

How to Formally Model Human in Collaborative Robotics

Mehrnoosh Askarpour

McMaster University
Canada

askarpom@mcmaster.ca

Human-robot collaboration (HRC) is an emerging trend of robotics that promotes the co-presence and cooperation of humans and robots in common workspaces. Physical vicinity and interaction between humans and robots, combined with the uncertainty of human behaviour, could lead to undesired situations where humans are injured. Thus, safety is a priority for HRC applications.

Safety analysis via formal modelling and verification techniques could considerably avoid dangerous consequences, but only if the models of HRC systems are comprehensive and realistic, which requires reasonably realistic models of human behaviour. This paper explores state-of-the-art solutions for modelling human and discusses which ones are suitable for HRC scenarios.

1 Introduction

A new uprising section of robotics is Human-Robot Collaboration (HRC), where human operators are not dealing with robots only through interfaces, but they are physically present in the vicinity of robots, performing hybrid tasks (i.e., partially done by the human and partially by the robot). These applications introduce promising improvements in the industrial manufacturing area by combining human flexibility and machine productivity [52], but they must assure the safety of human operators before being fully deployed to certify that interactions with robots will not cause any harm or injury to humans.

Formal methods have been widely used in robotics for decades in a variety of applications including mission planning [6, 20, 46, 53, 63], formal verification of properties [3, 48, 49, 55], and controllers [8, 24, 66]. They could be an effective means for the safety analysis of traditional and collaborative robotics due to their comprehension and exhaustiveness [32, 77, 78]. However, a formal model of an HRC system should also reflect the human factors that impact the state of the model without dealing with the additional details of human mental and emotional processes. It necessitates building a formal model of human behaviour that replicates their physical presence and the observable manifestation of their behaviour which includes both executing the required job following the expected instructions, and deviations from the expected behaviour (i.e., mistakes, errors, malicious use). Thus, the model of the human for each HRC scenario might be intertwined with the model of the executing job.

A 100% realistic formal model of human might not be a reasonable goal and, just like any other phenomena, any human model is subject to some level of abstraction and simplification. Moreover, unlike robots that perform only a limited set of activities, all possible activities of the human are not foreseeable. Besides, humans are non-deterministic, and a realistic algorithm for their behaviour is not easy to envision.

To tackle these issues, researchers have explored well-established cognitive investigations, task-analytic models, and probabilistic approaches [12]. These three tracks are not mutually exclusive; for example, there are instances of cognitive probabilistic models in the literature that will be discussed later in the paper. The rest of this paper, reports on the state-of-the-art on each of these possibilities and examines their compatibility for HRC scenarios, following a snowballing literature review. Moreover, for each

track there are several instances that adhered to the normative human behaviour, while other instances considered erroneous behaviour too; hence, a separate section is dedicated to examples that also model errors.

In addition to the three highlighted possibilities, Bolton et al. [11] considered human-device interface models, which are out of the scope of this paper. They do not consider the physical co-presence of humans and robots and focus only on their remote communication; thus, human physical safety has never been an issue in these studies [10].

The rest of this paper is structured as follows: Section 2 explores the cognitive approaches; Section 3 reviews task analytic approaches; Section 4 discusses the probabilistic techniques; Section 5 specifically discusses the difficulties of modelling human errors; finally Section 6 draws a few conclusions on the best-suited solution for HRC scenarios.

2 Cognitive Models

Cognitive models specify the rationale and knowledge behind human behaviour while working on a set of pre-defined tasks. They are often incorporated in the system models and contain a set of variables that describe the human cognitive state, whose values depend on the state of task execution and the operation environment [12]. Famous examples of well-established cognitive models follow.

SOAR [42] is an extensive cognitive architecture, relying on artificial intelligence principles, that reproduce human reasoning and short and long term memories. However, it only permits one operation at a time which seems not to be realistic in HRC scenarios (e.g., human sends a signal while moving towards the robot).

ACT-R [2] is a detailed and modular architecture that depicts the learning and perception processes of humans. It contains modules that simulate declarative (i.e., known facts like $2 + 2 = 4$) and procedural (i.e., knowledge of how to sum two numbers) aspects of human memory. An internal pattern matcher searches for the procedural statement relevant to the task that the human needs to perform (i.e., an entry in declarative memory) at any given time. SAL [38, 45] is an extended version of ACT-R, enriched with a neural architecture called Leabra.

SOAR and ACT-R highlight the cognition behind erroneous behaviours of the human [21] that could impact safety, such as over-trusting or a lack of trust in the system. It could be used to generate realistic models that also reflect a human deviation from the correct instructions. On the other hand, they both lack a formal definition and cannot be directly inputted to automated verification tools. Hence, they must be transformed into a formal model, which is a cumbersome and time-consuming process and requires extensive training for modellers [69]. Moreover, they are detailed and heavy models and, therefore, must be abstracted before formalization to avoid a state-space explosion. There are examples of formal transformation of ACT-R in [15, 30, 43] and of SOAR in [36]. However, they remain dependent on their case-studies or use arbitrary simplifications, and therefore, cannot be re-used as a general approach. Thus, providing a trade-off between abstraction and generality of these two cognitive models is not an easy task.

Programmable User Models (PUM) define a set of goals and actions for humans. The model mirrors both human mental actions (i.e., deciding to pick an object) and physical actions (i.e., pick an object). These models have a notion of a human *mental model* [31, 67] and separate the machine model from the user's perception of it [16], that avoids mode confusions which happen when the observed system behaviour is not the same as the user's expectation [17]. Additionally, Moher et al. [56] assign a *certainty* level to mental models whose different adjustments reveal various human reactions in execution situa-

tions. Curzon et al. [23] introduce two customized versions of mental models as naive or experienced.

Since PUM has been around for so long, simplified formal versions of it are defined for a variety of applications such as domestic service robots [72] Formalised with [71], interactive shared applications [18] Formalised with logic formulae, and Air Management System [79] Formalised with Petri-nets. PUM models have a general and accurate semantics and could be well-suited inputs for automated formal verification tools upon simplification. On the other hand, they are very detailed and large and risk the state-space exploration phenomena. As mentioned above, there are simplified versions of PUM models which again are not generic enough to be used for different scenarios including HRC applications. Hence, the effort and time required for customization and simplification remain as high as for SOAR and ACT-R.

3 Task-Analytic Models

Task-analytic models, as their name suggests, analyze human behaviour throughout the execution of the task. Therefore, they study the task as a hierarchy of atomic actions. By definition, these models reflect the expected behaviour of the human, which leads to correct execution of the task, and do not focus on reflecting erroneous behaviour [11]. Recently, however, Bolton et al. [13] put together a task-based taxonomy of erroneous human behaviour that allows errors to be modelled as divergences from task models; Li et al. [47] uses an analytic hierarchy process to identify hazards and increase the efficiency of the executing task.

Examples of task-analytic models follow. Paterno et al. [60] extend ConcurTaskTrees (CTT) [1] to better express the collaboration between multiple human operators in air traffic control; Mitchell and Miller [54] use function models to represent human activities in a simple control system (e.g., system shows information about the current state and operator makes relative control actions); Hartson et al. [33] introduces User Action Notation, a task and user-oriented notation to represent behavioural asynchronous, direct manipulation interface designs; Bolton et al. [14] establishes an Enhanced Operator Function Model, a generic XML-based notation, to gradually decompose tasks into activities, sub-activities and actions.

Task-analytic models depict human-robot co-existence and highlight the active role of humans in the execution of the task. However, they do not always reduce the overall size of the state spaces of the model [10], especially when multiple human operators are involved in the execution. Moreover, they do not offer re-usability (i.e., dependent on the case-study scenario and not generalizable) and generation of erroneous human behaviour.

4 Probabilistic Human Modelling

Another approach to reproduce human non-determinism and uncertainty more vividly are probabilistic models. Unlike deterministic models that produce a single output, stochastic/probabilistic models produce a probability distribution. In theory, a probabilistic human model could be very beneficial for model-based safety assessment in terms of a trade-off between cost and safety. For example, given that human manifests activity A with probability $P \ll 0.001\%$, and that A causes hazard H which is very expensive to mitigate, system engineers can save some money by not installing expensive mitigation for H that might practically never happen and settle for a more cost-effective mitigation. But we must analyse the challenges of probabilistic models too, so let us discuss a few examples.

Tang et al. [74] propose a bayesian probabilistic human motion model and argue that human mobility behaviour is uncertain but not random, and depends on internal (e.g., individual preferences) and external (e.g., environment) factors. They introduced an algorithm to learn human motion patterns from collected data about human daily activities from GPS and mobile phone data, and extract a probabilistic relation between current human place and past places in an indoor environment. The first issue about this work, and probabilistic models in general, is collecting a big enough dataset for extracting the correct parameters for a probabilistic distribution. The second issue is that the dataset gathered by GPS and cameras is very coarse-grained for modelling delicate HRC situations; they might capture human moving from one corner of a small workspace to another, but do not capture a situation such as human handing an object to the robot gripper or changing the tool-kit installed on the robot arm.

Hawkins et al. [34] present an HRC-oriented approach, based on a probabilistic model of humans and the environment, where the robot infers the current state of the task and performs the appropriate action. This model is very interesting but is more suited for simulation experiments, thus needs adaptive changes to be compatible with formal verification tools. The model also assumes that human performs everything correctly which is not quite realistic.

Tenorth et al. [75] uses Bayesian Logic Networks [37] to create a human model and evaluated their approach with TUM kitchen [58] and the CMU MMAC [76] data sets, both focusing only on full-body movements that exclude many HRC-related activities (e.g., assembling, screw-driving, pick and place, etc).

The Cognitive Reliability and Error Analysis Method (CREAM) [35] is a probabilistic cognitive model that is used for human reliability analysis [26], therefore, focuses on correct and incorrect behaviour both; it defines a set of modes to replicate different types of human behaviour which have different likelihoods to carry out certain errors. Similarly, Systematic Human Error Reduction and Prediction Analysis (SHERPA) [27] is a qualitative probabilistic human error analysis with a task-analytic direction which contains several failure modes: action errors, control errors, recovery errors, communication errors, choice errors. De Felice et al. [28] propose a combination of CREAM and SHERPA that uses empirical data to assign probabilities to each mode; therefore, the results are strongly dependent on the case study domain and the reliability and amount of data.

In recent years, machine-learning algorithms have been used quite extensively to create probabilistic human models [25, 39, 50], however having a reliable and large-enough dataset to learn probabilities from remains an issue. The communities (i.e., providers, users) shall develop such datasets for different domains by storing log histories and integrating them into a unique dataset (e.g., a dataset for industrial assembly tasks, a dataset for service robot tasks, etc.). The larger these datasets become, the more reliable they get, the better the human models extracted from them will be.

5 Human Erroneous Behavior

Human operators are prone to errors—an activity that does not achieve its goal [64], and one significant source of hazards in HRC applications is human errors. Reason [65] metaphorically states that the weaknesses of safety procedures in a system allow for the occurrence of human errors. Hence, a realistic human model for safety analysis shall be able to replicate errors, too.

The previous sections explained different possibilities to model human behaviour. They all could be used to model human errors, too, but are not always used as such; some of the papers discussed above, consider errors and others entirely skip them. This challenge is discussed in a separate section, but all of the papers discussed below could be listed in at least one of the previous three sections.

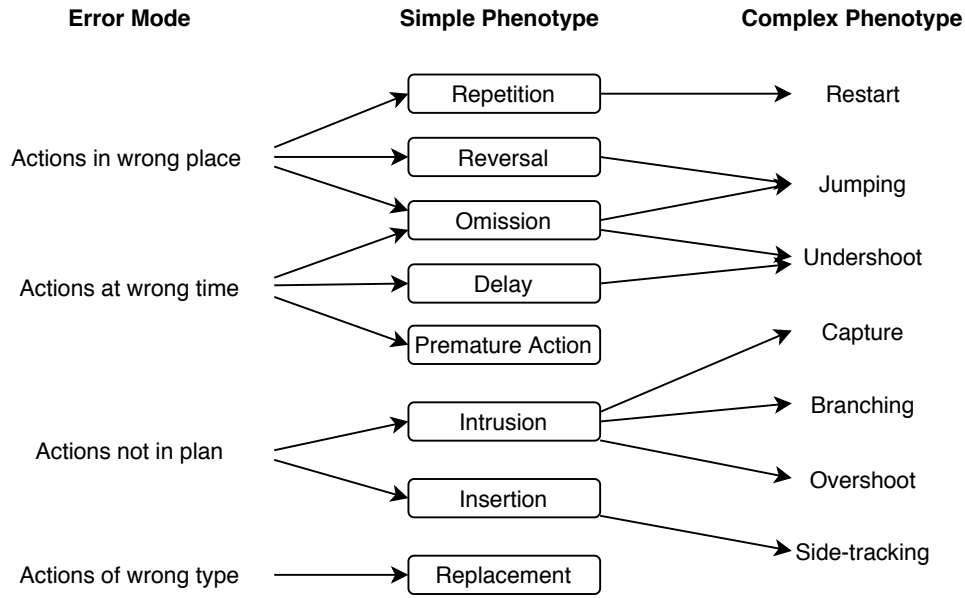


Figure (1) A taxonomy of erroneous actions phenotypes, taken from [29]. Wrong place refers to the action's temporal position in the execution sequence, not to a location in the layout; undershooting, which occurs when the action stops too early; sidetracking occurs when a segment of unrelated action is carried out, then the correct sequence is resumed; capturing occurs where an unrelated action sequence is carried out instead of the expected one; branching is where the wrong sequence of actions is chosen; overshooting happens where the action carries on past its correct endpoint by not recognizing its post-conditions.

One must first recognize errors and settle for a precise definition for them to model them. Although there is no widely used classification of human errors, some notable references follow.

Reason [64] classifies errors as behavioral (i.e., task-related factors), contextual (i.e., environmental factors) and conceptual (i.e., human cognition). Shin et al. [70] divide errors into two main groups; location errors happen when humans shall find a specific location for the task, and orientation errors happen due to humans' various timing or modality to perform a task. Hollnagel [35] classifies human errors in eight simple phenotypes: repeating, reversing the order, omission, late or early execution, replacement, intrusion of actions and inserting an additional action from elsewhere in the task; the paper then combines them to get complex phenotypes, as shown in Figure 1, and manually introduces them in the formal model of the model which produces too many false negatives. This issue might be resolved by introducing a probability distribution on phenotypes.

Kirwan has conducted an extensive review on human error identification techniques in [41] and suggested the following as the most frequent version:

- Slips and lapses regarding the quality of performance.
- Cognitive, diagnostic and decision-making errors due to misunderstanding the instructions.
- Maintenance errors and latent failures during maintenance activities.
- Errors of commission when the human does an incorrect or irrelevant activity.
- Idiosyncratic errors regarding social variables and human emotional state.
- Software programming errors leading to malfunctioning controllers that put the human in danger.

Cerone et al. [19] uses temporal logic to model human errors that are defined as: fail to observe potential conflict, ineffective or no response to observed conflict, fail to detect the criticality of the conflict.

As explained above, the error definition is context-dependent and might vary from one domain to another. It seems that the taxonomy presented in [29] is comprehensive enough for possibilities in HRC scenarios. Nevertheless, the discussion above was a general description of the violation of Norms; we must explore also papers that model the above definitions.

Many of the works on modelling human errors are inspired by one of the two main steps of human reliability analysis [26] that are error identification and error probability quantification [57]. For example, Martinie et al. [51] proposes a deterministic task-analytic approach to identify errors, while CREAM, HEART [80], THERP [73] and THEA [62] are probabilistic methods for modelling human. However, the models generated by these methods need formalization and serious customization because they strongly depend on their case study. Examples of other works with the same issues follow.

[9, 59] study the impacts of miscommunication between multiple human operators while interacting with critical systems; [7, 61] explore human deviation from correct instructions using ConcurTaskTrees; [68] upgrades the SAL cognitive model with systematic errors taken from empirical data.

On the other hand, several works developed formal human models. For example, Curzon and Blandford [22] propose a cognitive formal model in higher-order logic that separates user-centred and machine-centred concerns. It uses the HOL proof system [44] to prove that following the proper design rules, the machine-centred model does not allow for errors in the user-centred model. They experimented with including strong enough design rules in the model, so to exclude cognitively plausible errors by design [21]. This method clearly ignores realistic scenarios in which human actually performs an error. In another example, Kim et al. [40] map human non-determinism to a finite state automaton. However, their model is limited to the context of prospective control. Askarpour et al. [4, 5] formalize simple phenotypes introduced by Fields [29] with temporal logic, and integrate the result in the overall system model which is verified against a physical safety property.

6 Conclusions

This paper reviews the state-of-the-art on modelling human in HRC applications for safety analysis. The paper explored several papers, excluding those on interface applications, and grouped them into three main groups which are not mutually exclusive: cognitive, task-analytic, and probabilistic models. Then it reviewed state-of-the-art on modelling human errors and mistakes which is absolutely necessary to be considered for safety. Human error models expand over the three main groups but have been discussed separately for more clarity. Table 1 summarizes the observations from which the following conclusions are drawn:

- Cognitive models are very detailed and extensive. They often originate from different research areas (e.g., psychology), therefore have been developed with a different mentality from that of formal methods practitioners. So, they must be Formalised; but their formalization requires a lot of effort for abstraction and HRC-oriented customization. They also require specialist training for modellers to understand the models and be able to modify them.
- Among cognitive models, PUM seems to be the one with more available formal instances. It also could be tailored to different scenarios with much less time, effort and training, compared to ACT-R and SOAR.

Table (1) A summary of the discussed papers, extracted by a snowballing literature review. × means that the feature is not included, ✓ means that the feature is included, ? means that the feature has not been explored by the papers of the row but potentially could be addressed by their proposed approach (regardless of the required effort and the resulting efficiency), “semi” means that the model has partially formal semantics but is not in a form to be fed to an automated verification tool. When a reproducible model or approach has been used, the name of the approach is mentioned instead of a ✓. The acronyms used here are explained in Table 2.

| modelling Approach | Cognitive | Task-Analytic | Probabilistic | Modelling Errors | Formalised | HRC Compatible |
|--------------------|-----------|---------------|------------------|------------------|--------------------|----------------|
| [42] | SOAR | × | × | ? | × | × |
| [2] | ACT-R | × | × | ? | × | × |
| [15, 30] | ACT-R | × | × | ? | ✓ | × |
| [36] | SOAR | × | × | ? | ✓ | × |
| [31, 67, 72] | PUM | × | × | ? | ✓ | ? |
| [18, 79] | PUM | × | × | ? | ✓ | × |
| [13] | × | EOFM | × | ✓ | × | ✓ |
| [47] | × | Graphs | × | × | × | × |
| [60] | × | CCT | × | ? | semi | × |
| [54] | × | FSA | × | ✓ | ✓ | × |
| [33] | × | UAN | × | × | semi | × |
| [14] | × | EOFM | × | × | × | × |
| [74] | × | × | Bayesian | ✓ | × | × |
| [34] | × | × | DBN | × | × | ✓ |
| [75] | × | × | BLN | ? | × | × |
| [28] | × | × | CREAM and SHERPA | × | × | ? |
| [25, 39, 50] | × | × | ML | ? | × | ? |
| [29] | ✓ | ✓ | × | ✓ | × | ? |
| [19] | × | OCM | × | ✓ | ✓ | ? |
| [51] | × | ✓ | × | ✓ | ✓ | × |
| [80] | × | ✓ | CREAM and HEART | ✓ | ✓ | ? |
| [73] | × | ✓ | THERP | ✓ | ✓ | ? |
| [62] | × | ✓ | THEA | ✓ | ✓ | ? |
| [59] | ✓ | EOFMC | × | ✓ | LTL | × |
| [9] | SAL | EOFM | × | ✓ | × | × |
| [7] | × | CCT | × | ✓ | semi | ? |
| [61] | × | CCT | × | ✓ | × | ? |
| [68] | SAL | × | × | ✓ | × | × |
| [22] | PUM | ✓ | × | ✓ | Higher-Order Logic | × |
| [21] | PUM | ✓ | × | ✓ | Temporal Logic | ? |
| [40] | [26] | × | × | ✓ | FSA | × |
| [5] | × | ✓ | × | ✓ | Temporal Logic | ✓ |

Table (2) Acronyms Explained.

| | |
|--------|--|
| PUM | Programmable User Model |
| OCM | Operator Choice Model |
| EOFM | Enhanced Operator Function Model |
| EOFMC | Enhanced Operator Function Model with Communications |
| UAN | User Action Notation |
| CCT | ConcurTaskTrees |
| FSA | Finite State Automaton |
| CREAM | The Cognitive Reliability and Error Analysis Method |
| SHERPA | Systematic Human Error Reduction and Prediction Analysis |
| LTL | Linear Temporal Logic |
| DBN | Dynamic Bayes Network |
| BLN | Bayesian Logic Networks |
| ML | Machine Learning |
| THERP | Technique for Human Error-rate Prediction |
| THEA | Technique for Human Error Assessment Early in Design |
| HEART | Human error assessment and reduction technique |
| SHERPA | Systematic Human Error Reduction and Prediction Analysis |

- Task-analytic models offer little reusability; they are so intertwined with the task definition that a slight change in the task might cause significant changes to the model. They also must be first defined with a hierarchical notation for the easy and clear decomposition of the task and then be translated to an understandable format for a verification tool.
- Probabilistic models seem to be an optimal solution; they combine either task-analytic or cognitive models with probability distributions. However, the biggest issue here is to have reliable and large-enough data sets to extract the parameters of probability distributions from. It requires the robotics community to produce huge training datasets from their system history logs (e.g., how frequently human moves in the workspace, the average value of human velocity while performing a specific task, the number of human interruptions during the execution, the number of emergency stops, the number of reported errors in an hour of execution, ...).
- None of the three approaches above is enough if they exclude human errors. There are several works on outlining and classifying human errors. The best way to have a unified terminology would be for standards organizations in each domain to introduce a list of the most frequent human errors in that domain. It would be possible only if the datasets mentioned above are available. Modelling all of the possible human errors might not be feasible, but modelling those that occur more often is feasible and considerably improves the quality of the final human model.
- Error models might introduce a huge amount of false positive cases (i.e., incorrectly reporting a hazard when it is safe) during safety analysis. Thus, probabilistic error models might be the best combination to resolve it.
- In the safety analysis of HRC systems, the observable behaviour of humans and its consequences (i.e., how it impacts the state of the system) are very important. However, the cognitive elements behind it are not really valuable to the analysis. One could use them as a black-box that derives the observable behaviour with a certain probability but the details of what happens inside the box

do not really matter. Therefore, the cognitive model seems to contain a huge amount of detail that does not necessarily add value to the safety analysis but makes the model heavy and the verification long.

- Table 1 suggests that a combination of probabilistic and task-analytic approaches that model erroneous human behaviour is the best answer for a formal verification method for safety analysis of HRC systems.

References

- [1] (2012): *Concur Task Trees (CTT)*. available from w3.org. Accessed: 2017-07-06.
- [2] John R Anderson (1996): *ACT: A simple theory of complex cognition*. *American Psychologist* 51, doi:10.1037/0003-066X.51.4.355.
- [3] Mehrnoosh Askarpour (2016): *Risk Assessment in Collaborative Robotics*. In: *Formal Methods Doctoral Symposium, FM-DS, CEUR-WS 1744*.
- [4] Mehrnoosh Askarpour, Dino Mandrioli, Matteo Rossi & Federico Vicentini (2017): *Modeling Operator Behavior in the Safety Analysis of Collaborative Robotic Applications*. In Stefano Tonetta, Erwin Schoitsch & Friedemann Bitsch, editors: *Computer Safety, Reliability, and Security - 36th International Conference, SAFECOMP 2017, Trento, Italy, September 13-15, 2017, Proceedings, Lecture Notes in Computer Science 10488*, Springer, pp. 89–104, doi:10.1007/978-3-319-66266-4_6.
- [5] Mehrnoosh Askarpour, Dino Mandrioli, Matteo Rossi & Federico Vicentini (2019): *Formal model of human erroneous behavior for safety analysis in collaborative robotics*. *Robotics and Computer-Integrated Manufacturing* 57, pp. 465 – 476, doi:10.1016/j.rcim.2019.01.001.
- [6] Mehrnoosh Askarpour, Claudio Menghi, Gabriele Belli, Marcello M. Bersani & Patrizio Pelliccione (2020): *Mind the gap: Robotic Mission Planning Meets Software Engineering*. In: *FormaliSE@ICSE 2020: 8th International Conference on Formal Methods in Software Engineering, Seoul, Republic of Korea, July 13, 2020*, pp. 55–65, doi:10.1145/3372020.3391561.
- [7] Sandra Basnyat & Philippe Palanque (2005): *A task pattern approach to incorporate user deviation in task models*. In: *Proc. 1st ADVISES Young Researchers Workshop*. Liege, Belgium.
- [8] Marcello M. Bersani, Matteo Soldo, Claudio Menghi, Patrizio Pelliccione & Matteo Rossi (2020): *PuRSUE -from specification of robotic environments to synthesis of controllers*. *Formal Aspects Comput.* 32(2), pp. 187–227, doi:10.1109/TAC.2011.2176409.
- [9] Matthew L. Bolton (2015): *Model Checking Human-Human Communication Protocols Using Task Models and Miscommunication Generation*. *J. Aerospace Inf. Sys.* 12(7), pp. 476–489, doi:10.1177/1555343413490944.
- [10] Matthew L. Bolton & Ellen J. Bass (2010): *Formally verifying human–automation interaction as part of a system model: limitations and tradeoffs*. *Innovations in Systems and Software Engineering* 6(3), pp. 219–231, doi:10.1007/s11334-010-0129-9.
- [11] Matthew L. Bolton, Ellen J. Bass & Radu I. Siminiceanu (2012): *Generating Phenotypical Erroneous Human Behavior to Evaluate Human-automation Interaction Using Model Checking*. *Int. J. Hum.-Comput. Stud.* 70(11), pp. 888–906, doi:10.1016/j.ijhcs.2012.05.010.
- [12] Matthew L. Bolton, Ellen J. Bass & Radu I. Siminiceanu (2013): *Using Formal Verification to Evaluate Human-Automation Interaction: A Review*. *IEEE Trans. Systems, Man, and Cybernetics: Systems* 43(3), pp. 488–503, doi:10.1109/TSMCA.2012.2210406.
- [13] Matthew L. Bolton, Kylie Molinaro & Adam Houser (2019): *A formal method for assessing the impact of task-based erroneous human behavior on system safety*. *Reliab. Eng. Syst. Saf.* 188, pp. 168–180. Available at <https://doi.org/10.1016/j.ress.2019.03.010>.

- [14] Matthew L. Bolton, Radu I. Siminiceanu & Ellen J. Bass (2011): *A Systematic Approach to Model Checking Human-Automation Interaction Using Task Analytic Models*. *IEEE Trans. Systems, Man, and Cybernetics, Part A* 41(5), pp. 961–976, doi:10.1109/TSMCA.2011.2109709.
- [15] Fiemke Both & Annerieke Heuvelink (2007): *From a formal cognitive task model to an implemented ACT-R model*. In: *Proceedings of the 8th International Conference on Cognitive Modeling (ICCM)*, pp. 199–204.
- [16] Jan Bredereke & Axel Lankenau (2002): *A Rigorous View of Mode Confusion*. In: *Proceedings of the 21st International Conference on Computer Safety, Reliability and Security, SAFECOMP '02*, Springer-Verlag, London, UK, UK, pp. 19–31, doi:10.1016/0167-6423(95)96871-J.
- [17] Jan Bredereke & Axel Lankenau (2005): *Safety-relevant mode confusions modelling and reducing them*. *Reliability Engineering & System Safety* 88(3), pp. 229 – 245, doi:10.1016/j.res.2004.07.020.
- [18] Richard Butterworth, Ann Blandford & David Duke (2000): *Demonstrating the Cognitive Plausibility of Interactive System Specifications*. *Formal Aspects of Computing* 12(4), pp. 237–259, doi:10.1007/s001650070021.
- [19] Antonio Cerone, Peter A. Lindsay & Simon Connelly (2005): *Formal Analysis of Human-computer Interaction using Model-checking*. In Bernhard K. Aichernig & Bernhard Beckert, editors: *Third IEEE International Conference on Software Engineering and Formal Methods (SEFM 2005)*, 7-9 September 2005, Koblenz, Germany, IEEE Computer Society, pp. 352–362, doi:10.1109/SEFM.2005.19. Available at <https://ieeexplore.ieee.org/xpl/conhome/10529/proceeding>.
- [20] Matthew Crosby, Ronald P. A. Petrick, Francesco Rovida & Volker Krüger (2017): *Integrating Mission and Task Planning in an Industrial Robotics Framework*. In Laura Barbulescu, Jeremy Frank, Mausam & Stephen F. Smith, editors: *Proceedings of the Twenty-Seventh International Conference on Automated Planning and Scheduling, ICAPS 2017, Pittsburgh, Pennsylvania, USA, June 18-23, 2017*, AAAI Press, pp. 471–479. Available at <https://aaai.org/ocs/index.php/ICAPS/ICAPS17/paper/view/15715>.
- [21] Paul Curzon & Ann Blandford (2002): *From a Formal User Model to Design Rules*. In: *Interactive Systems. Design, Specification, and Verification, 9th International Workshop, DSV-IS 2002, Rostock Germany, June 12-14, 2002*, pp. 1–15, doi:10.1007/3-540-36235-5.1.
- [22] Paul Curzon & Ann Blandford (2004): *Formally justifying user-centred design rules: a case study on post-completion errors*. In: *International Conference on Integrated Formal Methods*, Springer, pp. 461–480, doi:10.1007/3-540-47884-1_12.
- [23] Paul Curzon, Rimvydas Rukšėnas & Ann Blandford (2007): *An approach to formal verification of human-computer interaction*. *Formal Aspects of Computing* 19(4), pp. 513–550, doi:10.1007/s00165-007-0035-6.
- [24] Lavindra de Silva, Paolo Felli, David Sanderson, Jack C. Chaplin, Brian Logan & Svetan Ratchev (2019): *Synthesising process controllers from formal models of transformable assembly systems*. *Robotics and Computer-Integrated Manufacturing* 58, pp. 130 – 144, doi:10.1016/j.rcim.2019.01.014.
- [25] R. Dillmann, O. Rogalla, M. Ehrenmann, R. Zöliner & M. Bordegoni (2000): *Learning Robot Behaviour and Skills Based on Human Demonstration and Advice: The Machine Learning Paradigm*. In John M. Hollerbach & Daniel E. Koditschek, editors: *Robotics Research*, Springer London, London, pp. 229–238, doi:10.1007/978-1-4471-1555-7.
- [26] E.M. Dougherty & J.R. Fragola (1988): *Human reliability analysis*. New York, NY; John Wiley and Sons Inc.
- [27] DE Embrey (1986): *SHERPA: A systematic human error reduction and prediction approach*. In: *Proceedings of the international topical meeting on advances in human factors in nuclear power systems*.
- [28] F. De Felice, A. Petrillo & F. Zomparelli (2016): *A Hybrid Model for Human Error Probability Analysis*. *IFAC-PapersOnLine* 49(12), pp. 1673 – 1678, doi:10.1016/j.ifacol.2016.07.821. 8th IFAC Conference on Manufacturing Modelling, Management and Control MIM 2016.
- [29] Robert E. Fields (2001): *Analysis of erroneous actions in the design of critical systems*. Ph.D. thesis, University of York.

- [30] Daniel Gall & Thom W. Frühwirth (2014): *A Formal Semantics for the Cognitive Architecture ACT-R*. In: *Logic-Based Program Synthesis and Transformation - 24th International Symposium, LOPSTR 2014, Canterbury, UK, September 9-11, 2014. Revised Selected Papers*, pp. 74–91, doi:10.1007/978-3-319-17822-6_5.
- [31] Wayne D. Gray (2000): *The Nature and Processing of Errors in Interactive Behavior*. *Cognitive Science* 24(2), pp. 205–248, doi:10.1207/s15516709cog2402_2.
- [32] Jérémie Guiochet, Mathilde Machin & Hélène Waeselynck (2017): *Safety-critical advanced robots: A survey*. *Robotics and Autonomous Systems* 94, pp. 43 – 52, doi:10.1016/j.robot.2017.04.004.
- [33] H. Rex Hartson, Antonio C. Siochi & D. Hix (1990): *The UAN: A User-oriented Representation for Direct Manipulation Interface Designs*. *ACM Trans. Inf. Syst.* 8(3), pp. 181–203, doi:10.1145/964967.801131.
- [34] K. P. Hawkins, Nam Vo, S. Bansal & A. F. Bobick (2013): *Probabilistic human action prediction and wait-sensitive planning for responsive human-robot collaboration*. In: *2013 13th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, pp. 499–506, doi:10.1109/HUMANOIDS.2013.7030020.
- [35] Erik Hollnagel (1998): *Cognitive reliability and error analysis method (CREAM)*. Elsevier.
- [36] Andrew Howes & Richard M. Young (1997): *The Role of Cognitive Architecture in Modeling the User: Soar's Learning Mechanism*. *Human-Computer Interaction* 12(4), pp. 311–343, doi:10.1007/BF01099424.
- [37] Dominik Jain, Stefan Waldherr & Michael Beetz (2011): *Bayesian Logic Networks*.
- [38] David J. Jilk, Christian Lebiere, Randall C. O'Reilly & John R. Anderson (2008): *SAL: an explicitly pluralistic cognitive architecture*. *J. Exp. Theor. Artif. Intell.* 20(3), pp. 197–218, doi:10.1007/10719871.
- [39] Been Kim (2015): *Interactive and interpretable machine learning models for human machine collaboration*. Ph.D. thesis, Massachusetts Institute of Technology.
- [40] Namhun Kim, Ling Rothrock, Jaekoo Joo & Richard A. Wysk (2010): *An affordance-based formalism for modeling human-involvement in complex systems for prospective control*. In: *Proceedings of the 2010 Winter Simulation Conference, WSC 2010, Baltimore, Maryland, USA, 5-8 December 2010, IEEE*, pp. 811–823, doi:10.1109/WSC.2010.5679107. Available at <https://ieeexplore.ieee.org/xpl/conhome/5672636/proceeding>.
- [41] Barry Kirwan (1998): *Human error identification techniques for risk assessment of high risk systems-Part 1: review and evaluation of techniques*. *Applied Ergonomics* 29(3), pp. 157 – 177, doi:10.1016/S0003-6870(98)00010-6.
- [42] John E Laird (2012): *The Soar cognitive architecture*. MIT Press, doi:10.7551/mitpress/7688.001.0001.
- [43] Vincent Langenfeld, Bernd Westphal & Andreas Podelski (2019): *On Formal Verification of ACT-R Architectures and Models*. In Ashok K. Goel, Colleen M. Seifert & Christian Freksa, editors: *Proceedings of the 41th Annual Meeting of the Cognitive Science Society, CogSci 2019: Creativity + Cognition + Computation, Montreal, Canada, July 24-27, 2019*, cognitivesciencesociety.org, pp. 618–624. Available at <https://mindmodeling.org/cogsci2019/papers/0124/index.html>.
- [44] Kung-Kiu Lau & Mario Ornaghi (1995): *Towards an Object-Oriented Methodology for Deductive Synthesis of Logic Programs*. In: *Logic Programming Synthesis and Transformation, 5th International Workshop, LOPSTR'95, Utrecht, The Netherlands, September 20-22, 1995, Proceedings*, pp. 152–169, doi:10.1007/3-540-60939-3_11.
- [45] Christian Lebiere, Randall C. O'Reilly, David J. Jilk, Niels Taatgen & John R. Anderson (2008): *The SAL Integrated Cognitive Architecture*. In: *Biologically Inspired Cognitive Architectures, Papers from the 2008 AAAI Fall Symposium, Arlington, Virginia, USA, November 7-9, 2008, AAAI Technical Report FS-08-04, AAAI*, pp. 98–104. Available at <http://www.aaai.org/Library/Symposia/Fall/2008/fs08-04-027.php>.
- [46] B. Li, B. R. Page, B. Moridian & N. Mahmoudian (2020): *Collaborative Mission Planning for Long-Term Operation Considering Energy Limitations*. *IEEE Robotics and Automation Letters* 5(3), pp. 4751–4758, doi:10.1109/LRA.2020.3003881.

- [47] Meng kun Li, Yong kuo Liu, Min jun Peng, Chun li Xie & Li qun Yang (2016): *The decision making method of task arrangement based on analytic hierarchy process for nuclear safety in radiation field*. *Progress in Nuclear Energy* 93, pp. 318 – 326.
- [48] Matt Luckcuck, Marie Farrell, Louise A. Dennis, Clare Dixon & Michael Fisher (2019): *Formal Specification and Verification of Autonomous Robotic Systems: A Survey*. *ACM Comput. Surv.* 52(5), pp. 100:1–100:41, doi:10.1007/s10458-010-9146-1.
- [49] Matt Luckcuck, Marie Farrell, Louise A. Dennis, Clare Dixon & Michael Fisher (2019): *A Summary of Formal Specification and Verification of Autonomous Robotic Systems*. In Wolfgang Ahrendt & Silvia Lizeth Tapia Tarifa, editors: *Integrated Formal Methods - 15th International Conference, IFM, Lecture Notes in Computer Science* 11918, Springer, pp. 538–541, doi:10.1145/1592434.1592436.
- [50] Andrea Mannini & Angelo Maria Sabatini (2010): *Machine Learning Methods for Classifying Human Physical Activity from On-Body Accelerometers*. *Sensors* 10(2), pp. 1154–1175, doi:10.3390/s100201154.
- [51] Célia Martinie, Philippe A. Palanque, Racim Fahssi, Jean-Paul Blanquart, Camille Fayollas & Christel Seguin (2016): *Task Model-Based Systematic Analysis of Both System Failures and Human Errors*. *IEEE Trans. Human-Machine Systems* 46(2), pp. 243–254, doi:10.1109/THMS.2014.2365956.
- [52] Björn Matthias (2015): *Risk Assessment for Human-Robot Collaborative Applications*. In: *Workshop IROS - Physical Human-Robot Collaboration: Safety, Control, Learning and Applications*.
- [53] Claudio Menghi, Sergio Garcia, Patrizio Pelliccione & Jana Tumova (2018): *Multi-robot LTL Planning Under Uncertainty*. In Klaus Havelund, Jan Peleska, Bill Roscoe & Erik de Vink, editors: *Formal Methods*, Springer International Publishing, Cham, pp. 399–417, doi:10.1177/0278364915595278.
- [54] Christine M. Mitchell & R. A. Miller (1986): *A Discrete Control Model of Operator Function: A Methodology for Information Display Design*. *IEEE Trans. Systems, Man, and Cybernetics* 16(3), pp. 343–357, doi:10.1109/TSMC.1986.4308966.
- [55] Alvaro Miyazawa, Pedro Ribeiro, Wei Li, Ana Cavalcanti, Jon Timmis & Jim Woodcock (2019): *RoboChart: modelling and verification of the functional behaviour of robotic applications*. *Software and Systems Modeling* 18(5), pp. 3097–3149, doi:10.1145/197320.197322.
- [56] Victor Moher, Thomas G. and Dirda (1995): *Revising mental models to accommodate expectation failures in human-computer dialogues*, pp. 76–92. Springer Vienna, Vienna.
- [57] A. Mosleh & Y.H. Chang (2004): *Model-based human reliability analysis: prospects and requirements*. *Reliability Engineering & System Safety* 83(2), pp. 241 – 253, doi:10.1016/j.res.2003.09.014. Human Reliability Analysis: Data Issues and Errors of Commission.
- [58] tum technische universitat munchen (2012): *The TUM Kitchen Data Set*. Available at <https://ias.in.tum.de/dokuwiki/software/kitchen-activity-data>.
- [59] Dan Pan & Matthew L. Bolton (2016): *Properties for formally assessing the performance level of human-human collaborative procedures with miscommunications and erroneous human behavior*. *International Journal of Industrial Ergonomics*, pp. –.
- [60] Fabio Paternò, Cristiano Mancini & Silvia Meniconi (1997): *ConcurTaskTrees: A Diagrammatic Notation for Specifying Task Models*. In Steve Howard, Judy Hammond & Gitte Lindgaard, editors: *Human-Computer Interaction, INTERACT '97, IFIP TC13 Interantional Conference on Human-Computer Interaction, 14th-18th July 1997, Sydney, Australia, IFIP Conference Proceedings* 96, Chapman & Hall, pp. 362–369.
- [61] Fabio Paternò & Carmen Santoro (2002): *Preventing user errors by systematic analysis of deviations from the system task model*. *International Journal of Human-Computer Studies* 56(2), pp. 225–245, doi:10.1006/ijhc.2001.0523.
- [62] Steven Pocock, Michael D. Harrison, Peter C. Wright & Paul Johnson (2001): *THEA: A Technique for Human Error Assessment Early in Design*. In Michitaka Hirose, editor: *Human-Computer Interaction INTERACT '01: IFIP TC13 International Conference on Human-Computer Interaction, Tokyo, Japan, July 9-13, 2001*, IOS Press, pp. 247–254.

- [63] Gayathri R. & V. Uma (2018): *Ontology based knowledge representation technique, domain modeling languages and planners for robotic path planning: A survey*. *ICT Express* 4(2), pp. 69 – 74, doi:10.1016/j.ict.2018.04.008. SI on Artificial Intelligence and Machine Learning.
- [64] James Reason (1990): *Human error*. Cambridge university press, doi:10.1017/CBO9781139062367.
- [65] James Reason (2000): *Human error: models and management*. *BMJ: British Medical Journal* 320(7237), p. 768, doi:10.1136/bmj.320.7237.768.
- [66] Pedro Ribeiro, Alvaro Miyazawa, Wei Li, Ana Cavalcanti & Jon Timmis (2017): *Modelling and Verification of Timed Robotic Controllers*. In Nadia Polikarpova & Steve Schneider, editors: *Integrated Formal Methods*, Springer International Publishing, Cham, pp. 18–33, doi:10.1007/BFb0020949.
- [67] FRANK E. RITTER & RICHARD M. YOUNG (2001): *Embodied models as simulated users: introduction to this special issue on using cognitive models to improve interface design*. *International Journal of Human-Computer Studies* 55(1), pp. 1 – 14, doi:10.1006/ijhc.2001.0471.
- [68] Rimvydas Ruksenas, Jonathan Back, Paul Curzon & Ann Blandford (2009): *Verification-guided modelling of salience and cognitive load*. *Formal Asp. Comput.* 21(6), pp. 541–569, doi:10.1007/s00165-008-0102-7.
- [69] Dario D. Salvucci & Frank J. Lee (2003): *Simple cognitive modeling in a complex cognitive architecture*. In Gilbert Cockton & Panu Korhonen, editors: *Proceedings of the 2003 Conference on Human Factors in Computing Systems, CHI 2003, Ft. Lauderdale, Florida, USA, April 5-10, 2003*, ACM, pp. 265–272, doi:10.1145/642611.642658.
- [70] Dongmin Shin, Richard A. Wysk & Ling Rothrock (2006): *Formal model of human material-handling tasks for control of manufacturing systems*. *IEEE Trans. Systems, Man, and Cybernetics, Part A* 36(4), pp. 685–696, doi:10.1109/TSMCA.2005.853490.
- [71] Maarten Sierhuis et al. (2001): *Modeling and simulating work practice: BRAHMS: a multiagent modeling and simulation language for work system analysis and design*.
- [72] Richard Stocker, Louise Dennis, Clare Dixon & Michael Fisher (2012): *Verifying Brahms Human-Robot Teamwork Models*. In: *Logics in Artificial Intelligence: 13th European Conference, JELIA*, doi:10.1023/A:1022920129859.
- [73] Alan D Swain & Henry E Guttmann (1983): *Handbook of human-reliability analysis with emphasis on nuclear power plant applications. Final report*. Technical Report, Sandia National Labs., Albuquerque, NM (USA).
- [74] Bo Tang, Chao Jiang, Haibo He & Yi Guo (2016): *Probabilistic human mobility model in indoor environment*, pp. 1601–1608. doi:10.1109/IJCNN.2016.7727389. Available at <https://ieeexplore.ieee.org/xpl/conhome/7593175/proceeding>.
- [75] Moritz Tenorth, Fernando De la Torre & Michael Beetz (2013): *Learning probability distributions over partially-ordered human everyday activities*. In: *2013 IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, May 6-10, 2013*, IEEE, pp. 4539–4544, doi:10.1109/ICRA.2013.6631222. Available at <https://ieeexplore.ieee.org/xpl/conhome/6615630/proceeding>.
- [76] Fernando de la Torre, Jessica K. Hodgins, Javier Montano & Sergio Valcarcel (2009): *Detailed Human Data Acquisition of Kitchen Activities: the CMU-Multimodal Activity Database (CMU-MMAC)*. In: *CHI 2009 Workshop. Developing Shared Home Behavior Datasets to Advance HCI and Ubiquitous Computing Research*.
- [77] Federico Vicentini, Mehrnoosh Askarpour, Matteo Rossi & Dino Mandrioli (2020): *Safety Assessment of Collaborative Robotics Through Automated Formal Verification*. *IEEE Trans. Robotics* 36(1), pp. 42–61, doi:10.1109/TRO.2019.2937471.
- [78] M.L. Visinsky, J.R. Cavallaro & I.D. Walker (1994): *Robotic fault detection and fault tolerance: A survey*. *Reliability Engineering & System Safety* 46(2), pp. 139 – 158, doi:10.1016/0951-8320(94)90132-5.
- [79] Bernd Werther & Eckehard Schnieder (2005): *Formal Cognitive Resource Model: Modeling of human behavior in complex work environments*. In: *2005 International Conference on Computational Intelligence for Modelling Control and Automation (CIMCA 2005), International Conference on Intelligent Agents,*

Web Technologies and Internet Commerce (IAWTIC 2005), 28-30 November 2005, Vienna, Austria, IEEE Computer Society, pp. 606–611, doi:10.1109/CIMCA.2005.1631535. Available at <https://ieeexplore.ieee.org/xpl/conhome/10869/proceeding>.

- [80] J. C. Williams (1988): *A data-based method for assessing and reducing human error to improve operational performance*. In: *Conference Record for 1988 IEEE Fourth Conference on Human Factors and Power Plants*, pp. 436–450, doi:10.1109/HFPP.1988.27540.