# Mental Models in Human Factors

'Mental Models' are internal, simplified representations of external systems. They enable individuals to interpret, predict, and interact with the world around them. Within the field of [[Human Factors]], mental models serve as a conceptual tool to understand and improve human-system interaction, particularly in technologically complex environments such as aviation, healthcare, and industrial control.

## Historical Background

The idea of mental models can be traced back to the work of [[Kenneth Craik]] in 1943, who suggested that the human mind builds internal models of reality in order to anticipate the future [16]. [[Philip Johnson-Laird]] developed this idea further in the 1980s, proposing a cognitive theory in which mental models form the basis of human reasoning and understanding [13]. The rise of [[Human-Computer Interaction]] (HCI) during this period prompted increased interest in applying cognitive theories to design problems. In 1989, Wilson and Rutherford published a comprehensive review aiming to clarify conceptual confusion and bridge theoretical psychology with human factors applications [1].

## Definitions and Core Concepts

A mental model, in the context of human factors, is generally understood as an internal, dynamic, and subjective representation of a system or process. These models are constructed through past experiences and current interactions, and they guide individuals in reasoning, decision-making, and problem-solving [1]. While this definition may appear straightforward, mental models are in fact complex cognitive entities that differ significantly depending on context, user experience, and the nature of the system being modeled.

One of the most influential contributions to the conceptual understanding of mental models in design comes from Donald Norman (1983, 1986), who framed mental models as central to system usability. He introduced a tripartite model composed of the designer’s conceptual model, the system image, and the user’s mental model [2][17]. The designer’s conceptual model refers to the intended functionality and internal logic of a system. The system image includes all external cues provided to the user, such as visual layout, controls, and documentation. The user’s mental model is based on how these cues are interpreted, often shaped by prior knowledge and expectations. A mismatch between these components can lead to confusion, errors, and inefficient system use.

Norman also emphasized that mental models are not static or fully accurate replicas of real-world systems. Instead, they are simplified, functional approximations designed to allow the user to operate effectively within a specific context. This means mental models may lack detail, be incomplete, or even contain incorrect in assumptions but still support acceptable performance. This property makes them both powerful and fragile: powerful in supporting action and reasoning, fragile in their susceptibility to error under novel or unexpected circumstances.

Beyond their structural components, mental models also encompass procedural knowledge. They allow users to simulate actions internally and anticipate outcomes, a quality that is particularly important in dynamic systems like aircraft cockpits or medical equipment. Mental models thus serve as internal simulators, enabling users to run “what-if” scenarios to predict system behavior.

The utility of mental models is further enhanced by their adaptability. They evolve with use, learning, and feedback. When users encounter new situations or when system behavior contradicts expectations, their mental models are revised sometimes incrementally, sometimes through more substantial restructuring. This capacity for model updating is a key characteristic and distinguishes mental models from more static knowledge forms such as schemas or frames.

From a human factors perspective, this plasticity is critical. It means that mental models can be shaped not only by direct experience but also through thoughtful design, training, and documentation. Therefore, system designers have the opportunity and responsibility to facilitate the construction of accurate and useful mental models through interface design, metaphors, tutorials, and consistent feedback mechanisms.

## Mental Models and Related Concepts

While mental models are central to understanding user behavior in complex systems, they are often conflated with other cognitive constructs. Schemas, for instance, refer to generalized knowledge structures that help individuals interpret information based on prior experiences [3]. Scripts are sequences of expected behaviors in familiar contexts, such as dining at a restaurant [4]. Frames, introduced in the context of artificial intelligence, are data structures that represent stereotypical situations with predefined slots and default values [5]. Related are "folk theories" and "naive theories" that users form to explain how systems work, even if those explanations are technically incorrect [18][19].

## Typologies and Structures

Wilson and Rutherford distinguish between physical and conceptual mental models. Physical models refer to representations of tangible systems, such as the internal mechanism of a gearbox. Conceptual models, in contrast, are more abstract and might involve representations like the logic of a thermostat or the structure of a financial system [1].

Further differentiation has been proposed by Rasmussen (1986), who introduced levels of functional abstraction [20]. These include form (what a system looks like), function (how it works), and purpose (why it exists). These layers help explain the cognitive strategies users employ when interacting with systems. Research has shown that expert users typically develop more structured, layered, and hierarchical models than novices, who often rely on simplified and fragmented representations [6].

## Mental Model Alignment

Successful interaction with complex systems depends heavily on the alignment between the mental model of the user and the designer's conceptual model as conveyed through the system image. When these models are consistent, users are more likely to operate the system effectively and safely. Misalignments, however, can lead to errors, confusion, and system misuse [21]. Wilson and Rutherford emphasize that understanding and bridging these gaps is a primary objective of human factors research [1][2].

## Practical Examples

One widely cited example is the [[London Underground map]], a schematic rather than geographic representation of the subway system. Though it distorts actual spatial relationships, it is extremely effective in supporting navigation decisions. Other examples include [[IKEA]] furniture instructions, which rely on visual schematics to convey assembly steps; smart home interfaces, where dashboards enable users to monitor and control complex systems; and thermostat interfaces, which are often misunderstood due to incorrect assumptions about how temperature regulation works [7][22].

## Applications

Mental models are foundational in a wide range of domains. In interface design, particularly in [[aviation]], [[nuclear power]], and [[healthcare]], they guide the development of user-friendly systems that align with user expectations. In training and simulation contexts, such as flight simulators or emergency response scenarios, mental models inform curriculum design and feedback strategies [23].

Emerging fields like [[Explainable Artificial Intelligence]] (XAI) also rely on mental models to bridge the gap between machine learning processes and human understanding. In such applications, designing interfaces that make algorithmic decisions comprehensible to users is crucial. Additionally, mental models are central to enhancing situational awareness in control rooms and collaborative environments [8][9][24].

## Measuring Mental Models

Since mental models are internal constructs, their measurement poses methodological challenges. Researchers have developed several indirect methods. Think-aloud protocols capture users' reasoning in real time, while concept mapping reveals how individuals organize and relate key concepts. Card sorting and cluster analysis techniques, such as those used by Baggett and Ehrenfeucht (1988), help visualize the structure and coherence of user models [11]. These methods are grounded in the assumption that the structure and behavior of externalized knowledge can give insight into internal representations.

Concept mapping, in particular, is valued for its ability to reflect the hierarchical and relational structure of mental models. Participants are asked to draw links between concepts, revealing not only what they know but how they connect their knowledge. Similarly, card sorting can show category structures by analyzing how users group or prioritize information, which can be especially useful in interface and menu design.

Think-aloud protocols, though rich in qualitative insight, require intensive analysis and are sensitive to task design and individual verbalization skills. These verbalizations can reflect the real-time use of a mental model, especially when users are faced with unexpected outcomes and must articulate their assumptions.

More recent innovations include the use of interactive simulations and structured reflection tools. For instance, the "Thought Bubbles" method developed by Mohaddesi et al. (2023) captures mental model development during real-time decision-making tasks, allowing researchers to observe how users revise and adapt their internal representations [12]. This method is based on Endsley’s theory of situation awareness, incorporating perception, comprehension, and projection phases, and it enables tracking changes over time.

Other modern approaches employ eye-tracking to identify focus areas and infer underlying mental representations. In simulation studies, system log data can also be mined to reveal behavioral patterns indicative of different model types. Techniques from machine learning have begun to support these analyses, particularly in large-scale environments like educational technology platforms or real-time control systems.

Despite these advancements, there remains no universally accepted standard for measuring mental models. The variety of available methods reflects the diversity of model types and their applications. Consequently, researchers often adopt multi-method approaches, combining qualitative and quantitative techniques to capture both the content and structure of mental models. The choice of method must align with the research context, user characteristics, and system complexity to ensure valid and meaningful results.

## Theoretical Controversies

Wilson and Rutherford highlight several ongoing debates. One key tension lies in the disciplinary divide: cognitive psychology tends to focus on the mental processes involved in forming and manipulating models, while human factors research is more concerned with the end product—the model as it manifests in user behavior. There is also a lack of consensus on operational definitions, making it difficult to compare studies or develop standard design practices [1].

Additionally, confusion often arises between "user models" (constructed by designers or systems to represent users) and "mental models" (held by users themselves). The tendency to reify mental models—treating them as fixed, external entities rather than dynamic internal constructs—has also been criticized [25]. Johnson-Laird’s theory offers a counterpoint by framing mental models as recursive, constructive, and flexible cognitive tools that underpin reasoning and action [13].

## Contemporary Research

Recent studies have expanded the scope of mental model research. Seppänen et al. (2017) examined mental models among operators in safety-critical domains, emphasizing the importance of clarity and shared understanding. Rupprecht and Schraagen (2021) focused on the dynamic updating of mental models in healthcare teams, highlighting how real-time collaboration affects model accuracy [14]. Gisick et al. (2018) proposed new ways to measure shared mental models in clinical settings, emphasizing intersubjective alignment [15]. Research by Halasz and Moran (1983) and by Kieras and Bovair (1984) further illustrates the role of instructional strategies in shaping user mental models [26][27].

## Critical Evaluation

Mental models offer a compelling framework for analyzing human interaction with technology. They support user-centered design by focusing attention on how users perceive and make sense of systems. They also facilitate effective training, enhance usability, and improve error management.

However, the concept remains theoretically and methodologically challenging. Definitions vary across fields, and empirical studies often rely on indirect or interpretive methods. As Wilson and Rutherford caution, the seductive simplicity of the term ‘mental model’ can lead to overgeneralization. For mental models to be truly useful in design, they must be grounded in robust theory and supported by rigorous empirical research [1][2][13].

## References

1. Wilson, J. R., & Rutherford, A. (1989). Mental Models: Theory and Application in Human Factors. Human Factors, 31(6), 617–634.
2. Norman, D. A. (1983). Some Observations on Mental Models. In Gentner & Stevens (Eds.), Mental Models.
3. Rumelhart, D. E. (1980). Schemata: The Building Blocks of Cognition.
4. Schank, R., & Abelson, R. (1977). Scripts, Plans, Goals, and Understanding.
5. Minsky, M. (1975). A Framework for Representing Knowledge.
6. Gentner, D., & Stevens, A. (1983). Mental Models.
7. Kempton, W. (1986). Two Theories of Home Heat Control. Cognitive Science.
8. Mohseni, S., et al. (2021). A Survey of Human-Centered Evaluations in XAI. ACM Transactions on Interactive Intelligent Systems.
9. Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems.
10. Novak, J. D., & Cañas, A. J. (2008). The Theory Underlying Concept Maps and How to Construct Them.
11. Baggett, P., & Ehrenfeucht, A. (1988). Conceptualizing in Assembly Tasks. Human Factors.
12. Mohaddesi, O., et al. (2023). Thought Bubbles: A Proxy into Players’ Mental Model Development. arXiv.
13. Johnson-Laird, P. N. (1983). Mental Models. Cambridge University Press.
14. Rupprecht, L., & Schraagen, J. M. (2021). Dynamic Mental Model Updating in Healthcare Teams. Frontiers in Psychology.
15. Gisick, L. M., et al. (2018). Measuring Shared Mental Models in Healthcare. Journal of Patient Safety and Risk Management.
16. Craik, K. J. W. (1943). The Nature of Explanation. Cambridge University Press.
17. Norman, D. A. (1986). Cognitive Engineering. In User Centered System Design.
18. McCloskey, M. (1983). Naive Theories of Motion. In Gentner & Stevens (Eds.).
19. Kempton, W. (1987). Folk Models in Thermostat Use. Cognitive Science.
20. Rasmussen, J. (1986). Information Processing and Human-Machine Interaction. North-Holland.
21. Zhang, K., & Wickens, C. D. (1987). Mental Models of a Complex Dynamic System. In Human Factors Society Proceedings.
22. Kieras, D. E. (1984). A Simulation Model for Procedure Inference. University of Arizona.
23. Moray, N. (1987). Intelligent Aids, Mental Models, and Theory of Machines. International Journal of Man-Machine Studies.
24. Hollnagel, E. (1988). Mental Models and Model Mentality. In Goodstein et al. (Eds.).
25. Bainbridge, L. (1988). Types of Representation. In Goodstein et al. (Eds.).
26. Halasz, F. G., & Moran, T. P. (1983). Mental Models and Problem Solving in Using a Calculator. CHI Proceedings.
27. Kieras, D. E., & Bovair, S. (1984). The Role of a Mental Model in Learning to Operate a Device. Cognitive Science.