INSIGHTS · WHAT AN INSIGHT IS

The Role of Mental Models in Facilitating or Blocking Insight Formation

Mental models, the internal pictures people hold of how things work, are what make insight possible, and also what most often prevents it. This article explains what mental models are, how they guide attention and reasoning, how expertise and shared team models help and hinder, and how research teams can work with them deliberately.

ModQu research article · 2,655 words · 24 references (APA 7th edition)

Every analyst approaches data with a picture of how the world works. A brand manager carries a model of why customers choose the brand, a pricing analyst a model of how demand responds to price, a service director a model of what makes customers satisfied. These internal pictures are called *mental models*. They allow people to interpret data quickly, to notice when something is unusual and to predict what will happen next. They also determine what people do not see. An insight, in most cases, is a change in a mental model. Understanding how mental models work therefore helps explain both why insight is possible and why it is so often blocked.

What a mental model is

The idea goes back to the Scottish psychologist Kenneth Craik (1943), who proposed that the mind constructs "small-scale models" of external reality and uses them to try out possible actions, anticipate events and react to situations before they happen. A person who knows how a door works can predict what will happen if they push it, without having to push it first.

Philip Johnson-Laird (1983) developed the idea into a theory of human reasoning. People, Johnson-Laird argued, reason not by applying formal rules of logic but by constructing models of the possibilities consistent with what they know and checking whether a conclusion holds in all of them. Errors arise because people tend to construct only one model, usually the most obvious, and do not consider alternatives that would show their conclusion to be wrong (Johnson-Laird, 2010).

Researchers in engineering and human factors gave a practical definition. Rouse and Morris (1986) described mental models as the mechanisms that allow people to describe the purpose and form of a system, explain how it works and why it is in its current state, and predict its future states. Jones, Ross, Lynam, Perez and Leitch (2011) reviewed the concept across disciplines and concluded that mental models are personal, internal representations of external reality that are incomplete, simplified and constantly revised through experience. They are not accurate copies of the world; they are working tools.

Where mental models come from

Mental models are built from experience, instruction and the stories people hear. A manager who began their career in sales may carry a model of the market built around relationships and persuasion; a colleague from finance may carry one built around costs and margins. Both models are partly right, and each makes some features of the data stand out while hiding others.

Organisations add their own layer. Training, planning templates, standard reports and the metrics that are tracked every month all encode assumptions about what matters. A company that reports customer satisfaction by product line, but not by stage of the customer journey, is quietly teaching its staff to think about satisfaction as a property of products. Over time, these organisational routines become part of each person's mental model, and they are rarely examined, because they appear to be simply "how things are".

Mental models are also shaped by success. Approaches that worked in the past are remembered and reused, while the conditions that made them work are often forgotten. This is the organisational equivalent of Luchins's water-jar experiments: a method that solved yesterday's problems is applied to today's, even when the problem has changed. Because successful models are reinforced and rarely tested, the models most in need of review are often the ones held with most confidence.

How mental models make insight possible

Mental models are not only obstacles. Without them, insight would be impossible.

They make anomalies visible. A finding can be surprising only against an expectation. Someone with no model of how customers behave cannot be surprised by anything they do. The data/frame theory of sensemaking describes how a frame tells people what to look for, and how data that do not fit the frame can trigger a search for a better one (Klein, Moon, & Hoffman, 2006).

They organise knowledge for rapid recognition. Chi, Feltovich and Glaser (1981) asked physics experts and novices to sort problems into groups. Novices grouped them by surface features, such as whether they involved inclined planes or springs. Experts grouped them by the underlying physical principle, such as conservation of energy. Experts' mental models allowed them to see the structure beneath the surface, which is exactly what insight requires.

They supply analogies. Many insights come from recognising that a new situation has the same structure as a familiar one. Gentner (1983) showed that useful analogies map relationships between elements, not their surface features. A rich store of mental models from different domains provides the material for such mappings.

They allow prediction and testing. A mental model that implies a prediction can be tested. When the prediction fails, the model must change, and that change is often where insight lies.

How mental models block insight

The same features that make mental models useful also create the main obstacles to insight.

Selective attention

A mental model directs attention to the information it considers relevant. Information outside the model is often not noticed at all. This makes analysis efficient, but it means that the data most likely to change the model are the data least likely to be seen.

Mental set and fixation

Once a model has worked, people tend to keep applying it. Luchins (1942) showed that participants who had learned a complicated method for solving water-jar problems continued to use it when a simpler method was available, and often failed problems that only the simple method could solve. Duncker (1945) showed that people fail to see new uses for objects because their model of the object's function is fixed. Knoblich, Ohlsson, Haider and Rhenius (1999) found that insight problems are hard to the extent that they require people to relax constraints built into their initial representation, and that the deeper the constraint, the harder the problem.

Bilalić, McLeod and Gobet (2008) studied this effect with chess experts using eye-tracking. After finding a familiar solution to a position, experts said they were searching for a better one. Their eye movements showed that they kept looking at the squares relevant to the familiar solution. The first model they had activated continued to steer their attention even while they believed they were looking elsewhere.

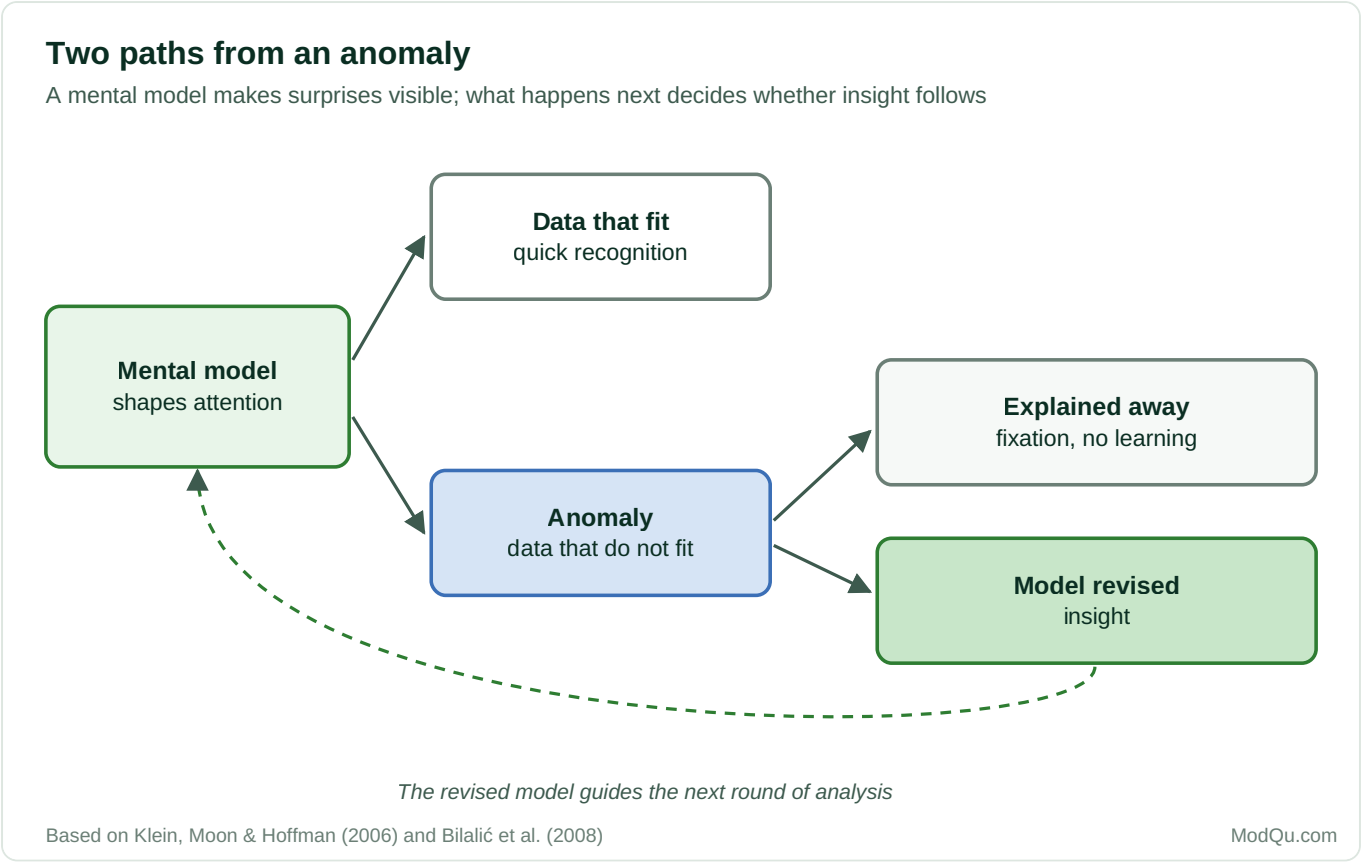


Figure 1. Two paths from an anomaly. A mental model makes surprising data visible, but whether the surprise leads to insight depends on whether the model is revised or the anomaly is explained away. Diagram by ModQu, based on Klein, Moon and Hoffman (2006) and Bilalić et al. (2008).

The illusion of understanding

People also overestimate how complete their mental models are. Rozenblit and Keil (2002) asked people to rate how well they understood how everyday devices such as zips and flush toilets work. When they were then asked to write a step-by-step explanation, their ratings of their own understanding dropped sharply. The researchers called this the *illusion of explanatory depth*. In organisations, managers may feel that they understand why customers behave as they do until they are asked to explain the mechanism in detail. The feeling of understanding can prevent the search for evidence that would improve it.

Expertise and entrenchment

Expertise strengthens mental models, which usually helps. But it can also make them harder to change. Wiley (1998) found that people with extensive knowledge of baseball were misled on word-association problems where baseball-related associations pointed toward wrong answers; novices were not. Ricks, Turley-Ames and Wiley (2007) later showed that this effect occurred regardless of working memory capacity. Dane (2010) described the risk as *cognitive entrenchment*: as expertise grows, the mental models that underpin it become more stable, which can reduce flexibility in novel situations. Experts are better than novices at seeing what their model predicts and sometimes worse at seeing what it does not.

Misperceiving dynamics

Mental models of complex systems are often especially poor. Sterman (1989) studied people managing a simulated supply chain, the "beer game". Participants consistently ignored the delays between placing orders and receiving goods, and the stock already on order. As a result, they over-ordered, then cut orders too sharply, creating large swings in inventory. Their mental models did not include the feedback loops and delays that drove the system. When the outcome was poor, they tended to blame outside factors rather than revise their models.

Mental models in organisations

Mental models are not only individual. Organisations and industries develop shared models that shape how their members interpret information.

Porac, Thomas and Baden-Fuller (1989) studied knitwear manufacturers in a small Scottish town. The firms shared a mental model of their competitive position: they saw themselves as makers of high-quality, classic garments competing mainly with each other, not with the mass-market producers that held most of the market. This shared model guided their strategies and their attention. It also limited what they considered possible. Walsh (1995) reviewed research on managerial and organisational cognition and found that such knowledge structures consistently shape how managers notice, interpret and respond to information, for better and worse.

Shared models can help organisations perform. Gary and Wood (2011) found that managers in a business simulation whose mental models of the environment were more accurate adopted better decision rules and achieved better results. In teams, shared mental models support coordination. Mathieu, Heffner, Goodwin, Salas and Cannon-Bowers (2000) found that teams whose members shared similar models of their task and of how they worked together coordinated better and performed better. DeChurch and Mesmer-Magnus (2010) confirmed in a meta-analysis that shared team cognition is positively related to team performance.

But shared models can also block change. When a new framing of the market challenges the dominant model, it may become the subject of a political contest. Kaplan (2008) studied strategy-making in a communications technology firm facing uncertainty and found that managers with different frames engaged in "framing contests", trying to persuade others to adopt their interpretation. Which frame won depended not only on the evidence but on how actively its supporters promoted it. Hodgkinson and Healey (2011) argued that adapting to change requires managers to combine reflection, deliberate analysis, with an awareness of the automatic, emotional responses that tie them to existing models.

Can mental models be observed?

If mental models are internal, how can anyone examine them? Rouse and Morris (1986) warned that mental models cannot be observed directly and that researchers must infer them from what people say and do, which introduces uncertainty. People may not be able to describe their models accurately, and the act of describing a model can change it.

Despite these limits, several practical methods exist. Jones and colleagues (2011) reviewed approaches used across disciplines, including interviews in which people explain how a system works, cognitive mapping exercises in which participants draw the concepts they consider important and the causal links between them, and structured comparisons of

how different people rank or group the same items. In a business setting, a simple version is to ask each member of a project team, before any data are analysed, to sketch the main reasons they believe customers behave as they do and to predict the key results. Comparing these sketches often reveals that team members hold quite different models, and that some assumptions are shared so widely that no one has thought to question them. The sketches also provide a record against which the findings can later be compared, which makes it much harder to claim afterwards that a surprising result was expected all along.

An illustration

A software company believes that small businesses cancel its accounting product mainly because of price. The belief is part of the shared mental model of the sales and product teams: price comes up in sales conversations, competitors advertise lower prices, and discounts have occasionally saved accounts. When a churn survey is analysed, the team's attention goes first to the price questions, and the results seem to confirm the model. Customers who cancelled rate value for money lower than those who stayed.

One analyst notices an anomaly. Many customers who cancelled had never connected their bank account to the product, a step the onboarding process treats as optional. The finding does not fit the price model, and the first reaction in the team is to explain it away: perhaps customers who were going to leave anyway did not bother to connect. This is the "explained away" path in Figure 1.

The analyst instead asks what would have to be true for the finding to matter, and constructs an alternative model: customers who never connect their bank account never experience the product's main benefit, automatic reconciliation, so the product seems expensive *for what it does for them*. Low ratings of value for money would then be a symptom of an onboarding failure, not evidence of price sensitivity. The two models make different predictions. If price is the cause, discounts should reduce churn equally among connected and unconnected customers. If onboarding is the cause, prompting customers to connect should reduce churn more than a discount. A small test supports the second model. The team's mental model is revised, and the insight changes the company's retention strategy from discounting to onboarding.

The example is illustrative, but the pattern is typical. The original model was not foolish; it fitted much of the evidence. It became an obstacle only when it was used to dismiss the data that could have improved it.

Working with mental models deliberately

Since mental models both enable and block insight, the goal is not to eliminate them but to manage them. Research suggests several practical steps for research and insight teams.

1. **Write the model down.** Before analysing data, record the team's current model: the assumed reasons for customer behaviour, the expected relationships and the predicted results. Explicit models can be checked; implicit ones cannot. Asking people to explain the mechanism in detail also reveals gaps, as the illusion-of-explanatory-depth research shows (Rozenblit & Keil, 2002).
2. **State predictions in advance.** If the model is correct, what should the data show? Predictions made before the analysis make it harder to explain away surprises afterwards.

3. **Hunt for anomalies.** Treat findings that do not fit as the most valuable part of the analysis, not as noise. Ask whether they can be explained without changing the model, and if so, whether that explanation is convincing or merely convenient.
4. **Construct alternative models.** Johnson-Laird's work suggests that reasoning errors come from considering only one model. Deliberately building two or three competing explanations, and asking what evidence would distinguish them, counters this tendency.
5. **Bring in different models.** People from other departments, industries or disciplines carry different mental models. Their questions can reveal assumptions the core team cannot see. This is one reason why analogies from distant domains are a common source of insight (Gentner, 1983).
6. **Use experts and novices together.** Experts see the structure the model predicts; novices may notice what it ignores. Combining them reduces both the novice's errors and the expert's entrenchment.
7. **Map dynamics explicitly.** Where outcomes depend on feedback and delays, as in customer retention, supply chains or word-of-mouth, simple diagrams of the causal loops can expose missing elements of the team's model (Sterman, 1989).
8. **Review after outcomes.** When a decision based on an insight produces unexpected results, revisit the model rather than attributing the gap to bad luck. Repeated revision is how mental models improve.

Key takeaways

Mental models are simplified internal representations of how things work, used to describe, explain and predict. They make insight possible by making anomalies visible, organising expert knowledge, supplying analogies and generating testable predictions. They block insight by directing attention away from inconvenient data, creating mental set and fixation, giving an illusion of understanding, becoming entrenched through expertise and leaving out important dynamics. In organisations, shared models support coordination but can resist change and become the subject of framing contests. Research teams can turn mental models from obstacles into tools by making them explicit, predicting before analysing, seeking anomalies, constructing alternatives and drawing on people with different models.

References

Bilalić, M., McLeod, P., & Gobet, F. (2008). Why good thoughts block better ones: The mechanism of the pernicious Einstellung (set) effect. *Cognition, 108*(3), 652–661. https://doi.org/10.1016/j.cognition.2008.02.003

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2

Craik, K. J. W. (1943). *The nature of explanation*. Cambridge University Press.

Dane, E. (2010). Reconsidering the trade-off between expertise and flexibility: A cognitive entrenchment perspective. *Academy of Management Review, 35*(4), 579–603. https://doi.org/10.5465/amr.35.4.zok579

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). The cognitive underpinnings of effective teamwork: A meta-analysis. *Journal of Applied Psychology, 95*(1), 32–53. https://doi.org/10.1037/a0017328

Duncker, K. (1945). On problem-solving. *Psychological Monographs, 58*(5), i–113. https://doi.org/10.1037/h0093599

Gary, M. S., & Wood, R. E. (2011). Mental models, decision rules, and performance heterogeneity. *Strategic Management Journal, 32*(6), 569–594. https://doi.org/10.1002/smj.899

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*(2), 155–170. https://doi.org/10.1207/s15516709cog0702_3

Hodgkinson, G. P., & Healey, M. P. (2011). Psychological foundations of dynamic capabilities: Reflexion and reflection in strategic management. *Strategic Management Journal, 32*(13), 1500–1516. https://doi.org/10.1002/smj.964

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences, 107*(43), 18243–18250. https://doi.org/10.1073/pnas.1012933107

Jones, N. A., Ross, H., Lynam, T., Perez, P., & Leitch, A. (2011). Mental models: An interdisciplinary synthesis of theory and methods. *Ecology and Society, 16*(1), Article 46. https://doi.org/10.5751/ES-03802-160146

Kaplan, S. (2008). Framing contests: Strategy making under uncertainty. *Organization Science, 19*(5), 729–752. https://doi.org/10.1287/orsc.1070.0340

Klein, G., Moon, B., & Hoffman, R. R. (2006). Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems, 21*(5), 88–92. https://doi.org/10.1109/MIS.2006.100

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(6), 1534–1555. https://doi.org/10.1037/0278-7393.25.6.1534

Luchins, A. S. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs, 54*(6), i–95. https://doi.org/10.1037/h0093502

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85*(2), 273–283. https://doi.org/10.1037/0021-9010.85.2.273

Porac, J. F., Thomas, H., & Baden-Fuller, C. (1989). Competitive groups as cognitive communities: The case of Scottish knitwear manufacturers. *Journal of Management Studies, 26*(4), 397–416. https://doi.org/10.1111/j.1467-6486.1989.tb00736.x

Ricks, T. R., Turley-Ames, K. J., & Wiley, J. (2007). Effects of working memory capacity on mental set due to domain knowledge. *Memory & Cognition, 35*(6), 1456–1462. https://doi.org/10.3758/BF03193615

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin, 100*(3), 349–363. https://doi.org/10.1037/0033-2909.100.3.349

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26*(5), 521–562. https://doi.org/10.1207/s15516709cog2605_1

Sterman, J. D. (1989). Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Science, 35*(3), 321–339. https://doi.org/10.1287/mnsc.35.3.321

Walsh, J. P. (1995). Managerial and organizational cognition: Notes from a trip down memory lane. *Organization Science, 6*(3), 280–321. https://doi.org/10.1287/orsc.6.3.280

Wiley, J. (1998). Expertise as mental set: The effects of domain knowledge in creative problem solving. *Memory & Cognition, 26*(4), 716–730. https://doi.org/10.3758/BF03211392