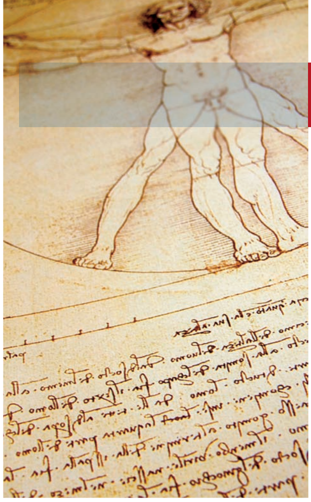

Editors: **Robert R. Hoffman, Jeffrey M. Bradshaw, and Kenneth M. Ford,**
Institute for Human and Machine Cognition, rhoffman@ihmc.us

Beyond Simon's Slice: Five Fundamental Trade-Offs that Bound the Performance of Macrocognitive Work Systems

Robert R. Hoffman, *Institute for Human and Machine Cognition*
David D. Woods, *Ohio State University*

Articulating the laws of cognitive work has been a continuing theme in this department. A number of the articles represent an effort to move toward a unified theory of "macrocognitive work systems."[1] These are complex adaptive systems designed to support near-continuous interdependencies among humans and intelligent machines to carry out functions such as sensemaking, replanning, mental projection to the future, and coordination.

The effort to identify empirical generalizations and use them to construct a formal theory[2] has led us to the identification of a number of fundamental trade-offs that place boundary conditions on all macrocognitive work systems. This article presents five trade-offs identified to date that define these boundary conditions. It also illustrates how the known empirical generalizations about the performance of human work systems can be systematically organized by the trade-offs.

Laws of Cognitive Work

Research in cognitive systems engineering has supported a considerable number of generalizations about macrocognitive work that have held up under empirical scrutiny across work domains.[3] For example, people always develop mental models of the agents, processes, and devices they deal with, including apparently intelligent technology. There is no "cognitive vacuum." The force of this generalization is that the developers bear the responsibility to design the technology such that that people can form accurate mental models of how the device or process works and the capabilities and limitations of that device or process relative to different situational factors.[4] If the design does not do this, one can be certain that people will form a variety of inaccurate explanations.

This inescapability makes the generalization law-like. For illustrative purposes, we present a few additional examples of the established empirical generalizations:

- *Law of Coordinative Entropy.* Coordination costs, continuously. The success of new technology depends on how the design affects the ability to manage the costs of coordinating activity and maintaining or repairing common ground.[5]
- *Law of Systems as Surrogates.* Technology reflects the stances, agendas, and goals of those who design and deploy the technology. Designs, in turn, reflect the models and assumptions of distant parties about the actual difficulties in real operations. For this reason, design intent is usually far removed from the actual conditions in which technology is used, leading to costly gaps between these models of work and the "real work."
- *Mr. Weasley's Law.* Based on their experiences, people develop unjustified trust and unjustified mistrust in their work system and its technology. As Mr. Weasley states in the *Harry Potter* series, "Never trust anything that can think for

1541-1672/11/$26.00 © 2011 IEEE
Published by the IEEE Computer Society

itself if you can't see where it keeps its brain." Understanding the intent of others, tracking and adjusting intent as situations change, and maintaining common ground across agents are critical in systems of interdependent agents, roles, and centers of control.[6]

- *The Law of the Kludge.* Work systems always require workarounds, with resultant kludges that attempt to bridge the gap between the original design objectives and current realities or to reconcile conflicting goals among workers. Sets of algorithms, plans, and procedures cannot account for inevitable variability and ongoing changes in the world. Thus, someone has to act responsibly to help plans match situations in order to meet mission goals.
- *The Law of Stretched Systems.* Every system is stretched to operate at its capacity. As soon as there is some improvement or new technology, some stakeholders will identify the opportunities that the change makes possible to achieve some of their goals. The process of exploiting these opportunities will result in a new and greater intensity and tempo of activity as the work system moves toward the edge of its competency envelope.[7]
- *The Law of Fluency.* Well-adapted cognitive work occurs with a facility that belies the difficulty of resolving demands and balancing dilemmas. The adaptation process hides the factors and constraints that are being adapted to or around. Uncovering the constraints that fluent performance solves, and therefore seeing the limits of or threats to fluency, requires a contrast across perspectives.[3]
- *The Reductive Tendency.* Agents at all scales develop and use simplifications, such as relying on decomposition to cope with interdependencies and decoupling to cope with dynamic interactions. Reductive understandings help workers manage what would otherwise be overwhelming complexity.[8]

Looking over the emerging set of empirical laws, now numbering approximately 30, several themes stand out, including interdependence and adaptation. Many of the laws seem to be a consequence of limits on systems that carry out cognitive work—for example, limits on building accurate models of devices (their capabilities and boundaries) and limits on the extent to which plans can match actual situations. Laws often include an interdependency relationship, either across different agents, perspectives, or roles or between system activities and the demands that arise in the environment.[9] Several laws express regularities about adaptation processes, capturing how people are active in a struggle to reach for goals despite the complications that arise regularly.

These observations led us to realize that the empirical generalizations might arise from constraints imposed by a set of more fundamental tradeoffs, or "covering laws."

It's Always about Trade-Offs

Discussions in cognitive science make much of a constraint introduced by Herbert Simon, which he called "bounded rationality."[10] Simon argued that there is a limit on the knowledge available to any agent or system and a limit on how that knowledge can be brought to bear in actual situations. Systems scientists have identified another trade-off that limits the performance of complex adaptive systems. Based on studies of biological and physical systems, John Doyle argued that the pursuit of increases in optimality with respect to some criteria guarantees an increase in brittleness with respect to changes or variations that fall outside of those criteria—a trade-off between optimality and fragility.[1] Work on proactive safety management and resilience engineering has identified two trade-offs that bound the performance of organizations that carry out risky activities: a tradeoff between acute and chronic goals, and a tradeoff between efficiency and thoroughness.[11,12] In addition, work on robotics and networks has illustrated limits on perspective taking and coordination across multiple agents.[6,13]

The Five Fundamental Bounds

We realized that the trade-offs organize subsets or "families" of laws as originally proposed by David Woods[7] and discussed in previous articles in this department. The end result is that five fundamental trade-offs seem to define five families of laws. The title of this article comes from our realization that Herbert Simon had found only "one slice" through the trade spaces governing the effectiveness of macrocognitive work systems.

Bounded Ecology

A work system can never match its environment completely; there are always gaps in fitness, and fitness itself is a moving target. There is always a struggle to adapt, which can ease or intensify as events unfold. This problem has been addressed by efforts to develop resilience and avoid brittleness, allowing the work system to either bounce back or gracefully degrade in the face of surprise.[14]

This is an optimality-fragility tradeoff—or stated in a way to balance two positives, the *Optimality-Resilience of Adaptive Capacity Trade-Off.* Increasing the scope of the routine increases the opportunities for surprise at the boundaries. Optimizing over some demands leads to brittleness when encountering other demands. As a result, resilience requires a capacity to adapt to surprising events, understanding that the ability to anticipate surprise requires additional resources whose contribution might be missed and might be mistaken for inefficiencies.

Bounded Cognizance

Limited resources and inevitable uncertainties lead to unavoidable gaps in knowledge. There is always "effort after meaning," although the struggle to acquire and deploy knowledge might temporarily ease or intensify. There are always gaps in plans, models, and procedures relative to the situations where they would be implemented to achieve goals. These gaps lead to an impetus to learn and adjust plans to fit the situations at hand. The process of testing the fit between plans and situations leads to a trade-off between being thorough and being efficient—the *Efficiency-Thoroughness of Situated Plans Trade-Off*.[11] Efficient plans mark well-worn paths, but they become cumbersome as the need to incorporate more contingencies and variations grows. Thoroughness expands the scope of the plans, expanding assessments, decisions, and ambiguities, but it constrains the ability to put plans into action and modify plans in progress.

We recast Simon's notion of bounded rationality as bounds on cognition, however, that might be embodied in an agent or agent architecture. We are agnostic with regard to the proposition that computer programs are models of cognition, and we do not adopt a rationalist-normative view of cognition. We therefore sought an alternative phrase to refer to Simon's slice: *bounded cognizance*.

Bounded Perspectives

Every perspective both reveals and hides certain aspects of the scene of interest, and work systems are limited in their ability to shift their perspective cost-effectively. Apprehension gaps can widen because situations differ in how strongly they signal the need to shift perspectives to reveal what has been hidden. Thus, there is always an invitation for reflection—that is, to step out of the current perspective. The work system is compelled to continuously devote resources to the integration of additional perspectives. But this comes at a cost, because the integration of different perspectives always requires an effort to translate or create a shared language to bridge or constructively contrast the perspectives.

Hence, there is a *Revelation-Reflection on Perspectives Trade-off*. Disambiguation for sensemaking arises from the ability to shift and contrast perspectives[13,15] or to rely on multimethod approaches that involve cycles of coactive emergence and convergence among humans and automation.[16] Modeling of complex adaptive systems has benefitted by including the concept of perspective as a basic parameter.[17] The ability to shift and contrast perspectives has proven to be essential to coordinated activity and collaborative work.[18]

Bounded Responsibility

Work systems divide up roles and responsibilities for subsets of goals; there are always gaps in authority and responsibility across the various subgoals. This means that all systems are simultaneously cooperative over shared goals and potentially competitive when goals conflict. Fundamental or chronic goals (such as safety and equity) tend to be sacrificed with increasing pressure to achieve acute goals ("faster-better-cheaper").[12] This, in turn, leads macrocognitive work systems to become blind to risks and sources of brittleness. Acute goals can be assessed through short-run tabulations, but chronic goals such as safety can only be assessed in the long run. They are more difficult to measure, and they function like values.

As a result, we term this bound as reflecting an *Acute-Chronic Goal Responsibility Trade-Off*. Work systems must continuously devote resources to manage responsibility across roles and ensure reciprocity. Without reciprocity across roles, the different agents or centers of control will tend to work at cross purposes in the face of goal conflicts.[18,19]

Bounded Effectiveness

Macrocognitive work systems are restricted in the ways they can act and influence situations. Agents understand that they are not omnipotent. This entails a trade-off between distributing autonomy, initiative, and authority across echelons, versus the more typical approach that concentrates autonomy, initiative, and authority in single centers of control. Distributing activities that define progress toward goals can increase the range of effective action, but increasing the distribution of activities entails difficulty of keeping them coherent and synchronized. Concentrating activities in single roles can produce more immediate and definitive progress toward landmarks, but it also reduces the range of effective action. Coordinating activity across distributed agents, units, or centers expands the scope and scale of factors that can be considered, but distributing macrocognitive work increases the cost of managing coordination as changes occur.[6,18,19] Concentrating the potential for action reduces the ability to consider potentially important interdependencies, whereas distributing the potential for action does just the opposite.

Ultimately, the challenge of the *Concentrated-Distributed Action Trade-Off* is to balance micromanagement with delegation over echelons to insure continuity and avoid fragmentation.

Covering Laws

These five bounds serve as covering laws that organize the first-order empirical generalizations of macrocognitive work. They place those generalizations in a larger theoretical and metatheoretical framework. For example, the Law of Stretched Systems is entailed by bounded ecology—that is, adaptation to become more optimal with respect to some goals and criteria will leave the system poised more precariously than its designers and managers realize. The Reductive Tendency law is entailed by bounded cognizance—that is, adaptation to finite resources inevitably leads to a reliance on simplifications.

Five Bounds and Adaptive Macrocognitive Work Systems

The bounds place inescapable limits on the performance of all macrocognitive work systems. This is seen in the frequency of unintended consequences following technology insertions: change directed only at one part within the system often inadvertently triggers deleterious effects on other aspects of the system that cancel out or outweigh the intended benefits. Technological interventions always produce a mix of desired and undesired effects because the change affects the system's positioning in the five trade spaces. For example, consider brittle but valuable drones that require large numbers of people to make up for the adaptive shortfall, or a system responding to faster, better cheaper pressure keeps eroding its margins and discounting warning signs until a sudden collapse occurs in the form of an accident (such as the Columbia space shuttle mission).[12]

The drones are part of a system that can be designed to be resilient rather than brittle, and organizations can develop cross-checking mechanisms that overcome the tendency to rationalize away warning signs. What are the basic architectural principles that allow systems to improve performance across the interdependencies captured by the five bounds rather than simply trade beneficial effects in one area for deleterious effects in others? Studies and models of complex adaptive systems have begun to identify principles that can be used to manage macrocognitive work systems, ensuring their capacity to maneuver in the trade-off spaces as evidence of risks of different types of adaptive breakdown grow—that is, to be resilient in the face of surprise.[14,20]

Grounding the modeling of human-machine work systems on the five trade-off functions provides a starting point for solving the perennial problem of how to evaluate the performance and adaptive capacity of macrocognitive work systems on principled formal grounds, rather than on the basis of ad hoc tabulations chosen to maximize local tractability and minimize immediate cost. We now see how efforts at measurement and prediction might evolve beyond the traditions of utility and the limits on measuring an individual person's performance.

The trade-offs take the bounds a crucial step closer to things we can actually measure at the system level. Assessing work-system performance requires at least two measurables for each of the five bounds. Numerical scales must address interactions across data types and must be interpretable as measurement scales that address the system's ability to adapt to events.[21] Positions along the bounds represent different solutions to the trade-offs, and as conditions change, the relative costs and benefits of different positions will change. Macrocognitive work systems might improve how they perform with respect to one of the trade-offs, but they cannot escape the inherent risks and the unintended consequences that can propagate across the other trade spaces. This means the set of trade-offs can be used to explain the adaptive history of past insertions of technology into complex fields of practice. Although design intent might attempt to move a work system toward the maximum limit on one of the trade-offs, the intervention can have effects that cascade across all of the trade-offs. Tracing those effects might let us make sense of the multiple and often unintended effects that actually result from technological interventions.

Prospects

The trade-offs can be used to project the potential reverberations or unintended consequences of any proposed injection of technology. We have used the five bounds to chart adaptive histories and reverberation paths for a number of specific cases—for example, the unintended consequences of introducing cockpit automation, networking technology for distributed intensive care in medicine, and nuclear power technologies following the Three Mile Island incident. Such exercises are suggestive of the potential value of the five bounds for modeling macrocognitive work systems. A next step is prospective analysis, tracing paths of reverberation across the linked trade-offs for proposed insertions of new technology, which should allow stakeholders to anticipate unintended consequences in advance of deployment.

One way of understanding the advance that we present here is to say that Herbert Simon only had one slice of the problem when he set out his framework for interpreting computer programs as theories of cognition. Broadening from the context of "one person, one machine" to the context of macrocognitive work systems, we have found additional bounds on performance. The trade-offs provide the theoretical underpinnings that can produce meaningful measures and metrics at the system level.

Acknowledgments

A preliminary presentation on these ideas was given at the 10th International Conference on Naturalistic Decision Making, sponsored by the University of Central Florida, in June 2011. We are indebted to Stephen M. Fiore and Jeffrey M. Bradshaw for their comments.

References

1. G. Klein et al., "Macrocognition," *IEEE Intelligent Systems*, vol. 18, no. 3, 2003, pp. 81–85.
2. R.R. Hoffman and D.D. Woods, "Steps Toward a Theory of Complex and Cognitive Systems," *IEEE Intelligent Systems*, vol. 20, no. 1, 2005, pp. 76–79.
3. D.D. Woods and E. Hollnagel, *Joint Cognitive Systems: Patterns in Cognitive Systems Engineering*, Taylor and Francis/CRC Press, 2006.
4. D.A. Norman, "Cognitive Engineering," *User-Centered System Design: New Perspectives on Human Computer Interaction*, D.A. Norman and S.W. Draper, eds., Erlbaum, 1986, pp. 31–61.
5. P. Feltovich et al., "Toward an Ontology of Regulation: Support for Coordination in Human and Machine Joint Activity," *Engineering Societies for the Agents World VII*, G. O'Hare et al., eds., Springer-Verlag, 2011, pp. 175–192.
6. G. Klein et al., "Ten Challenges for Making Automation a 'Team Player' in Joint Human-Agent Activity," *IEEE Intelligent Systems*, vol. 19, no. 6, 2004, pp. 91–95.
7. D.D. Woods, "Steering the Reverberations of Technology Change on Fields of Practice: Laws that Govern Cognitive Work," *Proc. 24th Ann. Meeting of the Cognitive Science Soc.*, Cognitive Science Soc., 2002, pp. 1–14.
8. P.J. Feltovich, R.R. Hoffman, and D. Woods, "Keeping it too Simple: How the Reductive Tendency Affects Cognitive Engineering," *IEEE Intelligent Systems*, vol. 19, no. 3, 2004, pp. 90–95.
9. M. Johnson et al., "Beyond Cooperative Robotics: The Central Role of Interdependence in Coactive Design," *IEEE Intelligent Systems*, vol. 26, no. 3, 2011, pp. 81–88.
10. H.A. Simon, *The Sciences of the Artificial*, MIT Press, 1969.
11. E. Hollnagel, *The ETTO Principle: Efficiency-Thoroughness Trade-Off: Why Things that Go Right Sometimes Go Wrong*, Ashgate, 2009.
12. D.D. Woods, "Essential Characteristics of Resilience," *Resilience Engineering: Concepts and Precepts*, E. Hollnagel, D.D. Woods, and N. Leveson, eds., Ashgate, 2006, pp. 19–30.
13. A. Morison et al., "Integrating Diverse Feeds to Extend Human Perception into Distant Scenes," *Advanced Decision Architectures for the Warfighter: Foundation and Technology*, P. McDermott, ed., Alion Science, 2009, pp. 177–200.
14. D.L. Alderson and J.C. Doyle, "Contrasting Views of Complexity and their Implications for Network-centric Infrastructures," *IEEE Systems, Man and Cybernetics*, *Part A*, vol. 40, 2010, pp. 839–852.
15. D.D. Woods and N.B. Sarter, "Capturing the Dynamics of Attention Control from Individual to Distributed Systems: The Shape of Models to Come," *Theoretical Issues in Ergonomics*, vol. 11, no. 1, 2010, pp. 7–28.
16. J.M. Bradshaw et al., *Sol: An Agent-Based Framework for Cyber Situation Awareness*, Kuenstliche Intelligenz, to be published in May 2012.
17. S.E. Page, *Diversity and Complexity*, Princeton Univ. Press, 2011.
18. P.J. Smith, A.L. Spencer, and C. Billings, "The Design of a Distributed Work System to Support Adaptive Decision Making Across Multiple Organizations," *Informed by Knowledge: Expert Performance in Complex Situations*, K.L. Mosier and U.M. Fischer, eds., Taylor and Francis, 2010, pp. 139–152.
19. E. Ostrom, "Toward a Behavioral Theory Linking Trust, Reciprocity, and Reputation," *Interdisciplinary Lessons from Experimental Research*, E. Ostrom and J. Walker, eds., Russell Sage Foundation, 2003.
20. D.D. Woods and M. Branlat, "Essential Characteristics of Resilience," *Resilience Engineering in Practice*, E. Hollnagel et al., eds., Ashgate, 2011, pp. 127–143.
21. R.R. Hoffman et al., "Measurement for Evaluating the Learnability and Resilience of Methods of Cognitive Work," *Theoretical Issues in Ergonomic Science*, 2011, doi:10.1080/14639220903386757.

Robert R. Hoffman is a senior research scientist at the *Institute for Human and Machine Cognition*. Contact him at rhoffman@ihmc.us.

David D. Woods is a professor at The Ohio State University in the Institute for Ergonomics. Contact him at woods.2@osu.edu.

Erratum

The name of the fourth author was misspelled in the Sept./Oct. 2011 issue's Human-Centered Computing Department, titled "Franchise Experts" (doi:10.1109/MIS.2011.85). That author's name is Steven M. Fiore, not Steven F. Fiore.