Pathfinder Networks for Measuring Operator Mental Model Structure with a Simple Autopilot System

Michael S. Politowicz[1,2], Tetsuya Sato[2], Eric T. Chancey[1], and Yusuke Yamani[2]
[1]*National Aeronautics and Space Administration (NASA), Langley Research Center, Hampton, VA*
[2]*Old Dominion University, Norfolk, VA*

Pathfinder networks are a method to represent mental models from empirically generated pairwise relatedness ratings. This study examined the effects of training exposure on mental model structures based on relatedness ratings collected using the Target Rating method. Forty-eight participants read instruction slides with or without explicit information on the functionality of an autopilot system (Advanced Mental Model or Basic Mental Model groups, respectively). Participants provided relatedness ratings and completed a comprehension test. The Advanced Mental Model group had more common links with the expected model, higher within-group network similarity scores, and higher mental model assessment questionnaire scores than the Basic Mental Model group. Both groups had coherence scores above the minimum threshold for internal consistency. Pathfinder network analysis was sensitive to changes produced by a simple exposure training intervention. In practice, a simple training program may effectively influence operator mental models in novel technological environments such as Advanced Air Mobility.

Increasingly autonomous systems are fundamentally changing the ways humans interact with automated systems. Automation, in its traditional role, is assumedly understandable with appropriate training and experience (see Lee & Seppelt, 2012, for review). However, as automation gives way to autonomy (i.e., nondeterministic systems), the ability for a human to fully understand the underlying functionality of a system will diminish (Endsley, 2017). An important concept to consider when the human has only a basic understanding of a system's functionality is mental models (Norman, 1983). The current work outlines the role of mental models in understanding human-automation interactions, a method for operationalizing mental models (i.e., Pathfinder networks), and the results of an initial study examining the effects of minimal training exposure on mental model structure (see Chancey & Politowicz, 2020, for discussion).

Mental Models

Mental model research has taken many forms. Despite a lack of consensus in the literature, the disparity in approaches is likely a result of the context-dependent nature of its application and operationalization (Moray, 1999; Norman, 1983; Rouse & Morris, 1986). Researchers in the field have used targeted representations of mental models to advance the understanding of cognitive functions (e.g., learning, problem solving, decision making) and generate novel approaches to relevant research areas (e.g., system design, training, human-system interaction tool development; see Allen, 1997). Mental models can be defined as "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" (Rouse & Morris, 1986, p. 351).

Many methods are available to elicit and represent mental models (see Rouse & Morris, 1986, for review). For example, Rowe and Cooke (1995) evaluated four mental model measurement techniques: think-aloud, laddering interview, diagramming, and relatedness rating. Using a system troubleshooting task, both the laddering interview and relatedness rating techniques independently predicted performance. The laddering interview required participants to list the components and connections of a system from memory. The relatedness ratings technique required participants to rate the functional relatedness of each pair of 11 predefined components. Both techniques were regarded as appropriate choices; however, the relatedness rating method allows for Pathfinder network analysis (Schvaneveldt et al., 1989), which enables mathematical comparison between models and produces graphical model representations that convey information not directly apparent from the ratings alone (Cooke et al., 1986).

Pathfinder Networks

Pathfinder network analysis is a psychometric scaling method that produces networks to visualize and analyze the relationships of concepts based on proximity data, which can be derived from empirically-generated pairwise comparison ratings (Schvaneveldt et al., 1989). The networks consist of a set of *nodes*, which correspond to predefined keywords or concepts. The nodes are connected by weighted *links*, which represent the distance between (or relatedness of) two nodes. Pathfinder analysis uses the relatedness ratings to reduce the set of links in the network to only the "strongest" connections. This is accomplished by removing "weak" links between nodes if there is a stronger connection available through some combination of other links. Thus, the distance between two nodes is the weight of the shortest path between those nodes (Dearholt & Schvaneveldt, 1990).

Pathfinder networks have been used to analyze cases ranging from evaluating broad relatedness of concepts (Cooke et al., 1986) to examining meaningful characteristics of relationships between specific nodes (Furlough & Gillan, 2018). Furthermore, Pathfinder networks have been used to study many topics, including team mental models (Lim & Klein, 2006), system troubleshooting (Rowe & Cooke, 1995), and video game expertise (Furlough & Gillan, 2018). However, in all cases, a list of concepts that represent nodes must be created prior to collecting proximity data from participants.

Two parameters are required for a Pathfinder analysis: the Minkowski *r*-metric and the *q* parameter. The *r*-metric determines how weights between nodes are computed. Specifically, the weight, *W*, of a path, *P*, is defined as:

$$W(P) = \left(\sum_{i=1}^{k} w_i^r \right)^{\frac{1}{r}},$$

where w_i is an individual link weight along the path of *k* links and *r* is the order of the Minkowski distance. The *q* parameter refers to the maximum number of links allowed in a path between two nodes. Networks become sparser (i.e., have fewer links) as either parameter increases. The most used values, which together yield the fewest number of links, are $r = \infty$ and $q = (n - 1)$, where *n* is the number of nodes in the network (Dearholt & Schvaneveldt, 1990). Useful in mental model research, network similarity, coherence, and internal consistency are common Pathfinder network analyses.

Network similarity. Calculating the similarity of two individual networks (i.e., network similarity score), or *C statistic*, is an evaluation of shared links between two networks, which is calculated using the formula, $X/[T - X]$, where *X* is the number of common links in the two networks and *T* is the total number of links across both networks (Goldsmith & Davenport, 1990; Lim & Klein, 2006). The scale of this metric is 0 (completely unrelated) to 1 (identical); thus, a score of 0.3 means 30% of the mental model structure is shared. This metric is a common approach for measuring the similarity of two networks (Cooke et al., 2000; Lim & Klein, 2006; Zhou, 2017). Additionally, comparing an individual network to a group of networks can be accomplished by calculating multiple *C* statistic values for the individual network (one value for each of the other networks in the group) and then taking the arithmetic mean of those values to get a single *C* statistic, which represents the group similarity score for that single network. This approach can then be repeated for each individual network resulting in a group similarity score for each network, a technique commonly used in team mental model research (Lim & Klein, 2006).

Coherence. The *coherence* measures the reliability of an individual's relatedness ratings, as determined by the transitive consistency of those ratings (Interlink, 2017a). For example, if ratings indicate that Concept A is similar to Concept B, and Concept B is similar to Concept C, then Concept A should also be similar to Concept C. A higher coherence value indicates that these transitive relationships are more reliable. The range of this measure is -1 (inverse coherence) to 1 (perfect coherence).

Internal Consistency. The *internal consistency* measure is a binary value derived from the coherence value to determine the overall reliability of an individual Pathfinder network. Coherence greater than or equal to 0.15 is considered adequate internal consistency, whereas coherence less than 0.15 may indicate a poor understanding of the concepts being rated (Interlink, 2017a). If coherence values are on average lower than 0.15, the data should not be analyzed (see Sanchez, 2021, for discussion).

Target Rating Method

Pathfinder networks are derived from pairwise comparison ratings, which are typically collected by rating individual node pairs in a random order until all possible pairs have been rated. Tossell et al. (2010) developed a novel method for efficiently collecting comparison ratings called the Target Rating method. This method uses concentric circles resembling a target to represent levels of relatedness to the concept in the center circle (Figure 1). Participants move all other concepts from the set into the circles that correspond to degree of relatedness with the primary concept. When the participant completes ratings for a concept, the target view resets, and a new primary concept appears in the center circle. This repeats until all concepts have cycled through the center circle position.

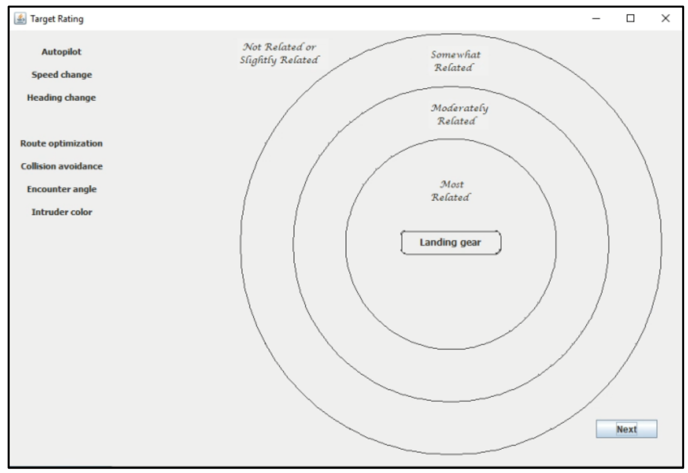

Figure 1. Target rating method user interface (JTarget software).

Study Purpose and Research Question

The present exploratory experiment was conducted to determine the effectiveness of Pathfinder network analysis and the Target Rating method for detecting differences in mental model structures introduced by brief training sessions in the context of a simple autopilot system. Specifically: How does a simple exposure training intervention affect mental model structures, operationalized as Pathfinder networks?

METHOD

Participants

Forty-eight undergraduate students (30 females, mean age = 20.27 years, *SD* = 3.92) were recruited to participate in the study from the community of a large public university in Eastern Virginia. All participants were screened for normal color perception and reported normal or corrected-to-normal visual acuity. Participants received course credit for participating. This research was approved by a University Institutional Review Board.

Design

The experiment employed a between-subjects design. Participants were randomly assigned to Basic Mental Model and Advanced Mental Model groups, which differed in system knowledge exposure during training.

Dependent Variables

The metrics corresponding to Pathfinder networks require only a single rating score for each pair of nodes. However, the Target Rating method collects two ratings per node pair. For this study, the duplicate ratings were averaged to provide a single rating score for each pair. A Pathfinder network was created to represent an ideal model based on the Advanced Mental Model training content (i.e., Expected Mental Model, Figure 2). This model is not intended to reflect an expert's mental model but rather the expected model based on the training.

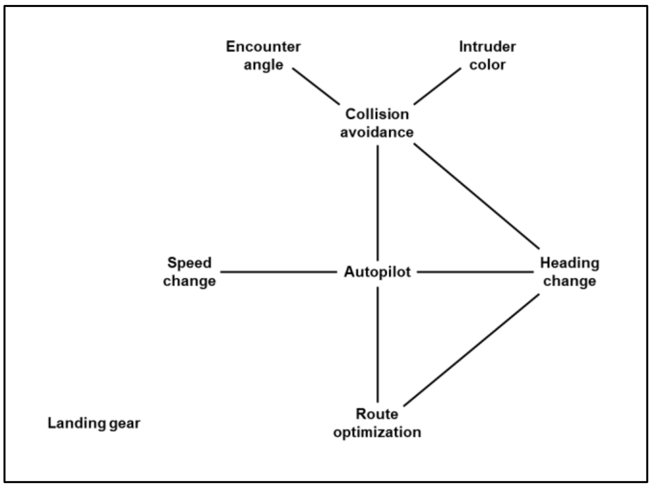

Figure 2. Expected mental model.

Pathfinder Analyses. Within-group similarity values were calculated for individual networks, which is the average similarity score (i.e., *C* statistic) of all possible within-group comparisons with that individual network. Similarity to the expected model values were calculated, which are the similarity scores of an individual model and the expected model. Coherence and internal consistency were also calculated (see Pathfinder Networks section above).

Common links with expected model. The number of common links between each participant's mental model and the expected mental model measures the extent to which the expected model is contained within the participant's model. In contrast to the similarity metric, this measure does not account for excess links beyond those shared between the two models.

Total links. The total number of links contained within a participant's mental model was evaluated to assess the overall complexity of the model.

Double rating mismatch %. The Target Rating method collects duplicate rating scores for each pair of nodes. This metric indicates the percentage of node pairs that did not have equivalent duplicate ratings.

Double rating average difference. This metric indicates the average difference between duplicate rating scores across all node pairs within each participant. Higher values indicate larger inconsistencies within a participant's ratings.

Mental model assessment score. A 10-question, true/false mental model assessment questionnaire was developed to test for training material content retention. Four questions referenced content that was only present in the Advanced Mental Model group training.

Pathfinder Nodes

Pathfinder analysis requires a list of keywords (or concepts) to serve as nodes in the Pathfinder network structure. This study used a total of eight nodes that were derived from concepts introduced in the training material: *autopilot*, *speed change*, *heading change*, *landing gear*, *route optimization*, *collision avoidance*, *encounter angle*, and *intruder color*. The training material and nodes were generated by the authors prior to the experiment. Note that *landing gear* was not mentioned in either training group.

Training Material

Both groups were introduced to the concept of Advanced Air Mobility (AAM; see National Academies of Sciences, Engineering, and Medicine, 2020, for detailed explanation of AAM) and the high-level functionality of an autopilot. The autopilot discussed in the training was fictitious and simplified.

For the Basic Mental Model group, the training document was a 13-slide presentation. This group was exposed to the following autopilot concepts: *heading change*, *route optimization*, and *collision avoidance*. All autopilot-related information was presented in the first 10 slides. The final 3 slides for the Basic Mental Model group simply provided additional details about AAM, such as economic impact of AAM technology, which were not directly relevant to the mental model concepts measured post-training.

For the Advanced Mental Model group, the training document was also a 13-slide presentation, and the first 10 slides were identical to the Basic group slides. However, the final 3 slides for the Advanced Mental Model group introduced additional information about the autopilot, specifically *speed change*, *intruder color*, and *encounter angle*. Both intruder color and encounter angle were presented as possible sources of autopilot failures related to collision avoidance.

Apparatus

JTarget (Interlink, 2017b) is a software tool with a user interface for collecting pairwise relatedness ratings using the Target Rating method. The software randomizes the order of target keywords presented to the participant. JTarget was used to collect pairwise ratings from participants during the study.

Procedure

Participants completed the study in a room with up to three other participants present (participants were isolated and worked independently). Participants in each experimental session were assigned to the same group to avoid possible contamination of subjects. The study took approximately 18 minutes to complete.

Participants first read an informed consent document and indicated whether they agreed to participate in the study. Upon their consent, participants then read a 13-slide training presentation that introduced the concept of Advanced Air Mobility (AAM) and described the general functionality of a simple, fictitious autopilot system. Participants received either the Basic Mental Model or Advanced Mental Model training.

Training took approximately 4 minutes to complete. The training presentation was not available to participants following this initial reading.

After the training, participants read instructions on using the JTarget software. Participants then provided pairwise relatedness ratings using JTarget. Finally, participants completed a 10-question mental model assessment questionnaire. At completion, participants were debriefed, thanked for their participation, and dismissed.

RESULTS

Pathfinder (Interlink, 2021) software was used for network analysis with $r = \infty$ and $q = 7$. Mental model aggregate networks for both groups are shown in Figures 3 and 4. Two-tailed, independent samples *t*-tests were used for this analysis. An alpha level of $p < .05$ was established to indicate statistical significance.

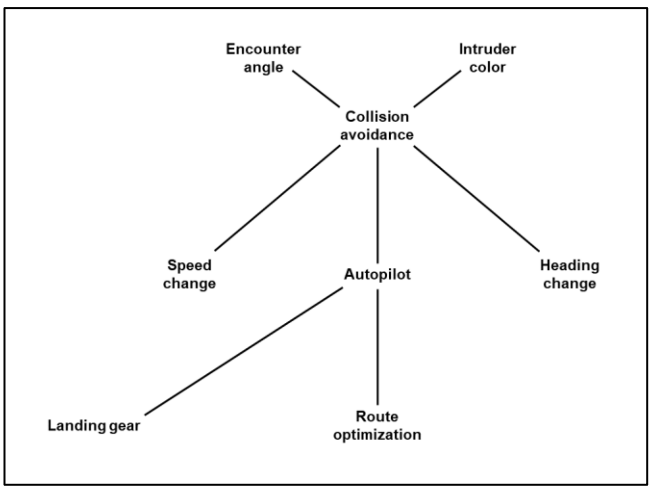

Figure 3. Advanced Mental Model group mean aggregate network.

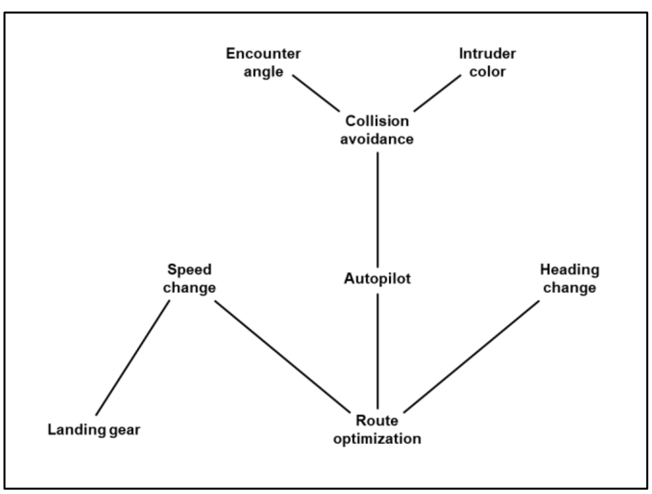

Figure 4. Basic Mental Model group mean aggregate network.

The Advanced Mental Model group ($M = 6.33$, $SD = 1.55$) scored significantly higher than the Basic Mental Model group ($M = 5.33$, $SD = 1.63$) on the mental model assessment questionnaire, independent-samples $t(46) = 2.18$, $p = .035$, $d = 0.63$, indicating that the training material was effective at manipulating mental models.

Coherence values were not significantly different between the Basic ($M = 0.25$, $SD = 0.37$) and Advanced ($M = 0.37$, $SD = 0.32$) Mental Model groups, independent-samples $t(46) = 1.20$, $p = .238$, $d = 0.35$. The proportions of participants with internal consistency in Basic (58.3%) and Advanced (79.2%) Mental Model groups did not differ, $\chi^2(1, 48) = 2.42$, $p = .119$.

The Advanced Mental Model group ($M = 0.35$, $SD = 0.05$) had significantly higher within-group similarity values than the Basic Mental Model group ($M = 0.30$, $SD = 0.04$), independent-samples $t(46) = 3.51$, $p = .001$, $d = 1.01$. Similarity to the expected model values were not significantly different between Basic ($M = 0.32$, $SD = 0.11$) and Advanced ($M = 0.37$, $SD = 0.10$) Mental Model groups, independent-samples $t(46) = 1.75$, $p = .087$, $d = 0.50$.

The Advanced Mental Model group ($M = 5.21$, $SD = 1.25$) had significantly more common links with the expected model than the Basic Mental Model group ($M = 4.42$, $SD = 1.21$), independent-samples $t(46) = 2.23$, $p = .031$, $d = 0.64$. The total number of links was not significantly different between Basic ($M = 10.54$, $SD = 2.30$) and Advanced ($M = 11.50$, $SD = 2.67$) Mental Model groups, independent-samples $t(46) = 1.33$, $p = .190$, $d = 0.38$.

Double rating mismatch percentages were not significantly different between Basic ($M = 0.52$, $SD = 0.10$) and Advanced ($M = 0.53$, $SD = 0.16$) Mental Model groups, independent-samples $t(46) = 0.16$, $p = .877$, $d = 0.05$. Double rating average difference values were also not significantly different between Basic ($M = 0.82$, $SD = 0.19$) and Advanced ($M = 0.80$, $SD = 0.27$) Mental Model groups, independent-samples $t(46) = 0.22$, $p = .829$, $d = 0.06$.

DISCUSSION

The goal of this study was to determine how a simple exposure training intervention affects mental model structures, operationalized as Pathfinder networks. The results showed that Pathfinder networks were affected by the manipulation on two measures. Notably, the within-group similarity value indicated that the Basic Mental Model group had more within-group variability in mental model structure than the Advanced Mental Model group. This result implies that the additional training content modulated their mental models to approach a common structure.

The second responsive measure was the number of common links with the expected model, which showed that, compared to the Basic group, the Advanced group's mental models contained more elements of the expected model. Yet, group differences in expected model similarity were not detected in another measure that accounted for the number of unshared links, implying that the Advanced Mental Model group may have had a better understanding of the autopilot but was not necessarily converging on the expected model any more than the Basic group. The lack of convergence may be due to limited information in the current training material. It is also important to note that the expected model was derived from the Advanced training material by the authors, so the expected model may not represent the expert model. However, based on the simplicity of the training content and few number of nodes, it is assumed to be an adequate representation.

Though the difference in coherence values between groups was not significant, both Basic and Advanced groups exceeded the minimum threshold for internal consistency (i.e., 0.15), implying that participants had an adequate understanding of the concepts being rated. Previous research has shown that experts tend to have higher coherence values than novices (Tossell et al., 2010). However, this finding was not observed in the present study, presumably because the training manipulation was relatively weak.

Mental models must be inferred from other measures because the models cannot be observed directly. To address this, the present study used a simple assessment questionnaire to offer some level of validation for mental model changes observed through Pathfinder network analysis. The Advanced group scored higher on the 10-question, true/false mental model assessment questionnaire, confirming the effect of training on mental model content retention. Yet, despite a medium-to-large effect size, the difference in mean scores between groups was lower than expected. Thus, the methodology may need to be revised for future studies. Additionally, future studies should incorporate other measures for mental model inference, such as performance.

The Target Rating method was an additional focus of this study, and two measures were explored to evaluate the inherent duplicate ratings and the potential for extracting additional useful information from these data. Results indicated no difference in the number of unequal (i.e., inconsistent) duplicate ratings between groups. Additionally, there was no difference in the magnitude of duplicate rating differences. Yet, the duplicate ratings were, on average, unequal for more than half of the comparisons in a single network. Thus, averaging the two values likely provides a more accurate relatedness rating.

Only eight nodes in the Pathfinder network were sufficient to detect the effect of the training intervention on mental model structure for the present study. Yet, future studies could be improved by employing more immersive training interventions. Overall, there is substantial research relating mental models and expertise in the context of automated systems, but the goal of this approach is to provide more structural resolution using a data collection method that does not require interviews by the experimenter.

Conclusion

Pathfinder network analysis, using only eight nodes, was sensitive to changes enacted by a simple exposure training intervention. Additionally, the Target Rating method functioned effectively in this research environment. The present study adds to the literature on Pathfinder's use in studying mental model structures. Future research should use this method to assess the effect of mental model structure on other measures, such as task performance.

This research was partially supported by NASA's Autonomous Systems Enduring Discipline under the Transformational Tools and Technologies Revolutionary Aviation Mobility Sub-project. The views expressed are those of the authors and do not necessarily reflect the official policy or position of NASA or the U.S. Government.

REFERENCES

Allen, R. B. (1997). Mental models and user models. In *Handbook of Human-Computer Interaction* (pp. 49–63). Elsevier. https://doi.org/10.1016/B978-044481862-1.50069-8

Chancey, E. T., & Politowicz, M. S. (2020). Designing and training for appropriate trust in increasingly autonomous Advanced Air Mobility operations: A mental model approach; Version 1 (NASA/TM–20205003378). *NASA technical memorandum*. https://ntrs.nasa.gov/citations/20205003378

Cooke, N. J., Salas, E., Cannon-Bowers, J. A., & Stout, R. J. (2000). Measuring team knowledge. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *42*(1), 151–173. https://doi.org/10.1518/001872000779656561

Cooke, N. M., Durso, F. T., & Schvaneveldt, R. W. (1986). Recall and measures of memory organization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(4), 538–549. https://doi.org/10.1037/0278-7393.12.4.538

Dearholt, D. W., & Schvaneveldt, R. W. (1990). Properties of Pathfinder networks. In R. W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization.* (pp. 1–30). Ablex Publishing.

Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *59*(1), 5–27. https://doi.org/10.1177/0018720816681350

Furlough, C. S., & Gillan, D. J. (2018). Mental models: Structural differences and the role of experience. *Journal of Cognitive Engineering and Decision Making*, *12*(4), 269–287. https://doi.org/10.1177/1555343418773236

Goldsmith, T. E., & Davenport, D. M. (1990). Assessing structural similarity of graphs. In R. W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization.* (pp. 75–87). Ablex Publishing.

Interlink. (2017a). *JPathfinder user manual*. http://interlinkinc.net/

Interlink. (2017b). *JTarget* [Computer software]. Interlink. http://interlinkinc.net/

Interlink. (2021). *Pathfinder* (Version 9.0) [Computer software]. Interlink. http://interlinkinc.net/

Lee, J. D., & Seppelt, B. D. (2012). Human factors and ergonomics in automation design. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 1615–1642). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118131350.ch59

Lim, B.-C., & Klein, K. J. (2006). Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy. *Journal of Organizational Behavior*, *27*(4), 403–418. https://doi.org/10.1002/job.387

Moray, N. (1999). Mental models in theory and practice. In *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application.* (pp. 223–258). The MIT Press.

National Academies of Sciences, Engineering, and Medicine. (2020). *Advanced aerial mobility: A national blueprint*. National Academies Press. https://doi.org/10.17226/25646

Norman, D. A. (1983). Some observations on mental models. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 7–14). Erlbaum.

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

Rowe, A. L., & Cooke, N. J. (1995). Measuring mental models: Choosing the right tools for the job. *Human Resource Development Quarterly*, *6*(3), 243–255. https://doi.org/10.1002/hrdq.3920060303

Sanchez, S. M. (2021). *Concept sourcing for mental models: A study of mental models and expertise using a video game context* [Master's Thesis, North Carolina State University]. NC State Repository.

Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1989). Network structures in proximity data. In *Psychology of Learning and Motivation* (Vol. 24, pp. 249–284). Elsevier. https://doi.org/10.1016/S0079-7421(08)60539-3

Tossell, C. C., Schvaneveldt, R. W., & Branaghan, R. J. (2010). Targeting knowledge structures: A new method to elicit the relatedness of concepts. *Cognitive Technology*, *15*(2), 11–19.

Zhou, Z. (2017). *Planning and team shared mental models as predictors of team collaborative processes* [Doctoral dissertation, Old Dominion University]. ProQuest Dissertations Publishing. https://doi.org/10.25777/T3CH-2521