]
   1. A course of reasoning aimed at demonstrating truth or falsehood: presented a careful argument for extraterrestrial life.
   2. A fact or statement put forth as proof or evidence; a reason: The current low mortgage rates are an argument for buying a house now.
   3. A set of statements in which one follows logically as a conclusion from the others.

4. *proposition* (n.)
   1. Proposition is a statement that expresses a concept that can be true or false.
   2. For example, consider the following proposition: "All men are created equal"; what is its *truth value*?

5. *syllogism* (n.)
   1. Formal arguments.
   2. A form of deductive reasoning consisting of a major premise, a minor premise, and a conclusion;
   3. For example: All humans are mortal (the major premise), I am a human (the minor premise), therefore, I am mortal (the conclusion).

6. *propositional logic*: Deductive reasoning from theory (the premise) to data (the conclusion) (cf. inductive reasoning), containing antecedents and consequents: If p (antecedent), then q (consequent).

7. *Modus Ponens* (confirmatory): If p, then q. p. Therefore, q.

8. *Modus Tollens* (disconfirmatory): If p, then q. Not q. Therefore, not p.

   (Logical Fallacies: Affirming the Consequent and Denying the Antecedent).

1.2 Thinking Critically

I strongly urge you to present your research questions and theses in terms of a formal argument (syllogism), for reasons that I hope will be clear in the following. I also urge you to reformulate scholarly articles you read in terms of formal syllogisms. For that matter, it would be a really good idea to reformulate *everything* you are asked to believe (e.g., advertising, political messages) as formal syllogisms.

This method allows for an easy check of whether you should believe the argument, or whether your audience should believe the argument you are making. An argument that is *necessarily truth-preserving* (NTP) is called *valid*. NTP means that

1. the truth of the premises guarantees the truth of the conclusion, or

2. it is impossible for the premises all to be true and the conclusion not to be true, or

3. there is no way for the premises all to be true without the conclusion being true.

An argument is *sound* if it is valid and, in addition, has premises that are all in fact true, that is, valid + all premises true = sound [3]. You should make sound arguments, as well as believe sound arguments and reject unsound arguments.

Hence, you should make these three checks of any deductive argument:

1. All terms must be clear and unambiguous;

2. All the premises must be true;

3. The argument must be valid and logically sound.

So, the three questions you should habitually ask of yourself when writing or speaking, and of others when reading and listening to them, are:

1. Are the terms clear and unambiguous?

2. Are the premises all true?

3. Is the reasoning valid?

1.3 The MAGIC criteria

Finally, consider the MAGIC criteria or Abelson's [4] as checks of research papers you read, and as guidelines for your own writing:

1. Magnitude: Statistical effect size, but also practical size of the difference between groups, or experimental conditions, as measured by the dependent variable(s).

2. Articulation: How much detail, in plain English, does the statement of the conclusions reveal? A "happy medium" is probably a good guideline, i.e., not so much detail that the point is lost and grammar threatened, but enough to allow the reader to understand what the results mean.

3. Generality: This answers (partially) the inevitable "So what?" question. How can the results of this research be applied? Do the findings generalize to settings outside the laboratory where the experiment was conducted? To populations other than the participants? To tasks other than the experimental one?

4. Interestingness: A second part of the answer to the "So what?" question. If the generality addressed the applied significance of research, interestingness refers to its theoretical significance. Both are important; research should be based on sound theory (see also credibility below) and also advance our knowledge of the phenomenon under investigation in terms of theory, as well as make practical difference.

5. Credibility: Research claims are believable if the research has been methodologically sound and theoretically coherent.

1.4 Scientific Truths

Science makes four claims about scientific truth [5]:

1. Rationality: Rational methods of inquiry use reason and evidence correctly to achieve substantial and specified success in finding truth, and rational actions use rational and true beliefs to guide good actions.

2. Truth: True statements *correspond* with *reality*; in other words, there is *correspondence* between the external physical world of objects and events and the internal mental world of perceptions and beliefs.

3. Objectivity: Objective beliefs concern external physical objects that can be tested and verified so that consensus will emerge among knowledgeable persons and they do not depend on controversial presuppositions or special worldviews (cf. *operational definitions*).

4. Realism: Realism is *correspondence* of human thoughts with external and independent reality, including physical objects.

5. I should add one more criterion for scientific truths: Theories should be *coherent*, both within themselves as well as with other relevant theories. *Coherence* is defined as (a) systematic or logical connection or consistency, and (b) integration of diverse elements, relationships, or values [6]. In most scientific endeavors, the "Holy Grail" is considered to be "a theory of everything", that is, a theory that explains every phenomena and shows how they are related, in a *coherent* manner.

2 An Example

The following is a very simple example (simple enough to be demonstrated in class), but I hope you see the value of expressing research problems very systematically using formal logic. The example is based on the PEL model: Presuppositions (P) + Evidence (E) + Logic (L) $\rightarrow$ Conclusions. [5]

1. Flip a coin; if heads, place the coin in the cup and cover with a lid, if tails, place the coin in pocket and cover the empty cup with a lid. Question: Is there a coin in the cup?

2. A hypothesis set, (list of all possible answers):

   H1: There is a coin in the cup.
   H2: There is not a coin in the cup.

   Note that these are mutually exclusive hypotheses; the truth of either implies falsity of the other. They are also jointly exhaustive; they cover all of the possibilities.

3. How to test the hypotheses? Peek in the cup!

   Premise: We see a coin in the cup.
   Conclusion: There is a coin in the cup.

   Formally, if "S" is seeing and "E" is the coin's existence in the cup, then S; therefore E. But this is a non sequitur; the conclusion does not follow from the premise. Another premise is required, that seeing implies existence (or that seeing is believing). A valid argument is therefore S; S implies E; therefore E.

4. We may now present the argument completely: If seeing implies existence, and we see a coin in the cup, there is a coin in the cup.

   Premise 1 (Presupposition): Seeing implies existence.
   Premise 2 (Evidence): We see a coin in the cup.
   Premise 3 (Logic): Modus ponens.
   Conclusion: There is a coin in the cup.

3 Conclusion

In conclusion, I hope that this handout prepares you to answer two questions you should ask yourselves as you do research or read about others' research:

1. "Is that true?" and
2. "How do you know that it is true, or that it is not?"

Additional questions about applied research, or engineered systems:

3. "Does it work?" and
4. "How do you know that it works, or that it does not?"

You should think of the above questions and habitually ask them of yourself when writing or speaking, and of others when reading and listening to them.

References

[1] P. Kreeft. *Socratic Logic*. St. Augustine's press, South Bend, IN, 3rd edition, 2008.

[2] *The American Heritage Dictionary of the English Language*. Houghton Mifflin Harcourt Trade, 5th edition, 2011.

[3] N. J. J. Smith. *Logic: The laws of truth*. Princeton University Press, 2012.

[4] R. P. Abelson. *Statistics as principled argument*. Psychology Press, 2012.

[5] H. G. Gauch. *Scientific method in practice*. Cambridge University Press, 2003.

[6] Merriam-Webster. Merriam-Webster Online Dictionary. https://www.merriam-webster.com.

DEPARTMENT OF PSYCHOLOGY, COLLEGE OF LIBERAL ARTS
ROCHESTER INSTITUTE OF TECHNOLOGY

PSYC 719—HUMAN FACTORS IN AI

HANDOUT: VISUAL PERCEPTION

PROF. RANTANEN

September 20, 2022

1 A Primer on Neuroanatomy

Before we get to the particular topics in human sensation, perception, and cognition, we need to briefly review the biology behind them, that is, the human brain. The brain is part of the central nervous system, which also includes the spinal cord. Consider what is known about the anatomy of the human nervous system:

- There are 2 major types of cells in the nervous system: **glia** and **neurons**. The main function of glia cells is to support the neurons by supplying them with nutrients and removing waste material. In the human brain, there are about ten glia cells for every neuron. **Neurons** are cells that receive, integrate, and transmit information.
- Neurons may be further classified into sensor neurons, motor neurons, and interneurons. In the human nervous system, the vast majority are **interneurons**, that is, neurons that communicate with other neurons. **Sensory (afferent) neurons** receive signals from outside the nervous system. **Motor (efferent) neurons** carry messages from the nervous system to the muscles that move the body.
- The **soma**, or cell body, contains the cell nucleus and much of the chemical machinery common to most cells.
- **Dendrites** form a branched structure called a **dendritic tree**. Each individual branch is a dendrite. Dendrites are specialized to receive information.
- The **axon** is a long fiber like structure specialized in transmitting information to other neurons or to muscles or glands. Most human axons are wrapped in a **myelin sheath**. Myelin is a white, fatty substance that serves as an insulator around the axon and speeds the transmission of signals.
- The axon ends in a cluster of **terminal buttons**, which are small knobs that secrete chemicals called **neurotransmitters**. These chemicals serve as messengers that may activate neighboring neurons.
- The points at which neurons interconnect are called **synapses**. The neural impulse is a signal that must be transmitted from a neuron to other cells. This transmission takes place at special junctions called synapses, where terminal buttons release chemical messengers.
- The two neurons are separated by the **synaptic cleft**, a microscopic gap between the terminal button of one neuron and the cell membrane of another neuron. Signals have to cross this gap for neurons to communicate.
- As an neural impulse is transmitted from a neuron to other cells by neurotransmitters at terminal buttons, neurotransmitters are stored in small sacs, called **synaptic vesicles**.

- The neurotransmitters are released when a vesicle fuses with the membrane of the presynaptic cell and its contents spill into the synaptic cleft. After their release, neurotransmitters diffuse across the synaptic cleft to the membrane of the receiving cell.When a neurotransmitter and a receptor molecule combine, reactions in the cell membrane cause a postsynaptic potential, or PSP, a voltage change at the receptor site on a postsynaptic cell membrane.

2 Taking Stock

Consider the complexity anatomy and physiology of a single neuron, as described above. Now *multiply* this complexity with the following numbers:

- The adult human brain is estimated to contain 86 ± 8 billion[1] neurons and about an equal number (8 ± 10 billion) of non-neuronal cells [1]. Some textbooks still estimate the number of neurons in the human brain to exceed 100 billion [2]. The bottom line is that we *do not know* much about the microanatomy of the human brain.

- There are several different *types* of neurons. Some types of neurons in the spinal cord may be identified by their specific function, e.g., *afferent* or sensory neurons neurons that convey information from tissues and organs into the central nervous system, *efferent* or motor neurons that transmit signals from the central nervous system to the effector cells (e.g., muscle cells), and *interneurons* that connect neurons within specific regions of the central nervous system. However, in the brain itself it is not so simple, and by some estimates there may be as many as 10,000 *specific types* of neurons in the human brain.

- Neurons communicate with each other through hundreds of trillions[2] of *synapses*. The human brain is estimated to have up to 500 trillion (5×10^{14}) synapses [3]. At synapses an electrical impulse traveling along one neuron is relayed to another, either enhancing or inhibiting the likelihood that the second nerve will fire an impulse of its own. One neuron may make as many as tens of thousands of synaptic contacts with other neurons [4].

- Synapses may be electrical or chemical. At a chemical synapse, one neuron releases neurotransmitter molecules into a small space called the synaptic cleft that is adjacent to another neuron. The neurotransmitters are kept within small sacs called synaptic vesicles, and are released into the synaptic cleft by exocytosis. These molecules then bind to neurotransmitter receptors on the postsynaptic cell's side of the synaptic cleft. Finally, the neurotransmitters must be cleared from the synapse.

- Neurotransmission implies both a convergence and a divergence of information. Convergence of input refers to one neuron being influenced by many others. Divergence refers one neuron sending a signal to many other neurons.

- The total number of different neurotransmitters is not known, but it is estimated to be well over 100 [5]. The main *classes* of neurotransmitters are amino acids (glutamate, aspartate, D-serine, γ-aminobutyric acid (GABA), glycine); gasotransmitters (nitric oxide, carbon monoxide, hydrogen sulfide); monoamines (dopamine, norepinephrine, epinephrine, histamine, serotonin); trace amines (phenethylamine, N-methylphenethylamine, tyramine, 3-iodothyronamine, octopamine, tryptamine, etc.); peptides (somatostatin, substance P, cocaine and amphetamine regulated transcript, opioid peptides); purines (adenosine triphosphate, adenosine); and others (acetylcholine, anandamide, etc.).

[1] A billion = a thousand million, or 1,000,000,000, or 10^9.

[2] A trillion = a million million, or 1,000,000,000,000, or 10^{12}.

- One synapse may contain on the order of 1,000 molecular-scale switches. Thus, a single human brain, which weighs between 1.2 and 1.4 kg (2.6–3.1 lbs, or about 2% of the total body weight, and a volume of around 1,260 cm^3 in men and 1,130 cm^3 in women, or about 1.2 quarts, with substantial individual variation) has more switches than all the computers and routers and Internet connections on Earth combined (but please check my math!).

So, what are we to make of the awesome numbers cited above? The first, and most appropriate, response I am looking is indeed *awe*. The second response that should immediately follow from the first is deep *humility* as we continue to study human factors in this course.

Neuroscience attempts to provide *compositional* explanations of human behavior and cognition. Causal and transitional explanations, which are familiar from experimental psychology, are an alternative approach to the science of the brain. Another approach has been use of *functional* explanations. This approach was driven by the developments in computer science and artificial intelligence (AI) in the 1970s and 1980s and the view that the mind is analogous to software. If one could understand the characteristics and rules of the software, the "rules of the mind", one could explain human cognition. It would not be necessary to consider the hardware (i.e., the biological brain) that runs the software. "Accordingly, the rules of mind could as easily be studied in a computer as they could in a human. The actual hardware, the structures and mechanisms of the brain, were deemed unimportant" [6]. Given the immense complexity of human brain, however, it should be clear that none of these explanations alone will suffice to account for all of human cognition, behavior, and performance in all circumstances.

Although metaphorical thinking (e.g., the brain as a human information processor) can be very useful in understanding complex things, one should be careful not to mistake a model for the reality being modeled. There are powerful arguments *against* the use of information processing metaphors in cognitive psychology, but we simply do not have better language or vocabulary at our disposal [7]. Therefore, whenever I use words like "processing", "storage", "representation", or "retrieval" in my lecture or these handouts, keep in mind that I am using metaphorical language that is removed from reality. Also, always remember this: "All models are wrong, but some are useful [8]."

3 Perception Preliminaries

Perception is a *huge* topic that could easily occupy an entire semester-long course. In fact, *visual* perception alone would be worthy of its own course. In the context of human factros in AI, a much shorter review of the main concepts and phenomena will have to suffice. Let us first examine three fundamental concepts that are common to *all* perception, across *all* senses.

The first is the *perceptual continuum* (Fig. 1). The first stage implies that perception is indeed dependent of an external, objective world (i.e., there must be something to be perceived). The second stage emphasizes the *physics* of perception. In the example in Figure 1 that is light, which is necessary for visual perception. The third stage emphasizes the *anatomy and physiology* of the organism, that is, the sensory organs (in this case the eye). Finally, the last stage contains the *perceptual object*, or the distal, real, object as it is perceived.

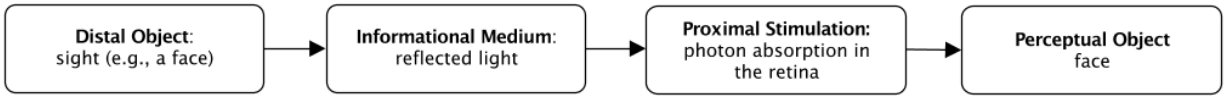


Figure 1. The perceptual continuum for vision.

The other fundamental concepts concern the perceptual *cycle*, which may also be divided into *bottom-up* and *top-down* parts. The perceptual cycle in Figure 2(a) [**?**] suggests that our perception does not happen in isolation but is closely integrated with the environment we find ourselves in as well as "higher" cognition. We can also see that the percpetual cycle can be divided in two parts in Figure 2(b). *Bottom-up* processing is driven by the external stimulus-world; for example, hearing some loud noise behind us we cannot help turning our heads to the direction of the noise. *To-down* processing is driven by our cognition, intentions, routines we follow, &c.; for example, if you want to know what time it is as you read this handout, you will move your eyes away from these words to the nearest timepiece.

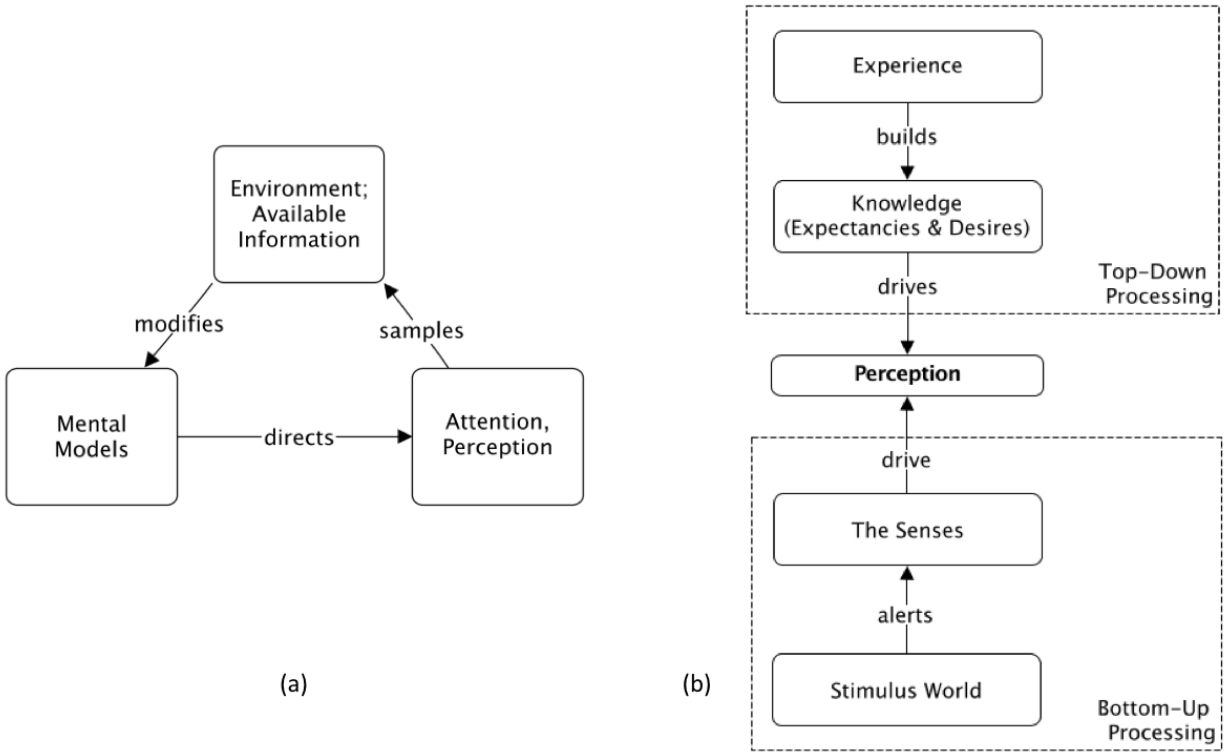


Figure 2. The perceptual cycle (a) and the bottom-up and top-down views of perception (b).

We also need a few initial definitions:

- **Exteroception**: Information about the layout of the environment, its objects and events;
- **Proprioception**: Information about the positions, orientations, and movements of body parts relative to each other;

- **Exproprioception**: Information about the position, orientation, and movement of the body as a whole relative to the environment;

4 Visual Perception

4.1 Physics of Light

The informational medium necessary for our vision is light. Therefore, we need to understand some basics about the physics of light. What is light? Visible light occupies small part (range of 380 nm–760 nm) of the electromagnetic spectrum, which extends from very long radio waves (1 mm–100 km) to γ rays ($< 10^{-12}$ m). The wavelength of visible light determines the color of light, with red the longest and violet the shortest wavelength.

Measurement of light is also complex. Here are some of the common measures:

- **Luminous flux** measures the total light output of a lamp or luminaire. Unit = lumen (lm). For example, GLS 100W = 1400 lm, SOX 70W = 6000 lm.
- **Luminous intensity** is a measure to describe the power of a light source to emit light in a given direction. Unit = candela (cd). For example, a candle = 1 cd, 100W GLS =110 cd, sun = 3×10^{27} cd.
- **Illuminance** defines the luminous flux that is received per unit area of a surface. Unit = lumen meter^{-2} (lm/m^2) or lux (lx). For example, sunlight = 100,000 lx, overcast = 10,000 lx = office 500 lx = street 10 lx = moon 0.5 lx = stars 0.2 lx. Here, pause for a minute to marvel the fact that our visual system works so well across such a vast range of illuminance; you can manage quite well in moonlight and even starlight!
- **Luminance** (L) is defined as the luminous intensity of a surface in a specific direction, divided by the projected area as viewed from that direction and it depends on on the reflective characteristics of the material. Unit = candela meter^{-2} (cd/m^2). Luminance is what finally determines our visual perception.

Note, too, that we do not perceive brightness consistently. Perceived brightness, B, is given by an equation knows as *Weber's Law*:

$$B = aI^{0.33} \tag{1}$$

where a is a constant and I is light intensity. If you plot the equation, it looks like in Figure 3 below:

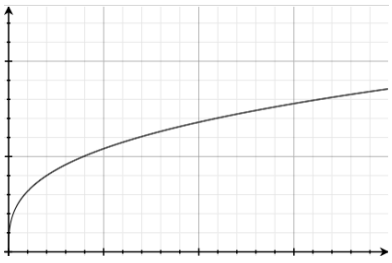


Figure 3. Weber's Law, where perceived intensity is on the y-axis and light intensity on the x-axis. The law states that we become less and less sensitive to changes in light intensity at higher intensities.

4.2 The Anatomy of the Eye

In the next stage on the perceptual continuum (proximal stimulation) we need to understand the anatomy and physiology of the visual sense organ, that is, the eye. The major parts of the eye are:

- **Pupil**, which egulates the amount of light entering the eye (c.f., aperture of a camera);
- **Lens**, which is a focusing system (c.f., focusing a camera). The lens changes its shape to accommodate, or focus the image on the retina;
- **Retina** in the back of the eyeball contains two types of photoreceptors: rods and cones. Some unique properties of rods and cones and their non-uniform distribution across the retina have important implications for visual sensory processing. **Fovea** is a small depression in the retina where visual acuity is highest and where the center of the field of vision is focused.

4.3 The Retina

The cones and rods correspond to two very different visual systems, *photopic* under good luminance and *scotpic* (under low luminance. Here are some of the most important properties of the cones in the retina, and their implications to *photopic* vision:

- Cones have three different photopigments $\rightarrow$ color sensitivity;
- Cones are much less sensitive to light than rods $\rightarrow$ photopic vision is useful only under good luminance conditions;
- Cones have direct individual connections to the brain $\rightarrow$ they can resolve fine detail, or have good *acuity*;
- Most importantly, cones are non-uniformly distributed, concentrated in the fovea $\rightarrow$ the area of high visual acuity is very narrow, only 1-2 degrees of visual angle, necessitating moving the eye from detail to detail (e.g., from word to word or sullable to syllable when reading).

Compare the above to the most important properties of the rods in the retina, and their implications to *scotopic* vision:

- All rods contain the same photopigment (rhodopsin) $\rightarrow$ the rods cannot distinguish colors;
- The rods are very sensitive to light, and absorption of light results in insensitivity (bleach) that requires regeneration $\rightarrow$ our scotopic vision is largely useless under conditions of high luminance;
- The rods are also grouped together in the primary neural networks $\rightarrow$ our scotopic vision is *in*sensitive to fine detail, but sensitive to motion;
- Most importantly, rods are on-uniformly distributed in the retina, and nearly absent in fovea $\rightarrow$ our peripheral vision is dominated by rods.

Some further properties of the photoreceptors is that while rods and cones have different sensitivity to color, as seen above, cones, too have different sensitivity to color: They are more sensitive to red light than to blue light, meaning that blue light must have higher intensity to be perceived as bright as red light. What are the implications of this property of cones to the design of aircraft instrument lighting for night flight?

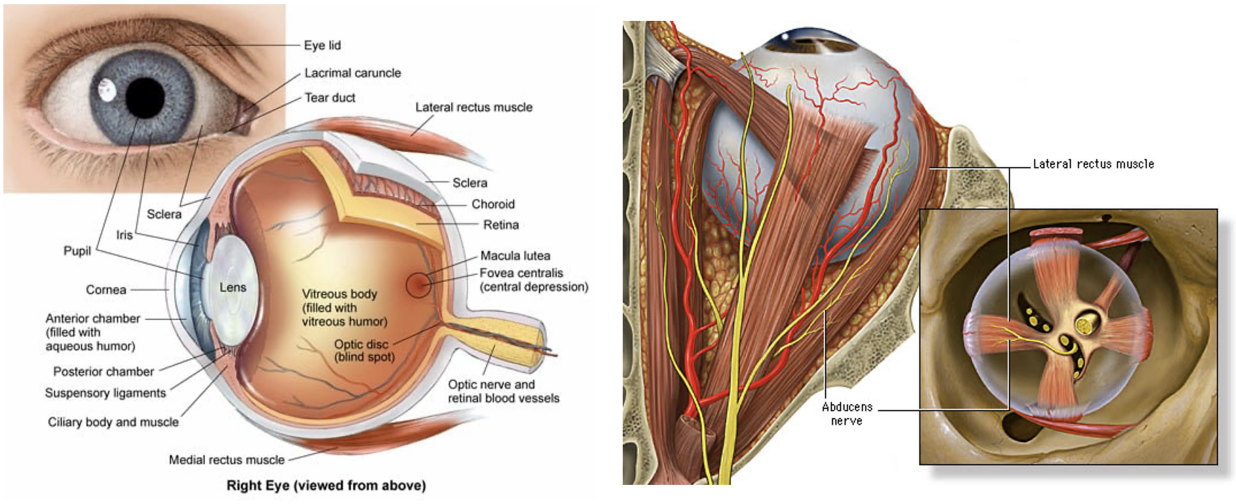


Figure 4. The anatomy of the eye.

4.4 Visual Acuity

The visual angle subtended by an object, VA, is given by the equation

$$VA = arctan\left(\frac{H}{D}\right) \tag{2}$$

where H is the height of the object to be seen, or a detail to be resolved (e.g., the gap in a Landolt ring) and D is the viewing distance.

Visual acuity is a *relative* measure. A person with an acuity of 20/15 can see as well at 20 feet as a person with 20/20 acuity at 15 ft. For example, a person with 20/15 vision is looking at an object from 100 feet away; where would a person with 20/20 vision need to stand to see the object just as well? A: 75 feet away from the object (15/20 $\times$ 100 feet = 75 ft). Contrast Sensitivity ($CS = 1/C_M$) depends on spatial frequency, contrast, and illumination. The CS function has a shape of an inverted U.

4.5 Eye Movements

Because of the non-uniform distribution of cones and rods in the retina, and the concentration of cones with high resolution in the fovea, we must have the ability to point our eyes at objects we want to see, or to "foveate" the object. The eyes are moved in their sockets by no fewer than 6 *extraocular muscles*: Medial and Lateral Rectus muscles, Superior and Inferior Rectus muscles, and Superior and Inferior Oblique muscles.

Because muscles only produce force by *contracting* (i.e., they can only "pull"), these pairs of muscles allow the eye to move up and down and left to right, with the oblique muscles providing additional movement and stability. The extraocular muscles produce the following *classes* of eye movements:

- **Saccadic** eye movements are rapid movement of the eye from a point of interest to another.
- **Smooth pursuit** eye movements are when we maintain fixation on some object and move our heads around, or track a bird or an airplane, or the Superman flying across the sky while keeping our heads stationary.

- **Microsaccades** and small saccadic eye movements while maintaining fixation; their purpose is to allow photopigments in cones to be regenerated during fixations.
- **Convergent** eye movements are when the eyes are moving to *opposite* directions, for example, when fixating on something very close to the eyes.
- **Nystagmus** refers to rapid saccadic eye movements and are of two types: *Optokinetic* nystagmus is when we try to track objects, for example, from a fast-moving vehicle on a roadside; our eyes keep jumping ahead every so often as he details pass by our visual field. ,phVestibular nystagmus occurs when we are spun around rapidly and suddenly stopped; our inner ear will interpret the stopping as spinning in thje opposite direction, with the eyes attempting to fixate on the surroundings.

4.6 3-D Vision

How is it that although the image of the external world lands on a two-d surface on the retina, we nevertheless perceive it as three-dimensional as the world is. We do this because of a number of *cues* available to us. There are three *classes* of cues for depth perception:

- **Object-centered** cues are linear perspective, interposition height in the plane, light and shadow, relative size, textural gradients, and proximity-luminance covariance. These cues are also used in arts, to create an illusion of a three-dimensional image from a painting on a two-dimensional canvas. Art students are certainly very familiar with these cues! Note, too, that these cues are *top-down* cues, that is, they have to be learned.
- **Observer-centered** cues include binocular disparity (steroscopic vision), convergence (see above), and accommodation (see also above). These are in a sense "hard-wired", or *bottom-up* cues.
- **Egomotion** cues include *motion parallax*, that is, when objects at different distances seem to move at different speeds and different directions when the viewer is moving him- or herself, and *optic flow*, which refers to the movement of details in the visual scene of the environment as the viewer is moving through the environment, emanating from the point the viewer is moving towards. These cues, too, are top-down.

5 Taking Stock, Again

Please compare and contrast the human visual system to any camera, or computer vision (please do; I look forward to learning much about computer vision from you!). Moreover, make sure you think about the human visual system *in context* (see Fig. 5), and as a part of a much larger *system*.

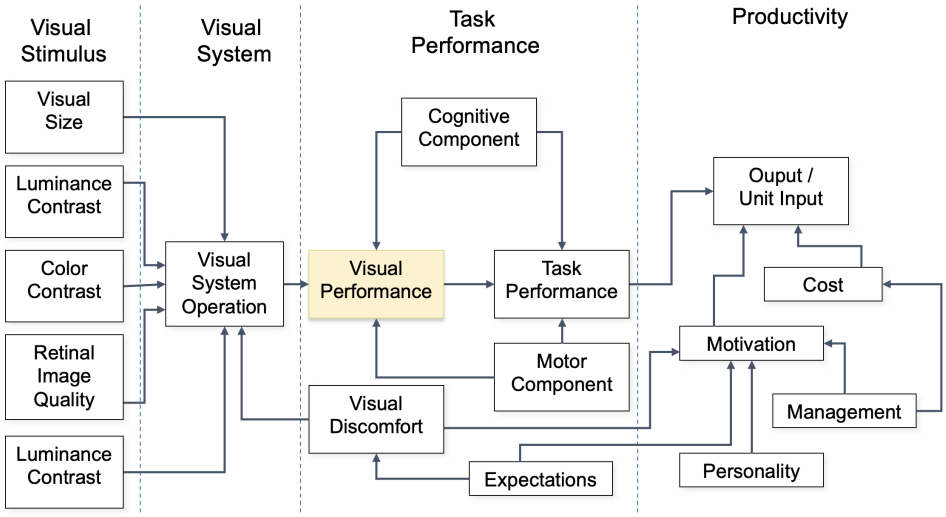


Figure 5. Human visual performance in context.

References

[1] F. A. C. Azevedo, L. R. B. Carvalho, L. T. Grinberg, J. M. Farfel, R. E. L. Ferretti, R E. P. Leite, R. Lent, S. Herculano-Houzel, et al. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5):532–541, 2009.

[2] J. R. Anderson. *Cognitive psychology and its implications*. Worth Publishers, New York, NY, 6th edition, 2005.

[3] D. A. Drachman. Do we have brain to spare? *Neurology*, 64(12):2004–2005, 2005.

[4] K. D. Micheva, B. Busse, N. C. Weiler, N. O'Rourke, and S. J. Smith. Single-synapse analysis of a diverse synapse population: proteomic imaging methods and markers. *Neuron*, 68(4):639–653, 2010.

[5] Purves D., Augustine G. J., Fitzpatrick D., et al. *Neuroscience, Chapter 6: Neurotransmitters (*`https://www.ncbi.nlm.nih.gov/books/NBK10795/`*)*. Sinauer Associates, Sunderland, MA, 2nd edition, 2001.

[6] R. Parasuraman. Neuroergonomics: Research and practice. *Theoretical issues in ergonomics science*, 4(1-2):5–20, 2003.

[7] R. Epstein. The empty brain. *Aeon.com*, 18, 2016.

[8] G. E. Box. Robustness in the strategy of scientific model building. In R. L. Launer and G. N. Wilkinson, editors, *Robustness in Statistics*, pages 201–236. Academic Press, New York, 1979.

DEPARTMENT OF PSYCHOLOGY, COLLEGE OF LIBERAL ARTS
ROCHESTER INSTITUTE OF TECHNOLOGY

PSYC 719—HUMAN FACTORS IN AI

HANDOUT: AUDITORY AND OTHER SENSORY PROCESSES

PROF. RANTANEN

September 28, 2022

1 Auditory Processes

Despite the different sensory modalities, auditory perception has very much in common with visual (and for that matter, other senses, too: haptic, olfactory, taste). Therefore, the review below repeats the format of the review of visual processes above.

1.1 Physics of Sound

As was the case with vision, auditory perception starts with the *distal object*, i.e., the source sound, and we need an *informational medium* to bring the sound to our sensory organs (i.e., ears). What is sound? Sound waves are moment-to-moment fluctuations in air pressure about the atmospheric level. Therefore, the informational medium is air (or water, or other solid sound waves can travel through). Sound cannot travel through a vacuum. *Pitch* is given by the *frequency* of oscillations; sound *intensity* by the amplitude of the oscillations.

Sound intensity, I, is measured by a pressure ratio:

$$I = 10log\left(\frac{p_1}{p_2}\right) \tag{1}$$

where $p_2 \approx 1$ kHz at 20 N/m^2. The unit is decibel, dB, which has three common scales, with different frequency weights: dB(A,B,C); dB(A) corresponds best to the human ear. Pitch, or frequency is measured by cycles per second, and the unit is Herz (Hz). Loudness, L, is a *subjective* measure, at it follows the Weber's Law: $L = aI^{0.6}$.

1.2 Physiology of the Ear

Our outer ear (what you can see) consists of *pinna* (the earlobe and the auditory canal. The middle ear lies between the outer ear and the inner ear. It consists of an air-filled cavity called the tympanic cavity and includes the three *ossicles* and their attaching ligaments, the auditory tube, and the round and oval windows. The ossicles are three small bones that function together to receive, amplify, and transmit the sound from the eardrum to the inner ear. *Malleus* (Latin for "hammer") s attached to the eardrum and connected to *stapes* (Latin for "stirrups") by incus (Latin for "anvil"). Stapes is attached to the oval window (a membrane).

The inner ear consists of the *cochlea* and the *hair cells* that provide neural input to the brain. The cochlea forms the vestibular system is designed to detect the position and motion (acceleration) of the head in space. In addition to hearing, it also provides the sense of balance. See Figure 1.

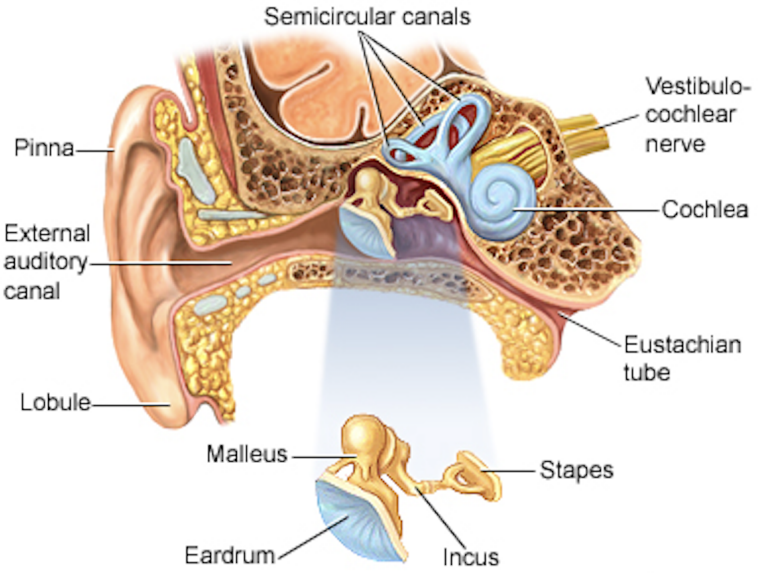


Figure 1. The anatomy of the human ear.

1.3 Psychophysical Indices

When it comes to measurement of sound as we *perceive* it, things will get quite a bit more complicated. There are two primary subjective measures of sound:

- **Phon** is the subjective equality of various sounds.
- **Sone** measures psychological experience of loudness, or relative loudness of different sounds; 1 sone = loudness of a 1000 Hz tone of 40 dB. A sound that is *judged* (i.e., regardless its intensity) twice as loud = 2 sones. Loudness measured in sones approximately doubles at every 10 dB increase in intensity.

Figure 2 illustrates the human perception of sound. Note that the human ear is most sensitive to sound between 2,000 and 4,000 Hz. Masking effects of noise depend both on intensity and frequency. Female voice (high pitch) is more vulnerable to masking effects than male (low pitch) voice. Similarly, consonants are masked more easily than vowels.

1.4 Measuring Speech Communications

There are two primary measures of speech communication: *Articulation Index* (AI), which is bottom-up measure, and measures the signal-to-noise ratio, and *Speech Intelligibility Level* (SIL), which is a top-down measure [1]. Consider also the following examples in radiotelephony communication:

- Standardized phraseology (top-down effect);
- Limited vocabulary (top-down effect);
- Pronunciation guidelines, for example, 9 is pronounced “niner” and 5 “fife” to avoid confusion between then (bottom-up effect);
- Redundancy, or readback and hearback (bottom-up effect).

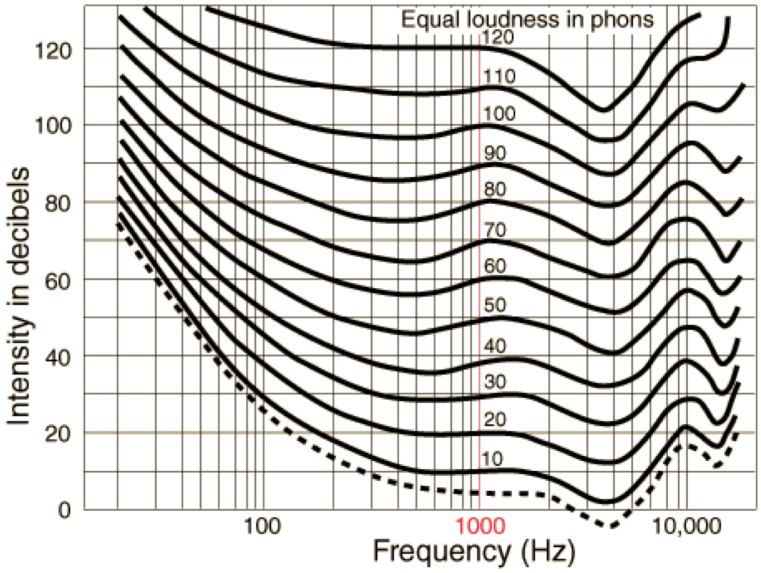


Figure 2. Equal loudness curves.

1.5 Noise

Noise is potentially very dangerous to hearing. Exposure to loud noise may results in Temporary Threshold Shift (TTS), for example, after a loud rock concert, or Permanent Threshold Shift (PTS) after prolonged working with noisy machines. Exposure time is hence critical. Consider these different measures of noise: *Time Weighted Average* (TWA) is the most important measure by the Occupational Safety and Health Administration (OSHA); *Noise Dose* is the time spent at sound level/max. permissible time at sound level; and a myriad of other measures, depending on the purpose of the mesurement (e.g., *mean annoyance*!)

A few examples help illustrate the concept of noise: Sound levels above 85 dB are considered harmful, 120 dB is unsafe, and 150 dB causes physical damage to the human body; windows break at about 163 dB; jet airplanes cause A-weighted levels of about 133 dB at 33 m, or 100 dB at 170 m; eardrums rupture at 190 dB to 198 dB; shock waves and sonic booms cause levels of about 200 dB at 330 m; sound levels of around 200 dB can cause death to humans and are generated near bomb explosions (e.g. 23 kg of TNT detonated 3 m away); even louder are nuclear bombs, earthquakes, tornadoes, hurricanes and volcanoes.

2 Tactile and Haptic Senses

While the visual and auditory senses get rightly the most attention in human factors literature, it is important not to forget the other senses (and yes, there are more than five!). Following the perceptual continuum framework, the *distal object* is anything we touch, the *informational medium* is direct contact with our skin, the *proximal stimulation* consists of sensory receptors just under the skin that respond to pressure on the skin, and the *perceptual object* what we make of the thing we are touching, its texture, weight, temperature, hardness/softness, and other tactile and haptic properties.

Consider how much information we receive through the sense of touch. We routinely judge how things respond to our touch (e.g., keystrokes, mouse clicks, braille alphabet). Virtual reality applications rely on

artificial electrical stimulation to the fingers as people manipulate "virtual objects". There are also haptic *informational displays*, such as control handles in aircraft for landing gear and wing flaps, which have distinctively different shapes so that pilots know which control they are grabbing without having to look at it and preventing them from retracting the landing gear when the aircraft is on the ground (and when one would operate the flaps) [1].

3 Proprioception and Kinesthesis

There is a vast set of receptor systems within the muscles, tendons, and joints in the body that convey an accurate representation of muscle contractions and stretches, joint angles, and limb positions in space to the brain. This *proprioceptive channel* is closely coupled with the *kinesthetic channel*, which conveys information about limb motion to the brain (speed, acceleration, direction) [1].

4 The Vestibular Senses

In the inner ear there are two sets of receptors, in the semicircular canals and the vestibular sacs. These receptors convey to the brain information about angular and linear accelerations of the body. Hence, even when blindfolded, a person can reliably maintain balance and be aware of body motions [1]. The vestibular senses work beautifully in Earth gravity and when we move at “natural” speeds under our own power.

Trouble arises when humans are traveling in machines (e.g., aircraft) that subject them to sustained accelerations (lulling the vestibular system to cease sensing acceleration, resulting in opposite sensation when the acceleration stops or changes) and orientations to which the system is not naturally adapted. This results in *spatial disorientation* which can be deadly when piloting an aircraft, for example.

Senses also work together. Hence, if visual sensation conflicts with vestibular sensation, the person develops *motion sickness*. This happens especially in moving vehicles without a view to the outside world, which could provide visual motion cues consistent with vestibular cues. “Simulator sickness” is a big problem in research involving driving simulators, where the projected scenery moves but the fixed simulator platform does not provide corresponding vestibular sensations.

5 Smell and Taste

Human senses of smell and taste are closely coupled. Although the human tongue can distinguish only five distinct qualities of taste, mastication of food in the mouth releases odorants that may be smelled during exhale, contributing to the overall sense of flavors. Although human sense of smell pales in comparison with many animals (e.g., dogs), it still performs amazingly well. For example, some wine experts can identify the wine’s grape varieties, growth region, and even the year the wine was produced, matching the bouquet and taste of the wine with the same from memory. Smells also readily evoke other memories, often automatically.

The sense of smell can also serve as a channel for *information displays*. Natural gas is colorless and odorless, so odorizers are often added to it to inform users of gas leaks in houses.

References

[1] J. D. Lee, C. D. Wickens, Y. Liu, and L. N. Boyle. *Designing for people: An introduction to human factors engineering*. CreateSpace, Charleston, SC, 2017.

DEPARTMENT OF PSYCHOLOGY, COLLEGE OF LIBERAL ARTS
ROCHESTER INSTITUTE OF TECHNOLOGY

PSYC 719—HUMAN FACTORS IN AI

HANDOUT: HUMAN INFORMATION PROCESSING

PROF. RANTANEN

October 5, 2022

1 Model of Human Information Processing

First, a few caveats. When we speak of human information processing, we must first define our terms. I think that "human" is clear enough, but what is "information" and what do we mean by "processing"? The so-called *Information Theory* originally referred to the *Mathematical Theory of Communication* by Claude E. Shannon [1] (April 30, 1916–February 24, 2001). Shannon developed his theory to to find fundamental limits on signal processing operations such as compressing data and on reliably storing and communicating data. Shannon's theory defines information as reduction of uncertainty, and information may be quantified by the number of true/false questions required to learn a new fact, or *bits* of information.

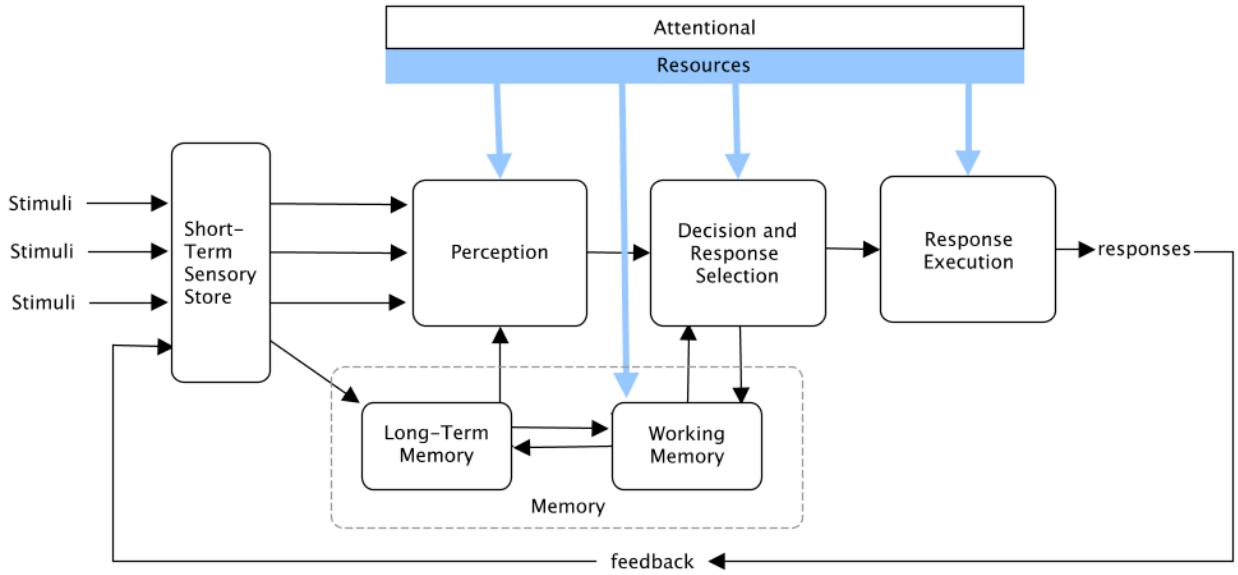


Figure 1. The model of human information processing.

It is worth repeating: "All models are wrong, but some are useful [2]." It is far from certain, perhaps even far from probable, that the "information" we take in through our senses and store in experiences look anything like bits. Yet, Shannon's theory has proved very useful in human factors. The model in Figure 1 is a composite of the works by several researchers [3, 4, 5, 6, 7]. Note the boxes from left to right in the middle of the illustration, reflecting the idea of separate processing stages in the information processing system. Information enters from the left and passes through the different stages, transformed in different ways along the way, each transformation consuming some time.

There are also representations of various supporting components, most importantly long-term- and working memory, and attention. The arrows show the relationships the different components have with each other. The depiction of attention as a "reservoir" imply limited attentional resources and the "drain pipes" show where these resources are consumed as information is processed.

2 Response Time

One of the most important dependent variables in cognitive psychology research has been—and remains—response time (RT). Timing data (e.g., reaction times) are relatively easy to obtain under both experimental and naturalistic conditions, and time is a variable that is common to the human, the task, and the environment. Time offers a common unit of measurement of human performance in the context of the task, and can be used to infer mental processes relevant to the task performance [8]. Two seminal experimental paradigms are reviewed next.

2.1 Donders' Reactions

Consider the following reaction time tasks [9]:

$$S \rightarrow R \tag{1}$$

In this *simple reaction time* task, or Donders' A reaction, whenever the stimulus (S; e.g., a light, or sound) is presented, the research participant is required to respond (R; e.g., by pressing a key on a keyboard). Reaction time from stimulus onset to response is measured.

$$S_1 \rightarrow R_1, S_2 \rightarrow R_2 \tag{2}$$

This is an example of a *choice reaction time* task, or Donders' B reaction, where different stimuli (S_1, S_2) require different responses (R_1, R_2, repectively). Response time is again measured.

$$S_1, S_2, S_3 \rightarrow R \tag{3}$$

Here, out of multiple different stimuli only one requires a response. This is Donders' C reaction.

Now, we can use *subtractive logic* to estimate the time required for mental operations in each case. Thus, if the simple response time in (1) is a, and the choice response time in in (2) is c, then $c - a$ is the *identification time*. Similarly, if the response time in (3) is b, then $b - c$ is the *selection time*. Such subtractive methods strongly suggest the existence of stages in human information processing.

2.2 Additive Factors

An alternative technique to the subtractive method, the additive factors technique [3] makes three assumptions:

1. RT equals the sum of a series of nonoverlapping independent processing stages, each of which performs some transformation of information;

2. The nature of the transformation is independent from its and other stages' duration (i.e., processing of degraded stimulus may take longer to process than a clear stimulus, but the output, or stimulus identity, is the same in each case);

3. Experimental manipulations that influence the same stage will produce *interactive* effect on RT, whereas manipulations that influence separate stages will have *additive* effects (this is helpful in distinguishing between different stages).

This technique, or an experimental paradigm, has been a powerful tool in discovery and modeling of many aspects of human cognition. See Figure 2 for an example of the additive factors technique. The experimenter will orthogonally manipulate several variables (i.e., the variables may be treated as statistically independent) that hypothetically may impact different stages in human information processing (Fig. 2a). For example, perceptual encoding may depend on masking the stimulus, or on stimulus discriminabiity; response selection may depend on the number of alternative responses or stimulus-response (S-R) compatibility. The point is that analysis of the RT data can reveal whether the independent variables *interact* or are *additive*.

The results from a hypothetical experiment in Figure 2b indicate the the variables A and B indeed do interact (the difference between the two levels of B is much greater in level 2 than level 1 of A. This implies that the variables A and B impacted the same stage in the information processing system. On the other hand, results from another hypothetical experiment in Figure 2c show no interaction between variables B and C, but that their separate impacts on RT is additive. Hence, we may conclude that variables B and C impacted different stages in the information processing system.

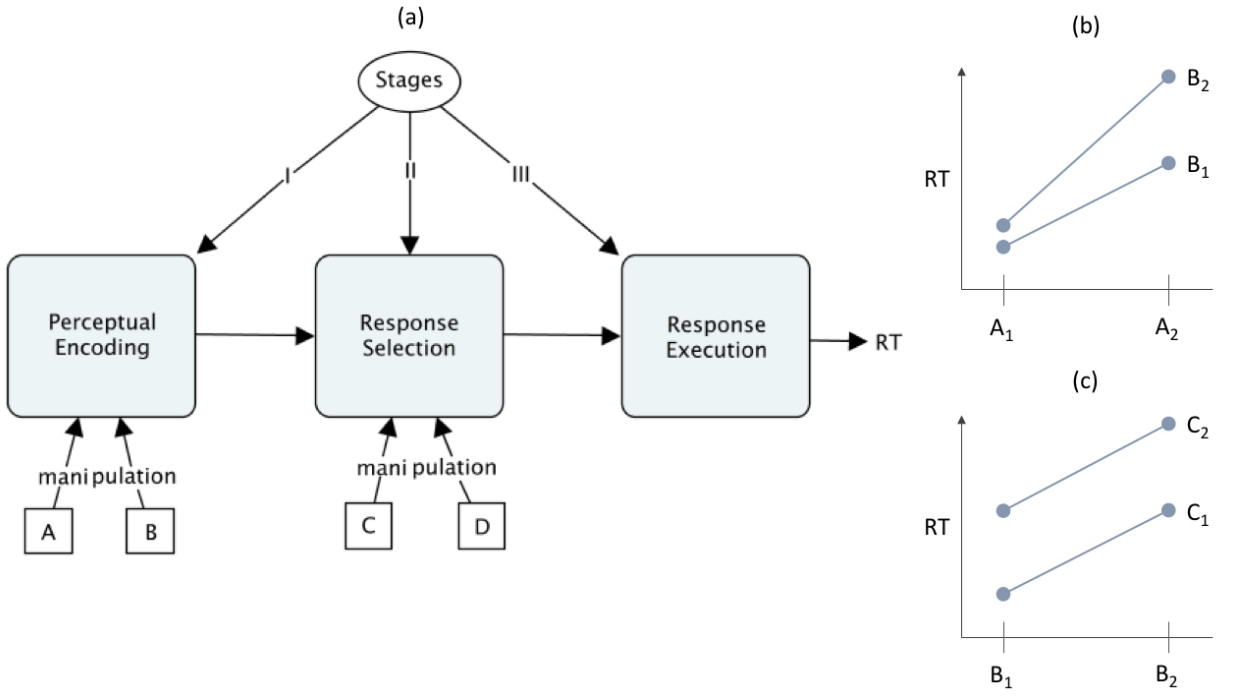


Figure 2. The additive factors experimental paradigm (a). .

3 Dual Processes

The idea that there are two different *systems* of reasoning is not new. The table below [10] shows some of the seminal papers written on this two-system view, organized in a chronological order. See the titles of the papers cited here to get a picture of the kinds of research where this view has emerged. Note also the distinct

characteristics of each system

Dual-Process Theories:	System 1	System 2
Posner & Snyder (1975, 2004) [11, 12]	automatic activation	conscious processing system
Shiffrin & Schneider (1977) [13]	automatic processing	controlled processing
Johnson-Laird (1983) [14]	implicit inferences	explicit inferences
Evans (1984; 1989) [15, 16]	heuristic processing	analytic processing
Pollock (1991) [17]	quick and inflexible modules	intellection
Reber (1993) [18]	implicit cognition	explicit learning
Epstein (1994) [19]	experiential system	rational system
Levinson (1995) [20]	interactional intelligence	analytic intelligence
Sloman (1996) [21]	associative system	rule-based system
Hammond (1996) [22]	intuitive cognition	analytical cognition
Evans & Over (1996) [23]	tacit thought processes	explicit thought processes
Klein (199) [24]	recognition-primed decisions	rational choice strategy

4 The Two-System View in Decision Making

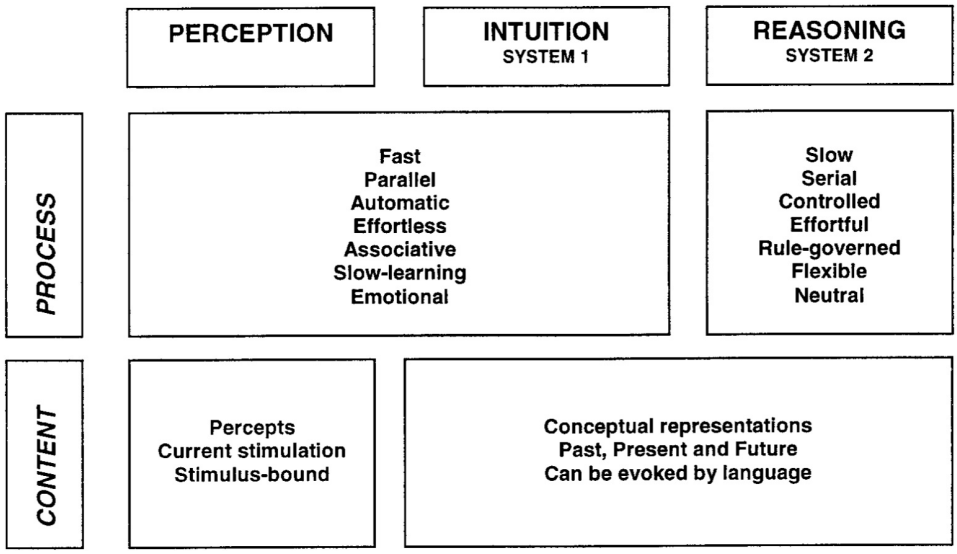


Figure 3. The 2-system view from Kahneman's Nobel Prize lecture.

The 2-system view has featured prominently in Amos Tversky's and Daniel Kahneman's [25] work to discover why people make "wrong" decisions in simple gambling problems and what kinds of strategies led to their decisions. The point is to explain deviations from "rational", normative decisions without labeling such behavior "irrational". The alternative term adopted by Kahneman and Tversky was "bounded rationality", a term first coined by Herbert Simon [26]. Other term used are "reluctant rationality" [27] or "cognitive strain" [28].

The point is that "rational" thinking, or reasoning, is hard and effortful, and people prefer mental shortcuts to reasoning whenever possible. These mental shortcuts also tend to be highly reliable and successful, *most* of the time, which further reinforces them and make them seem valid and robust. In other words, why

would a person engage in effortful and slow reasoning if a handy rule provides the same result. In a very regular world we live in this is a very reasonable strategy. The problem is that when circumstances are *un*usual, rules of thumb may result in poor or wrong decisions.

5 Schematic and Attentional Control Modes

Stanovich and West [10] labeled the two systems simply as System 1 and System 2. Other researchers have used different terminology to describe essentially the same phenomenon. In the cognitive science tradition, behavior is said to be controlled by two distinct *modes*: The *attentional* control mode, or the *schematic* control mode .

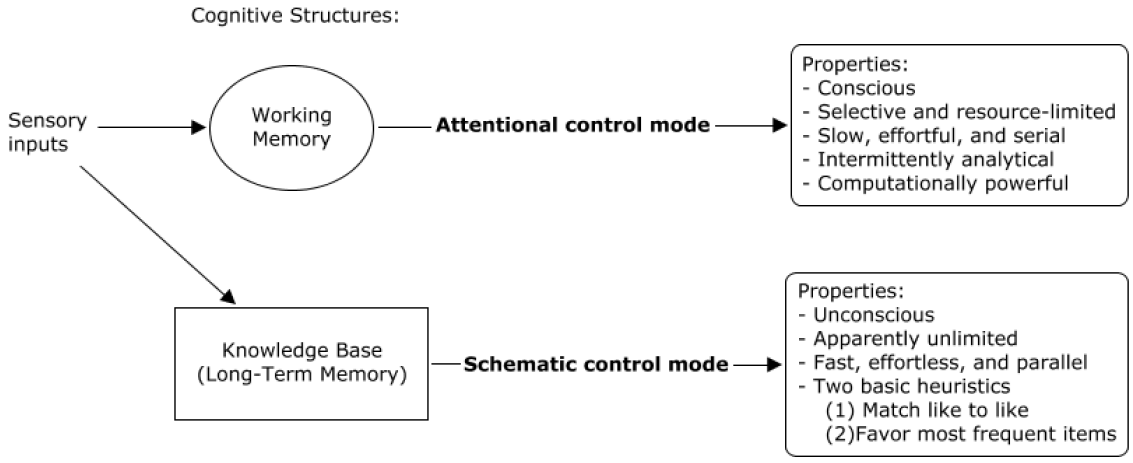


Figure 4. The schematic and attentional control modes. Looking at the properties of each mode, it is clear that schematic control mode corresponds to System 1 and attentional control mode to System 2.

5.1 Attentional Control Mode

Attentional control mode is associated with working memory (WM), a "workbench" of the mind. In WM information is gathered from senses as well as the long-term memory (LTM), processed, integrated, and acted upon or sent to LTM for storage. Information processing in WM demands attentional resources (hence the name attentional control), it is conscious, resource-limited (WM has serious capacity limitations) and typically slow, effortful, and serial in nature. This is where the "heavy" thinking happens.

5.2 Schematic Control Mode

In contrast, schematic control mode allows for direct matches to be made between incoming sensory information (e.g., a person's face, smell of food, sound of words or music) and existing representations of them in the LTM. For example, if you like coffee, the smell of a coffee will immediately and *automatically* retrieve all kinds of associations (schemas about coffee) from the LTM to consciousness. The schematic control mode itself is unconscious (one does not need to think), automatic, fast, effortless, and parallel (capable of retrieving multiple schemas simultaneously (Figure 5).

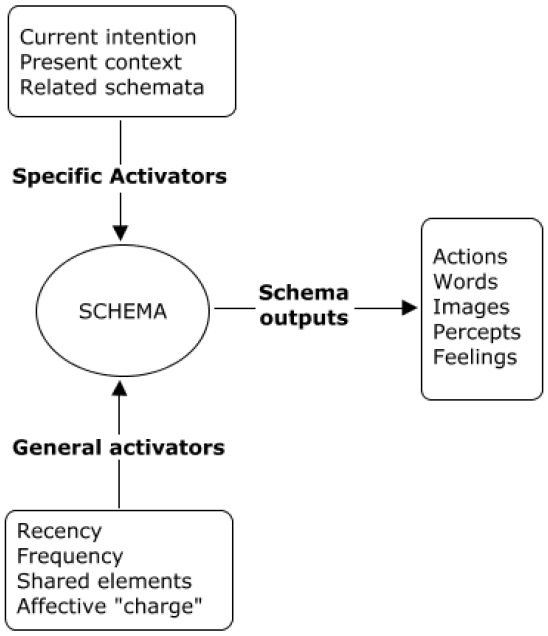


Figure 5. Specific and general activators of schema and outputs in the schematic control mode.

6 The SRK Framework

Coming from a different domain, that of engineering, yet another model captures the characteristics of multiple modes of reasoning. The skills-, rules-, and knowledge (SRK) framework (Figure 6) by Rasmussen [29] is an immensely influential model of human behavior and performance and foundational to many other models (e.g., the Generic Error Modeling System [27]). It also has much in common with the 2-system view of human performance.

6.1 Skill-Based Behavior

Skill-based behavior consists of sensory-motor performance during acts that (after a statement of intention) take place without conscious control as smooth, automated, and highly integrated patterns of behavior. Diagnostic troubleshooting is done by a direct match between the features of the problem observed and patterns previously experienced and stored in the LTM. SB performance is fast, requires little cognitive activity, and accurate.

6.2 Rule-Based Behavior

Rule-based behavior is controlled by a stored (in LTM) rule. Performance is goal-oriented, although the goal may not be explicitly formulated but is found implicitly in the situation that released the stored rules. Diagnosis is done by applying sets of rules stored in LTM, e.g., sequence of steps and the procedures for doing so. RB behavior also utilizes mental "checklists" and mental simulation of "what if" scenarios.

6.3 Knowledge-Based Behavior

Knowledge-based behavior is explicitly goal controlled performance in unfamiliar situations where no rules or know-how are available. Functional reasoning is done based on mental models. KB processes are Iterative

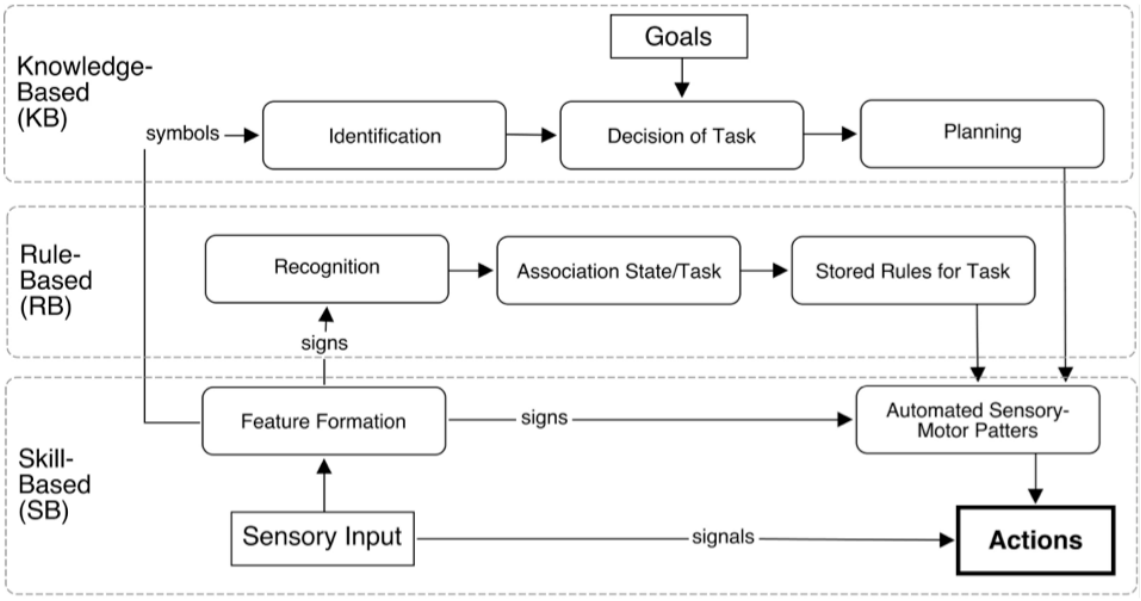


Figure 6. The SRK framework. Note that in Rasmussen's original formulation stimuli may be labeled signals, signs, or symbols, depending on at what level they are used.

diagnostic testing and subsequent analyses necessary for problem-solving.

7 The SRK Framework and Decision Making

The SRK framework has also much in common with different decision making models. For example, the knowledge- and rule-based behaviors correspond to the System 2 in Kahneman's 2-system view, being slow, serial, controlled (by attention) effortful, *rule-governed* and flexible. Skill-based performance corresponds to the System 1, being fast, parallel, automatic, effortless, associative, and slow-learning. In Figure 7 the 2-system view is superimposed on the SRK framework.

8 Cognitive Continuum Theory

The cognitive continuum theory (CCT) [30] posits a continuum between intuitive and analytical decision making. Decisions along the continuum between the end points is said to be *quasi-rational*, or have characteristics of both intuitive and analytical reasoning.

The CCT is based on two *metatheories*: The *coherence* theories concern rationality of judgments and decisions in terms of logical or mathematical consistency, or absence of contradictions. The *correspondence* theories are concerned about correspondence of judgment with empirical facts (i.e., empirical accuracy of judgment; e.g., physicians' diagnoses) and factors leading to empirical accuracy or inaccuracy of the judgment.

The psychological underpinnings of competence of decision making come in two flavors as well. *Correspondence competence* focuses on the remarkable perceptual abilities of humans (and other species) in perceiving a 3D world projected on 2D retina and constancy of shape, size, color, etc. perception. *Coher-*

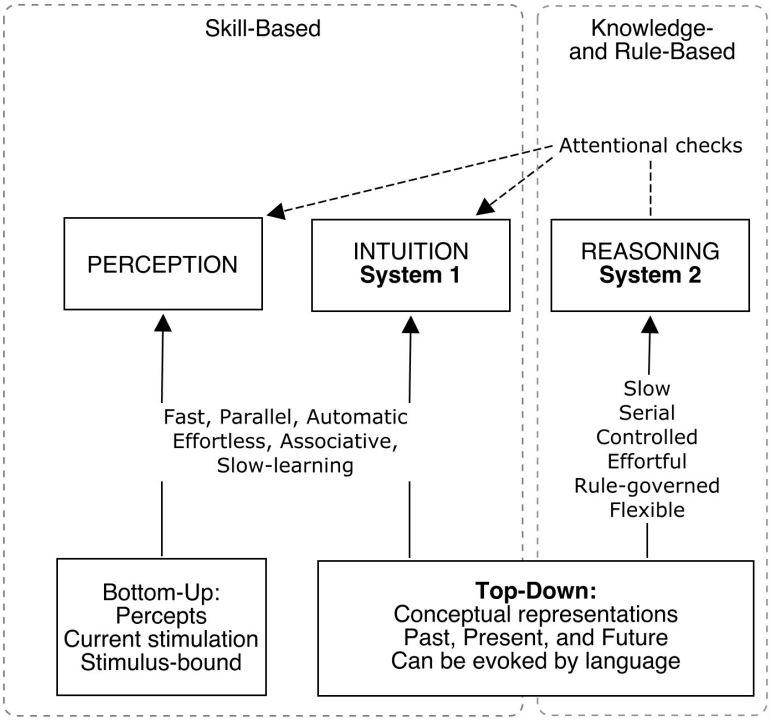


Figure 7. The 2-system view superimposed on the SRK framework. System 1 corresponds to skill-based performance and System 2 to rule- and knowledge-based performance.

ence competence, on the other hand, appears to be unique to humans; it is an acquired (taught!) skill (story telling). Different tasks demand coherence competence than correspondence competence.

There are two aspects of cognitive competence: Subject matter (or, domain) competence, including learning, memory, and deduction, and judgment and decision-making competence, consisting of observation and inference. Both can be achieved by both intuition and analysis.

The CCT is based on five premises:

1. Various modes, or forms, of cognition can be ordered in relation to one another on a continuum with intuitive cognition in one end and analytical cognition in the other.

2. The forms of cognition that lie on the continuum between intuition and analysis contain elements of both, and is termed quasi rationality (a.k.a. common sense).

3. Cognitive tasks can be ordered on a continuum with regard to their capacity to induce intuition, quasi rationality, or analytical cognition.

4. Cognitive activities may move along the intuitive-analytical continuum over time; successful condition in stable environments require stability along the continuum, but changing environments necessitate movement along it (dynamic cognition).

5. Human cognition is capable of both pattern recognition and the use of functional relations.

The point of CCT is that decision making style along the cognitive continuum should match the task demands. The matrix in Figure 8 combines a cognitive continuum index (intuitive–quasirational–analytical)

with a similar task continuum index. Best performance is achieved along the diagonal across the matrix where the cognitive- and task continuum indices are matched.

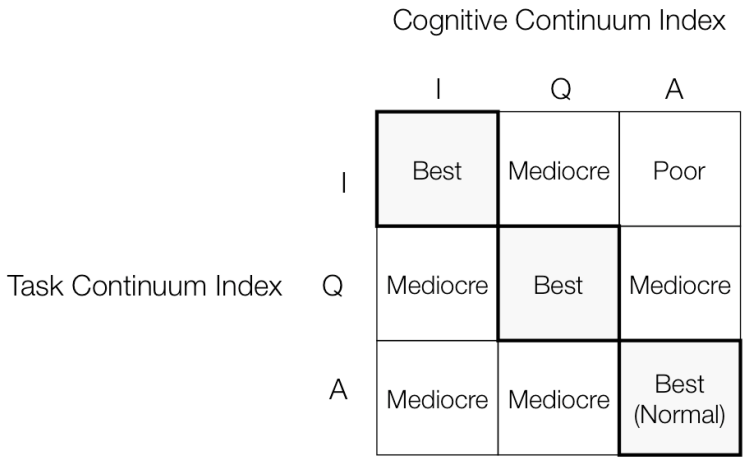


Figure 8. A cognition-task interaction matrix combines a cognitive continuum index (Intuitive–Quasirational–Analytical) with a similar task continuum index..

References

[1] C. E. Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[2] G. E. Box. Robustness in the strategy of scientific model building. In R. L. Launer and G. N. Wilkinson, editors, *Robustness in Statistics*, pages 201–236. Academic Press, New York, 1979.

[3] S. Sternberg. The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 30:276–315, 1969.

[4] D. Broadbent. *Perception and communication*. Pergamon Press, New York, NY, 1958.

[5] E. E. Smith. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin*, 69(2):77, 1968.

[6] A. T. Welford. *Skilled performance: Perceptual and motor skills.* Scott & Foresman, Glenview, IL, 1976.

[7] C. D. Wickens. *Engineering Psychology and Human Performance*. Charles E. Merrill, Columbus, OH, 1984.

[8] B. R. Levinthal and E. M. Rantanen. Measurement of taskload and performance in a dynamic multi-task experiment. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 48, pages 567–570. SAGE Publications Sage CA: Los Angeles, CA, 2004.

[9] B. H. Kantowitz and R. D. Sorkin. *Human factors: Understanding people-system relationships*. John Wiley & Sons Inc, 1983.

[10] K. E. Stanovich and R. F. West. Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23(05):645–726, 2000.

[11] M. I. Posner and C. R. R. Snyder. Attention and cognitive control. In RL Solso, editor, *Information processing and cognition: The Loyola Symposium*. Psychology Press, 1975.

[12] M. .I Posner and C. R. R. Snyder. Attention and cognitive control. *Cognitive psychology: Key readings*, page 205, 2004.

[13] R. M.. Shiffrin and W Schneider. Controlled and automatic human information processing: Ii. perceptual learning, automatic attending and a general theory. *Psychological review*, 84(2):127, 1977.

[14] PN Johnson-Laird. *Mental models: Towards a cognitive science of language, inference, and consciousness*. Number 6. Harvard University Press, 1983.

[15] J. S. B. T. Evans. Heuristic and analytic processes in reasoning*. *British Journal of Psychology*, 75(4):451–468, 1984.

[16] J. S. B. T. Evans. *Bias in human reasoning: Causes and consequences.* Lawrence Erlbaum Associates, Inc, 1989.

[17] J. L. Pollock. Oscar: A general theory of rationality. 1991.

[18] A. S. Reber. *Implicit learning and tacit knowledge*. Oxford University Press, 1996.

[19] S. Epstein. Integration of the cognitive and the psychodynamic unconscious. *American psychologist*, 49(8):709, 1994.

[20] S. C. Levinson. Interactional biases in human thinking. *Social intelligence and interaction*, pages 221–260, 1995.

[21] S. A. Sloman. The empirical case for two systems of reasoning. *Psychological bulletin*, 119(1):3, 1996.

[22] K. R. Hammond. Human judgment and social policy, 1996.

[23] J. S. B. T. Evans and D. E. Over. *Rationality and reasoning*. Psychology Press, 2013.

[24] G. A. Klein. *Sources of power: How people make decisions*. MIT press, 1999.

[25] D. Kahneman. A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58(9):697, 2003.

[26] H. A. Simon. Theories of bounded rationality. *Decision and organization*, 1:161–176, 1972.

[27] J. Reason. *Human error*. Cambridge university press, 1990.

[28] J. S. Bruner and G. A. Goodnow, J. J .and Austin. *A study of thinking*. Science Editions, New York, 1956.

[29] J. Rasmussen. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man and Cybernetics*, (3):257–266, 1983.

[30] K. R. Hammond. *Judgments under stress*. Oxford University Press, 2000.

DEPARTMENT OF PSYCHOLOGY, COLLEGE OF LIBERAL ARTS
ROCHESTER INSTITUTE OF TECHNOLOGY

PSYC 719—HUMAN FACTORS IN AI

HANDOUT: HUMAN MEMORY

PROF. RANTANEN

October 11, 2022

1 Preliminaries

1.1 Metaphorical Thinking

As we are trying to understand the construct of memory this week, it is useful to consider, and important to be aware of, metaphorical thinking. Metaphors and analogies are helpful in making sense of the world and shaping mental models of reality. They are particularly useful in making abstract concepts (such as memory) more tangible and providing a common image and language for discussion of problems and research paradigms.

Metaphorical thinking is particularly important as much of behavioral sciences deal with theoretical constructs. A construct (n.) is a concept, model, or schematic idea. It is not a "real thing", but a mere representation of the phenomenon or mechanism or structure of interest. Metaphors may make these representations easier to understand, and to think and talk about. Just do not think that they *are* the real thing.

The "brain as a computer" metaphor has a relatively long history that parallels the invention and development of the digital computer [1]. This kind of thinking is still in vogue in some circles, for example [2, 3]. This metaphor appears attractive on the surface. After all, human memory seems to have components that closely correspond to a computer, for example, a hard drive = long-term memory, the RAM = working memory, and the CPU = the central executive. The problem is, however, that the brain does not work at all like a computer and that simplistic metaphors like this are therefore highly misleading:

Computer *store* words or the *rules* that determine how to *manipulate* them. Computers create *representations* of visual stimuli, them in a short-term *memory buffer*, and then *transfer* the representation into a *long-term memory device*. Computers *retrieve* information or images or words from *memory registers*. However, *living organisms* do not do any such things [4].

Alas, there is no way of avoiding such misleading metaphorical language when we talk about memory, for such language is *all we have*!. Therefore, please keep the above in mind as we try to make sense of the little we can know about memory.

1.2 Memory Research

Consider the following different *types* of tasks used to study memory [5]:

- Explicit memory tasks: Conscious recall of particular information (e.g., "Who wrote *Hamlet*?")
- Declarative knowledge tasks: Conscious recall of facts (e.g., the three memory processes above).

- Recall tasks: Producing a fact, a word, or any other item from memory (e.g., fill-in-blank tests).
- Serial-recall tasks: Repeating items in a list in the exact order they were given (e.g., reading back a phone number.
- Free-recall tasks: Repeating items in a list in any order they were given (e.g., names of people invited to a party.
- Cued-recall tasks: After memorizing paired items, given on of the items recalling its pair.
- Recognition tasks: Selecting or otherwise identifying an item that has been learned before (e.g., multiple-choice exams).
- Implicit memory tasks: Drawing on information in memory without conscious realization of doing so (e.g., word-completion tasks).
- Tasks involving procedural knowledge: Being able to know *how* to do something (e.g., experience in solving puzzles, doing skiiled tasks)

Note that the differences between the above tasks seem small, but they nevertheless allow distinguishing between different memory processes, and thus help in understanding what memory is and how it works. Several models of memory have been proposed over the years of research. We will review the most important models next.

1.3 Memory in Context

Please refer to Figure 1, a model of human information processing that serves as useful "roadmap" to much of (cognitive) psychology [6]. Note the central role of memory in nearly all of the other components and processes in this model. Note also the role of attention (focusing awareness on a narrowed range of stimuli). Attention is crucial to memory. If you are not paying attention to lectures in class you will not remember what the professor was trying to teach you! Memory, and encoding of information into memory in particular, is usually an *effortful* process (i.e., it required attention; but see for the memory processes below).

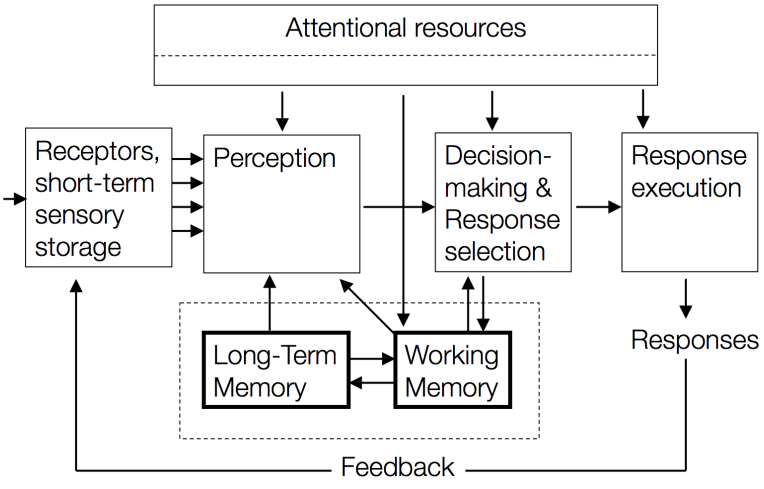


Figure 1. Memory in the context of the model of human information processing.

2 Memory Processes: Encoding

To remember things, or "to put things into memory", requires that the information is *encoded* in a proper way. An illustrative test of encoding is to try to remember what the U.S. penny coin looks like *exactly*. See Figure 2 and pick out the correct penny among the distracters (or counterfeits, as if anyone would counterfeit pennies!). Why is this task so hard? The answers is: Because we only need to encode the size (small) and the color (copper) of the coin to recognize it among some change; Lincoln's face is also familiar enough so the we know that it is his image on the penny, but that is really all. Other details, like the placement of the various text elements ("In God We Trust", "Liberty", "E Pluribus Unum", "United States of America", and "One Cent") and the year on the coin are usually *not* encoded and therefore *not available* to us as we try to identify the right penny.

Figure 2. Which is the right penny?.

Encoding refers to forming a memory code from some stimulus. The levels-of-processing theory holds that qualitative differences in attending affect how well we remember things. This theory distinguishes between 4 *types* of encoding:

1. In the levels-of-processing theory, the most basic type is *structural* encoding. Structural encoding is shallow and emphasizes the physical structure of the stimulus. For example, encoding only the shapes that form letters in the alphabet; one would be able to answer the question "Is the word written in all capital letters?"

2. *Phonemic* encoding emphasizes what a word sounds like, for example, when reading aloud or to oneself. Phonetic encoding would allow for answering a question "Does the word rhyme with...?"

3. *Semantic* encoding includes the *meaning* of the word. This would allow for answering the question "Does the word fit in a sentence...?"

4. Finally, *self-referent* encoding involves deciding how or whether information is personally relevant. Information that is personally meaningful is more memorable.

Elaboration means that a stimulus is linked to other information at the time of encoding. For example, you are studying this handout and see how it matches the textbook description of memory processes (am I adding anything new, or explaining things differently?). Elaboration often consists of thinking of examples, and self-generated examples seem to work best (how can you use this information to improve your retention of course materials?).

Creation of visual images to represent the words to be remembered helps encoding the information. Concrete words are much easier to create images of than abstract concepts. For example, "juggler" is easier to visualize than "truth". The dual-coding theory holds that memory is enhanced by forming semantic or visual codes, since either can lead to recall.

3 Memory Processes: Storage

Storage of information is determined by the *structural* characteristics of memory. At least three main structures have been identified: sensory memory, short-term/working memory, and long-term memory.

3.1 Sensory Store

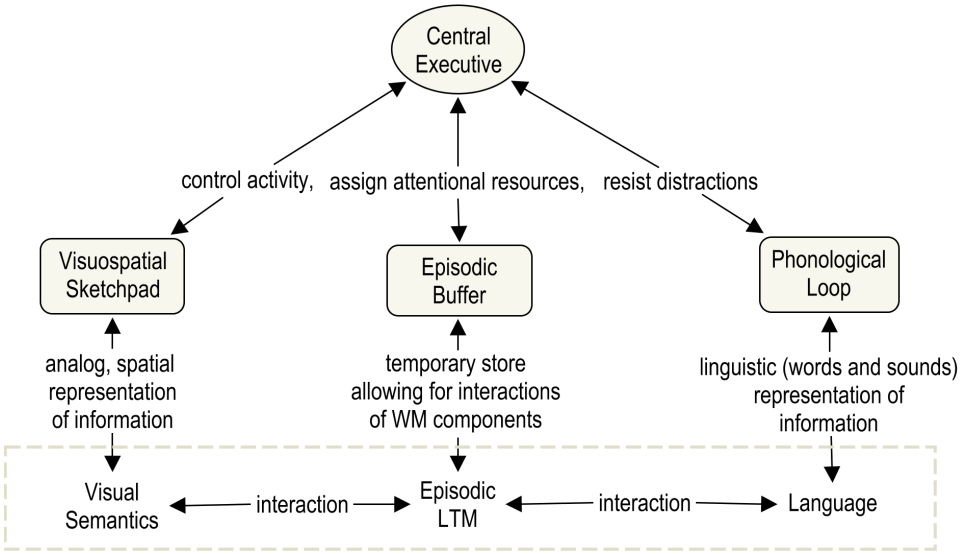
Each sensory modality has a mechanism that prolongs the sense impression. Visual sensory store, or *iconic memory* holds visual stimuli for about 250–500 ms. Auditory sensory store, or *echoic memory* holds auditory stimuli for up to 4 seconds. It is assumed to make hearing possible, acting a buffer to allow for attention to be directed to auditory information. These stores are *preattentive* (i.e., they work without conscious attention) and *veridical* (i.e., they preserve most of the physical details of the stimuli). However, they decay rapidly without attention.

3.2 Working Memory

Working memory (WM) has been associated with consciousness. It is a memory register that holds current and recently attended information, from both sensory stores and long-term memory. Information integration also takes place in WM. Information in WM is maintained through rehearsal (sustained attention); without rehearsal decay occurs within about 15 to 20 seconds.

The *capacity* of WM is also seriously limited, thought to be only about 7 ± 2 *units* of information [7]. This "magical" number does not mean that the WM capacity is fixed, for "chunking" allows for substantial increase of information that may be held in WM at any one time. For example, memorizing the 12 letters FBINBCCIAIBM would exceed the WM capacity. However, it is easy to see that there are 4 familiar acronyms, FBI, NBC, CIA, and IBM, which do not pose any problem to our WM!

WM is further thought to have at least 4 distinct parts, each with a unique function (Fig. 3.2) [8]. Note that the many interactions within this model may be used to explain and *predict* interferences in WM and their consequences for both human performance and the design of interfaces. See [6] pp. 200–210 for a detailed discussion on design implications.



3.3 Long-Term Memory

Information processed in WM is designated for permanent storage in long-term memory (LTM). LTM, too, may be understood to have different structural components. Distinction may be made between *event* memory (about personally experienced and remembered concrete events) and *semantic* memory (about generic world knowledge, such as vocabulary and grammar, and meaning of things and abstract concepts). Event memory may be further divided into *episodic* memory about *past* events (e.g., your first day at RIT) and *prospective* memory about *to-be-remembered* things (e.g., the next lab report due in our class). Semantic memory may be divided further into *declarative knowledge* (i.e., knowledge of *what* is; most of what you learn in college is—unfortunately—declarative knowledge that is hard to remember) and *procedural knowledge* (i.e., knowledge of *how* to do things, like to ride a bicycle, or navigate a now familiar campus). See Figure 3 for an illustration of LTM organization.

Information is structured as it is stored in LTM. One way to organize information is through clustering. People have tendency to remember similar or related items in groups.hen possible, factual information may be organized into conceptual hierarchies. A conceptual hierarchy is a multilevel classification system based on common properties among items. For example, consider what you know about minerals: There are metals and stones. Metals can be classified into rare, common and alloys. Stones may likewise be classified into precious and masonry. We will return to the topic of memory organization when we examine memory retrieval mechanisms.

4 Memory Retrieval Mechanisms

Memory search may be convergent or divergent. Convergent memory search matches specific cues to a uniquely specified knowledge item. The general heuristic in convergent search is *similarity-matching*. In

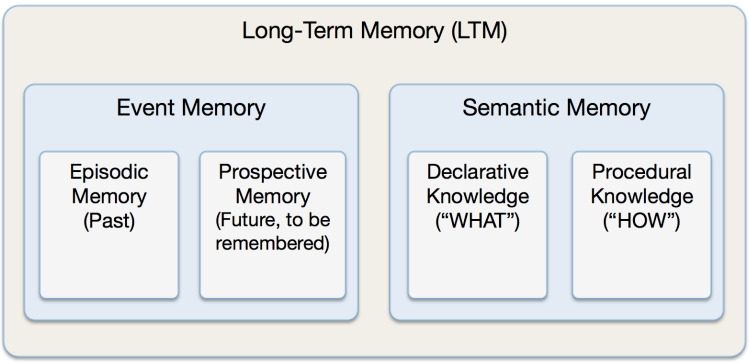


Figure 3. The structure of long-term memory (LTM).

divergent memory search category exemplars belonging to a given category are recalled. The general heuristic in divergent memory search is *frequency-gambling*, where most frequently encountered exemplars are easiest—and first–recalled. Similarity-matching and frequency-gambling are adaptive and optimal strategies for an uncertain but highly regular world.

Another dichotomy is between serial and parallel memory search process. Associative, serial search is attention-demanding, slow, and indirect (cf., convergent and divergent searches above). Direct search, on the other hand, is fast, achieved by activating well-established memory pathways. When cognitive operations are underspecified, they tend to default to contextually appropriate (similarity-matching) and high frequency (frequency-gambling) response. This makes responses highy *predictable*, which is a very good thing, as we will see later in the course.

In the cognitive science tradition, behavior is said to be controlled by two distinct *modes*: The *attentional* control mode, or the *schematic* control mode (see prevous week's handout).

We have often used the computer metaphor when discussing human information processing. Note, however, that human memory does not seen to work like the RAM or hard drive on a computer, where whatever you saved into the had drive or were working on in the RAM can be recalled just as it was saved. In contrast, it seems that human memories are *constructed* as they are recalled. Simply recalling may distort memory, as do simple suggestions. Prior experience influences how we recall information. Having retrieval cues can help us recall more information, but cues can also lead to errors. Consider the Schacter's "Seven Sins of Memory" [9]:

1. Memories are transient (fade with time)
2. We do not remember what we do not pay attention to
3. Our memories can be temporarily blocked
4. We can misattribute the source of memory
5. We are suggestible in our memories
6. We can show memory distortion (bias)
7. We often fail to forget the things we would like not to recall (persistence of memory)

False eyewitness memory is the single greatest cause of wrongful convictions nationwide, playing a role in more than 75% of convictions overturned through DNA testing (`http://www.innocenceproject.org/`). Consider the classic experiment by Loftus and Palmer [10] where participants were all shown the same video of an accident between two cars. Participants were asked "How fast were the cars going when they smashed/collided/bumped/hit/contacted each other?" The choice of the word in the question resulted in speed estimates varying between 32 mph ("contacted") and 41 mph ("smashed").

Problems with lineups include the assumption that perpetrator is in lineup, who are selected as distractors, and police behavior. Children's eyewitness memory is particularly problematic, and may be influenced by repeated questioning and leading questions distorting memory. Younger children are more suggestible than older.

5 Prospective Memory

Prospective memory (PM) is such an important topic in terms of human performance that it deserves its own section. In general, our ability to "remember to remember" things in the future is very poor. This has resulted in numerous tragic accidents. For example, the Northwest flight 255 crash in Detroit on August 16, 1987, was attributed to a PM failure., The National Transportation Safety Board determined "that the probable cause of the accident was the flightcrew's failure to use the taxi checklist to ensure that the flaps and slats were extended for takeoff" [11].

PM is itself a complex construct. Research has identified several classes of PM by, for example, simple or complex prospective activity [12], event-based (remembering cued by the environment) or time-based (monitoring the passage of time) remembering [13], whether the task is habitual or infrequently performed [14], or by time-scale, whether the action is to be performed in the near-term or the long-term [15]. Variables affecting PMinclude retrieval context, individual strategies for remembering, and the importance of the to-be-remembered activity.

5.1 Paradigms and Theories

The Einstein-McDaniel paradigm of PM research has experimental participants engaged in an ongoing activity. They must then recognize events (e.g., key words) that are relevant to earlier established intentions (e.g., press a key). In this "noticing plus search" paradigm environmental cues elicit a feeling of familiarity (noticing), which prompts memory search.

PM research is not uncontroversial. For example, it is no clear whether PM processes are automatic or controlled (conscious) or some combination of both. PM performance seems to depend on WM, and in particular on attentional resources within the central executive, which is involved in memory search stage of "noticing plus search" model.

5.2 Applied Research

Well-controlled laboratory experiments are essential for understanding of the cognitive processes underlying human performance, but field studies identify crucial phenomena and sources of variance in the real world. For example, consider the following PM demands in airliner cockpits [16]:

- Episodic tasks: Remembering to perform a task at a time it is not habitually performed (e.g., request to report passing 10,000 ft while still at 15,000 ft)

- Habitual tasks: Sub-tasks or steps in a procedure implicit in the action schema.
- Atypical actions substituted for habitual actions: Necessary deviations from well-established procedural sequences.
- Interrupted tasks: Remembering to resume the original task after the interruption.
- Interleaving tasks: Multitasking situations.

The above represent diverse situations, but may share a common conceptual framework. Some theoretical accounts include automatic retrieval of stored intention by environmental cues or physical stimuli and strategic monitoring of opportunities to perform a delayed task. Also, strength of a cue's association with intention, number of intentions associated with the cue, and number of intermediate links between cue and the intention are factors in PM performance.

Note that an intention will have to be encoded in memory. Intentions can be implicit and explicit. Interruptions discourage forming of explicit intentions; remembering then depends on cues. There may also be a mismatch between cues and intentions (cues fail to trigger recognition), and end of an interruption may be followed by other task demands that prevent retrieval of associated intentions.

6 Mental Models

A functional definition of mental models is as follows [17]: "Mental models are the mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future system states."

Mental models reside in the LTM. Elements of representation are aggregated into larger units, chunks, within the same model category as familiarity with the context increases. The representation of properties of a system or an environment is transferred to a model category at a higher level of abstraction. Mental models also include analogies and use of ready-made solutions: The representation is transferred to a category of model for which a solution is already known or rules are available to generate the solution.

Mental models have a long history in psychological research. According to Craik [18] knowledge consists of a model of the world formed by humans in their nervous system (cf., behaviorism). Control engineering models, especially the optimal control theory, included mathematical representations of operators' mental models of the systems they were controlling.

Norman [19] has 3 uses of the term: (1) designer's mental model of the system being designed, (2) user's mental model of the device or system, and (3) researcher's model of the operator's mental model. When the designer's and user's models do not overlap we have a problem, for the user will operate the system based on an incorrect understanding of the system's functioning (e.g., setting the temperature in the freezer and fresh food compartments in a refrigerator [20], or attempting to heat a room faster by cranking a thermostat all the way up).

References

[1] J. Von Neumann. *The computer and the brain*. Yale University Press, 1958.

[2] R. Kurzweil. *The singularity is near: When humans transcend biology*. Penguin, 2005.

[3] R. Kurzweil. *How to create a mind: The secret of human thought revealed*. Penguin, 2013.

[4] R. Epstein. The empty brain. *Aeon*, 18:1–13, May 2016.

[5] R. J. Sternberg. *Cognitive Psychology*. Wadsworth, 5th edition, 2009.

[6] C. D. Wickens, J. G. Hollands, R. Parasuraman, and S. Banbury. *Engineering Psychology and Human Performance*. Pearson, 4th edition, 2012.

[7] G. A. Miller. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2):81–97, 1956.

[8] A. D. Baddeley. *Working Memory*. Clarendon Press, Oxford, 1986.

[9] D. L. Schacter. The seven sins of memory: Insights from psychology and cognitive neuroscience. *American psychologist*, 54(3):182, 1999.

[10] E. F. Loftus and J. C. Palmer. Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, 13(5):585–589, 1974.

[11] NTSB. Aircraft Accident Report–Northwest Airlines, Inc., McDonnell Douglas DC-9-82, N312RC, Detroit Metropolitan Wayne County Airport, Romulus, Michigan, August 16,1987. NTSB/AAR-88/05, NTSB, 1988.

[12] G. O. Einstein, L. J. Holland, M. A. McDaniel, and M. J. Guynn. Age-related deficits in prospective memory: The influence of task complexity. *Psychology and Aging*, 7(3):471, 1992.

[13] G. O. Einstein and M. A. McDaniel. Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(4):717, 1990.

[14] J. E. Harris. Memory aids people use: Two interview studies. *Memory & Cognition*, 8(1):31–38, 1980.

[15] E. F. Loftus. Memory for intentions: The effect of presence of a cue and interpolated activity. *Psychonomic Science*, 23(4):315–316, 1971.

[16] R. K. Dismukes and J. L. Nowinski. Prospective memory, concurrent task management, and pilot error. In A. F. Kramer, D. A. Wiegmann, and A. Kirlik, editors, *Attention: From theory to practice*, chapter 16, pages 225–236. Oxford University Press, 2006.

[17] W. B. Rouse and N. M. Morris. On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, 100(3):349, 1986.

[18] K. Craik. *The Nature of Explanation*. Cambridge University Press, 1943.

[19] D. A. Norman. Some observations on mental models. In D. Gentner and A. L. Stevens, editors, *Mental Models*, pages 7–14. Psychology Press, 1983.

[20] D. A. Norman. The design of everyday things. *Doubleday Currency*, 1990.

DEPARTMENT OF PSYCHOLOGY, COLLEGE OF LIBERAL ARTS
ROCHESTER INSTITUTE OF TECHNOLOGY

PSYC 719—HUMAN FACTORS IN AI

HANDOUT: ATTENTION

PROF. RANTANEN

October 18, 2022

1 What is Attention?

"Everyone knows what attention is."

"It is the taking possession by the mind, in clear and vivid form, of one out of what seems several simultaneous possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal of some things in order to deal effectively with others." (William James, 1890).

Several *varieties* of attention may be identified [1]:

1. Selective attention: A serial "searchlight" on selected elements in the external world. Selective attention involves sampling of right information at a right time. Failures of selective attention may result in cognitive "tunneling", such as was the case in the Eastern Airlines accident in Fla. Everglades on December 29, 1972, where three pilots were so focused on troubleshooting a landing gear down-light that they did not notice their aircraft descending into the swamp.
2. Focused attention: Goal-directed orientation of the "searchlight".
3. Divided attention: Parallel processing of multiple channels of information or carrying on multiple tasks simultaneously.
4. Sustained attention: Concentration of effort over time

The searchlight metaphor is nicely illustrated by eye movements, if one thinks of the eye and the gaze direction as a the searchlight. Because the very narrow (about 2°) area of foveal vision (i.e., where we can resolve fine details), it is necessary to move the eyes to collect information about the external world. The eye is controlled by no less than 6 extraocular muscles (the medial and lateral rectus, superior and inferior rectus, and superior and inferior oblique muscles) that allow for 6 *kinds* of very accurate eye movements (saccadic, smooth pursuit, microsaccades, convergence, and optokinetic and vestibular nystagmus). The connection between attention and eye movements is captured in a poem by John W. Senders: [2]

More on the Eye (J. W. Senders, 1980)

"The eye is both a servomechanism and a mécanisme de cerveau.
And sometimes it does its own thing and sometimes it goes
Where the brain wants it to go.

The eyes are the window to the mind and the mind's window
To the scene
So that one is never quite sure whether it is the world or
The mind that makes the eye shift to where it's going
From where it's been.
You can watch the eyes and catch the thought
While it's so hot that even the mind hasn't had it yet.

With a mind of its own the eye looks at the place best calculated
To let the mind's eye see what the mind wants to see;
And then all the world rushes in to be reduced
To common sense and percept before the next saccade is loosed."

2 The Model of Human Information Processing

The human information processing model also shows the central role attention plays in *most* other processes of the model. However, it is worth repeating: "All models are wrong, but some are useful [3]." The model in Figure 1 is a composite of the works by several researchers [4, 5, 6, 7, 8]. Note the boxes from left to right in the middle of the illustration, reflecting the idea of separate processing stages in the information processing system. Information enters from the left and passes through the different stages, transformed in different ways along the way, and each transformation consuming some time.

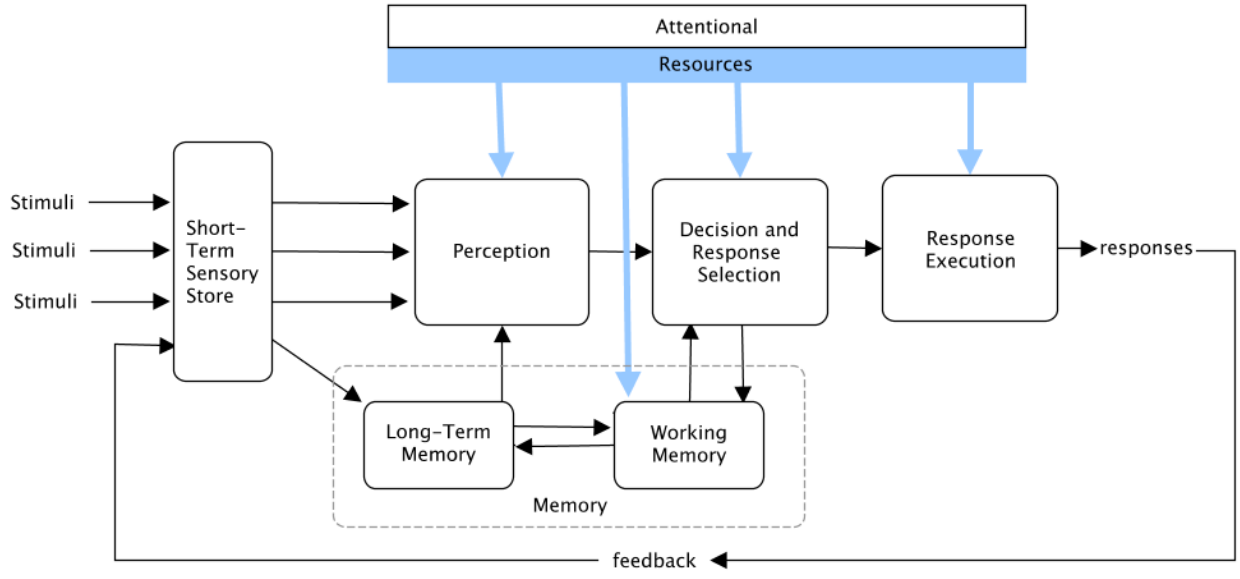


Figure 1. The model of human information processing.

There are also representations of various supporting components, most importantly long-term- and working memory, and attention. The arrows show the relationships the different components have with each other. The depiction of attention as a "reservoir" imply limited attentional resources and the "drain pipes" where these resources are consumed as information is processed.

3 A (Really) Brief History of Theories of Attention

The filter theory of Broadbent [9] posits three stages of attention: (1) selective filter, (2) limited capacity channel, (3) detection device. Stimuli are stored in sensory registers and subjected to preattentive analysis that determines physical characteristics. According to these analyses, a stimulus is selected for further analysis. Selected stimuli shunted along a limited capacity (in terms of number of stimuli) channel to a detection device. If multiple events must be attended to, selection filter switches rapidly between channels in sensory register.

A bottleneck theory supposes that more information is stored simultaneously in sensory register than can be transmitted for further processing, but that the meaning of stimuli are known at the detection stage. This theory has some problems, however. It has been shown that subjects could recognize their name in the non-shadowed channel [10]. Therefore the meaning of the stimulus must have been analyzed, which contradicts the filter and bottleneck theories.

That subjects are able to follow meaningful message in a dichotic listening task when it switched ears led to a "new and improved" theory by Treisman [11], the *attenuation theory*. According this theory, Incoming stimuli go through three tests: (1) physical properties, (2) linguistic properties that group into syllables and words, and (3) recognition of words and meanings. All test are not necessarily completed. According to this theory, non-shadowed messages are not completely tuned out, but attenuated, and preattentive analysis is much more complete than in the filter theory

Another view [12], which supplements the "bottleneck" models, posits that *attentional resources* are not completely fixed, but depend on level of arousal. Furthermore, attention may be *allocated* in proportions to different tasks. Allocation *policy* determines which stimuli get attentional resources. Attention allocation policy is flexible and can be altered to meet demands, but *attentional inertia* results in a tendency to shift attention to a distracter. The model predicts that we are able to do two things at once, as long as the total resources are not exceeded. One of the tasks will suffer if sum of total demands exceeds total capacity. See Figure 2 for an illustration.

A model that nicely illustrates the attentional resources theory is the performance operating characteristic (POC) curve (Fig 3). The idea behind POC is that attention may be *divided* between two tasks, but proportionally. The POC model makes very few assumptions, namely, a complete *complementarity* of processing resources and only weak assumptions about the *mechanisms* of human information processing system (i.e., no need to consider stages or levels of processing). On the other hand, the POC model does not account for multiple resources (more about that later i the course) and it is critical to selected task pairs.

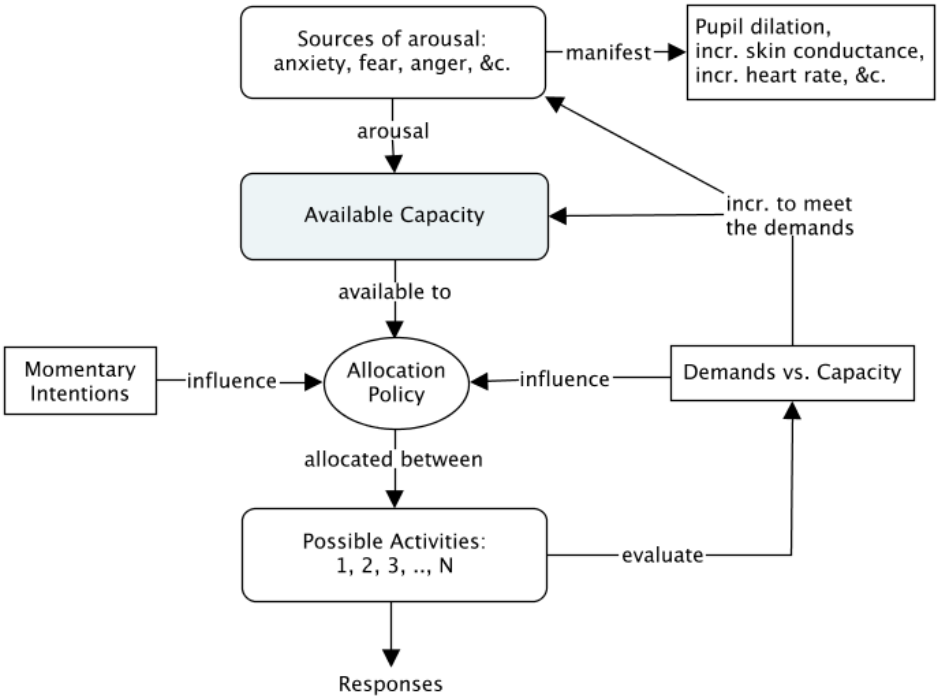


Figure 2. The capacity model of attention; adapted from [12, 8].

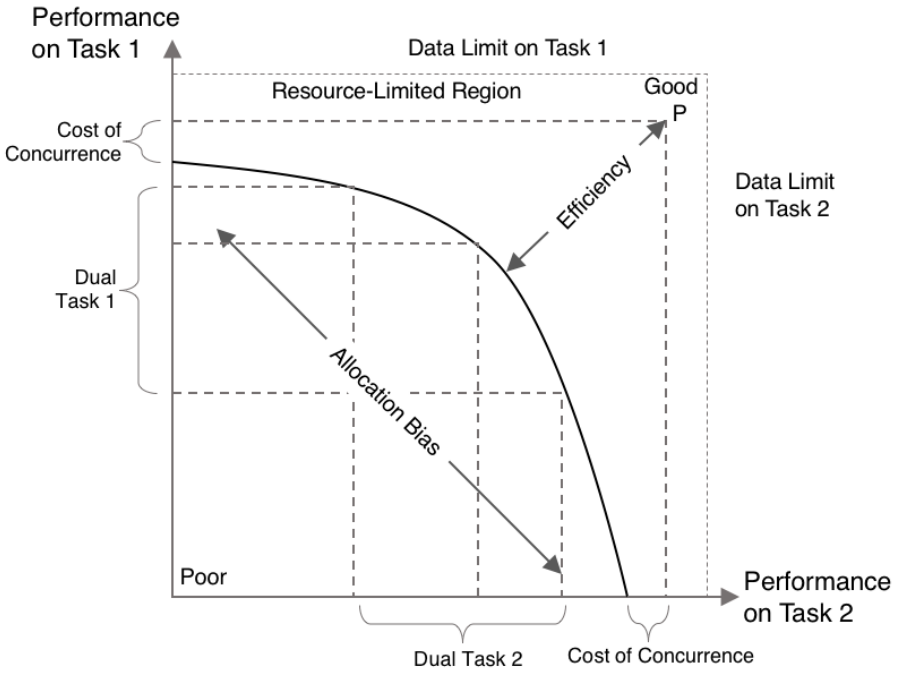


Figure 3. The POC curve (adapted from [8]). The curve in the figure represents performance tradeoffs between the two tasks performed simultaneously; increasing attention allocation to Task 2 (on the x-axis) improves Task 2 performance but at a cost of performance decrement on Task 1 (on the y-axis).

Note that in Figure 3 a mere *presence* of another task in the dual-task paradigm comes at a performance cost of *concurrence* on the other task. Although the POC model does not directly account for multiple resources, the shape of the curve can be used to infer the degree of competition for resources between two tasks: a straight diagonal line between Task 1 and Task 2 suggests *perfect complementarity* of resources, that is, anything towards one task is immediately away from the other; the more concave the curve the less competition for resources there is between the two tasks, and the point marked with "P" would indicate successful dual-task performance

4 More About Selective and Divided Attention

4.1 Controlled and Automatic Processing

Divided attention means simultaneous processing of multiple channels. Successful multi-tasking usually implies high levels of *automaticity* in one task. Automaticity refers to learned sequence of elements in the long-term memory (LTM), initiated by external or internal stimuli, and without conscious control, capacity limits, or attentional demands [13, 14]. In contrast, *controlled* processing is characterized as attention-demanding, capacity-limited, under conscious control. We will return to the automatic and controlled processing later in the course.

4.2 The Stroop Effect

There are several clever experimental paradigms that allow for investigation of divided attention. The Stroop effect [15] is a demonstration of interference between the name of a color (e.g., "red") and the color of ink it is printed that is not denoted by the name (e.g., the word "red" printed in blue ink instead of red ink). The task is to say out loud the color of the *ink*. Reading, (i.e., word recognition) is a largely automatic process but color recognition is not; therefore, saying the color of the ink requires selective attention and is slower than reading the printed word. Furthermore, the automatic response of reading the printed word out loud will *interfere* with the color recognition, further slowing response times.

4.3 Visual Search

Visual search refers to searching for a given *target* in the visual field (e.g., "Where's Waldo"). Visual search can be serial or parallel. Imagine a scenario where you have agreed to meet a friend at an RIT hockey game and your friend is already sitting in the bleachers. If your friend is a true RIT fan, she will be wearing orange, just like everybody else in the ice arena. How hard it will be to find your friend in the packed arena? However, what if your friend had come straight from a formal dinner wearing a dark suit and a tie? Now how hard is to spot him in the sea or orange in the ice arena?

The first scenario above describes *serial* visual search requiring selective attention. In other words, you would need to foveate on each face in the breachers to find your friend. The latter scenario describes *parallel* visual search, or divided attention, where your friend's dark suit would simply "pop out" from the audience requiring no attentional resources. The serial visual search may be modeled by the following equation:

$$T = \frac{(N \times I)}{2} \tag{1}$$

where $T =$ search time, $N =$ number of elements in the search field, and $I =$ the average inspection time for each element. We divide the right side by 2 because *on average* one would find the target after searching *half* of the elements. Parallel visual search is done "in a single glance" and search time does not depend on

the number of items in the search field or the presence or absence of a target. Several target properties help induce parallel search: Discriminability, simplicity, and automaticity. There are also human properties that induce parallel search: Practice, experience, and expectations.

The feature-Integration theory (FIT) [16] posits that individual feature processing is done in parallel. In other words, simultaneous processing is done on the whole display and if feature is present, we detect it automatically. Conjunctive searching requires attention to the integration or combination of the features. Attention to particular combination of features must be done sequentially to detect presence of a certain combination.

There are other theories attempting to explain the observations from visual search experiments. The *similarity theory* [17] disagrees with Treisman's FIT theory [16] and offers alternative explanation to Treisman's data. The theory says that similarity between targets and distracters is important, not number of features to be combined. The more shared features among items in display, the more difficult to detect a particular target.

4.4 Divided Attention

How many tasks can you do at once (e.g., driving and talking on a phone, manipulating radio controls, &c.)? Not that many! Almost 80% of crashes and 65% of near-crashes involved some form of driver *inattention* within three seconds of the event. Data from naturalistic observation of cell phone use and driver behavior showed that that 74.5% of drivers talking on a cell phone failed to stop, while only 25.5% stopped properly. These numbers are reversed for drivers *not* on cell phone while driving, with 21.5% failed to stop and 78.5% stopped properly [18]. Cell-phone conversation led to *inattention blindness*: even if participants looked at an object, they did not remember the object. In the now-famous series of experiments Chabris and Simons demonstrated many cases of such inattention blindness [19].

5 Models of Attention in Instrument Scanning Tasks

There are several seminal experiments that have examined selective attention in realistic instrument scanning tasks. Note how in all the following studies *both* the task demands and human performance are integral parts of the modeling efforts.

5.1 Fitts' Study of Pilot Instrument Scanning

Fitts, Jones, and Milton (1949–1950) [20] collected vast amounts of data on pilots' eye movements during instrument approaches to landing. No instrument readings were recorded. Large variation in dwell times between different instruments were discovered. Variation in dwell times were explained by the *difficulty* of reading the instrument Variation in fixation frequency was explained by the *importance* of the instrument. Results of this study were used to develop the optimal layout of aircraft instrument panels, minimizing the distance of eye movements between frequently viewed instruments and to place the most important instruments right in front of the pilot.

5.2 Sender's Queuing Model of Instrument Sampling

The queuing model of Senders [21] is premised by several key Ideas:

1. Uncertainty grows with time

2. Instruments should be sampled at rates proportional to their bandwidths. Recall that bandwidth = the amount of data that can be passed along a communications channel in a given period of time and that it is measured in bits/s.

3. It is assumed that the observer is a classical Shannon communication channel.

4. It is further assumed that the observer will try to extract all the information from the signal, such that the observations would be necessary and sufficient to reconstruct the signal.

5. A random signal with a limited bandwidth *W* Hz (2W radians; 1 rad = 360°/2) must be sampled at least 2*W* times per second to be reconstructed.

6. Because the response latency is linearly related to the information content of the stimulus (signal), fixation duration (dwell time) should also be proportional to the instrument bandwidth.

   The rate a signal generates information in bits/s is given by the equation

$$\bar{H} = W_i \log_2 \frac{A_i^2}{E_i^2} \tag{2}$$

where W_i = maximum frequency (cutoff frequency), A_i = root mean square (RMS) of amplitude, and E_i = permissible RMS error

In Senders et al. experiment the observers' task was to monitor 4 instruments and press a key at the moment any needle (pointer) exceeded a specified limit. The instruments were driven by a sum of sinusoids quasi-random forcing functions. A camera recorded the observers' eye movements. The performance of the human observers was remarkably similar to that of an ideal observer (i.e., fulfilling all assumptions).

5.3 Carbonnell's (1968) Queuing Model

This model by Carbonell et al. [22] made several important assumptions about the task:

1. Flight instruments can be observed only one at a time; multiple instruments form a queue to be observed;

2. Queue is formed to minimize the risk of not observing critical information at a critical time;

3. Risk = Cost(non-observation) x P(displayed value exceeding threshold);

4. Visual sampling of instruments is part of a feedback loop closed by pilot control actions;

5. P(displayed value exceeding threshold) increases with time.

It is important to note that this is an *optimal, prescriptive, engineering* model. It also contains only expectancy (P) and value (Cost)

5.4 The SEEV Model of Visual Attention Allocation

This model by Wickens [23] not only predicts eye-movements (dwell times at instruments, or areas on interest) but also describes underlying attentional mechanisms. The model adds two parameters to Carbonnell's two, Value (V) and Expectancy (E): Salience (S) of information, and Effort (E) required to access the information. The probability of attending a particular area of interest (AOI) is given by the equation:

$$P(A) = sS - efEF + exEX + vV \tag{3}$$

The letters represent properties of salience, effort, expectancy and value. Upper case denote physical characteristics, whereas lower case denote relative influence of factor on human scanning. Figure 4 illustrates the difference between the queuing and SEEV models.

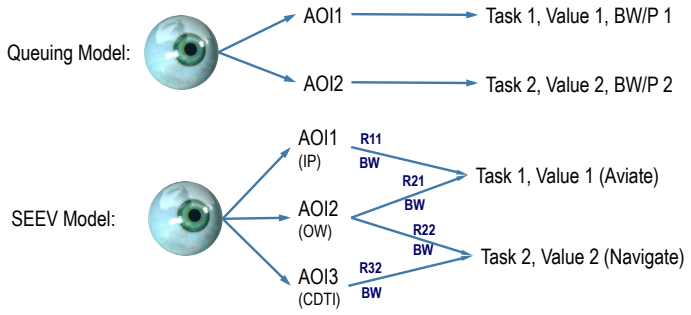


Figure 4. The difference between the queuing and SEEV models.

References

[1] R. Parasuraman and D. R. Davies, editors. *Varieties of Attention*. Academic Press, 1984.

[2] J. W. Senders. Visual scanning processes. University of Tilburg, the Netherlands. Unpublished doctoral dissertation, 1980.

[3] G. E. Box. Robustness in the strategy of scientific model building. In R. L. Launer and G. N. Wilkinson, editors, *Robustness in Statistics*, pages 201–236. Academic Press, New York, 1979.

[4] S. Sternberg. The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 30:276–315, 1969.

[5] D. Broadbent. *Perception and communication*. Pergamon Press, New York, NY, 1958.

[6] E. E. Smith. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin*, 69(2):77, 1968.

[7] A. T. Welford. *Skilled performance: Perceptual and motor skills.* Scott & Foresman, Glenview, IL, 1976.

[8] C. D. Wickens. *Engineering Psychology and Human Performance*. Charles E. Merrill, Columbus, OH, 1984.

[9] D. E. Broadbent. *Perception and communications*. Pergamon, New York, 1958.

[10] N. Moray. Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly journal of experimental psychology*, 11(1):56–60, 1959.

[11] A. M. Treisman. Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, 12(4):242–248, 1960.

[12] D. Kahneman. *Attention and Effort*. Prentice-Hall, Englewood Cliffs, NJ, 1973.

[13] W. Schneider and R. M. Shiffrin. Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84(1):1–66, 1977.

[14] R. M. Shiffrin and W. Schneider. Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, 84(2):127–190, 1977.

[15] J. R. Stroop. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6):643–662, 1935.

[16] Anne Treisman. Search, similarity, and integration of features between and within dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, 17(3):652–676, 1991.

[17] J. Duncan and G. W. Humphreys. Visual search and stimulus similarity. *Psychological Review*, 96(3):433–458, 1989.

[18] D. L. Strayer and F. A. Drews. Cell-phone–induced driver distraction. *Current Directions in Psychological Science*, 16(3):128–131, 2007.

[19] C. Chabris and D. Simons. *The invisible gorilla: And other ways our intuitions deceive us*. Harmony, 2010.

[20] P. M. Fitts, R. E. Jones, and J. L. Milton. Eye fixations of aircraft pilots. III. Frequency, duration, and sequence fixations when flying air force ground-controlled approach system (GCA). Technical Report No. 5967, Wright Patterson Air Force Base, Dayton OH, 1949.

[21] J. W. Senders, J. I. Elkind, M. C. Grignetti, and R. Smallwood. An investigation of the visual sampling behavior of human observers. NASA Contractor Report CR-434, NASA, Washington, DC, 1966.

[22] J. R. Carbonell, J. L. Ward, and J. W. Senders. A queueing model of visual sampling experimental validation. *IEEE Transactions on Man-Machine Systems*, 9(3):82–87, 1968.

[23] C. D. Wickens, J. Helleberg, J. Goh, X. Xu, and W. J. Horrey. Pilot task management: Testing an attentional expected value model of visual scanning. Technical Report No. ARL-01-14/NASA-01-7, University of Illinois, Aviation Research Lab, Savoy, IL, 2001.

DEPARTMENT OF PSYCHOLOGY, COLLEGE OF LIBERAL ARTS
ROCHESTER INSTITUTE OF TECHNOLOGY

PSYC 719—HUMAN FACTORS IN AI

HANDOUT: DECISION MAKING

PROF. RANTANEN

October 25, 2022

1 Preliminaries

There are three features of decision making:

1. Uncertainty and risk: Decisions by definition pertain to the future, which is always uncertain. "Prediction is very difficult, especially about the future" (quote attributed to Niels Bohr). Uncertainty always involves risk.
2. Expertise: Expert and novice decisions are very different. We will get to expertise later.
3. Time: Decisions are often (always?) irreversible due to passage of time. Time pressure is also a critical factor in decision making, especially in situations where correct decisions count most.

We may also distinguish between three classes of decision making research:

1. Normative, *prescriptive* models show how decisions *ought* to be made.
2. An information processing model of decision making is focused on human limitations (memory, attention), and resulting heuristics and biases. The heuristics and biases tradition is a *descriptive* approach to decision making.
3. Naturalistic decision making models are are also descriptive models, and focus on decision making by domain *experts* in the operational context (i.e., outside the laboratory).

2 Normative Decision Models

2.1 Why Norms?

There are many reasons why norms are necessary. Scientific research necessitates measurement, and measurement necessitates some standard. Our need to make value judgments, i.e., "good" or "bad" judgments or decisions also necessitate some standard.

Norms are also needed as means to improve human judgment and decision-making, measurement of systematic errors in judgment and decision-making (biases). This view of norms leads to *prescriptive* models of judgment and decision making (i.e., how people *ought* to make decisions).

Normative models are properly the task of philosophy: They depend on philosophical views of what sort of creatures humans are, as well as views on morality and moral values. But there are problems of distinction between facts and values in social sciences. For example the naturalistic *fallacy* treats "good" as if it were a

natural property.

In judgment and decision making (JDM) literature, norms are closely related to *instrumental rationality*. According to this view, humans are goal-oriented, and "good" or *utility* depends on the extent to which we achieve our goals. This definition bypasses the value problem. The measurement problem remains, however: "Good" or utility must be able to be measured on the interval scale to make statements about "more" or "less good".

2.2 Assumptions (Axioms) of Normative Decision Models

Normative decision models depend on several assumptions:

1. Utility must be transitive: $A > B \wedge B > C \Rightarrow A > C$
2. Utility must be connected: For any A and B, either $A > B$, or $A < B$, or $A = B$
3. Past preferences are valid indicators of present and future preferences.
4. People correctly perceive the values of the uncertainties that are associated with the outcomes of decision alternatives.
5. People are able to assess decision situations correctly, and the resulting decision situation structural model is well formed and complete.
6. People make decisions that accurately reflect their true preferences over the alternative courses of action, each of which may have uncertain outcomes.
7. People are able to process decision information correctly.
8. Real decision situations provide people with decision alternatives that allow them to express their true preferences.
9. People accept the axioms that are assumed to develop the various normative theories.
10. People make decisions without being so overwhelmed by the complexity of actual decision situations that they would necessarily use suboptimal decision strategies.

Even a casual review of the above axioms reveal them to be highly unrealistic in everyday decision making.

2.3 Probability

Because decisions are always done under uncertainty, probability is an integral factor in all decisions. Probability is a very difficult concept, however. Let's look at some definitions.

A dictionary definition of probability is

1. The quality or condition of being probable; likelihood. (This is not helpful, for it only gives us a *synonym* for probability, i.e., "likelihood").
2. A probable situation, condition, or event: Her election is a clear probability. (This is just an example of the use of the word).

3. a. The likelihood that a given event will occur: little probability of rain tonight. (Again, just a synonym and an example; we still do not know what "probability" means).

   b. Statistics. A number expressing the likelihood that a specific event will occur, expressed as the ratio of the number of actual occurrences to the number of possible occurrences. (Finally, an actual definition and a useful computational formula. But, as we small see, this definition applies to only one *kind* of probability).

A better *general* definition of probability is *quantification of uncertainty*. Note, however, that this definition does not tell us *how* to quantify probability (i.e., calculate probabilities.

There are several different approaches to probability and quantification of uncertainty. The main approaches are classical, frequency, propensity, logical (Bayesian), subjective (Bayesian), and mathematical. For the purposes of this class, we only consider the classical, frequentist, and Bayesian positions.

2.3.1 Classical Probability

Classical theory is restricted to equipossible cases, that is, it only applies to equally probable events. Examples are outcomes of tossing a coin, or rolling dice (i.e., equipossible or equiprobable events. Recall the wording of all probability problems in any introductory textbook. The problem *always* started with a statement "A *fair* coin is tossed", or "A pair of *fair* dice are rolled". For a coin of die to be fair, the outcomes (head or tail, numbers 1 to 6) must be *equally probable*. The probability is given by:

$$P = \text{number of favorable equipossibilies}/\text{total number of relevant equipossibilities} \tag{1}$$

There are several problems with the classical approach that limit its applicability to situations that are of interest from decision making perspective. For events to be equipossible, equal probability must have already been assumed, which is circular reasoning. It is also not always appropriate to assign equal probabilities, and alternatives are not always obviously finite and definite in number.

2.3.2 Probability Based on Frequency

Probability is defined as the ratio of times something happens to times it might happen. For example, if the proportion of smokers who die of cancer remains steady at 10 per cent then the probability of smokers dying of cancer is 10 per cent. Probability is defined as the mathematical limit to which the frequency tends in the long run.

The frequentist approach has multiple problems, too. First of all, what defines a "long run"? John Maynard Keynes has been quoted saying "In the long run we are all dead." Frequency theory is restricted to repeatable, random phenomena and limited to classes that have well-defined limits. But what about single, unique events, or events that have never happened yet?

2.3.3 Subjective Probability (Bayesian)

Thomas Bayes (1702–1761) was an English country clergyman, an amateur mathematician, and a gambler. His theory of probability is described in his essay "Towards Solving a Problem in the Doctrine of Chances", published posthumously in 1763. Bayes defined probability as the degree of belief of a particular individual. He made no assumptions that all rational human beings with the same evidence will have the same degree of belief in a hypothesis or prediction (cf. assumptions behind other normative models); differences of opinion are allowed in Bayesian formulation.

Bayes' theorem states that probability (P) of two events (A and B) happening—$P(A, B)$—is equal to the conditional probability of one event occurring given that the other has already occurred—$P(A|B)$—multiplied by the probability of the other event happening—P(B). In other words,

$$P(A, B) = P(A|B) \times P(B) = P(B|A) \times P(A) \tag{2}$$

and

$$P(A|B) = P(B|A) \times P(A)/P(B) \tag{3}$$

Consider the following example of Bayes' Theorem. Data from a large number of mammograms is presented in the table below:

	True State of the World	
X-Ray Results	Cancer	No Cancer
Positive	.792	.096
Negative	.208	.904

In addition, the physician estimated (based on long experience) that a lump in a breast in benign with 99% probability. The Bayes' formula for calculating the probability that *given* a positive test result the lump is cancerous is

$$P(ca|pos) = \frac{P(pos|ca)P(ca)}{P(pos|ca)P(ca) + P(pos|no-ca)P(no-ca)} \tag{4}$$

Plugging in the values from the example table makes things a bit clearer:

$$P(ca|pos) = \frac{(.792)(.01)}{(.792)(.01) + (.096)(.99)} = .077 = 8\% \tag{5}$$

However, when this example was presented to real physicians, most (95 out of 100) estimated $P(ca|pos)$ to be about 75%! This is because the physicians failed to account for the *base rate* of only 1% chance of cancer.

Note that Bayes' theorem is a normative model, that is, is shows how decisions *ought* to be made. Bayesian reasoning is that P(hypothesis) = prior belief $\times$ strength of evidence. The prior is the degree of belief before viewing the data. For example, given my prior beliefs, it would require a great deal more evidence to convince me that astrology works than it would require to convince me that it does not.

In classical statistics, the null hypothesis is that no difference exists and it is presumed to be true initially. Note that from the outset the onus lies with those who hypothesize inequality, rather than those who support the more surprising hypothesis. The conventions of classical statistics (setting alpha less than beta) ensure that scientists bias their decision-making systems towards accepting the null hypothesis when it is false (Type II error, i.e., failing to observe a difference when in truth there is one).

2.4 Subjective Expected Utility (SEU) and Expected Value (EV) Models

2.4.1 Certain Choice

This is a very useful algorithm for making important decisions (such a buying a car) that involve substantial subjective preferences and also very easy to do (see also Fig. 1 below):

1. Rank order the importance of each attribute you consider in a product (highest number = greatest importance), for example, the reliability, fuel efficiency, and cargo capacity in a car;

2. Assess the value of each object on each attribute (e.g., highest number = least expensive), that is, if you are considering, say, five potential makes and models of cars, rank them from best to worst on reliability record, fuel economy, and other important attributes;

3. Assess the sum of the products (Value $\times$ Importance) for each product;

4. Choose to purchase the object with highest sum of products.

		Attributes		
		Price	Usability	&c.
Importance:		1	4	...
Object	Apple	1	5	...
	Dell	3	1	...
	&c.	...	...	...

Figure 1. The compensatory decision process in certain choice cases.

2.4.2 Choice Under Uncertainty

Choice under uncertainty is also known as the expected value model. Consider the example of null hypothesis significance testing, familiar from statistics:

		Null hypothesis (H_0 –no effect)	
		True	False
Decision	True	Correct acceptance ($1-\alpha$)	Type I Error (α)
	False	Type II Error (β)	Correct rejection ($1-\beta$)

Figure 2. A decision matrix in null hypothesis significance testing.

Now consider another example, the costs and benefits of doing research. Imagine that you are a grant officer, sitting on a pile of money (a small pile, enough to fund just one study), and you have to choose a proposal to fund among, say, a 100 proposal you have received. How do you decide which proposal should win?

See Figure 3 for a decision matrix. Note the similarity with SDT. In this case, however, the probabilities of α andβ will have to be estimated, possibly based on your experience and knowledge of the proposers' past work, resources, etc. Each correct decision has some value (V) and each error has some cost (C), which also must be estimated. According to the normative decision theory, you should calculate the *expected value* (EV) of each decision and choose the proposal that has the highest EV. The EV may then be calculated as

$$EV = (1-\alpha)P(S)V(CA) + (1-\beta)P(N)V(CR) - (\beta)P(N)C(T_{II}) - (\alpha)P(S)C(T_I) \quad (6)$$

See where the values come from in Figure 3.

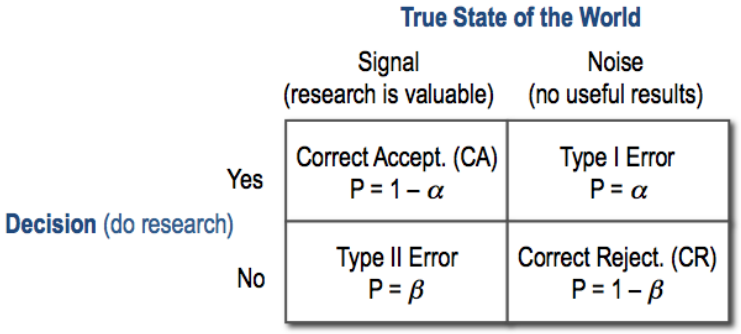


Figure 3. A decision matrix in a hypothetical example.

In normative decision making models each alternative (A) must be assigned a utile, i.e., a numeric value or worth, and each outcome (O) must be assigned a probability (P). The decision rule is simple: Maximize Expected Utility. For example, if A1: U x P = .9, and A2: U x P = .01, then O1 ought to "win". There a some additional assumptions that go with this method:

1. Order alternatives: compare and rank order 2 alternatives (prefer 1 or indifferent)
2. Transitivity: consistent rank order preferences, prefer A>B>C (not C more than A) (utile)
3. Invariance: Decision maker is not affected by way alternatives presented, i.e., by framing effects
4. Dominance: Select alternative (outcome) that has greater utility (i.e., no other strategy—weakly or strongly—dominates on any or all attributes)
5. Cancellation: If 2 outcomes are identical and have equal probabilities ignore utility of outcomes (optimal choice is to leave it to chance!)
6. Continuity: Prefer a gamble between a best and a worst outcome if the odds of the best outcome are good enough
7. Probabilities and utiles can be calculated
8. The decision maker will have complete information about P and U

3 Paradoxes of Rationality

Normative theories show how people *ought* to make decisions. The norms are provided by formal logic, probability theory, decision theory, etc. Normative theories are also know as theories of *rational* decisions. Yet, there is plenty of evidence that people do not make "rational" choices. These exempts are know as *paradoxes of rationality* and some even have names. For example, consider Bernoulli St. Petersburg paradox: A normative model would predict that utility increases linearly with wealth, but this is obviously not true. A sum of $ 1,000 has much utility for someone with no or very little money than for a millionaire. Therefore,

additional assumptions are needed, such as that additional value of money diminishes with wealth, and subjective utility.

The following example illustrates the Allais Paradox. Consider the following two problems; your task is to choose either option A or B based on the information presented:

Problem 1	Winnings	Odds	Problem 2	Winnings	Odds
A:	\$1 million	1.00	A:	\$1 million	0.11
B:	\$2.5 million	0.10		\$0	0.89
	\$1 million	0.89	B:	\$2.5 million	0.10
	\$0	0.01		\$0	0.90

In problem 1, most people choose A even though B has much larger EV (do the math and see for yourself). In problem 2, most people choose B; there is not much difference in odds, but payoff is much greater for B. However, choosing differently between the two problems violates the Cancellation Principle. All these paradoxes violate the assumptions of the Expected Utility theory (or -ies). So, which is right: People making decisions in the context or normative models of how decisions ought to me made. Perhaps a good answer is provided by my favorite quote, by the statistician George Box [1]:

"All models are wrong, but some are useful." (7)

3.1 About Rationality

Are deviations from normative models indications of irrational behavior? Perhaps we first need to find out what is rational:

rational: "Positive term used to commend beliefs, actions, processes as appropriate. In the case of beliefs this means likely to be true, or at least likely to be true from within the subject's perspective. Cognitive processes are rational insofar as they are reliable, and actions rational at least insofar as they provide likely means to an agent's ends."[1]

We must make a few more distinctions:

1. Epistemic rationality and rationality of action; distinction between rationality of belief and inference and rational action;

2. Theoretical and practical reasoning: Theoretical reasoning concerns acquisition of rational beliefs about the world; practical reasoning concerns "judgments" and choice of rational actions;

3. Instrumental rationality: Research on judgment and decision making presupposes a definition of rationality that is basic, or primary. Rational action to achieve goals is the primary notion, rational belief and inference are secondary, or derived notions.

The answer to the question posed above, then, is that the opposite, or alternative, to rational decision making is *intuitive* decision making. This notion also leads to the *information processing* approach to decision making.

[1] *The Oxford Dictionary of Philosophy*. Oxford University Press, 2005.

4 Heuristics and Biases

A heuristic (n., from Gk. "heuriskein", "to find") is defined as process, such as trial and error, for solving a problem for which no algorithm exists. A heuristic for a problem is a rule or method for approaching a solution [2]. Albert Einstein has been quoted to define a heuristc as incomplete (due to limits in knowledge) but useful idea. In computer science Herbert Simon used heuristic searches for faster performance that what could be accomplished by algorithms.

The origins of the heuristics and biases research tradition are in human limitations in clinical judgment. Inferiority of human clinical judgment was deemed due in part to systematic errors, such as the consistent neglect of the base rates of outcomes in discussion of individual cases [3]. Informal judgments were contrasted with statistical techniques, which always turned out to be more accurate and more often "correct". Kahneman called this "illusion of validity", or the unjustified sense of confidence in clinical judgment [4]. Tversky and Kahneman [5] had also shown that sophisticated methodologists and statisticians performed poorly about the sample sizes they considered appropriate in different situations and concluded that "Faulty statistical intuitions survive both formal training and actual experience".

4.1 Properties of Heuristics

Heuristics exploit evolved capacities; they are simple compared to the learned or evolved capacities of organisms. Several important properties of heuristics can be identified:

- Simplicity make for fast, frugal, transparent, and robust judgments;
- Heuristics exploit structures of environment; they are based on ecological rather than rational logic;
- Heuristic are not *optimal* in the sense of optimization algorithms;
- Heuristics *work* (i.e., they are practical)!

The last property above warrants emphasizing. Think how a simple rule could become a heuristic: By working time after time after time. In other words, heuristics have a strong track record of working, that is, producing a correct or at least a "close enough" solution so often that their use makes statistical sense. In other words, the odds of a heuristic working must be so good that it makes more sense to use them rather than elaborate calculations of optimal decisions.

4.2 Examples

There are innumerable heuristics, but a few are so common that they have been named:

The Gaze Heuristic: How to catch a flyball? A simple algorithm guarantees success: (1) keep your eye on the ball; (2) start running; (3) adjust running speed so that the angle between the eye and the ball remains constant. While one cannot compute where the ball will land, this heuristic guarantees that when the ball lands, the player will be there also.

Recognition Heuristic: Goldstein and Gigerenzer [6] asked American and German students about which city has more inhabitants, San Diego or San Antonio (San Diego at the time of the study in 2002). About two thirds of American students got the question right, but 100% of German students who knew little about U.S. cities answered the question correctly, too. How come? The answer is that the German students used a heuristic that bigger cities are better known, and because San Diego is better known in Germany than San

Antonio, it must be bigger.

The Representativeness Heuristic: Consider the following "thumbnail" bio and a question: "Bill is 34 years old. He is intelligent, but unimaginative, compulsive and generally lifeless. In school, he was strong in mathematics but weak in social studies and humanities." Which statement is more probable: A: Bill is an accountant that plays jazz for a hobby, or B: Bill plays jazz for a hobby?

A was erroneously selected by 92% of subjects including those who were informed in matters of statistics. The probability of a conjunction $P(A \wedge B)$ cannot exceed the probability of either of its constituents, $P(A)$ or $P(B)$. Hpwever, conjunction is *more representative* of its class than either of its constituents, or more available in some way, and therefore judgments of its probability are subject to one of the representativeness or availability heuristics.

The representativeness heuristic is easy to demonstrate by listing, say, results of 10 coin tosses. People typically generate 5 heads and 5 tails, because they know that P(Heads) = P(Tails) = 0.5 and make the outcome representative of those probabilities. Such results are too regular to be produced by a random process of coin tosses, however.

Descriptive "thumbnail" bios and knowledge of base rate have resulted in correct estimates of probabilities of randomly chosen person profession, but non-descriptive bio masked the base rate and resulted in erroneous 50% estimates. Generic information is about the frequency of events of that type (e.g., information about the prevalence of the disease). Specific information is about the case in question (e.g., information about the patient revealed by an examination or tests). When contrasted with specific information, generic information is called "base rate" information. People who have only generic information tend to use it to judge probabilities. However, when people have both types of information, they tend to make judgments of probability based entirely upon specific information, leaving out the base rate. Yet, one should use)all available information in decision-making.

The Availability Heuristic. Consider the following example:
Suppose that you wish to buy a new car and have decided that on grounds of economy and longevity you want to purchase one of those solid, stalwart, middle-class Swedish cars, either a Volvo or a Saab. As a prudent and sensible buyer, you go to Consumer Reports, which informs you that the consensus of their experts is that the Volvo is mechanically superior, and the consensus of the readership is that the Volvo has the better repair record. Armed with this information, you decide to go and strike a bargain with the Volvo dealer before the week is out. In the interim, however, you announce this intention to an acquaintance. He reacts with disbelief and alarm: "A Volvo! You've got to be kidding. My brother-in-law had a Volvo. First, that fancy fuel injection computer thing went out. Had to replace it. Then the transmission and the clutch. Finally sold it in three years for junk." Would you still buy the Volvo?"

This is called the availability heuristic, a.k.a. the nnecdotal fallacy, a.k.a. the "Volvo Fallacy". It is based on a biased sample (e.g., "Why, some of my best friends are..."). People tend to judge the probabilities of types of event by using what is called the "availability", or "ease of representation" heuristic. The easier it is to remember, or to imagine, a type of event, the more likely it seems that an event of that type will occur. The anecdotal fallacy occurs when a recent memory, an unusual event, or a striking anecdote leads one to overestimate the probability of events of that type occurring, especially if one has access to better evidence of the frequency of such events.

Fischoff et al. (1978) presented subjects with various versions of a diagram describing ways in which a

car might fail to start. These versions differed in how much a full diagram had been pruned. When asked to judge how complete the diagrams were, the subject were very insensitive to the missing parts. Even omission of major, commonly known components such as ignition or fuel system were barely detected.

5 Information Processing Approach to Decision Making

Recall the 2-system view of human cognition as well as the SRK framework. Rasmussen also developed a decision ladder model of decision making. The model is analogous to an A-ladder, where one climbs up on one side and down on the other. Heuristics serve as shortcuts between "rungs" on the two sides. It is easy to superimpose also this model on the SRK framework (Figure 4).

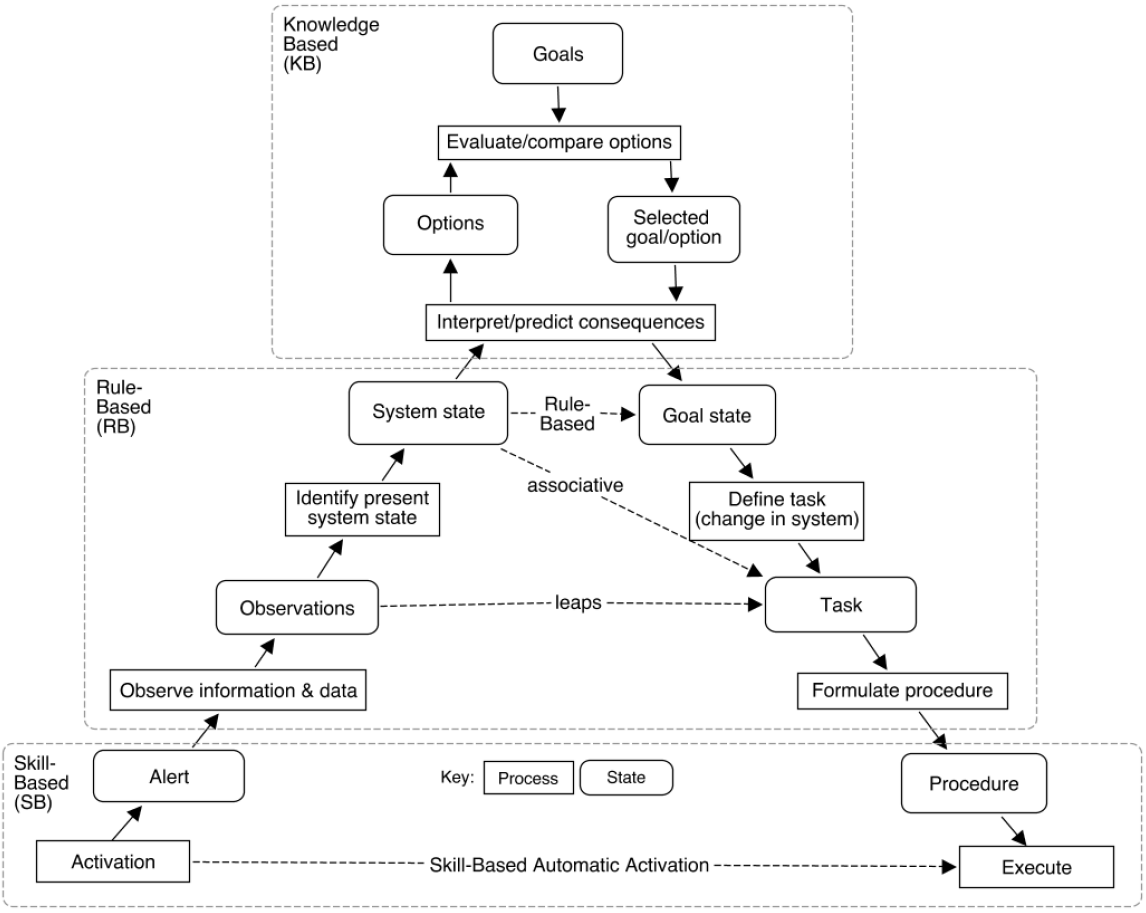


Figure 4. A decision ladder. The decision maker climbs up the ladder on the left-hand side and down on the right-hand side, but may also take "shortcuts" between the "rungs" on the two sides by employing heuristics. The levels on the ladder clearly correspond to the skills-, rule-, and knowledge-based levels in the SRK framework.

6 Naturalistic Decision Making (NDM)

Our final take on decision making is the relatively new approach called *naturalistic decision making* (NDM) [7].

6.1 Features of NDM Settings

There are several features that distinguish NDM from other decision making models and research traditions:

1. Time pressure: 80% fireground commanders make their decisions in $<$ 1 min; often decisions take only seconds.
2. High stakes: Life and death decisions or high monetary stakes (*real* money!).
3. Expertise: The decision-makers are experienced (i.e., typically, only experienced people make high-stakes decisions); experience and expertise in decision-making is a *focus* of research.
4. Uncertainty: Information often missing, ambiguous, or unreliable in naturalistic settings.
5. Goals are unclear: See above.
6. Procedures are poorly defined or altogether missing; in contrast, laboratory studies separate decision-making from problem-solving (hence, problem-solving procedures must be explicit and detailed).
7. Cue learning; Decision makers need to perceive patterns and make distinctions, in contrast to unambiguous stimuli in laboratory studies (e.g., 20% chance to win \$1m of *imaginary* money).
8. Context: Larger than in laboratory studies: higher-level goals, different tasks with different requirements, plus ambient conditions.
9. Dynamic conditions: Changing—often rapidly—situations.
10. Teams: Naturalistic decisions are made in, or about, teams of people.

The NDM approach offers *descriptive* views of how people perform decision making and cognitive activities (including situation conception and problem definition through choice of action) given real-world situations, goals and constraints. It involves models, processes, and characteristics of human decision making and cognition, alone and in conjunction with other individuals and/or with intelligent systems. Decisions are made in the context of ill-structured problems in dynamic, uncertain environments, with shifting, ill-defined, and competing goals, multiple event-feedback loops, time constraints, high stakes, multiple players, and organizational settings [8]. In other words, NDM is the way people use their experience to make decisions in field settings [9]

Contrast between inexperienced and experienced decision makers offer insights into the decision making processes of the latter. Task settings involving ill-structured problems, uncertainty, dynamic environments, complex systems, time stress, risk, multiple, changing goals, multiple individuals, organizational influences, and technologies for assisting, modifying or supplanting human decision making. Training strategies and programs may be devised for influencing or assisting human decision making and cognitive tasks.

The NDM approach is very "hot" presently. For example, the HFES Cognitive Engineering and Decision Making Technical Group is the largest TG of the society. Members of the TG are focused on the study of human decision making and cognition and the application of this knowledge to the design and development of systems and training programs in variety of domains.

The HFES also has a dedicated publication for NDM, *The Journal of Cognitive Engineering and Decision Making*, which "is the premier journal of the Society for peer-reviewed original papers of scientific

merit examining how people engage in cognitive work in real-world settings and how that work can be supported through the design of technologies, operating concepts and operating procedures, decision-making strategies, teams and organizations, and training protocols. Thus, the journal publishes rigorous approaches to the observation, modeling, analysis, and design of complex work domains in which human expertise is paramount and multiple aspects of the work environment may drive performance."

6.2 Recognition Primed Decision (RPD) Model

The NDM approach is formalized in the recognition primed decision (RPD) model, which is based on the many interviews Klein and collaborators did with firemen [7]:

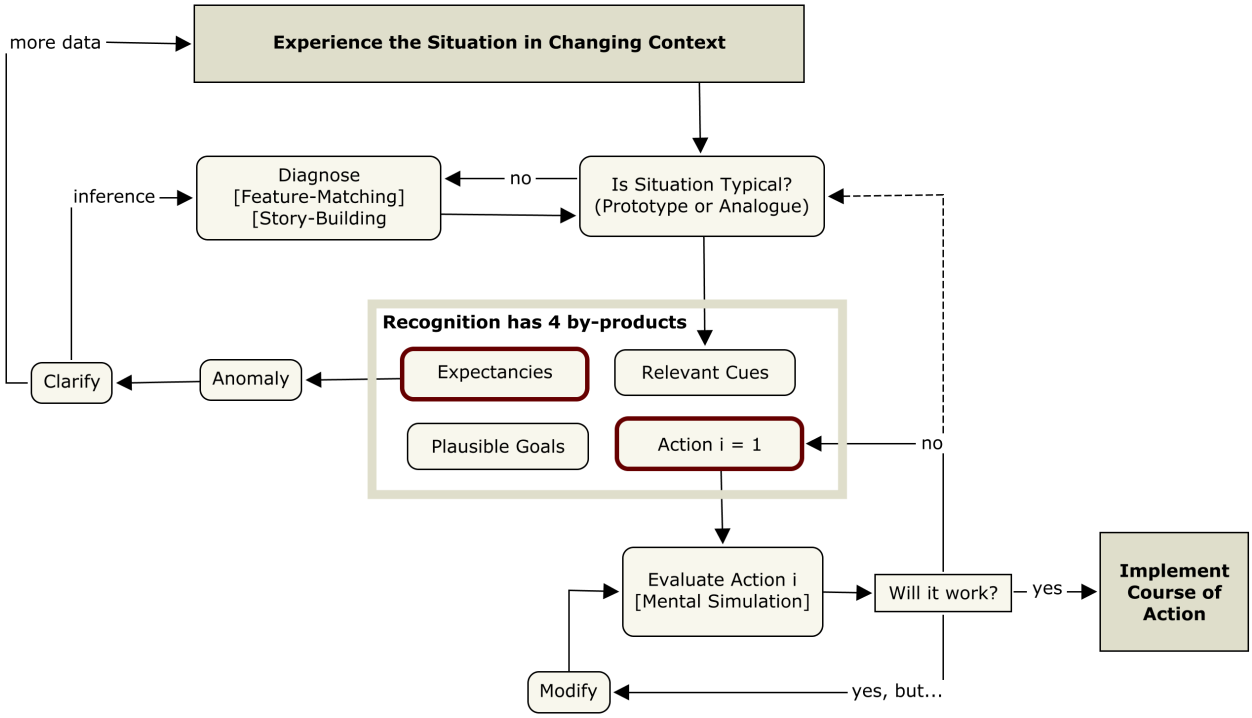


Figure 5. The Recognition Primed Decision (RPD) Model.

6.2.1 Empirical Support for the RPD Model

Can people use experience to generate a plausible option as the first one they consider? Chess experiments show that first moves by experts are stronger than if they were randomly chosen from a set of legal moves. First moves are also rated strong by grand masters. Time pressure should not matter; consider, for example blitz chess (6 s/move).The number of blunders (bad moves) increased for "class B" players in blitz conditions but not for chess experts.

There are also accounts of decision maker adopting a course of action without comparison of possible alternatives and "hypervigilance" by Navy officers. In diagnostic decision making most common diagnostic strategy is feature matching, having a strong resemblance of the *representativeness heuristic*, or identification of relevant features of a situation to categorize it. Minority of cases studied involved *story building*, or

attempts to synthesize the features of the situation into a causal explanation.

In problem solving the traditional prescriptive decision analysis paradigm emphasized "filtering down" of options to arrive at the best one, whereas RPD model emphasized the role of expertise to generate options in the first place.

7 Contrasts Between Decision Research Traditions

It is useful to examine the decision making literature and research traditions historically. The classical decision making tradition (incl. normative decision models) has its roots in the 18th century and extend to the 1940s and 1950s. It was followed by the behavioral decision theory (BDT) and related *judgment and decision making* (JDM). The *heuristics and biases* approach is also part of this era, extending into the 1970s. This in turn was followed by what had become known as the *naturalistic decision making* (NDM) approach from the 1980s on, started by Klein et al. [10] study of fire ground commanders. It is also illuminating to compare these traditions side-by-side:

7.1 Contrasts Between CDM and JDM/BDT

CDM	JDM/BDT
DM conceptualized as choice between concurrently available alternatives	Bounded rationality and adaptive behavior
Input-output orientation (preferences of the decision-maker)	Loosely coupled real-world problems can be approached sequentially
Comprehensive information processing and analysis	Effective adaptation does not require comprehensive analysis
Formal, context-free models	Systematic deviations from the rational choice models (bit retained these as standards)

7.2 Synthesis of HB and NDM (?)

Intuitive judgments can arise from genuine skill (NDM) but also from inappropriate application of the heuristic processes (HB). Skilled judges are often unaware of the cues that guide them, and individuals whose intuitions are not skilled are even less likely to know where their judgments come from. True experts know when they don't know. Non-experts do not know when they don't know. Subjective confidence is an unreliable indication of the validity of intuitive judgments and decisions.

The determination of whether intuitive judgments can be trusted requires an examination of the environment in which the judgment is made and of the opportunity that the judge has had to learn the regularities of that environment. Task environments are "high-validity" if there are stable relationships between objectively identifiable cues and subsequent events or between cues and the outcomes of possible actions. Medicine and firefighting are high validity environments. Predictions of the future value of individual stocks and long-term forecasts of political events are made in a zero-validity environment.

7.3 Contrasts Between JDM/HB and NDM

JDM/HB	NDM
Compare expert performance with performance by formal models or rules.	Admiring stance toward experts.
Expect that experts will do poorly in such comparisons.	Identify critical features of the situation that are obvious to experts but invisible to novices.
Bias towards replacing informal judgment with algorithms.	Critical of formal approaches because these are hard to impose on judgments and choices made in complex contexts.
Favor well-controlled experiments in the laboratory.	Favor an ecological approach, question the relevance of laboratory experiments to real-world situations, methods include cognitive task analysis and field observation of complex conditions that would be difficult to recreate in the laboratory.
"Expert" is an optimal decision-maker.	Experts are those who have been recognized within their profession as having the necessary skills and abilities to perform at the highest level.

References

[1] G. E. Box. Robustness in the strategy of scientific model building. In R. L. Launer and G. N. Wilkinson, editors, *Robustness in Statistics*, pages 201–236. Academic Press, New York, 1979.

[2] S. Blackburn. *The Oxford Dictionary of Philosophy*. OUP Oxford, 2005.

[3] P. E. Meehl. *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* University of Minnesota Press, 1954.

[4] D. Kahneman, P. Slovic, and A. Tversky. *Judgment under uncertainty: Heuristics and biases*. Cambridge university press, 1982.

[5] A. Tversky and D. Kahneman. Belief in the law of small numbers. *Psychological Bulletin*, 76(2):105–110, 1971.

[6] D. G. Goldstein and G. Gigerenzer. Models of ecological rationality: the recognition heuristic. *Psychological review*, 109(1):75–90, 2002.

[7] G. A. Klein. *Sources of power: How people make decisions*. MIT press, 1999.

[8] J. Orasanu and T. Connolly. The reinvention of decision making. In G. A. Klein, J. Orasanu, R. Calderwood, and C. E. Zsambok, editors, *Decision making in action: Models and methods*. Ablex Publishing, Westport, CT, 1993.

[9] C. E. Zsambok. Naturalistic decision making research and improving team decision making. In C. E. Zsambok and G. A. Klein, editors, *Naturalistic decision making. Expertise: Research and applications*, pages 111–120. Lawrence Erlbaum Associates, Hillsdale, NJ, 1997.

[10] G. A. Klein, R. Calderwood, and A. Clinton-Cirocco. Rapid decision making on the fire ground. Technical Report Tech. Rep. No. TR-8546-12, Alexandria, VA, 1985.

DEPARTMENT OF PSYCHOLOGY, COLLEGE OF LIBERAL ARTS
ROCHESTER INSTITUTE OF TECHNOLOGY

PSYC 719—HUMAN FACTORS IN AI

HANDOUT: SITUATION AWARENESS

PROF. RANTANEN

November 1, 2022

1 Why is Situation Awareness Such an Important Construct?

The August 14, 2003, blackout in Northeast US/Canada left 50 million people without power and had an estimated cost between $4 and $10 *billion*. The subsequent investigation cited "inadequate situation awareness" in control room for grid operations: "Training deficiencies, ineffective communications, and inadequate reliability tools and backup capabilities all contributed to a lack of situation awareness (SA) for the operators involved." [1]

In aviation, as much as 88% of human errors are due to problems with SA. Pilots do not get information that is needed (78%), do not correctly understand information they do get (17%), and do not project what will happen in the future (5%).

2 Definitions

- Construct: (n) an idea or theory containing various conceptual elements, typically one considered to be subjective and not based on empirical evidence.
- Situation awareness (SA): The first and still most common and generally accepted definition of SA is by Endsley [2]: "Situation awareness is the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future."

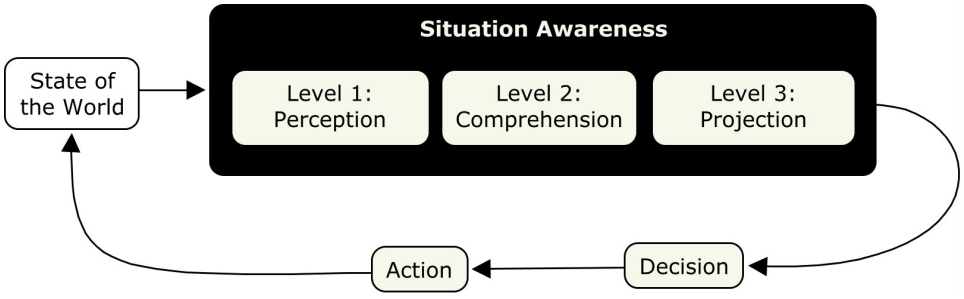


Figure 1. Situation awareness as a basis of human performance.

Note the three distinct *levels* of SA in this definition. Figure 1 illustrates the role of SA in human performance. According to this view, SA is fundamental to decision making and subsequent response selection (actions). A particular critique of SA is that it is just a "buzzword of the 90s" (and even today), a name for something we know little about, or a "black box" model of phenomena that are not understood.

3 Levels of Situation Awareness

Perception (Level 1), comprehension (Level 2) and projection (Level 3) are not necessarily linear stages. Levels 2 and 3 can be used to drive the search for Level 1 information (Fig. 2. Also, default values from the mental model can provide reasonable values, even when no Level 1 info has been perceived on an element

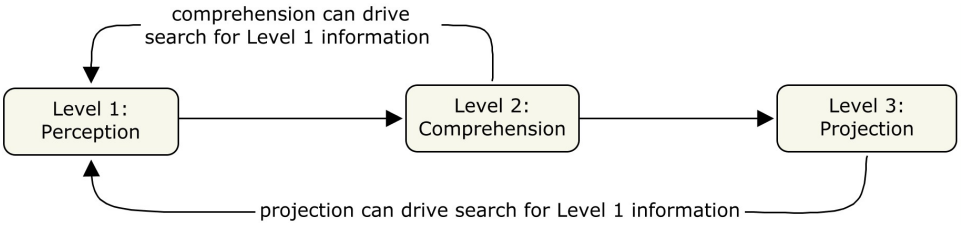


Figure 2. Interactions of the three levels of SA.

4 Mental Models and Situation Awareness

Recall the distinction between bottom-up and top-down processing. Bottom-up, or *data driven processing* depends on external stimuli or cues in the environment. Salient cues "catch" attention, are interpreted, options what to do about the cues are generated and evaluated, an option is selected, and appropriate actions are taken. In contrast, top-down or *goal driven processing* is characterized by goals that direct attention, determine development of Level 2 SA, and determine selection of a model for interpreting information (Fig. 3).

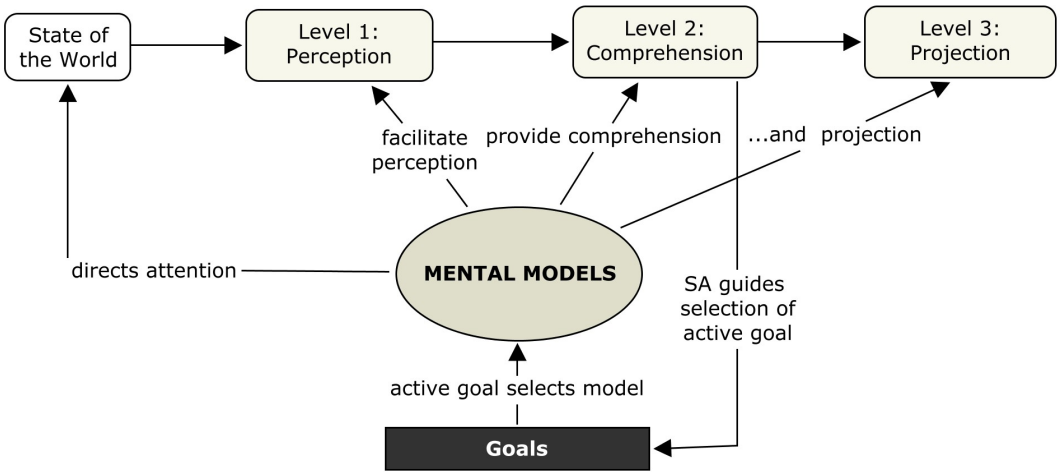


Figure 3. The role of mental models in SA.

Goals are key organizing feature for cognitive work. SA ties goals to mental models and explains dynamic goal switching and dynamic prioritization of Information. SA also facilitates selection of mental models and development of more refined schema and models through homomorphisms.

5 Measurement of SA

As is the case with any theoretical construct, methods for *measurement* of SA is critical to validation of SA and refining the theory with empirical data.

5.1 The Situational Awareness Rating Technique (SART)

The simplest way to measure SA is to ask the participant or operator to rate their own perception of their SA in a given task or situation. SART is a multi-dimensional rating technique (3-D SART) with three primary rating dimensions, corresponding to the three clusters of the original constructs elicited from military aircrew: (1) Demand on attentional resources, (2) supply of attentional resources, and (3) understanding. Naturally, subjective rating scales have several drawbacks, such as accessibility of SA by introspection and criteria used to assess SA [3].

5.2 Situation Awareness Global Assessment Technique (SAGAT)

SAGAT is an objective method of measuring SA. Mission or task simulations are frozen at randomly selected times, the system displays are blanked and the simulation is suspended while operators quickly answer questions about their current perceptions of the situation. The questions correspond to their situation awareness requirements as determined from the results of an SA requirements' analysis. Operator perceptions are then compared to the real situation, based on simulation computer databases, to provide an objective measure of SA [4].

5.3 The Situation Present Assessment Method (SPAM)

SPAM is based on the assumption that SA involves simply knowing where to find a particular piece of information, as opposed to remembering what that piece of information is. For example, an air traffic controller might need to remember aircraft call signs if they are able to retrieve than information from the environment if required [5]. Figure 4 illustrates the SPAM procedure and products.

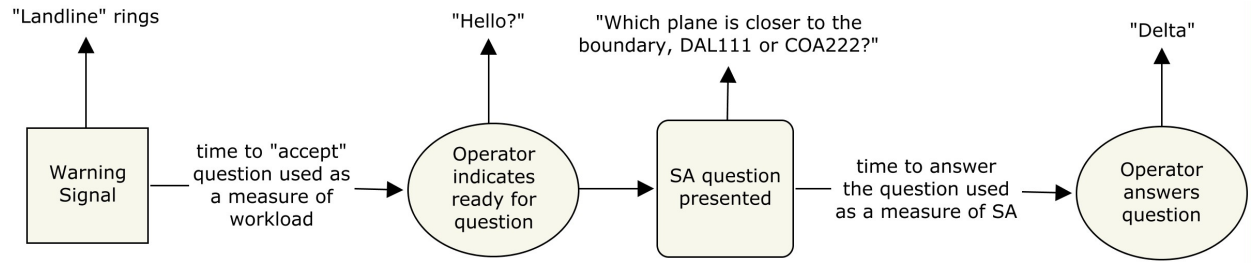


Figure 4. The SPAM procedure.

References

[1] U.S.-Canada Power System Outage Task Force. *Final report on the August 14, 2003 blackout in the United States and Canada: Causes and recommendations*. 2004.

[2] M. R. Endsley. SAGAT: A methodology for the measurement of situation awareness. Technical Report NOR DOC 87-83, Northrop Corp., Hawthorne, CA, 1988.

[3] R. M. Taylor. Situational Awareness Rating Technique (SART): The development of a tool for aircrew systems design. Technical Report SEE N 90-28972 23-53, AGARD, 1990.

[4] M. R. Endsley. Situation awareness global assessment technique (sagat). In *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference (NAECON)*, pages 789–795. IEEE, 1988.

[5] F. T. Durso, A. R. Dattel, S. Banbury, and S. Tremblay. SPAM: The real-time assessment of SA. In S. Banbury and S. Tremblay, editors, *A cognitive approach to situation awareness: Theory and application*, chapter 8. Ashgate, Aldershot, UK, 2004.

DEPARTMENT OF PSYCHOLOGY, COLLEGE OF LIBERAL ARTS
ROCHESTER INSTITUTE OF TECHNOLOGY

PSYC 719—HUMAN FACTORS IN AI

HANDOUT: CONTROLS AND DISPLAYS

PROF. RANTANEN

November 8, 2022

1 Preliminaries

To put control in context, consider the model of human information processing system and some system to be controlled (a nuclear power plant, an airplane, even your automobile); the human operator needs and *interface* between him or her and the system. The interface must necessarily have two parts, (1) *displays* to convey information about the system's state (e.g., the speed the car is traveling on a freeway) to the operator, and (2) *controls* by which the operator can change the state of the system (e.g., accelerator and steering in a car). See Figure 1 for an illustration.

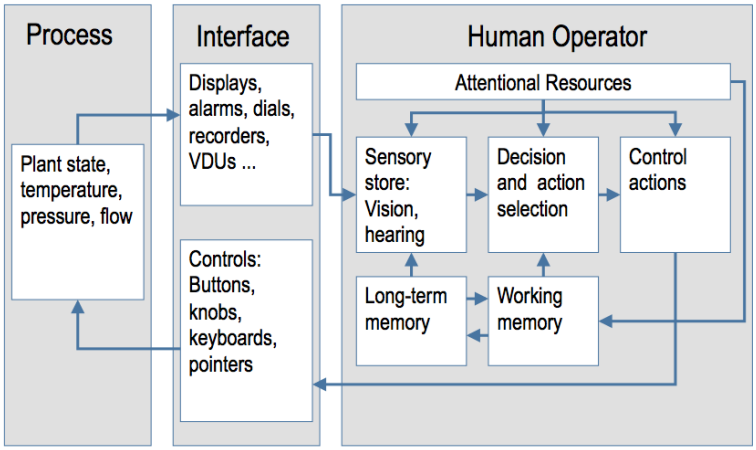


Figure 1. The model of human information processing in context.

2 Controls

There are several important factors that must be understood and considered when examining control systems:

- Discrete of continuous control
- Open- or closed-loop control
- Positive or negative feedback
- Stability
- Control order (zero-, first-, or second-order control)
- Time delays (lags)
- Gain
- Tracking displays (pursuit or compensatory tracking)

Another important concept to note is the *stimulus-response (S-R) compatibility* are the departures from the information theory assumptions. S-R compatibility is defined as "the expected relationship between the location of a control or movement of a control response and the location or movement of the stimulus or display to which the control is related." Good S-R compatibility reduces response time by reducing the uncertainty associated with appropriate movement of the control. Departures from the information theory include stimulus discriminability (e.g., consider 4 and 7 vs. 721834 and 721837; the former would result in faster responses because of the greater discriminability), the repetition effect (advantage over alternations), confusability and complexity of the stimuli, and practice effects.

3 Control Models

3.1 Fitts' Law

Fitts' idea was to determine the information capacity of the human motor system, relying on the metaphor of information transmission. According to this metaphoric account the difficulty of a task can be measured in bits, familiar from the communication theory of Shannon and Weaver (1949). In carrying out a movement task information is transmitted through a human channel.

A measure of difficulty of the task is the index of difficulty (ID), usually expressed as:

$$ID = \log_2 \left(\frac{2A}{W} \right) \tag{1}$$

where A is the movement amplitude and W is the target width. Thus, large movements to narrow targets are more difficult than small movements to wide targets. Fitts' index of performance (IP) is analogous to Shannon's channel capacity and is calculated as:

$$IP = \frac{ID}{MT} \tag{2}$$

where MT is the observed movement time. IP can also be calculated by regressing MT on ID, resulting in an equation of the regression line:

$$MT = a + b(ID) \tag{3}$$

where a and b are the regression coefficients. Hence, Fitts' law usually takes the form

$$MT = a + b \log_2 \left(\frac{2A}{W} \right) \tag{4}$$

where a is the y-intercept and b is the slope of the line.

Figure 2 illustrates the Fitts' task and the *law like* relationship between target width (accuracy requirement), movement amplitude, and movement time (speed).

Example: Consider two different controls: A trackball and a joystick; one could devise and perform the Fitts' task with different accuracy requirements (W) and thus IDs with each, and compute $MT = a + b(ID)$. Suppose that the slope b turns out much smaller for the trackball than for the joystick; what would that tell about the two different controls? Which would be *better*?

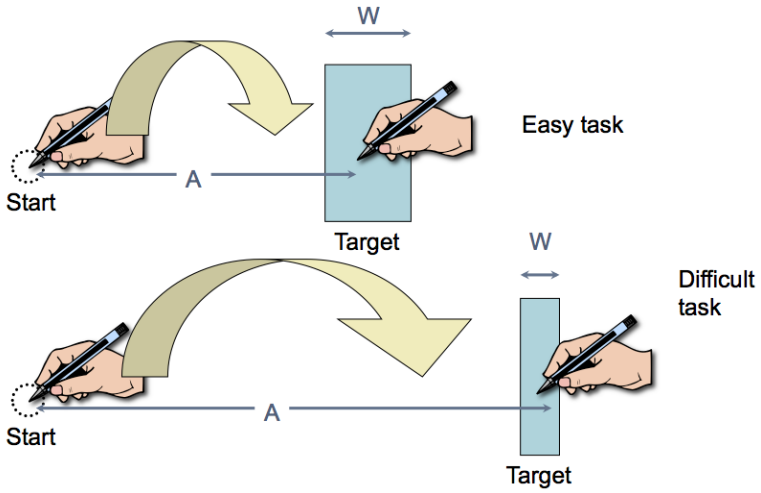


Figure 2. The Fitts' task and the *law like* relationship between target width (accuracy requirement), movement amplitude, and movement time (speed).

Despite its simplicity, Fitts' Law is immensely influential and useful in many aspects of human-machine interaction. Consider the common input device of alphanumeric keyboards. The common QWERTY keyboard is *difficult* to use, for it has over 50% within-hand consecutive keystrokes, over 60% of keystrokes away from the "home-row", and 10% of keystrokes that require moving a single hand two rows from the "home-row" (for typing in English). All these factors work to increase movement time, i.e., to slow typing speed (but then, that was the very *purpose* of the keyboard design!). QWERTY also has uneven distribution of strokes among the fingers resulting in overload of some fingers. The Dvorak keyboard has 5%–20% speed advantage over QWERTY and reduced hand and finger fatigue, but cost and effort of retraining the population (changing a standard) are important considerations, too.

Of the two kinds of numeric keypads ("telephone", which has 1-2-3 on the top row, and "calculator", which has 7-8-9 on the top row), the telephone arrangement has been shown to be slightly faster than calculator arrangement. Alternating between keypads resulted in poorest performance.

A specific application of Fitts' Law is in the GOMS (Goals, Operators, Methods, and Selection rules) Model. GOMS is an HCI model developed by Stuart Card, Thomas P. Moran, and Allen Newell [**?**]. *Goals* are what the user intends to accomplish, *operator* is an action performed in service of a goal, *method* is a sequence of operators that accomplish a goal, and *selection rule* is to choose among more than one method available to accomplish a goal (a.k.a. CMN-GOMS, where CMN = Card, Moran, Newell).

A Keystroke Level Model (KLM) GOMS is a simplified version CMN-GOMS, eliminating the goals, methods, and selection rules, leaving only 6 primitive operators: (1) pressing a key, (2) moving the pointing device to a specific location, (3) pointer drag movements, (4) mental preparation, (5) moving hands to appropriate locations, and (6) waiting for the computer to execute a command. The times for each of the six operations determined empirically by Fitts' Law. The operations for a compete task are arranged into a serial stream, and total task execution time is a simple calculation.

3.2 Continuous Control and Tracking

Continuous control means continuous tracking and correction of errors, that is, deviations from some set point or criterion state of the system (e.g., consider lane keeping while driving a car). Control systems are often automated (e.g., a thermostat, or even a toilet tank that automatically shuts off water when the tank is full). For illustration of simple control systems, see Figure 3.

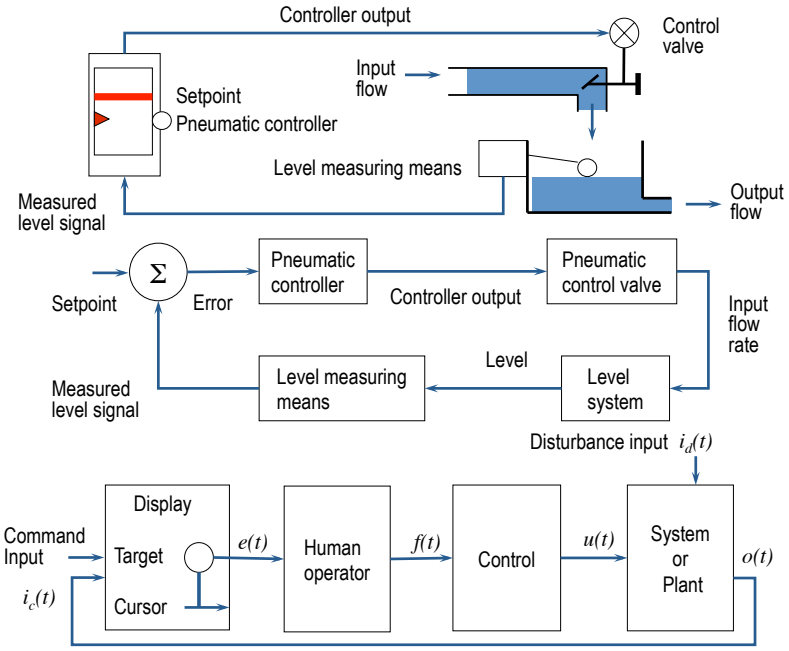


Figure 3. Block diagram of a simple control system. It is also easy to include the human operator in models like this and represent him or her by some appropriate transfer function.

3.2.1 Feedback

Examples of *positive* feedback include a booster rocket, or backing a tractor-trailer. Response error is to the same direction as the error. These systems are very difficult to control (as anyone who has tried to back up a tractor-trailer combination can attest) and typically not suitable for human control. In *negative* feedback systems response to error is opposite (= negative) to the error (e.g., lane-keeping while driving: if the car is drifting to the left, the driver will turn the steering wheel to the right). We quickly learn to control a negative feedback system without even thinking about it (i.e., automatically).

3.2.2 Stability

This concept should be familiar to everyone from high school physics. Figure 4 offers an illustration.

3.2.3 Control Order

Zero-order or *position* control refers to a direct change in system as a result of a change in some control. For example, movement of a computer mouse results in a corresponding movement of the cursor on the screen, or a turn of a steering wheel in a car results in the front wheels turning to some angle away from their straight position.

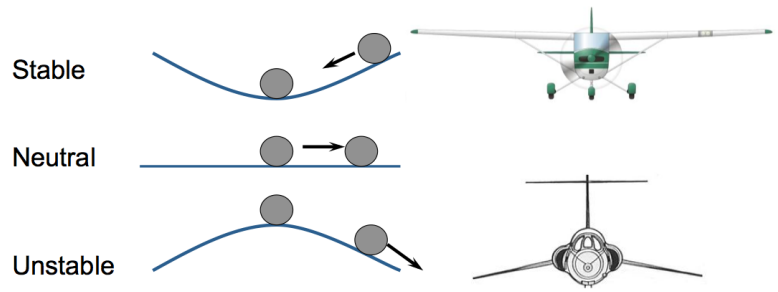


Figure 4. Which aircraft (Cessna C-150 on the top, Lockheed F-104 on the bottom) appears more stable and thus easier to control?.

First order, or *velocity* control has the control input result in a *start* of a change in the system. For example, changing the position of the steering wheel of a car (= input) results in a rate of change of the heading of the automobile (=output); to stop the car from turning, the steering wheel must be returned to the neutral (center) position. The scanner in a car radio works the same way, continually jumping from station to station and it must be stopped at a desired station. Also joysticks are 1st order controllers.

Second order or *acceleration* control results in *increasing* (i.e., accelerating) response of the system. For example, turning a steering wheel results in an increasing distance of the car relative to the lane. Balancing an empty soda can on a board is also an example of 2nd order control. First and second-order systems also have *inherent* time lags, making their control very difficult (See Fig. 5).

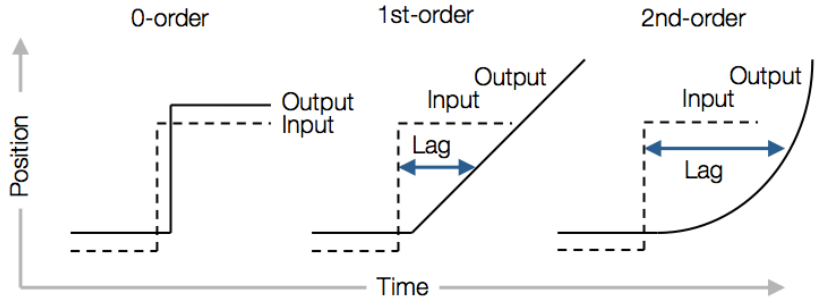


Figure 5. Different responses in 0-, 1st-, and 2nd order control systems. The dashed lines represent control input, solid lines system response. Note that *time lags* are *inherent* to 1st- and 2nd order systems.

3.2.4 Time Lags

Time delays are *always* harmful. The greater the delay, the poorer the performance. Time lags require *anticipation* of system responses, which in turn requires cognitive resources and is generally imperfectly done. Figure 6 illustrates the role of time lags. Note that also 0-order systems may have time lags, depending on the system design. Figure 7 illustrates the complex control inputs required for 1st and 2nd order systems.

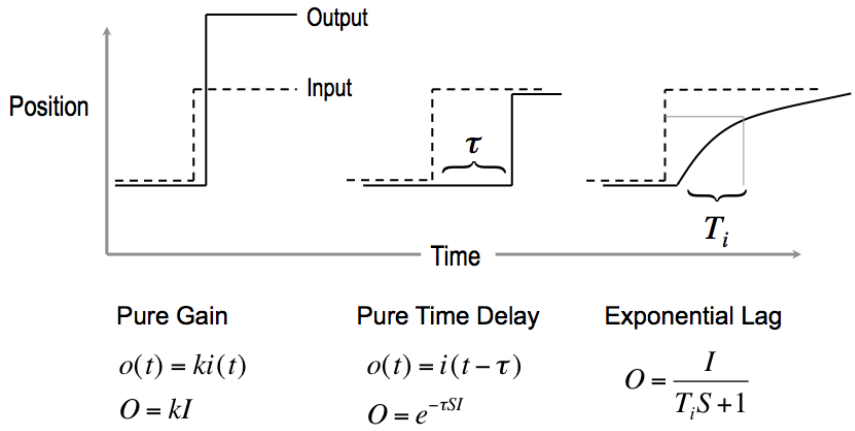


Figure 6. The role of time lags in control systems.

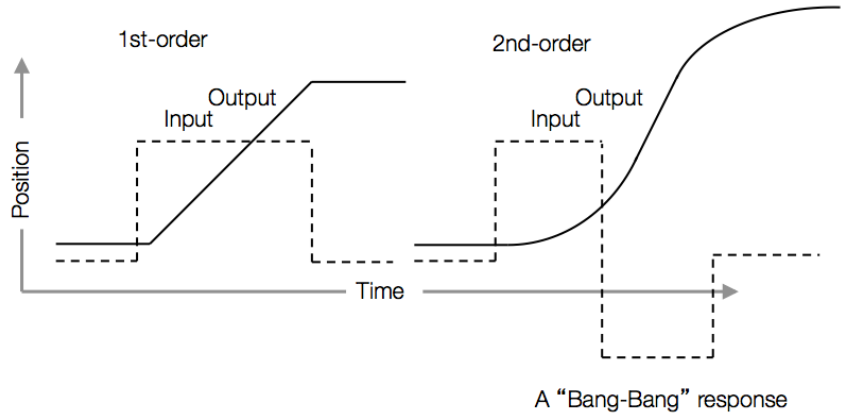


Figure 7. The required control inputs for 1st and 2nd order systems. Note the complexity of control of the 2nd order system and anticipation required to achieve the desired system state.

3.2.5 Gain

Gain is the ratio of system output to control input. High gain may results in overcorrection and loss of accuracy, whereas low gain makes control effortful and slow. An example is different steering ratios of different automobiles. Generally, a "happy medium" of gain affords best performance. Note that gain can also be variable, such as speed-sensitive steering in some cars of in computer track pads.

3.2.6 Tracking

There are two distinct types of tracking, *pursuit* tracking and *compensatory* tracking. Pursuit tracking has independent representation of movement of both target and cursor (control). In compensatory tracking only the movement of error relative to a fixed reference is displayed (e.g., flight instruments).

Tracking tasks can be aided by preview, allowing for prediction of the command input For example, preview of the road ahead in driving makes the lane keeping task quite easy, especially when compared to trying to maintain position on the lane by watching the lane edge marking from the side window of a car (do *not* try

this at home!). Output prediction and quickening may also be used in displays to aid operators. Predictive displays show the future state of the system; see for an example Figure 8 for an EHSI (Electronic Horizontal Situation Indicator) (bottom left).

Figure 8. A compensatory display (top left) only shows error relative to a reference, in this case a given radial from a radio beacon (the aircraft is to the right of the radial and should steer left to follow it). The display on the bottom left shows an EHSI; the dashed line emanating from the "ownship" symbol shows the *predicted* turning radius of the aircraft as it is following an alternative course, shown by the magenta line, and offering a *preview* to the pilots. The display on the right shows a "tunnel in the sky" where pilots are offered cues where they *should* fly as well as preview where the aircraft *will* go given the present control inputs.

3.2.7 Performance Measurement

Performance in tracking tasks is commonly measured in terms of error, or the difference of the desired and actual paths or system states. Position error error refers to frequent deviations around the desired state or path. Velocity error represents tracking "out of phase" (see Figure 9 for illustrations).

Note that if we simply *averaged* the error in the examples in Figure 9, it would be close to zero. However, tracking performance in these examples was far from perfect. To accurately represent error in tracking, we must eliminate the signs. This is commonly done by calculating the Root Mean Square Error (RMSE), given by the following formula:

$$RMSE = \sqrt{\frac{\sum_{n=1}^{i}(X_i - \bar{X})^2}{n}}$$

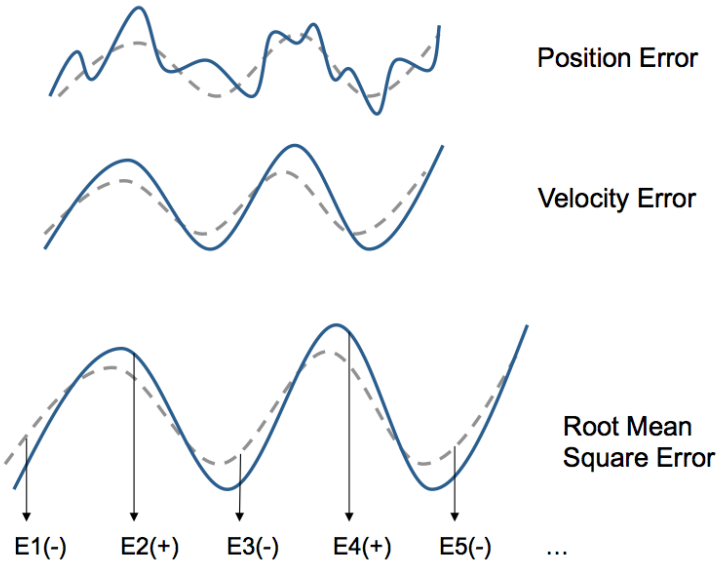


Figure 9. Measurement of error (performance) in tracking tasks.

3.3 Human Operator Limitations

Processing delay from perception of error to decision on a corrective action to execution of the correction (cf. the model of human information processing) is an inherent limiting factor in tracking performance. First order systems may be successfully tracked with 150-300 ms time delays. Second order systems exhibit 400-500 ms delays, reflecting complexity of decision-making.

The limit of information transmission in tracking is about 4–10 bits/s. The limit on frequency at which corrective decisions can be made, that is, max. bandwidth of random inputs, is therefore about 0.5–1.0 Hz.

Performance may be improved beyond these limitations by having a good mental model of system dynamics, which allows for accurate anticipation and even open-loop control, and compatibility of control and display relationship in tracking (S-R compatibility). Conversely, incompatible compensatory displays, e.g., left-moving error cursor requires right-moving response, or the tracking motion is not aligned with display motion, hurt performance.

4 Displays

In design of information displays it it very important to consider a multitude of factors and criteria. These criteria are perhaps best expressed in various taxonomies of displays. The designer should choose the best *kind* of display for a particular purpose. The main considerations are sensory mode and type of information to be displayed

1. Sensory mode:
   (a) Visual; perhaps the most common class of displays.
   (b) Auditory; the auditory channel should not be ignored, especially if the visual channel is becoming overloaded with too much information to be displayed.

(c) Haptic; examples include a cell phone ringer set to vibrate, keyboard response, touch screen response.

(d) Olfactory; an example of an olfactory display is the odor added to gas to make gas leaks perceptible.

2. Type of information:

(a) Static vs. dynamic: Does the information stay the same like in a road sign or does is change in time like in a car's speedometer?

(b) Qualitative vs. quantitative: Is an actual numerical value important, or does the operator only be aware of a limit or a safe operating range?

(c) Discrete vs. continuous: A status indicator such as a traffic light displays discrete information, a thermometer continuous.

4.1 Principles of Display Design

Wickens [1] developed 13 principles of display design based on the known human capabilities and limitations:

4.1.1 Perceptual Principles

1. Make displays legible (or audible). A display's legibility is critical and necessary for designing a usable display. If the characters or objects being displayed cannot be discernible, then the operator cannot effectively make use of them.

2. Avoid absolute judgment limits. Do not ask the user to determine the level of a variable on the basis of a single sensory variable (e.g. color, size, loudness). These sensory variables can contain many possible levels.

3. Top-down processing. Signals are likely perceived and interpreted in accordance with what is expected based on a user experience. If a signal is presented contrary to the user's expectation, more physical evidence of that signal may need to be presented to assure that it is understood correctly.

4. Redundancy gain. If a signal is presented more than once, it is more likely that it will be understood correctly. This can be done by presenting the signal in alternative physical forms (e.g. color and shape, voice and print, etc.), as redundancy does not imply repetition. A traffic light is a good example of redundancy, as color and position are redundant.

5. Similarity causes confusion: Use discriminable elements. Signals that appear to be similar will likely be confused. The ratio of similar features to different features causes signals to be similar. For example, A423B9 is more similar to A423B8 than 92 is to 93. Unnecessary similar features should be removed and dissimilar features should be highlighted.

4.1.2 Mental Model Principles

6. Principle of pictorial realism. A display should look like the variable that it represents (e.g. high temperature on a thermometer shown as a higher vertical level). We read English from left to right, and numbers on our rulers grow from left to right; low numbers should therefore be to the left of the scale and large numbers to the right there are multiple elements, they can be configured in a manner that looks like it would in the represented environment.

7. Principle of the moving part. Moving elements should move in a pattern and direction compatible with the user's mental model of how it actually moves in the system. For example, the moving element on an altimeter should move upward with increasing altitude. On a moving tape display (where the scale moves behind a small window) the movement indicating increase in values should be to the right. However, in the latter case the principle of pictorial realism would be violated (see Fig. 10)

Figure 10. A panel-mounted magnetic compass (left) has—necessarily, due to its design—the scale values in reverse order, larger numbers to the left. If a pilot, who in this case is on a heading 014, should fly heading 060 (indicated by number 6 on the scale), he or she must turn *right*, not left. This is very confusing and student pilots take a long time to lear to fly on magnetic compass, always turning the wrong way first. On the right is directional gyro, which is slaved to a separate magnetic compass. This display heeds both principles of pictorial realism and moving part. A pilot is in this case on heading 030 (read at the nose of the aircraft symbol) and to fly heading 060 should turn to the right—immediately apparent on the display.

4.1.3 Attentional Principles

8. Minimizing information access cost. When the user's attention is diverted from one location to another to access necessary information, there is an associated cost in time or effort. A display design should minimize this cost by allowing for frequently accessed sources to be located at the nearest possible position. However, adequate legibility should not be sacrificed to reduce this cost.

9. Proximity compatibility principle. Divided attention between two information sources may be necessary for the completion of one task. These sources must be mentally integrated and are defined to have close mental proximity. Information access costs should be low, which can be achieved in many ways (e.g. proximity, linkage by common colors, patterns, shapes, etc.). However, close display proximity can be harmful by causing too much clutter.

10. Principle of multiple resources. A user can more easily process information across different resources. For example, visual and auditory information can be presented simultaneously rather than presenting all visual or all auditory information.

4.1.4 Memory Principles

11. Replace memory with visual information: knowledge in the world. A user should not need to retain important information solely in working memory or retrieve it from long-term memory. A menu, checklist, or another display can aid the user by easing the use of their memory. However, the use of memory may sometimes benefit the user by eliminating the need to reference some type of knowledge

in the world (e.g. an expert computer operator would rather use direct commands from memory than refer to a manual). The use of knowledge in a user's head and knowledge in the world must be balanced for an effective design.

12. Principle of predictive aiding. Proactive actions are usually more effective than reactive actions. A display should attempt to eliminate resource-demanding cognitive tasks and replace them with simpler perceptual tasks to reduce the use of the user mental resources. This will allow the user to not only focus on current conditions, but also think about possible future conditions. An example of a predictive aid is a road sign displaying the distance to a certain destination.

13. Principle of consistency. Old habits from other displays will easily transfer to support processing of new displays if they are designed consistently. A user's long-term memory will trigger actions that are expected to be appropriate. A design must accept this fact and utilize consistency among different displays. If display layout is inconsistent, adequate alerts must be provided (for example, if an exit ramp from freeway is on the *left* side, it should be preceded by several signs over several miles prior to the exit signing "Exit [number] Left, [number of] miles").

4.1.5 Multiple Displays: Layout Principles

Layout principles of multiple displays follow directly from the above 13 display design principles:

- Frequency of use: Minimize information access cost (P8)
- Display relatedness or sequence of use: Proximity compatibility principle (P9)
- Consistency: Minimize information access cost (P8 and P13)
- Organizational grouping: Minimize information access cost (P8)
- Stimulus-Response (S-R) compatibility: Display should be close to its associated control (P9)

4.2 Configural Displays

These are a special case displays that combine two or more variables and display their values in such a way that a new feature *emerges*, which provides the operator a holistic picture of the system. Such emergent features include line (e.g., lining up needles on multiple dials to resemble a straight line across the dials), area and shape (e.g., Fig. 11, on the left), or symmetry (e.g., the octagonal nuclear safety parameter display, on the right in Fig. 11).

Configural, or or object, displays potentially improve decision-making performance by shifting the load from limited cognitive processes (e.g., working memory) to virtually unlimited processes (e.g., object perception and pattern recognition), but these benefits depend on task requirements. Tasks that require parallel processing benefit from object displays. However, if extraction of information from a single variable is required, object displays may hinder performance. This follows from the proximity compatibility principle: Tasks requiring information integration benefit from proximate displays, whereas tasks requiring independent processing or focused attention on individual variables benefit from separate displays.

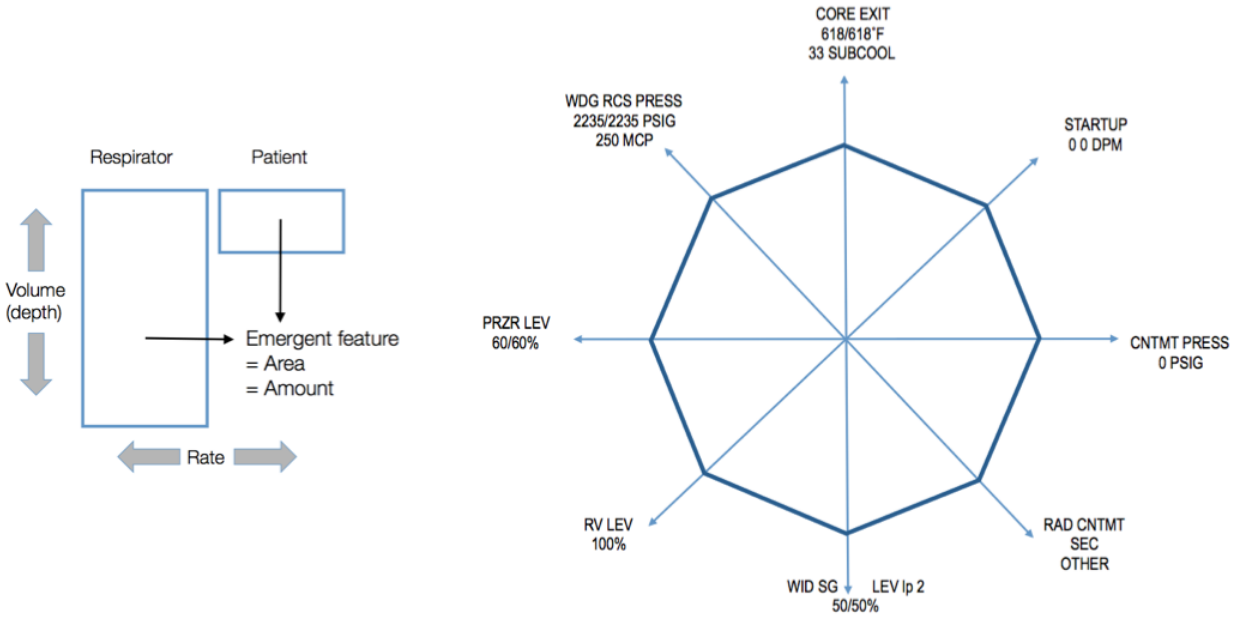


Figure 11. Examples of emergent features from object displays. On the left, the object is a rectangle, and its *area* represents the total *amount* of oxygen the percent is receiving. On the right is a safety parameter display for nuclear power plants; it combines 8 critical parameters on the 8 vertices and displays normal operation as a perfectly symmetrical octagon. Any deviations of any parameters will be immediately apparent in asymmetry of the figure.

References

[1] S. K. Card, T. P. Moran, and A. Newell. *The psychology of human-computer interaction*. Erlbaum, Hillsdale, NJ, 1983.

[2] C. D. Wickens, S. E. Gordon, and Y. Liu. *An Introduction to Human Factors Engineering*. Longman, New York, 1998.

DEPARTMENT OF PSYCHOLOGY, COLLEGE OF LIBERAL ARTS
ROCHESTER INSTITUTE OF TECHNOLOGY

PSYC 719—HUMAN FACTORS IN AI

HANDOUT: AUTOMATION

PROF. RANTANEN

November 19, 2022

1 Preamble

In this handout about automation, or rather, human-automation interaction, it is important to recognize two very different approaches, or angles, to the topic. The first one, which is what most of what follows is drawn, concerns automation in high-risk, complex, task environments with highly trained and professional human operators (e.g., aviation, power systems, manufacturing, healthcare). The two key features of this domain are their limited scope (i.e., limited to a particular industry) and lack of human variability due to the highly uniform training and experience of the operators.

The second angle to automation is particularly pertinent to the "Intelligent Infrastructure" (II), affecting large populations without training and with huge individual variability, that is, everybody. The challenge is to apply what has been learned over several decades of research into human-automation interaction in specialized, "high-tech" domains to "Everyman's" experiences with "everyday" automation.

We know from experience that there are unintended consequences from overreliance on automation. The primary problem with artificial intelligence (AI) and machine learning (ML) is that they are not transparent. "Clumsy Automation" [1] and "Automation Surprises" [2] terms were coined in the 80s and the concepts are applicable to AI/ML, too. Human operators (mostly pilots in the 80s and 90s) ran into trouble when they could not understand what autopilots were doing, and those were more straightforward, deterministic systems. Today's algorithms are more capable, dangerous, and can cause more operator confusion (as recently illustrated with the Boeing 737 flight accidents). Automation based on AI/ML will be completely opaque and its actions inscrutable by humans. [3]

Back in the 80s there were plenty of high-profile, tragic, aircraft crashes that were scrutinized and that brought up the problems with autopilots that were beyond pilots' training. If AI-based diagnostic systems misdiagnose one patient at a time at some (relatively low) rate, nobody will pay attention, or bother to collect such instances as data. There may be incident reporting systems in healthcare that could be accessed and analyzed, but that is just one domain of AI/ML applications.

There is no reason to believe that this will not continue to be true in the world of ML models that drive automation. In particular, as ML models are able more to mimic elements of human behavior and decision making, there may be an increased tendency to trust these models beyond a prudent level. Consider a recommender system that suggests the right additional products to solve someone's shopping needs. When the system starts acting like it has unique knowledge it becomes easier to trust, and harder to decide to counteract when needed.

As AI systems become increasingly ubiquitous and capable, and their widespread applications in myriad

domains and across all societal elements, the influence of AI, including its unintended and unforeseeable consequences, will also be felt on a larger, societal, scale than has been a case with earlier technologies. The most difficult research questions therefore pertain to scale. Current and future ML/AI-based automation will penetrate every aspect of people's lives, through networked devices (Internet of Things, IoT) and collection of vast amounts data through them, and sharing these data in myriad forms across the IoT and different government and private entities. Past approaches to human-system integration and cognitive systems engineering must be correspondingly "scaled up". As AI-based automation will touch the lives of everybody, increasing human variability across heterogeneous user groups presents additional challenges [4].

2 Definitions

Automation may be defined as "...execution by a machine agent (e.g., a computer) a function that was previously carried out by a human." However, this definition is so broad that is covers almost everything in our modern lives and blurs the line between automation and (mere) machine operation. Would you consider most household appliances as forms of automation? At least the autocorrect function on word processors and smartphones and automatic transmission in cars have the word "automatic" in them. Note, too, that automation of physical functions is very common, but automation of cognitive functions (still) relatively rare, although AI- and ML-based automation is changing this situation rapidly.

3 When and What to Automate

There are many areas where it is preferable to have a machine rather than a human perform certain tasks. Dangerous tasks are such, where protection of humans requires keeping them away from the immediate task environments, for example in hazardous materials handling. Tasks that exceed human capabilities are another candidate for automation, such as computational task, tasks requiring extreme vigilance, or tasks performed by disabled populations. Tedious and repetitive tasks are also frequently automated due to their regular nature (e.g., assembly jobs, routine or "mindless" tasks) and as humans do not thrive in such task environments. Automation is used to improve safety, reduce *human* error (but see below), improve surveillance (vigilance), improve weather data gathering and presentation, improve reliability of equipment (monitor equipment condition and reduce maintenance costs), prevent system overload, improve overall system efficiency, and to reduce delays.

4 Levels of Automation

The following 10-level classification of automation was proposed by [5] and is still very influential:

1. The computer offers no assistance; the human must do it all
2. The computer offers a complete set of alternatives,
3. ...and narrows the selection down to few,
4. ...or suggests one,
5. ...and executes the suggestion if the human approves,
6. ...or allows the human a restricted time to veto before automatic execution,
7. ...or executes automatically, then necessarily informs the human,
8. ...or informs him or her after execution only if he or she asks,
9. ...or informs him or her after execution if it, the computer, decides to do so.
10. The computer decides everything and acts autonomously, ignoring the human.

The last two levels always bring to my mind Stanley Kubrick's 1968 movie "Space Odyssey 2001", and the final dialog with the astronaut Dave Bowman and the computer running the spaceship, HAL:

Dave Bowman: Open the pod bay doors, HAL.
HAL: I'm sorry, Dave. I'm afraid I can't do that.
Dave Bowman: What's the problem?
HAL: I think you know what the problem is just as well as I do.
Dave Bowman: What are you talking about, HAL?
HAL: This mission is too important for me to allow you to jeopardize it.
Dave Bowman: I don't know what you're talking about, HAL.
HAL: I know that you and Frank were planning to disconnect me, and I'm afraid that's something I cannot allow to happen.

5 The Substitution Myth

The "Substitution Myth" [2] debunks the notion that human operators, error-prone and overworked as they may be, could simply be repaced by automation. Table 1 juxtaposes the putative benefits of automation with the real complexities introduced by it.

Table 1. *The putative benefits and the real complexities of automation according to [2]*
.

Putative Benefits	Real Complexities
Improves results with simple substitution	Transforms practice and creates new roles for humans
Offloads work	Creates new cognitive demands
Offloads requirements for attention	Requires active tracking and integration of multiple activities
Requires less knowledge	Requires different knowledge and new skills
Operates autonomously	Requires but does not support team-work between humans and machine
Integrates all necessary data	Requires new levels and types of feedback to recognize what is informative in context
Provides generic flexibility	Feature-rich designs result in new demands, error opportunities, and paths to failure
Reduces human error	Creates new kinds of failures

Additional problems with automation are unevenly distributed workload, where automation may reduce workload in low-workload situations but increase it in high-workload situations (e.g., cockpit automation). This is referred to as "clumsy automation", edistribution of workload over time instead of overall reduction [1]. Automation often changes in the quality of workload, merely shifting it from control to monitoring.

Automation brings about ew attentional and knowledge demands as operators need to learn not only how the system works, but also "how to work the system", requiring consideration of both the capabilities and limitations of the system, increasing demand on training and requiring lmuch learning "on-the-job". Consequently, operators may only use a subset of a system.

"Automation surprises" [2] refers to mode confusion, or breakdown in mode awareness. Multiple levels of automation result in mode-rich systems, increasing the complexity of mode interaction as well. Increased

autonomy of modern systems result in increased delays between user input and observable system response.

New coordination demands come in two-phase decompensation, where automation may compensate for abnormal situations to a point, but then "dumps" the problem to human operator when it encounters conditions it has not been programmed to handle. Communication and coordination between human and an automated system are especially necessary when automatic system is having trouble handling the situation or when automation is taking extreme action or moving toward extreme part of authority. Automation may infringe teamwork and communication between team members, mask human incompetence, and make monitoring between team members difficult.

6 Automation: Use, Misuse, Disuse, and Abuse

This section title is from a seminal article by [6]. Human attitudes toward automation are an important factor in human-automation interactions. Automation reliability largely determines if automated systems are used as intended. If a smoke alarm in your house or apartment goes off every time you cook, it is very likely that you will disable it. Some people, especially elderly and inexperienced users may suffer from "technofobia" and forgo automated aids altogether. There is no evidence that task difficulty would increase the use of automation, and so-called "cognitive overhead" refers to the cost-benefit ratio of use of automation (i.e., is the task easier to perform with or without automation? Think of examples from your experiences with "everyday automation"). Risk associated with use or not to use automation must also be evaluated separately, as they are different in many ways.

In sum, human-automation interaction involves myriad factors, multiple interactions between the factors, and large individual differences. This makes research into and modeling of human-automation interactions extremely difficult.

One of the most critical factor in human-automation interaction is *trust* [7]. Trust has multiple attributes. These attributes are easier to understand if we think about them in interactions between two humans, instead of human and machine.

- **Reliability**: Consistent functioning of automation (cf. consistent behavior of a friend);
- **Robustness**: Demonstrated or promised ability to function in a variety of circumstances (cf. so-called "fair-weather friends", who would not be "friends in need");
- **Familiarity**: System employs procedures, terms, and cultural norms that are familiar, friendly, and natural to the user. Consider forming a trusting relationship with someone from a foreign land, with very unfamiliar customs, manners, and culture;
- **Understandability**: Match between the system and the operator's mental model of the system (just how well do you *know* your friend);

  Explication of intention: The system explicitly displays that it will act in a certain way (cf. human friends letting you know of their plans);
- **Usefulness**: Is the system truly useful for the task it is to be used for, or will using an automated system introduce extra steps or delays to task performance?
- **Dependence**: This is different from dependability of an automated system and it means the degree you are dependent of automation, possibly because of lack of your own skills or time. Note that

you can be dependent on an unreliable system, just as we sometimes depend on unreliable people in relationships.

A key to *proper* trust is *calibration*, the matching of perceived and real automation reliability and avoidance of both mistrust and overtrust (complacency). Mistrust results from the "Cry Wolf" phenomenon (and I trust that you are all familiar with the story of the boy who cried wolf!). Overreliance in automation results in decision biases and failures of monitoring, premature cognitive commitment to automation, and loss of operator's skill

7 Ironies of Automation

This section title, too, is a title of a seminal article [8]. One of the ironies is that automation increases the proportion of designer's errors, i.e., *latent errors* embedded in the systems design. Another irony is that automation seeks to eliminate the human, and human error, yet leaves the human operator "do the tasks which the designer cannot think to automate". Deskilling results from automated systems offering little opportunity for skill maintenance; yet when automation fails, the human operator needs to be more skilled, not less, to deal with the situation. The "Catch-22" of human supervisory control may be stated thus:

Q: Why do we need human operators in complex systems?
A: To cope with emergencies and unforeseen, unimaginable situations.

Q: What will humans use to deal with such situations?
A: Stored routines based on previous interactions with a specific environment.

Q: What is their experience with automated systems?
A: Monitoring and occasionally adjusting the system.

Q: How do humans actually perform in case of automation failures?
A: Given the complexity of the system and alien task, terribly!

References

[1] E. L. Wiener and R. E. Curry. Flight-deck automation: Promises and problems. *Ergonomics*, 23(10):995–1011, 1980.

[2] N. B. Sarter, D. D. Woods, C. E. Billings, et al. Automation surprises. In G. Salvendy, editor, *Handbook of Human Factors and Ergonomics*, volume 2, pages 1926–1943. Wiley, 1997.

[3] D. Hannon, E. M. Rantanen, B. Sawyer, R. Ptucha, A. Hughes, K. Darveau, and J. D. Lee. A human factors engineering education perspective on data science, machine learning and automation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 63, pages 488–492. SAGE Publications, 2019.

[4] E. M. Rantanen, J. D. Lee, K. Darveau, D. B. Miller, J. Intriligator, and B. D. Sawyer. Ethics education of human factors engineers for responsible ai development. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 65, pages 1034–1038. SAGE Publications, 2021.

[5] T. B. Sheridan and W. L. Verplank. Human and computer control of undersea teleoperators. Technical report, Massachusetts Institute of Technology Cambridge Man-Machine Systems Lab, 1978.

[6] R. Parasuraman and V. Riley. Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2):230–253, 1997.

[7] R. Parasuraman and M. Mouloua. *Automation and human performance: Theory and applications*. Human Factors in Transportation. CRC Press, 1996.

[8] L. Bainbridge. Ironies of automation. *Automatica*, 19(6):775–779, 1983.