Review of Mental Models as a Method of Trust Facilitation for Human System Interaction

Abstract

A systematic mapping approach was utilised to explore the literature and critically review the use of mental model frameworks in analysing trust facilitation of automation and human system interactions, using a variety of knowledge databases and keywords to yield appropriate results.

The review highlights key trends of human automation interaction research relating to trust facilitation and mental models schemata. Principal outcomes suggest limiting factors of appropriate trust with automation include system transparency, cognitive dissonance and performance and system disuse/misuse. However, the literature suggests counteractive actions such as priming, training or educating operators can support development of robust mental models.

Keywords: *human mental models, human system interaction, automation, cognitive systems engineering, review*

Introduction

Automation is a critical future trend in military domains [1] and civilian industries. Increased levels of automation can provide greater system capability and assistance in decision making, data selection or process management, as well as performing tasks that are otherwise "*dangerous, time-consuming, or outside of human desire or capability*" [2] [3]. As automated systems advance, the role of the operator will change in nature from active control (i.e. 'Human-in-the-loop') to a supervisory management position (i.e. 'Human-on-the-loop') [4]. Understanding the difference between levels of automation in a predominantly human-oriented and safety- or mission-critical system is an important consideration as the utilisation of more advanced, sophisticated technologies becomes ubiquitous in both military and civilian domains. However, trust in automation (TIA) remains a diverse and broad research field as technology advances faster than literature can appropriately keep up with [5].

There are three main views throughout the literature regarding human factors and human system interaction; these are 'human-in-the-loop', 'on-the-loop' and 'out-of-the-loop' (Wogalter, 2006). The latter is problematic for human-automation interaction (HAI) research. It suggests the system may act independently of a human operator, or system integration has failed and operators are unsure on what the status of the system actually is, thus divorced from the interaction. Integration currently strives to achieve in-the-loop interaction, with human and system integrated well, and with both 'actors' aware of each other's influences. On-the-loop refers to the ideal for HAI system integration activities and development programmes, with human actors removed from the overall system functionality and allocated tasks that are predominantly focused on surveillance and maintenance of the autonomous actions of the system. Factors that influence trust are imperative for appropriate use and reliance on the system and subsequent decision making. In the context of this review Human Mental Models (HMM) are frameworks (schema/schemata) that individuals cognitively construct based on prior experience and knowledge to support their expectations and interpretations of their environment.

Review Methodology

A scoping review methodology was adopted to explore new and emerging research in the field of human system automation and mental models. This protocol was used to focus on the gaps in the literature as opposed to answering a narrow research question which systematic review methodology traditionally utilises [7]. However, systematic review procedure was applied where appropriate, such as inclusion criteria, quality filtering and data extraction where appropriate.

Scoping Review Framework

The search criteria for the databases included (Google Scholar, IEEE Xplore, Science Direct and Wiley Online Library databases) include keywords reflective of the broad scope of HAI research. Related articles and bibliographic citations were screened for applicable sources. To ensure relevance, a date range restriction of articles since 2005 was imposed. Figure 1 illustrates the search strategy adopted.

All articles were qualitatively assessed for relevancy. Such as human mental models as the focus of discussion, and/or experimentation with automation, systems interaction or computer interaction as secondary variables. Journals within critical workplaces, such as the military or healthcare were prioritised (with relevant age groups (working age, 18-60)). Clinical studies were excluded, as they lack ecological validity/environmental generalisability. Other exclusion criteria are unpublished or non-English literature, or where literature was not available in full text form.

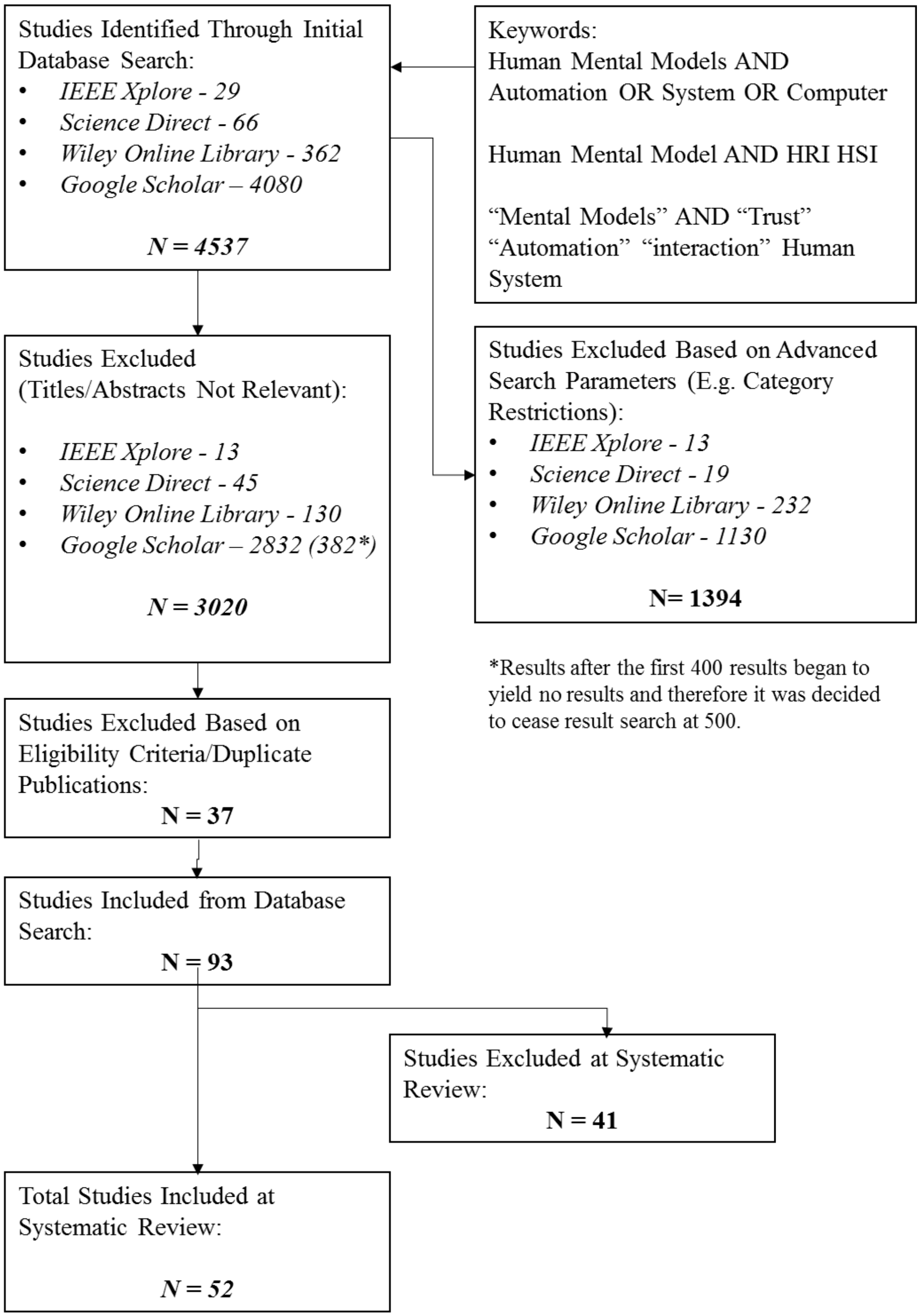

Figure 1- Eligibility Screening Identification

Study design and methodological information were extracted where possible. Primarily qualitative research materials that provided scope and context include discussion pieces, reviews and meta-analyses. Elements such as location, participant demographics, outcome measures and results were sought where possible to provide appropriate context and generalisability to the sources reviewed. Outcomes were clustered according to grouping criteria using a modified version of the Three Factor Human-Robot Trust Model outlined in Hancock *et al* (2011). [8]. Although this model focusses on Human-Robot[1] issues, these themes closely align to general system behaviour within sociotechnical domains

Quality assurance of the scoping review adhered to existing systematic review protocol such as the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocols (The PRISMA Group) (Moher & Liberati, 2009) [9] with regard to evaluation guidelines and inclusion criteria[2] and Meta-analysis of Observational Studies in Epidemiology (MOOSE) [10] guidelines.

The assessed literature are categorised into the relevant thematic categories. These are as follows: Human Related, subcategorised by Ability Based and Characteristics; System Related, subcategorised by Performance Based and Attribute Based; and Environmental, subcategorised by Team Collaboration and Tasking. Literature within these tables have been refined by author, alphabetically.

---

[1] 'robot' is under the umbrella of system in this review

[2] the risk of bias could not be conclusively assessed across all included sources in the recommended methodology [64]. High risk sources were excluded and medium risk literature was moderated.

Results

Search Results

The eligibility screening process in Figure 1 provided keyword and terminology parameters for each database search. The resulting data pool was relatively large for the field of research (*N=4537*) [11]. A portion of sources were excluded due to irrelevant titles or abstracts and a third excluded through search parameters (e.g. category restrictions available on electronic databases (e.g. full text versions/English only)). A number of duplicates and reproductions of identical studies to multiple sources (conference proceedings and subsequent journal article omitted. Of the remaining 93 eligible sources, a further 41 were excluded through review of content, leaving a total 52 sources reviewed in this document.

Table 1

Summary of Results

Study: Author(s), Country	Subdomain	Exp[3]	Obv	Lit/D	Actors
Human Related					
Ability Based: 12					
Aydoğan, Sharpanskykh, & Lo (2014) *Netherlands*	Aviation	–	✓	✓	–
Birkmeier, Korn, & Flemisch (2011, October) *Germany*	Aviation	–	✓	✓	ATCo (n=8)
Chua, Storey, & Chiang (2012) *North America*	Civilian	–	✓	–	High Skilled Engineer (n=14)
Fallon, Murphy, Zimmerman, & Mueller (2010, May) *North America*	Civilian/University	–	–	✓	–
Hawkley, Mares, & Giammanco (2005) *North America*	Military	–	–	✓	–
Hoffman & Woods (2011) *North America*	Civilian	–	–	✓	–
Lim, Dey, & Avrahami (2009, April) *North America*	Civilian	–	✓	–	51% female; 49% male; Mean Age 29.8 (n=55)
Neerincx, et al (2008, January) *Netherlands*	Naval	✓	–	–	Warfare Officers/Assistants (n=8)
Piccinini, Rodrigues, Leitão, & Simões (2015) *Portugal*	Automotive	✓	–	–	13 ACC users (age 42.2 (SD=9.9), 13 non-ACC users (age 26.7 (SD=9.9).(n=26)
RTO/NATO (2007) *Multinational*	Military	–	–	✓	–
Wilkison, Fisk, & Rogers (2007, October) *North America*	Civilian/University	–	✓	–	Undergraduate participants (58% F;42% M;Ages18-30)(n=12)

[3] Exp = Experimental study design / Obv = Observational study design / Lit/D = Literature review, discussion piece and other review based sources

Study: Author(s), Country	Subdomain	Exp[3]	Obv	Lit/D	Actors
Zhang, Kaber, & Hsiang (2010) *North America*	Automotive	✓	–	–	Male Participants, Mean Age: 25 (SD=2.4)(n=12)
Characteristics: 19					
Arkin, Ulam, & Wagner (2012) *North America*	Military	–	–	✓	–
Beggiato & Krems (2013) *Germany*	Automotive	✓	–	–	51% Female;49% Male; Mean Age: 24 (n=51)
Beggiato, Pereira, Petzoldt, & Krems (2015) *Germany*	Automotive	✓	–	–	47%: Female; 53% Male; Mean Age: 28 (SD=1.82); (n=15)
Bruemmer, Gertman, & Nielsen (2007) *North America*	Automotive (USAR)	–	✓	–	(n=153)
Bunt, Lount, & Lauzon (2012, February) *Canada*	Civilian/University	–	✓	–	Exp1 (n=21); Exp2: (n=14)
Dehais, Causse, Vachon, & Tremblay (2012) *France*	Military/Automotive	✓	–	–	Mean age: 27.84 (SD = 6.53) (n=13)
Groom & Nass (2007) *North America*	Military/Civilian	–	–	✓	–
Hancock, et al. (2011) *North America*	Military/Civilian	–	–	✓	–
Hoff & Bashir (2015) *North America*	Military/Civilian	–	–	✓	–
Hoffman, Johnson, Bradshaw, & Underbrink (2013) *North America*	Civilian	–	–	✓	–
Lee, Lau, Kiesler, & Chiu (2005, April) *North America; Hong Kong*	Civilian	✓	–	–	Exp 1: n=60); Exp 2: (n=48)
Nachtwei (2011) *Germany*	Civilian	–	–	✓	–
Nothdurft, Lang, Klepsch, & Minker (2013, April) *Germany*	Civilian	✓	–	–	(n=48)
Oleson, Billings, Kocsis, Chen, & Hancock (2011, February) *North America*	Military/Civilian	–	–	✓	–

Study: Author(s), Country	Subdomain	Exp[3]	Obv	Lit/D	Actors
Olson, Fisk, & Rogers (2009, October) *North America*	Civilian	✓	–	–	Older population (ages between 60-80) (n=19)
Sanders, Oleson, Billings, Chen, & Hancock (2011, September) *North America*	Military	–	–	✓	–
Schaefer, et al. (2014) *North America*	Military	–	–	✓	–
Talone, Phillips, Ososky, & Jentsch (2015, September) *North America*	Military	✓	–	–	(n=100)
System Related					
Performance Based: 6 Barg-Walkow (2013) *North America*	Civilian	✓	–	–	N= 60 (38% Female; 62% Male; Mean Age: 19.8 (0.21 SD)
Cassidy (2009) *North America*	Military/Naval	✓	–	–	26% Female;74% Male (n=42)
Dawson, Crawford, Dillon, & Anderson (2015, May) *North America*	Military/Civilian	✓	–	–	(n=40)
Mosier, et al. (2013) *North America*	Aviation	–	–	✓	–
Schaefer, Evans III, & Hill (2015) *North America*	Military	–	–	✓	–
Westin, Borst, & Hilburn (2016) *Netherlands*	Aviation/Civilian	–	–	✓	–
Attribute Based: 4					
Andersson (2010) *Sweden*	Civilian	–	✓	–	–
Sheridan & Nadler (2006) *North America*	Aerospace	–	–	✓	–
Shin, Busby, Hibberd, & McMahon (2005) *UK*	Civilian	–	–	✓	–
Silva & Hansman (2015) *North America*	Aviation	–	–	✓	–
Environment Related:					
Team Collaboration: 10					

Study: Author(s), Country	Subdomain	Exp[3]	Obv	Lit/D	Actors
Chen & Barnes (2013) *North America*	Military	–	–	✓	–
Hawley, Mares, & Giammanco (2006) *North America*	Military	–	–	✓	–
Joe, O’Hara, Medema, & Oxstrand (2014, June) *North America*	Civilian/Military	–	–	✓	–
Morita & Burns (2014) *Canada*	Civilian/University	–	✓	–	(n=200)
Ososky (2013) *North America*	Civilian/Military/University	✓	–	–	Undergraduate psychology student population. Mean Age: 18.78 (SD 1.61) (n=120)
Phillips, Ososky, & Jentsch (2014, September) *North America*	Military	–	–	✓	–
Phillips, Ososky, Grove, & Jentsch (2011, September) *North America*	Military	–	–	✓	–
Sætrevik (2013) *Norway*	Civilian	–	✓	–	–
Schaffernicht & Groesser (2011) *Multinational (South America/Switzerland)*	Civilian	–	–	✓	–
Smith, Borgvall, & Lif (2007) *UK*	Military	–	–	✓	–
Tasking: 1					
Clancey, Linde, Seah, & Shafto (2013) *North America*	Aviation/Aerospace	–	–	✓	–
TOTAL COUNT: 52					

Human-Automation Interactions

A common trend in HAI literature is priming (or knowledge of the system and its capabilities) can impact cognitive schema [12] [13] [14] [15]. Beggiato and Krems (2013) [12] and Beggiato *et al* (2015). [13] conducted studies of automated cruise control (ACC) and found initial information had an enduring effect on trust and acceptance and trust facilitation observed the Power Law of Learning [16]. Piccinini *et al.*(2015) [15] found automation error during critical situations had negative effects on behaviour and negative correlations between mental model and ACC operation. Zhang *et al* (2010) [17] also observed learnt mistrust behaviours from automation failure, as a result of over-estimation of system capability with SA and confidence ratings reflective of negative mental schema changes.

Conversely, Lo *et al.* (2015) [18] observed that prior knowledge or experience does not necessarily impact trust or performance with automation, but noted that schema differ with experience associated with goal orientated decision making. Dehais *et al* (2012) [19] observed improper perservation behaviour [6] as a result of psychosocial attitudes and socio-technical demands, decreased performance, proper use of automation and overall trust facilitation in the operators' mental models. Estimated knowledge or capability of a system and the effect on developing appropriate attitudes towards over- or under-reliance on automation also featured in Lee *et al.* (2005) [14].

The literature indicates TIA observes a positive trend when systems exhibit transparency and reliability, as these significantly influence construction of appropriate mental models and improvements to SA [20] [21] [22] [23]. Birkmeier *et al* (2011) [21] suggest increased trust in

[6] Perservation behaviour is the repetition or continuation of a particular response despite the absence or cessation of the initial stimulus. For example, an individual may solely focus on completion of a task to the exclusion of initial overall goal, time taken to complete, communication with team members or other critical tasks (such as DRI).

automation appropriate human-system interaction (HSI) is key for decision making and over-reliance is connected to LOA and out-of-the-loop system architecture.

Lim *et al* (2009) [23] found positive outcomes from transparency in experimental conditions with TIA, however found that priming did not have a statistically significant effect on mental models or perceptions of the system. Wilkinson *et al* (2007) [24] conversely reported primed mental models had improved task performance and TIA, stating participants with lower acquisition (weak schema of the system), compared to higher acquisition (robust mental models), demonstrated higher misuse and disuse with automation.

Three studies used qualitatively grounded psychological approaches [25] [26] [27]. Bruemmer *et al* (2007) [27] utilised metaphors as primers to modify existing narratives to manipulate existing schema. Similarly, Arkin *et al* (2012) [26] used human psychology, behaviour and attitudes to develop artificial intelligence to simulate HMM and schema regarding moral traits for high-level automation in future battlespaces.

The literature reviews and meta-analyses reviewed cover a diverse range and the following are grouped by the main outcome or theme explored in their respective sources. Education and training to promote appropriate HMM featured in five sources [28] [29] [30] [31] [11] in that schema can influence the operator's SA and behaviours through dispositional, situational and learned trust which impact facilitation with automation [11]. Augmented, Mixed and Virtual Environment (AMVE) technology suggest embedding training within operational equipment to maintain appropriate schema [28] as do critical environments within the military domain Oleson *et al* (2011) [30], through implementation of appropriate training interventions. Fallon *et al* (2010) [29] reports cognitive elements affect operator trust facilitation, such as sense-making to reframe schema after automation failure and error. Shaefer *et al* (2014) [31] reported error ($\bar{g} = +.44$) and communication feedback failure ($\bar{g}$ = +.45) had negative effects on trust development, whereas scenario training to support continued training has a large mediation effect ($\bar{g}$ = +.79).

Function appropriation is discussed in the research by Hawley *et al* (2005) [32] observing external factors (such as task allocation) and Hoffman and Woods (2011) [33] exploring internal factors (such as macro-cognition trade-offs between LOA and operator capability). Key outcomes of inappropriate task allocation [32] suggests fragmentation of workload (creating residuals) occurs between operator and system that cannot be properly framed by existing schema, may lead to interaction errors. Literature recommendations suggest clarification of task and suitable responses to automation error feedback, and increase fidelity and transparency to improve trust in operators. Hancock *et al.* (2011) [8] meta-analysis found a positive correlation ($r^2 = +26$) between factors improving TIA in operators, were supported with literature analysis within the field [34].

A primary problem in psychosocial narrative frameworks in the literature stem from viewing system-teammates as parallel to human-human interaction. Negative impact on appropriate trust development and reliance [35] [36] and narrative framework divergence between operators and system designers can also facilitate mistrust with automation [37] was observed in this review.

Human-System Interactions

Perceived reliability and actual reliability were not aligned in neither Barg-Walkow (2013) [38] nor Cassidy (2009) [39], which implies despite reliability increases, perception of automation capability may be limited to the operator schemata. Dawson *et al* (2015) [40] reported those with prior training or information had more stable perceptions of capability and were more inclined to trust automation. However, the study also indicated training did not significantly impact HMM in operators, and in some cases detrimental to developing TIA. SA played a role in system transparency and appropriate use, the authors reported spatial, temporal and environmental cues (STEC) are key for operators in developing appropriate understanding of autonomous agent behaviour.

Andersson (2010) [41] reported miscalibration exists between actual and perceived reality, when there is a dissonance between operator's schema and the technical capability of the automation. For example, when the mental model attributes larger functional capacity to the technical processes underpinning an automated system, there is an over-reliance on that system. The research implies that LOA may affect operator's mental models through degradation from non-continuous use (skill-fade) and creating cognitive dissonance.

Sheridan and Nadler (2006) [42] reported high occurrences of HAI error as a result of misuse of the system. The main findings found fidelity and transparency are key to maintaining appropriate schema, otherwise incorrect, inadequate or inaccurate feedback may negatively impact trust facilitation. Cognitive divergence and dissonance between human-automation capabilities is a common trend in accident analysis scenarios. Silva and Hansman (2015) [43] reported cognitive divergence as a failure in HSI through lack of transparency and feedback, such as the system changing state without operator input.

Shin, Busby, Hibberd and McMahon (2005) [44] report increased system complexity generates new mechanisms for design and human error. Their analysis supports cognitive dissonance errors in the mismatch between internal representations of system function and reality of operators' existing schemata with increased LOA. Shin *et al* (2005) [44] suggests there is only a partial overlap of mental models of system between designers and users regarding the underlying rationality of the system.

Environment and Social Influences

The environmental based research support the assertion that positive associations in operator interpretation of accuracy with congruent information, develop robust mental models which improve task performance and facilitate appropriate use of automated systems.

Morita and Burns (2014) [45] explore intuitive trust and HMM and frames group trust to explore the socio-cultural impact of external influences. The study investigates shared mental models (SMM), the impact of interpersonal trust and the regulating factors involved in best facilitating TIA with human-human teams. Sætrevik and Eid (2013) [46] similarly report SMM in human-human-automation teams facilitate appropriate team processes and performance. They observe shared information reflects higher degrees of SMM and appropriate SA. Furthermore, misinformed leadership (influenced by weak mental models and weak SMM with teammates) had a negative impact on team similarity indexes and reported lower accuracy and performance.

Clancey *et al* (2013) [47] explores the use of Brahms GÜM model to verify and validate a theoretical new assessment method for human-system simulations. The model uses cognitive framework to include interactions of pilots and air traffic control operators (ATCOs; modelled to represent human ontologies for different actors) – the research found distributed actants operating without knowledge of the other's actions (e.g. low transparency or increased cognitive divergence) create unexpected behaviour that is difficult to control and simulate.

Hawley *et al* (2006) [48] found better accretion with automation when appropriate mental model framework aids were utilised, and weak schema were accompanied with lower performance outcomes and error prone behaviour. In the Phillips *et al* (2014) [49] review, they reported increased levels of misuse, disuse and abuse where operators had weaker mental models and lower TIA and produced an overview of the many antecedents to facilitating TIA. Operator cognitive capacity (such as memory, task allocation, etc.) and HMM is discussed in another review by

Phillips *et al* (2011) [50] in which transference of human attributes from H-H teams onto H-S teams is discussed. Joe *et al* (2014) [51] reviews the mimicry of human behaviour in H-S teams and automation and suggests avoiding utilising human narratives and analogies as system parallels. Cognitive capacity is also reviewed by Chen (2013) [52] into the appropriateness of ecological interface design to HAI and the appropriateness of human-system information exchange impacting on the performance of H-H and H-S teams.

Schaffernicht and Groesser (2011) [53] and Smith *et al* (2007) [54] comprehensively explore metrics utilised in HMM research with a focal point towards individual and shared mental models in team communication and for mission success or goal attainment.

Limitations of the Review

The limitations are methodological as it is a scoping review and quality assessments for critical literature review require stricter guidelines. The synthesis of findings had challenges as guiding principles are ambiguous and undefined [7]. Some sources may have been missed in terminology selection mis-capture, article restrictions and unpublished technical reports.

There are differences in statistical significance and effect size in the trends and results, due to small populations of participants. The literature has limitations with lack of randomisation in experimental study protocols, self-reporting error and androcentric subject populations. The issue of self-reporting is accepted in the studies which utilise validated questionnaires (e.g. Trust in Automation Questionnaire) or appropriate qualitative design protocol. However, the review supports comparable outcomes in the existing literature.

Conclusions

The sources discussed support the use of HMM for exploring issues and limitations in facilitating TIA. The outcomes could be utilised in interface design recommendations, operator training and socio-technical bottlenecks.

The experimental design studies overall suggest priming and prior training can positively influence operators TIA through reduction of mistrust or inappropriate behaviour with the system, and appropriate reliance and knowledge of limitations of the system capability. However, cognitive dissonance between perceived reliability and actual reliability, may be a factor requiring more exploration regardless of fidelity of automated systems. Olson, Fisk, and Rogers (2009) [55] Ososky (2013) [56] and Wilkinson, Fisk and Rogers (2007) [24] suggest distrust and incorrect estimations of automation accuracy were still apparent at 100% precision. Transparency between human and system interaction is vital for providing operators with congruent information which provide accurate and appropriate interpretation of the system's capability and reliability, thereby robustly framing their schemata.

The observational studies support these findings through recommendations of improving transparency, which thus improves SA [20] [21] [45] [46], especially in environments with high LOA. However, priming operators through education and training has mixed reviews in the observational studies [23]. Nonetheless, the majority of studies reviewed throughout the literature summarised herein indicate that lower acquisition mental models can increase inappropriate behaviour with the system, whereby robust mental models support improved interaction and facilitation. Cognitive dissonance between actual and perceived reality (such as the capability of the system, or SA) and cognitive overloads impacts trust facilitation, performance and subsequent degradation of mental models if behaviour is not accurately framed by the operator's mental models.

Mental model framework divergence is key in studies reviewed that explored accident analysis of human-system teams, as incidents featured human-system interaction misuse and disuse heavily. Weak schema are associated with lower performance outcomes and error behaviour – together with transparency education and training, may have prevented the incidents and more robust models would re-converge appropriate framework. Training and transparency are recommended to facilitate HAI/HSI in human-human and human-automation team collaboration (both individual operators and groups).

The current literature seeks to explore underlying psychosocial impacts affecting performance and interaction within complex socio-technical environments through exploring mental models and schema. System transparency and operator priming through education can aid in facilitating appropriate trust as levels of automation rise. There are limitations within HAI research, as metrics used are varied and not cohesive across the literature and task performance outcomes are not necessarily representative of internal cognition and attitudes. Qualitative research and inquiry, although less objective, may yield rich contextual data to influence future HSI research in emerging and novel automation interaction and interfaces

The overall scope of current literature is the utilisation of mental models as a theoretical framework for inquiry, is an expanding field of research in the identification of the shifting limitations of TIA research as technology emerges at an ever-expanding pace. Transparency of automation behaviour within the socio-technical system is key to the improvement of appropriate reliance and performance of both system and operator. In addition, training has shown positive effects on trust as it aids in creating appropriate schema, SA and the priming of mental models. The socio-cultural context and environment of human-system interaction is also significant in improving performance and task outcome through team communication, collaboration and leadership.

Acknowledgements

This work was supported by EPSRC and BAE Systems.

References

[1] MoD U.K, "Global strategic trends out 2045.," 21 08 2014. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/348164/20140821_DCDC_GST_5_Web_Secured.pdf.

[2] R. Parasuraman, T. B. Sheridan and C. D. Wickens, "A Model for Types and Levels of Human Interaction with Automation," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans,* vol. 30, no. 3, pp. 286-297, 2000.

[3] J. D. Lee and K. A. See, "Trust in Automation: Designing for Appropriate Reliance.," *Human Factors: The Journal of the Human Factors and Ergonomics Society.,* vol. 46, no. 1, pp. 50-80, 2004.

[4] D. Atkinson, P. Friedland and J. Lyons, "Human-Machine Trust for Robust Autonomous Systems," in *4th IEEE Workshop on Human-Agent-Robot Teamwork,* 2012.

[5] G. Siemens, "Description of connectivism. Connectivism: A learning theory for today's learner," 2008. [Online]. Available: http://www.connectivism.ca/about.html.

[6] M. S. Wogalter, "Communication-human information processing (C-HIP) model.," in *Handbook of warnings*, CRC Press, 2006, pp. 51-61.

[7] H. Arksey and L. O'Malley, "Scoping studies: towards a methodological framework," *International journal of social research methodology,* vol. 8, no. 1, pp. 19-32, 2005.

[8] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. Chen, E. J. De Visser and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Human Factors: The Journal of the Human Factors and Ergonomics Society,* vol. 53, no. 5, pp. 517-527, 2011.

[9] D. Moher, A. Liberati, J. Tetzlaff and D. Altman, "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement," *BMJ,* vol. 339, no. 2535, 2009.

[10] D. Stroup, J. Berlin, S. Morton, I. Olkin, G. Williamson, D. Rennie, D. Moher, B. Becker, T. Sipe and S. Thacker, "Meta-analysis of observational studies in epidemiology: a proposal for reporting," *JAMA,* vol. 283, no. 15, pp. 2008-2012, 2000.

[11] K. A. Hoff and M. Bashir, "Trust in Automation Integrating Empirical Evidence on Factors That Influence Trust.," *Human Factors: The Journal of the Human Factors and Ergonomics Society,* vol. 57, no. 3, pp. 407-434, 2015.

[12] M. Beggiato and J. F. Krems, "The evolution of mental model, trust and acceptance of adaptive cruise control in relation to initial information.," *Transportation research part F: traffic psychology and behaviour,* vol. 18, pp. 47-57, 2013.

[13] M. Beggiato, M. Pereira, T. Petzoldt and J. Krems, "Learning and development of trust, acceptance and the mental model of ACC. A longitudinal on-road study.," *Transportation research part F: traffic psychology and behaviour,* vol. 35, pp. 75-84, 2015.

[14] S. L. Lee, I. Y. M. Lau, S. Kiesler and C. Y. Chiu, "Human mental models of humanoid robots. In . Proceedings of the (pp. 2767-2772). IEEE.," in *IEEE International Conference on Robotics and Automation, 2005,* 2005, April.

[15] G. F. B. Piccinini, C. M. Rodrigues, M. Leitão and A. Simões, "Reaction to a critical situation during driving with adaptive cruise control for users and non-users of the system," *Safety Science,* vol. 72, pp. 116-126, 2015.

[16] A. Newell and P. S. Rosenbloom, "Mechanisms of skill acquisition and the law of practice," in *Cognitive skills and their acquisition,* vol. 1, 1981, pp. 1-55.

[17] T. Zhang, D. Kaber and S. Hsiang, "Characterisation of mental models in a virtual reality-based multitasking scenario using measures of situation awareness," *Theoretical Issues in Ergonomics Science,* vol. 11, no. 1-2, pp. 99-118, 2010.

[18] J. C. Lo, K. R. Pluyter and S. A. Meijer, " Individual Markers of Resilience in Train Traffic Control The Role of Operators' Goals and Strategic Mental Models and Implications for Variation, Expertise, and Performance," *Human Factors: The Journal of the Human Factors and Ergonomics Society,* p. 1–12, 2015.

[19] F. Dehais, M. Causse, F. Vachon and S. Tremblay, "Cognitive conflict in human–automation interactions: a psychophysiological study," *Applied ergonomics,* vol. 43, no. 3, pp. 588-595, 2012.

[20] R. Aydoğan, A. Sharpanskykh and J. Lo, "A Trust-Based Situation Awareness Model.," *Multi-Agent Systems,* pp. 19-34, 2014.

[21] B. Birkmeier, B. Korn and F. O. Flemisch, "First findings on the controller's mental model in sectorless air traffic management," in *Digital Avionics Systems Conference (DASC), IEEE/AIAA 30th*, 2011, October.

[22] A. Bunt, M. Lount and C. Lauzon, "Are explanations always important?: A study of deployed, low-cost intelligent interactive systems.," in *ACM international conference on Intelligent User Interfaces*, 2012, February.

[23] B. Y. Lim, A. K. Dey and D. Avrahami, "Why and why not explanations improve the intelligibility of context-aware intelligent systems.," in *SIGCHI Conference on Human Factors in Computing Systems*, Boston, Massachusetts, USA., 2009, April.

[24] B. D. Wilkison, A. D. Fisk and W. A. Rogers, "Effects of mental model quality on collaborative system performance," in *Human Factors and Ergonomics Society Annual Meeting* , 2007, October.

[25] C. E. H. Chua, V. C. Storey and R. H. Chiang, "Deriving knowledge representation guidelines by analyzing knowledge engineer behavior.," *Decision Support Systems,* vol. 54, no. 1, pp. 304-315, 2012.

[26] R. Arkin, P. Ulam and A. Wagner, "Moral decision making in autonomous systems: enforcement, moral emotions, dignity, trust, and deception.," *Proceedings of the IEEE,* vol. 100, no. 3, pp. 571-589, 2012.

[27] D. J. Bruemmer, D. I. Gertman and C. W. Nielsen, "Metaphors to Drive By: Exploring New Ways to Guide Human- Robot Interaction.," *Open Cybernetics & Systemics Journal,* vol. 1, pp. 5-12, 2007.

[28] RTO/NATO, "Virtual Environments for Intuitive Human-System Interaction – Human Factors Considerations in the Design, Use, and Evaluation of AMVE-Technology," 2007.

[29] C. K. Fallon, A. K. Murphy, L. Zimmerman and S. T. Mueller, "The calibration of trust in an automated system: A sensemaking process.," in *Collaborative Technologies and Systems (CTS)*, 2010, May.

[30] K. E. Oleson, D. R. Billings, V. Kocsis, J. Y. Chen and P. A. Hancock, "Antecedents of trust in human-robot collaborations," in *First International Multi-Disciplinary Conference onCognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 2011, February.

[31] K. E. Schaefer, D. R. Billings, J. L. Szalma, J. K. Adams, T. L. Sanders, J. Y. Chen and P. A. Hancock, "A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Human-Robot Interaction," 2014.

[32] J. K. Hawkley, A. L. Mares and C. A. Giammanco, "The human side of automation: Lessons for air defense command and control (No. ARL-TR-3468)," Maryland, 2005.

[33] R. R. Hoffman and D. D. Woods, "Beyond Simon's slice: Five fundamental trade-offs that bound the performance of macrocognitive work systems.," *Intelligent Systems,* vol. 26, no. 6, pp. 67-71, 2011.

[34] T. Sanders, K. E. Oleson, D. R. Billings, J. Y. Chen and P. A. Hancock, "A model of human-robot trust theoretical model development," in *Human Factors and Ergonomics Society Annual Meeting* , 2011, September.

[35] V. Groom and C. Nass, "Can robots be teammates?: Benchmarks in human–robot teams.," *Interaction Studies,* vol. 8, no. 3, pp. 483-500, 2007.

[36] R. R. Hoffman, M. Johnson, J. M. Bradshaw and A. Underbrink, "Trust in automation," *Intelligent Systems,* vol. 28, no. 1, pp. 84-88, 2013.

[37] J. Nachtwei, "The many faces of human operators in process control: a framework of analogies.," *Theoretical Issues in Ergonomics Science,* vol. 12, no. 4, pp. 297-317, 2011.

[38] L. H. Barg-Walkow, "Understanding the role of expectations on human responses to an automated system.," 2013.

[39] A. M. Cassidy, "Mental models, trust, and reliance: Exploring the effect of human perceptions on automation use.," Naval Postgraduate School, Monterey, CA, 2009.

[40] S. Dawson, C. Crawford, E. Dillon and M. Anderson, "Affecting operator trust in intelligent multirobot surveillance systems.," in *IEEE International Conference on Robotics and Automation (ICRA)*, Washington, 2015, May.

[41] J. Andersson, "A conceptual model for analysis of automation usability problems in control room settings.," Department of Product and Production Development (Division Design & Human Factors) Chalmers University of Technology, Göteborg, Sweden, 2010.

[42] T. B. Sheridan and E. D. Nadler, "A review of human-automation interaction failures and lessons learned," Volpe National Transportation System Center, Cambridge, MA, 2006.

[43] S. S. Silva and R. J. Hansman, "Divergence Between Flight Crew Mental Model and Aircraft System State in Auto-Throttle Mode Confusion Accident and Incident Cases," *Journal of Cognitive Engineering and Decision Making,* vol. 9, no. 4, p. 312–328, 2015.

[44] I. J. Shin, J. S. Busby, R. E. Hibberd and C. A. McMahon, "A theory-based ontology of design induced error," in *15th International Conference on Engineering Design*, Melbourne, Australia, 2005.

[45] P. P. Morita and C. M. Burns, "Understanding 'interpersonal trust' from a human factors perspective: insights from situation awareness and the lens model," *Theoretical Issues in Ergonomics Science,* vol. 15, no. 1, pp. 88-110, 2014.

[46] B. &. E. J. Sætrevik, "The "Similarity Index" as an Indicator of Shared Mental Models and Situation Awareness in Field Studies.," *Journal of Cognitive Engineering and Decision Making,* vol. 8, no. 2, p. 119–136, 2013.

[47] W. J. Clancey, C. Linde, C. Seah and M. Shafto, "Work practice simulation of complex human–automation systems in safety critical situations: the Brahms Generalized Überlingen Model (Tech. Publ. 2013-216508).," NASA, Washington., 2013.

[48] J. K. Hawley, A. L. Mares and C. A. Giammanco, "Training for effective human supervisory control of air and missile defense systems (No. ARL-TR-3765)," Army Research Labratory, Texas, 2006.

[49] E. Phillips, S. Ososky and F. Jentsch, "Leveraging Features of Human—Technology Teams to Support Mental Models in Future Soldier—Robot Teams.," in *Human Factors and Ergonomics Society Annual Meeting*, 2014, September.

[50] E. Phillips, S. Ososky, J. Grove and F. Jentsch, " From Tools to Teammates Toward the Development of Appropriate Mental Models for Intelligent Robots," in *Human Factors and Ergonomics Society Annual Meeting* , 2011, September.

[51] J. Joe, J. O'Hara, H. Medema and J. Oxstrand, "Identifying requirements for effective human-automation teamwork," in *12th International Conference on Probabilistic Safety Assessment and Management*, 2014, June.

[52] J. Y. Chen and M. J. Barnes, "Human-Agent Teaming for Multi-Robot Control: A Literature Review (No. ARL-TR-6328).," Army Research Lab Aberdeen Proving Ground, Maryland, 2013.

[53] M. Schaffernicht and S. N. Groesser, "A comprehensive method for comparing mental models of dynamic systems," *European Journal of Operational Research,* vol. 210, no. 1, pp. 57-67, 2011.

[54] E. Smith, J. Borgvall and P. Lif, "Team and collective performance measurement.," Defence Science and Technology Lab, Bedfordshire, UK, 2007.

[55] K. E. Olson, A. D. Fisk and W. A. Rogers, "Collaborative Automated Systems: Older Adults' Mental Model Acquisition and Trust in Automation.," in *Human Factors and Ergonomics Society Annual Meeting*, 2009, October .

[56] S. J. Ososky, “Influence of Task-Role Mental Models on Human Interpretation of Robot Motion Behavior,” University of Central Florida Orlando, Florida, 2013.

[57] J. D. Lee and K. A. See, “Trust in automation: Designing for appropriate reliance.,” *Human Factors: The Journal of the Human Factors and Ergonomics Society,* vol. 46, no. 1, pp. 50-80, 2004.

[58] A. M. Collins and E. F. Loftus, “A spreading-activation theory of semantic processing.,” *Psychological review,* vol. 82, no. 6, p. 407, 1975.

[59] G. E. Moore, “Cramming more components onto integrated circuits,” *Electronic,* vol. 38, no. 8, p. 114, 1965.

[60] G. Abe and J. Richardson, “Alarm timing, trust and driver expectation for forward collision warning systems.,” *Applied Ergonomics,* vol. 37, no. 5, pp. 577-586, 2006.

[61] N. Ezer, A. D. Fisk and W. A. Rogers, “Reliance on automation as a function of expectation of reliability, cost of verification, and age.,” in *The Human Factors and Ergonomics Society Annual Meeting*, 2007.

[62] A. Mayer, “The manipulation of user expectancies: Effects on reliance, compliance, and trust using an automated system.,” Georgia Institute of Technology,, Atlanta, 2008.

[63] S. M. Merritt, H. Heimbaugh, J. LaChapell and D. Lee, “I trust it, but I don't know why: effects of implicit attitudes toward automation on trust in an automated system.,” *Human Factors: The Journal of the Human Factors and Ergonomics Society,* vol. 55, no. 3, pp. 520-534, 2013.

[64] J. P. Higgins and S. Green, Cochrane handbook for systematic reviews of interventions, vol. 5, Chichester: Wiley-Blackwell, 2008.

[65] F. Nothdurft and W. Minker, “Justification and transparency explanations in dialogue systems to maintain human-computer trust,” in *4th International Workshop on Spoken Dialogue Systems,* Napa, 2014, January.

[66] M. A. Neerincx, G. M. Te Brake, J. G. M. Van de Ven, H. F. R. Arciszewski, T. E. De Greef and J. Lindenberg, “Situated cognitive engineering: Developing adaptive track handling support for naval command and control centers.,” in *Human Centred Processes Third International Conference on Human-Centered Processes,* 2008, January.

[67] A. B. Talone, E. Phillips, S. Ososky and F. Jentsch, “An Evaluation of Human Mental Models of Tactical Robot Movement,” in *Human Factors and Ergonomics Society Annual Meeting* , 2015, September.

[68] D. N. Beede, T. A. Julian, D. Langdon, G. McKittrick, B. Khan and M. E. Doms, “Women in STEM: A gender gap to innovation,” Economics and Statistics Administration Issue Brief,, 2011.

[69] Y. J. Xu, “Gender disparity in STEM disciplines: A study of faculty attrition and turnover intentions,” *Research in Higher Education,* vol. 49, no. 7, pp. 607-624, 2008.

[70] F. Nothdurft, H. Lang, M. Klepsch and W. Minker, “Bases of human-computer trust and explanations.,” in *Human Factors in Computing Systems*, Paris, France., 2013, April.

[71] A. Cynkar, “The changing gender composition of psychology.,” *Monitor on Psychology,* vol. 38, no. 6, p. 46, 2007.

[72] K. L. Mosier, U. Fischer, D. Morrow, K. M. Feigh, F. T. Durso, K. Sullivan and V. Pop, “Automation, Task, and Context Features Impacts on Pilots’ Judgments of Human–Automation Interaction.,” *Journal of Cognitive Engineering and Decision Making,* vol. 7, no. 4, pp. 377-399, 2013.

[73] K. E. Schaefer, A. W. Evans III and S. G. Hill, “Command and Control in Network-Centric Operations: Trust and Robot Autonomy.,” U. S. Army Research Laboratory, 2015.

[74] C. Westin, C. Borst and B. Hilburn, “Strategic Conformance: Overcoming Acceptance Issues of Decision Aiding,” *IEEE Transaction on Human-Machine Systems,* vol. 46, no. 1, pp. 41--52, 2016.

[75] J. Wilson and A. Rutherford, “Mental models: Theory and application in human factors,” *Human Factors,* vol. 31, no. 6, pp. 617-634, 1989.