# An Integrated Model of Macrocognitive Work and Trust in Automation

**Robert R. Hoffman**
Institute for Human and Machine Cognition
rhoffman@ihmc.us

### Abstract

Conceptual models of trust in automation range from lists of factors that are believed or known to have a causal influence trust, to conceptual diagrams that express a hypothetical causal relation between trust and the operator's prediction of the automation's behavior. The unified theory presented here is more complete. Trusting is modeled as a phenomenon that emerges from multiple interacting and parallel processes. The integrated model is a reasonable starting point for computational modeling that would capture the richness of expert cognition and re-planning, and simultaneously allow for the modeling of reliance and trust in the automation that mediates the cognitive work.

## Introduction

The focus of this paper is on macrocognition, that is, the ways in which cognition adapts to complexity in sociotechnical work systems (Cacciabue and Hollnagel, 1995; Klein et al., 2003). This includes modeling the reasoning of domain experts, an enterprise that was stimulated initially by the advent of expert systems, entailing studies of domains spanning a great variety from weather forecasting to industrial process control to emergency management (Schragen, et al., 2008). The subject matter for macrocognitive studies is processes such as sensemaking, problem recognition, mental projection to the future, collaboration, identification of leverage points, adaptation, and maintaining common ground. This subject matter distinguishes macrocognitive research from "microcognitive" research (e.g., the laboratory study of millisecond shifts of attention, the retrieval of items from short-term memory). The study of cognition at both scales, and using diverse methodologies, is obviously necessary for a complete cognitive psychology. This paper concerns itself with two primary macrocognitive theories. One is a theory of sensemaking and the other a theory of replanning.

During the era of expert systems, psychologists started to study whether people tend to believe that a computer's outputs must always be correct. More recently, technology has increased in complexity, depending on higher levels of knowledge and skill on the part of operators. Technologies trigger automation surprises and spawned new forms of error (Koopman and Hoffman, 2003; Woods. and Hollnagel, 2006). People often mistrust their computer technology.

A number of conceptual and computational models of trust in automation have been proposed, having the goal of capturing the variables that might influence the dynamics and development of trust in automation. Some models are lists of causal factors and dimensions, some are mathematical models (e.g., Khasawneh, et al., 2003; Lee and See, 2004). Models typically regard trust as a state; the goal is to predict or estimate values or levels of trust, or of trust "calibration."

These classes of models, models of cognitive work and models of trust, have yet to be integrated. The integration presented here is intended to capture: (1) The reasoning sequences in which problems are solved and people make sense of situations as a part of their cognitive work, and (2) The operator's reasoning about the trustworthiness and reliability of the technologies that are mediating the work. I rely on the Data/Frame model of sensemaking and the Flexecution model of replanning to forge this integrated model. I show how the integrated model can be used as a new method for analyzing results from cognitive task analysis. In addition, the unified model is suggestive of architectures for computational emulations.

## Sensemaking

The Data/Frame model of sensemaking (Klein, Moon and Hoffman, 2006a,b; 2007) is presented in Figure 1 (below). In this model, "frames" can be understood as conceptual approximations to more formal notions of frames (e.g., Minsky and Papert, 1974), although no commitment is made here with regard to any particular formalism. A detailed discussion of how this model emerged from studies of the reasoning of domain experts is presented in Klein et al., (2003).

The Data/Frame model covers a host of models, such as Karl Duncker's classic model of hypothesis testing (Duncker, 1945) and models of expert reasoning developed in the era of expert systems (e.g., Einhorn and Hogarth, 1981). (For more examples see Hoffman and Militello, 2008). If one looks at those sorts of models one sees various loops (e.g., testing an hypothesis by searching for additional data) but the models are basically all causal chain, input-output models. A premise of the macrocognitive approach is that the functions of macrocognition are all parallel, essentially continuous, and interacting. Sensemaking often proceeds in fits and starts; there can be gaps, distractions, and multi-taskings. Beginnings and endings can be anything but clear-cut. While there may be instances in which reasoning seems to follow a sequence with a clear-cut starting point ("surprise" is often the trigger to problem solving) and an apparent stopping point (a decision is "made"), such causal chains are the exception and not the rule (see Hoffman and Yates, 2005). Thus, there are no input and output arrows in Figure 1; it is all closed loops. The sensemaking process can commence anywhere, though as I say, it is often triggered by surprise. But even in that case, some sort of frame must have "been in mind" earlier, otherwise there would be no anomaly to notice.

## The Flexecution Model of Replanning

The assumed nominal case is that a plan is made de novo and then followed lock-step, but this is widely understood to be the exception to the rule. The more typical case is continual re-planning (Klein, 2007a,b), as has often been expressed by military leaders such as Publius Syrus (~100 BC) who commented, *It is a bad plan that admits of no modification*. But more than this, the flexecution model recognizes that in complex reasoning and problem solving, people clarify and adapt their goals while pursuing them. This distinguishes the Flexecution model from other models of dynamic or adaptive planning:

> Planning-as-problem-solving generally relies on a set of well-understood representations and mechanisms, such as planning graphs, task networks, search in a state space, chaining, type hierarchies, and constraints on variables (Allen, Hendler and Tate, 1009: Bacchus, 2001; Weld, 1999). These mechanisms and approaches seem to run into difficulty when faced with the intractability of large, messy, real-world problems. For such planning problems, the state spaces aren't completely predefined, and planning requires human decision making based on knowledge and sensitivity to context (Klein, 2007a, p.79).

Within AI, a main planning strategy is to find provably correct sequences of actions that will accomplish the (well-specified) goal (Allen, Hendler, and Tate, 2000). Kambhampati (2007) has noted that this work assumes a complete domain model that is specified in advance, including goal utilities, an assumption that may sometimes be warranted but which is too restrictive for many situations.

The Flexecution model is depicted in Figure 2 (below). The Flexecution model does not assume that there are "start" and "stop" points. Studies of how people reason about complex indeterminate causation (Hoffman, Klein and Miller, 2011) have shown that there are often no clear-cut starting or halting points in reasoning about complex causation. The closed loop at the top is the counterpart to the topmost closed loop in the D/F model. Likewise, the other loops in the Flexecution model are counterparts to those in the D/F model. The two conceptual models are cut of the same cloth, one describing how people make sense of complex situations, and the other describing how people act on the basis of their understanding. But there is more.

## Working the Technology

Macrocognitive work involves observing and acting upon the world. These activities are mediated, in whole or in part, by the technology (sensors, computational systems, displays, etc.). When conducting macrocognitive work people have to devote time and effort to making sense of their technology as well as making sense of the observed or controlled world. A challenging aspect of trust in automation is that activities supporting the primary tasks are necessarily understood and managed *through* the mediating artifact—that is, the sensors, displays, computers, and so forth. Figure 3 (below) presents the most straightforward combination of the Data/Frame and Flexecution models specifically tailored to express the dynamics of trust in and reliance on automation.

## The Unified Model

Issues of *trust* reside in the interplay of sensemaking the world and sensemaking the technology, whereas issues of *reliance* reside in the interplay of flexecuting the work and flexecuting the technology. Sensemaking of the observed/controlled world depends on sensemaking of the technology. Flexecuting one's actions on the observed/controlled world depend on the ability to flexecute the technology. The full diagram showing the four models looped together is rather large (Figure 4, below). If we know anything about complex systems, we know that this complexity cannot be "reduced" or simplified (Feltovich, Hoffman and Woods, 2004; Hoffman, Norman and Vagners, 2009; Hoffman and Woods, 2011).

## Putting the Integrated Model to Work

The integrated model has a real use: It enables us to chart paths in a cognitive task analysis, tracing the worker's attention as it moves from activity to activity in the cognitive work. And along with it, both trust and reliance can morph. There are many methods of task analysis, some dating to the early 1900s, that trace reasoning processes in order to generate task or hierarchical goal decompositions (see Crandall, Klein and Hoffman, 2006; Endsley, Bolte and Jones, 2000; Rasmussen, Pejtersen, and Goodsetin, 1994; Shepherd, 2001). Methods involve observing and measuring performance at work, and asking workers questions about what they are doing and why. The results are used to describe reasoning sequences in terms of cognitive categories of states, procedures, and goals (e.g., "observe," "decide," "act", etc.). These are concatenated into hierarchies of tasks and sub-tasks and their associated goals. (For a history of task analysis, see Hoffman and Militello, 2008.) Cognitive task analysis is widely used to inform the creation of intelligent systems, work methods, web sites, and interfaces (Diaper and Stanton, 2004; Hollnagel, 2003; Jenkins, et al., 2009).

None of the process tracing methods has language that unifies the analysis of macrocognitive work, the analysis of trust in automation, and the analysis of reliance on automation. Using the integrated model, protocols can indeed be coded in terms of activities such as "questioning the frame for making sense of the observed world while flexecuting the interactions with the technology." A reanalysis of protocols of the reasoning of expert weather forecasters, originally collected by Hoffman, et al. (2006), has revealed interesting sequences and patterns, such as the tendency for preservation of a frame to be preceded by questioning the frame, as the model predicts. There are instances where action plans underwent major revision related to justified mistrust in the technology. There were many cases of sensemaking of the artifacts, especially understanding the weather radar. It might be assumed that if attention has to shift away from the quadrant of sensemaking the world—that is, attention shifts away from the primary task goals and has to focus instead on making sense of the technology—that the cognitive work would suffer due to distraction and increased mental workload. The re-analysis of the weather forecasting case suggests that this assumption may be incorrect. The forecaster's awareness of the capabilities and limitations of the technology, and methods for coping with such things as limited or sparse data, seem to mesh seamlessly as sensemaking of the observed world progresses. This hypothesis deserves further study.

The unification of the Data/Frame model of sensemaking and the Flexecution model of replanning permits the analysis of macrocognitive work and simultaneously the analysis of the dynamics of trusting. A conceptual architecture of this sort is perhaps a necessary foundation for attempts at computational modeling at the work system level.

## Acknowledgement

Development of the model presented in this article was with the support of the Office of Naval Research via Adaptive Cognitive Systems, LLC under contract No. N00014-12-M-0398, and with the support of the Intelligence Advanced Research Projects Activity (IARPA) via the Air Force Research Laboratory. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, AFRL, or the U.S. Government.

## References

Allen, J., Hendler, J., and Tate, A. (Eds.) (1990). *Readings in planning*. San Francisco: Morgan Kaufman.

Cacciabue, P.C., and Hollnagel, E. (1995). Simulation of cognition: Applications. In J.M. Hoc, P.C. Cacciabue, and E. Hollnagel (Eds.), *Expertise and technology: Cognition and human-computer cooperation*, (pp. 55–73.). Mahwah, NJ: Lawrence Erlbaum.

Crandall, B., Klein, G., and Hoffman R.R. (2006). *Working Minds: A Practitioner's Guide to Cognitive Task Analysis*. Cambridge, MA: MIT Press.

Diaper, D. and Stanton, N.A. (Eds.) (2004). *The handbook of task analysis for human-computer interaction*. Mahwah, NJ: Erlbaum.

Duncker, K. (1945). On problem solving. *Psychological Monographs*, *58*, 1-113 (Whole No. 270) (L. S. Lees, Trans.).

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology*, *32*, 53-88.

Endsley, M.R., Bolte, B., and Jones, D.G. (2003). *Designing for situation awareness: An approach to user-centered design.* London: Taylor and Francis.

Feltovich, P. J., Hoffman, R. R., and Woods, D. (May/June 2004). Keeping it too simple: How the reductive tendency affects cognitive engineering. *IEEE Intelligent Systems*, pp. 90-95.

Hoffman, R. R., Coffey, J. W., Ford, K. M. and Novak, J. D. (2006). A method for eliciting, preserving, and sharing the knowledge of forecasters. *Weather and Forecasting, 21*, 416–428.

Hoffman, R.R., Klein, G., and Miller, J. E. (2011). Naturalistic investigations and models of reasoning about complex indeterminate causation. *Information and Knowledge Systems Management, 10*, 397-425.

Hoffman, R.R. and Militello, L.G. (2008). *Perspectives on Cognitive Task Analysis*: *Historical Origins and Modern Communities of Practice*. Boca Raton, FL: CRC Press/Taylor and Francis.

Hoffman, R. R., Norman, D. O. and Vagners, J. (2009, May/June). Complex Sociotechnical Joint Cognitive Work Systems? *IEEE: Intelligent Systems*, pp. 82-89.

Hoffman, R.R. and Woods, D.D. (2011, November/December). Beyond Simon's slice: Five fundamental tradeoffs that bound the performance of macrocognitive work systems. *IEEE: Intelligent Systems*, pp. 67-71.

Hoffman, R. R. and Yates, J. F. (July/August, 2005). Decision(?)Making(?). *IEEE: Intelligent Systems*, pp. 22-29.

Hollnagel, E. (Ed.) (2003). *Handbook of cognitive task design*. Mahwah NJ: Erlbaum.

Jenkins, D.P., Stanton N.A. Salmon, P.M., and Walker, G.H. (2009). *Cognitive work analysis: coping with complexity*. Surrey, England: Ashgate.

Kambhampati, S. (2007). Model-lite planning for the web age masses: The challenges of planning with incomplete and evolving domain models. In *AAAI 2007 Senior Member Track.* Menlo Park, CA: AAAI.

Khasawneh1, M.T., Bowling, S.R. Jiang, X., Gramopadhye, A.K. and Melloy, B.J. (2003). A model for predicting human trust in automated system. In *Proceedings of the 8th Annual International Conference on Industrial Engineering – Theory, Applications and Practic*e (p. 216-222). Sponsored by the International Journal of Industrial Engineering [http://ijietap.org].

Klein, G. (2007a). Flexecution as a paradigm for replanning, Part 1: *IEEE Intelligent Systems, 22*, 79-83.

Klein, G. (2007b). Flexecution, Part 2: Understanding and supporting flexible execution: *IEEE Intelligent Systems, 22*, 108-112.

Klein, G., Moon, B. and Hoffman, R. R. (2006, November/December). Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems*, pp. 88-92.

Klein, G., Moon, B., and Hoffman, R. R. (2006, July/August). Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems*, pp. 22-26.

Klein, G., Ross, K. G., Moon, B. M., Klein, D. E., Hoffman, R. R., and Hollnagel, E. (May/June, 2003). Macrocognition. *IEEE: Intelligent Systems*, pp. 81-85.

Koopman, P. and Hoffman, R. R., (November/December 2003). Work-Arounds, Make-Work, and Kludges. *IEEE: Intelligent Systems*, pp. 70-75.

Lee, J.D. and See, K.A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46,* 50–80.

Minsky, M., and Papert, S. (1974). *Artificial intelligence.* Eugene, OR: Oregon State System of Higher Education.

Rasmussen, J., Pejtersen, A.M., and Goodsetin, L.P. (Eds.) (1994). *Cognitive systems engineering*. New York: John Wiley.

Schraagen, J.M., Militello, L.T. Ormerod, T. and Lipshitz, R. (Eds.) (2008), *Naturalistic decision making and macrocognition*. Aldershot, England: Ashgate.

Shepherd, A. (2001). *Hierarchical task analysis*. London: Taylor & Francis.

Woods, D. D. and Hollnagel, E. (2006). *Joint cognitive systems: Patterns in cognitive systems engineering*. Boca Raton, FL: CRC Press.

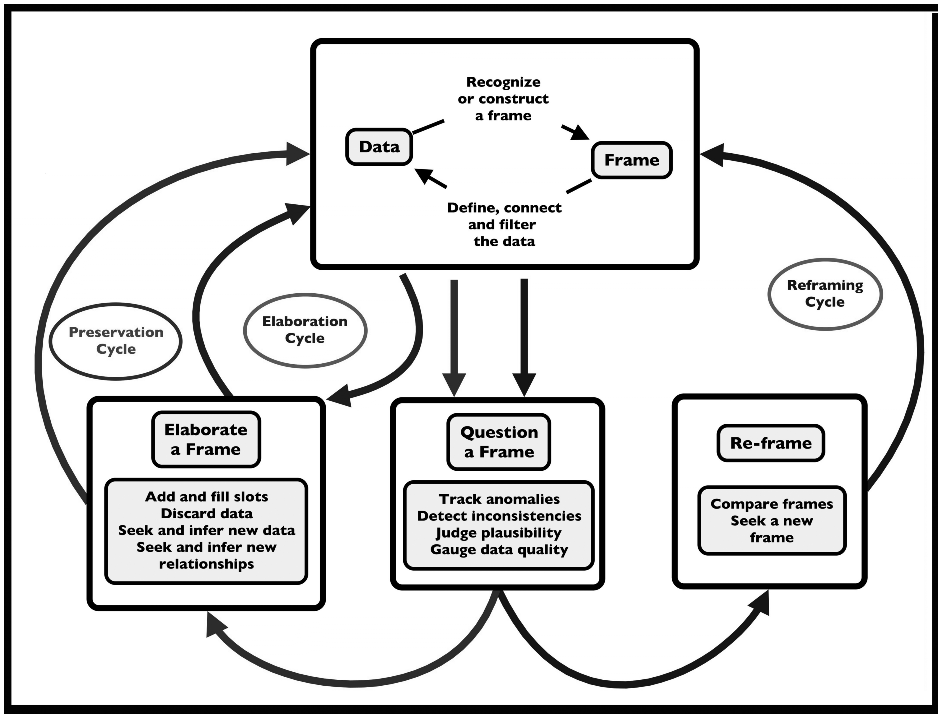


**Figure 1.**
The Data/Frame Model of sensemaking.

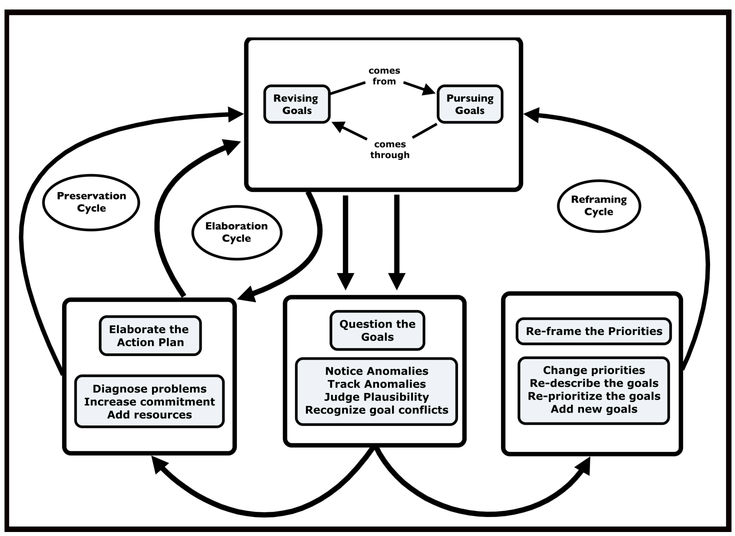


**Figure 2.**
The Flexecution model of replanning.

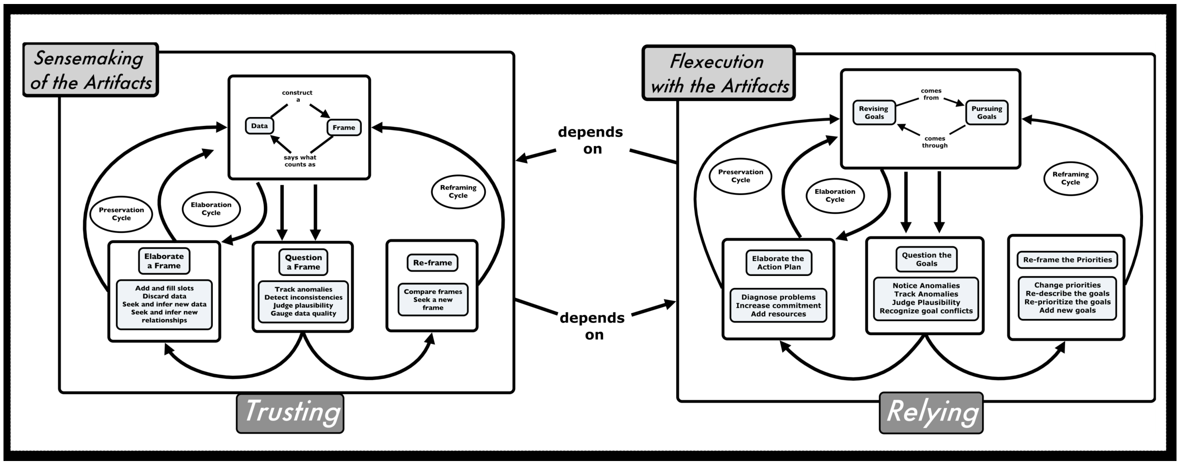


**Figure 3.**
Combined Data/Frame and Flexecution models that
express the dynamics of trust in and reliance upon automation.

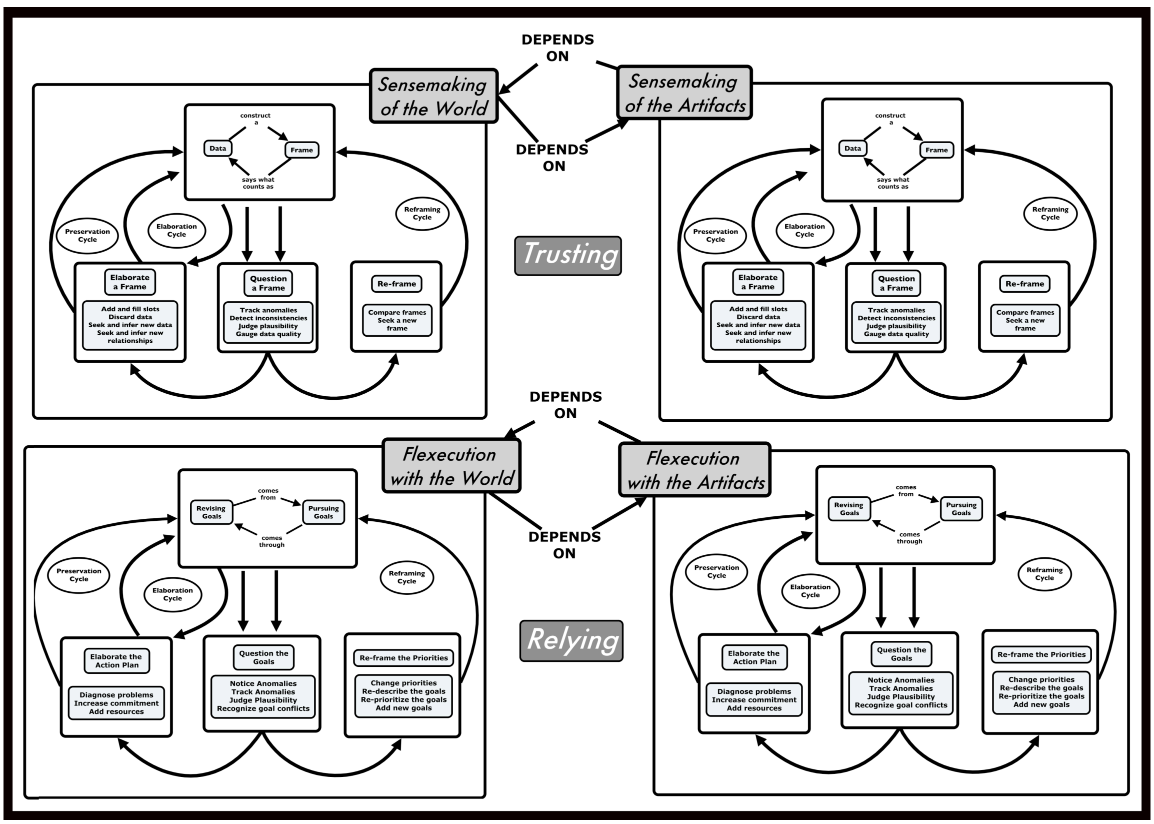


**Figure 4.**
A model that integrates sensemaking of the observed/controlled world and sensemaking
of the technology, with flexecuting in the task work and flexecuting with the technology.