Special Issue

Informing Autonomous System Design Through the Lens of Skill-, Rule-, and Knowledge-Based Behaviors

Mary (Missy) Cummings, Duke University

A framework is presented that relates skill-, rule-, and knowledge-based reasoning to expertise and uncertainty. This taxonomy is designed to help people from various technical backgrounds conceptualize functional allocation for autonomous systems that interact with human decision makers in order to better understand potential design implications.

Keywords: function allocation, human–system integration, human–robot interaction

In reflecting on the "levels of automation" (LOA) debate that Kaber (2017 [this issue]) addresses, I agree with his assertion that as a community, we are generally asking the right questions but, more importantly, we need better strategies that provide tangible design guidance for a much broader class of engineers than just those with a human factors background. With the explosion of autonomous technologies in the form of self-driving cars, drones for both hobbyists and commercial operators, and advanced medical technologies, more than ever before, engineers and computer scientists need a principled way to think about designing systems to promote human–automation and human–autonomy interactions. As others have noted (Woods, 2016), risks of autonomous systems are often downplayed, and there is a need to better communicate where and how uncertainty drives such risk.

Address correspondence to Mary (Missy) Cummings, Duke University, Box 90300 144 Hudson Hall, Durham, NC, USA, m.cummings@duke.edu.

Journal of Cognitive Engineering and Decision Making
201X, Volume XX, Number X, Month 2017, pp. 1–4
DOI: 10.1177/1555343417736461
Copyright © 2017, Human Factors and Ergonomics Society.

I, too, have struggled with the inexact nature and coarseness of the traditional LOA approach but also understand its appeal in terms of simplicity and categorical nature. However, my students, who are primarily engineers and computer scientists and are going to take only one class related to human factors, almost always immediately pick up on the fact that although helpful for perhaps describing a system at a high level, the use of LOAs, or any derivative model, leaves them frustrated with little understand of *how* a system should be designed.

Because it is not sufficient to consider what information-processing stage could be automated and the degree of that automation (the two dimensions highlighted by traditional LOA taxonomies; Endsley & Kaber, 1999; Parasuraman, Sheridan, & Wickens, 2000), I developed an alternative model to help them better understand what functions should or could be distributed either exclusively across or collaboratively between humans and autonomous agents. I call this model the SRKE (skill, rule, knowledge, expertise) model. This model serves as an indicator of how challenging it likely will be to develop an automated system, either fully or partially, that can reliably perform even when uncertainty is at its greatest.

Figure 1 represents this model that links information-processing behaviors and cognition to increasingly complex tasks, which is based on the SRK (skills-, rules-, and knowledge-based behavior) taxonomy (Rasmussen, 1983). It should be noted that the SRKs do not occur in discrete stages with clear thresholds but rather are on a continuum.

For Rasumussen (1983), skill-based behaviors (SBBs) are sensory-motor actions that are highly automatic, typically acquired after some period of training. Skills are characterized as

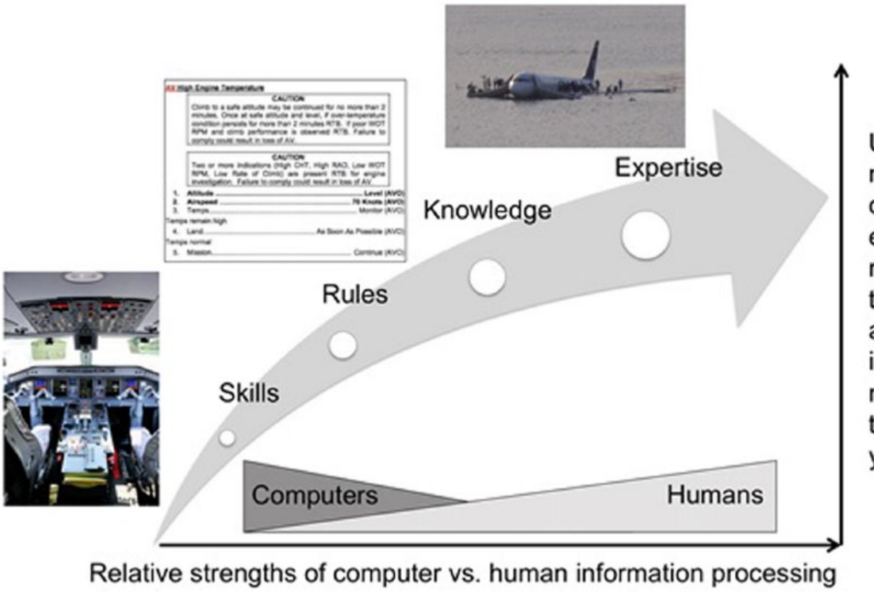

Figure 1. Role allocation for information-processing behaviors (skill, rule, knowledge, expertise) and the relationship to uncertainty (Cummings, 2014).

motor outputs that are "a response to the observation of an error signal representing the difference between the actual state and the intended state in a time-space environment" (Rasmussen, 1983, p. 259). Controls engineers design controls algorithms to minimize error signals in a very similar fashion. Piloting an aircraft is an example of skill-based control for humans as well as for automation. Once this set of skills is acquired, which often takes significant time, pilots can then turn their attention to higher cognitive tasks.

Assuming SBBs are mastered, humans can turn their attention to rule-based behaviors, which are actions guided by subroutines, stored rules, or procedures, often likened to following a cookbook recipe (Rasmussen, 1983). Humans can experience problems in the application of rule-based reasoning when they fail to recognize the correct goal, causing them to select an incorrect procedure or set of rules. Recognizing which procedure to follow is not always obvious, as in systems where one aural alert can indicate different alert or failure modes.

The highest level of cognitive reasoning in the SRK framework is that of knowledge-based behaviors, where mental models that are built over time aid in the formulation and selection of plans for an explicit goal (Rasmussen, 1983). The "Miracle on the Hudson," the 2009 landing of US Airways 1549 in the Hudson River in New York, as depicted in Figure 1, is an example of a knowledge-based behavior. Faced with a complete loss of thrust, the captain had to decide whether to ditch the aircraft or attempt to land at a nearby airport. Given his mental model, observations, and the significant uncertainty in the environment, his mental simulation led to his choosing the ditching option.

Although this example demonstrates the power of human reasoning under maximum uncertainty, this same accident highlights the importance of the need for a collaborative approach between humans and autonomous systems. When a complete engine failure occurs in an Airbus 320, as in the previous example, the fly-by-wire system automatically trims the plane, computes the ideal glide speed, and readjusts the landing pitch position far more rapidly and reliably than a human. Pressing the *Ditching* button seals the aircraft for water entry. This mutually supportive flight control environment was critical to the successful outcome of this potentially catastrophic event.

One of my additions to the SRK taxonomy is that of expertise (leading to the SRKE model) to demonstrate that knowledge-based behaviors are a prerequisite for gaining expertise. However, expertise cannot be achieved without significant experience in the presence of high uncertainty. So although a person can be knowledgeable about a task through repetition, they become true experts when they must exercise their knowledge under vastly different conditions in the presence of high uncertainty. For example, one pilot who has flown thousands of hours with no system failures is not as much of an expert as one who has responded to many system failures over the same time period. Moreover, judgment and intuition, which often make traditional engineers uncomfortable because these concepts lack a mathematical formal representation, are the key behaviors that allow experts to quickly assess a situation in a fast and frugal method (Gigerenzer, Todd, & ABC Research Group, 1999) without necessarily and laboriously comparing all possible plan outcomes.

The remaining key addition to Rasmussen's (1983) SRK taxonomy is my representation of uncertainty via the *y*-axis. Uncertainty is not just "unanticipated variability," which has been suggested in previous literature as a key attribute of a resilient system (Borst, Flach, & Ellerbroek, 2015; Roth, Bennett, & Woods, 1987). Uncertainty, which can have very clear mathematical representations (and thus is a term engineers feel comfortable with), occurs when a situation cannot precisely be determined, often due to missing or imperfect information with potentially many unknown (i.e., unanticipated) variables. However, both external and internal sources of uncertainty need to be considered, including missing or erroneous sensor readings, low-probability events, and human performance variability but also the use of probabilistic reasoning algorithms that can be biased in ways not obvious to their human designers (Briscoe & Feldman, 2011). The use of data-driven machine-learning algorithms, for example, to model a process can introduce additional uncertainty into the system if they or the underlying data are biased, because the resulting statistical models do not reflect reality.

For complex systems with embedded automation and autonomy, uncertainty can arise from exogenous sources, such as the environment, for example, late-afternoon low lighting conditions that cause sensor washout. However, uncertainty can also be introduced from endogenous sources—either from human behaviors, like distracted driving, or from computer/automation behaviors, like an erroneous algorithm in the Google car that caused it to hit a city bus (Shepardson, 2016).

As illustrated in Figure 1, characterizing uncertainty is the key to function allocation and provides a road map for when and why a human versus a computer could or should be part of the system. Thus understanding where the greatest uncertainty lies, both internally and externally, as well as understanding the capabilities of various sensors that are often found in autonomous systems, like GPS and lidar, provides engineers and computer scientists with a starting point for function allocation in design.

The bottom of Figure 1 indicates, generally, where technology is today in terms of automation or autonomy being able to deal with increasing uncertainty. If sensors are reliable and function across a myriad of conditions, and the uncertainty is low (much like an automated braking system), this function could and should be automated. Rule-based reasoning in moderately uncertain conditions likewise can be automated, as is seen in the relative ease that Teslas and other similar cars have in being able to track lanes and pass cars automatically on highways. However, until self-driving cars can master uncertainty in all conditions that require knowledge-based reasoning to at least the same degree as humans, we will have only partially capable, but potentially very dangerous, systems that cannot cope with uncertainty.

Another important point to note is that although Figure 1 may seem to represent that skill-based tasks are cognitively easier than knowledge-based tasks, for autonomous systems, the difficulty of these cognitive behaviors for automation will hinge directly upon the capability of the sensors that drive skill-based behaviors. For example, perception systems for driverless cars today are still immature, which ultimately influences rule- and knowledge-based reasoning. Many of these cars rely on lidar systems, which have known issues with moisture in the air. So if these cars cannot build an accurate world model in inclement weather due to flawed sensors, then they cannot execute correct higher cognitive behaviors.

Although the SRKE taxonomy can be used to describe a system, as I have done here, it is meant to provide guidance for those variables that need to be considered when designing a system for some degree of human–autonomy/automation interaction. It is meant to highlight those areas where sensors may not be up to the task of coping with uncertainty and understanding how, when, and why humans should be used to augment such systems. Conversely, this model can also be used to understand when automation may be able to help humans, particularly as tasks grow in complexity. Combing quickly through large data sets and seeing patterns in data not obvious to humans are strengths of machine learning, used in applications like predicting disease from electronic health care records (Miotto, Li, Kidd, & Dudley, 2016). Moreover, humans often struggle in decision making involving estimating likelihoods of uncertain events (Tversky & Kahneman, 1974), so the ideal system is one that is mutually supportive between autonomous systems and humans, particularly in knowledge-based reasoning tasks.

The SRKE framework in Figure 1, in conjunction with understanding what capabilities exist for various sensors and autonomous systems, can aid engineers and designers in determining both a feasible and a desirable system (which are not always the same). In its current form, the SRKE taxonomy is not a model in any mathematical sense. However, work is under way to determine how such a representation could be formulated and whether more specific quantitative design guidance could result. The goal is to help engineers and computer scientists recognize a potential mismatch between a proposed system design and that system's ability to exhibit the required information-processing behaviors at the various SRKE levels. Doing so is a step toward providing the tangible design guidance that Kaber recognizes is needed.

REFERENCES

Borst, C., Flach, J., & Ellerbroek, J. (2015). Beyond ecological interface design: Lessons from concerns and misconceptions. *IEEE Transactions on Human–Machine Systems*, *45*, 164–175.

Briscoe, E., & Feldman, J. (2011). Conceptual complexity and the bias/variance tradeoff. *Cognition*, *118*, 2–16.

Cummings, M. L. (2014). Man vs. machine or man + machine? *IEEE Intelligent Systems*, *29*(5), 62–69.

Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, *42*, 462–492.

Gigerenzer, G., Todd, P. M., & ABC Research Group (1999). *Simple heuristics that make us smart*. Oxford, UK: Oxford University Press.

Kaber, D. (2017). Issues in human–automation interaction modeling: Presumptive aspects of frameworks of types and levels of automation. *Journal of Cognitive Engineering and Decision Making, XX*, XX–XX.

Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. *Scientific Reports*, *6*(26094).

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans*, *30*, 286–297.

Rasmussen, J. (1983). Skills, rules, and knowledge: Signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, *13*, 257–266.

Roth, E. M., Bennett, K. B., & Woods, D. D. (1987). Human interaction with an "intelligent" machine. *International Journal of Man–Machine Studies*, *27*, 479–526.

Shepardson, D. (2016, February 29). Google says it bears "some responsibility" after self-driving car hit bus. *Reuters*. Retrieved from http://www.reuters.com/article/us-google-selfdrivingcar-idUSKCN0W22DG

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

Woods, D. D. (2016). The risks of autonomy: Doyle's catch. *Journal of Cognitive Engineering and Decision Making*, *10*, 131–133.

Mary (Missy) Cummings received her PhD in Systems Engineering from the University of Virginia in 2004. She is currently a professor in the Duke University Department of Mechanical Engineering and Materials Science, the Duke Institute of Brain Sciences, and the Duke Electrical and Computer Engineering Department. She is the director of the Duke Humans and Autonomy Laboratory and Duke Robotics.