Uncertainty, Action, and Interaction: In Pursuit of Mixed-Initiative Computing

Eric Horvitz
Microsoft Research

Recent debate has highlighted differing views on the most promising opportunities for user-interface innovation.[1] One group of investigators has expressed optimism about the potential for refining intelligent-interface agents, suggesting that research should focus on developing more powerful representations and inferential machinery for sensing a user's activity and taking automated actions.[2–4] Other researchers have voiced concerns that efforts focused on automation might be better expended on tools and metaphors that enhance the abilities of users to directly manipulate and inspect objects and information.[5] Rather than advocating one approach over the other, a creative integration of direct manipulation and automated services could provide fundamentally new kinds of user experiences, characterized by deeper, more natural collaborations between users and computers. In particular, there are rich opportunities for interweaving direct control and automation to create *mixed-initiative* systems and interfaces.

Computer scientists have used the term *mixed-initiative* in various ways. These include references to the automated control of turn taking in human-computer conversation,[6] and the coordinated application of a set of problem-solving methodologies.[7] I shall use the phrase to refer broadly to methods that explicitly support an efficient, natural interleaving of contributions by users and automated services aimed at converging on solutions to problems.[8] Taking a mixed-initiative approach promises to dramatically enhance human-computer interaction by allowing computers to behave more like associates, able to work with users to develop a shared understanding of goals and to contribute to problem solving in the most appropriate way. Achieving such a dream of fluid collaboration between users and computers requires solving several difficult challenges. In particular, we need to develop machinery for gathering information and making inferences about the intentions, attention, and competencies of users—and for ultimately making decisions about the nature and timing of automated services. Computers will often be uncertain about the goals and needs of users. Thus, methods for reasoning under uncertainty play a critical role in mixed-initiative interaction.

Supporting joint activity under uncertainty

People appear to be well adapted to mixed-initiative problem solving. In daily life, we continue to engage one another in efficient, tightly woven collaborations. We assume and rely on a rich interleaving of efforts to achieve goals through a sharing of beliefs, needs, and context. A common arena for exploring mixed-initiative interaction is conversation, centering on a collaboration to achieve the goal of communicating needs and information. However, mixed-initiative interaction extends beyond conversation to encompass a wide variety of interactions that rely on a collaborative interleaving of contributions by participants, some of which might include conversational interaction.

Intelligent Systems, September / October 1999, IEEE Computer Society, pp. 17-20.

Psychologists and computer scientists have referred to efficient collaborations converging on shared goals as *joint activity*.[9–11] Joint activity captures the behavior displayed in fast-paced, well-coordinated interactions among people who work together to solve a mutual goal. Examples of joint activity include the collaborative behaviors seen in conversation, dancing, and the familiar struggle of moving a large piece of furniture through cramped hallways. Participants in joint activity seek convergence on a shared set of beliefs about the setting, activity, goals, and the nature and timing of individual contributions. Uncertainties about goals and needs are resolved through a drive towards a mutual understanding or common ground in a process referred to by psychologists as *grounding*.[9,10,12]

Joint activity embodies an especially fluid and efficient form of mixed-initiative interaction. The pursuit of metaphors, designs, and reasoning machinery for supporting joint activity presents the most difficult challenges—and the greatest opportunities—for research on mixed-initiative interaction.

Beliefs, actions, and initiative

Mixed-initiative systems must consider a set of key decisions in their efforts to support joint activity and grounding. These include *when* to engage users with a service, *how* to best contribute to solving a problem, *when* to pass control of problem solving back to users for refinement or guidance, and when to query a user for additional information in pursuit of minimizing uncertainty about a task.

Systems that provide automated services rely on the ability to make good guesses about a user's needs by considering evidence obtained through the narrow keyhole of user interface events. A system's ability to understand users can be enhanced by coupling richer systems for monitoring user activity with more expressive knowledge representations and sophisticated grounding skills. However, even given more complete knowledge about a user's activity, mixed-initiative systems must still grapple with significant uncertainties. Thus, building effective mixed-initiative systems requires the consideration of key uncertainties both at design time and in real time.

A *Bayesian* approach to human-computer interaction provides a valuable perspective for the design of mixed-initiative systems. Bayesian agents maintain beliefs about such critical variables as a user's intentions and attention. Bayesian agents also update their beliefs continuously with probabilistic procedures that consider both passively observed and actively gathered evidence.

Recent work on the use of real-time Bayesian inference suggests that dynamic reasoning under uncertainty can be a valuable component of mixed-initiative interaction. Both hand-built and automatically learned probabilistic user models, including Bayesian networks, have been embedded as key components of user-interface prototypes. For example, in the Lumière system,[4] a Bayesian network model analyzes a stream of events generated by the user's interaction with Microsoft Excel. It continuously infers probability distributions over the user's goals and user's interest in receiving active assistance. When the user makes an explicit query for assistance, information about this query is added to the analysis. The bar graph in Figure 1 represents a snapshot of a probability distribution inferred by Lumière over a user's goals. My colleagues and I have developed prototypes that not only reason about a user's goals and needs,

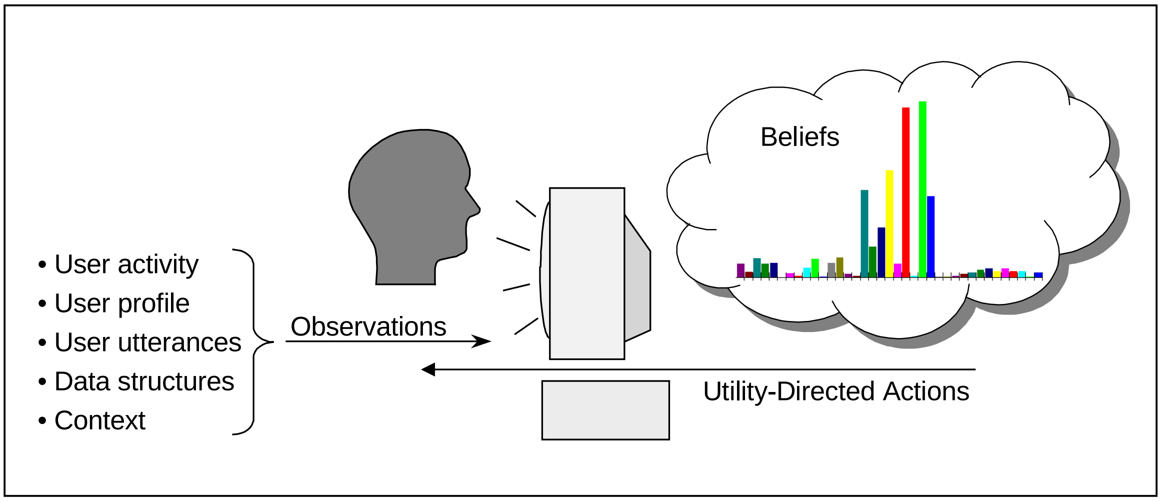

Figure 1. A Bayesian perspective on human-computer interaction. A probability distribution about a user's goals (bar graph) is computed from a set of observations and background information about the user. Actions are selected based on their expected utility. The probability distribution here, generated by the Lumière system, displays the likelihoods of a user's different goals while working with Microsoft Excel.

but additionally harness Bayesian networks to infer a probability distribution over the attentional focus of users.[13]

Guiding mixed-initiative action with expected utility

A system endowed with the ability to infer beliefs about the states of a user's intention and attention can make ideal decisions about how and when an automated service should step in to assist a user. More specifically, access to beliefs about a user's goals give a mixed-initiative system the ability to take information-gathering and problem-solving actions that have the highest *expected utility*, taking into consideration the expected benefits and costs of attempting to participate in problem solving. Expected benefits represent the gains in efficiency associated with offering a contribution under uncertainty. Expected costs capture the frustration and inefficiency associated with the distraction of presenting an otherwise valuable contribution—or of executing an inappropriate contribution. Beyond reasoning about goals, inferences about the attention of users are critical in making decisions about the best time to provide assistance. Significant costs may be associated with querying a user or providing a partial solution when the user is not ready to accept the intervention.

The policy of taking actions associated with maximum expected utility has a long tradition, founded on the axioms of utility, formulated originally by John von Neumann and Oskar Morgenstern over 50 years ago. Although expected utility has enjoyed a rich history of application in such fields as economics and decision analysis, it has only recently been applied in human-computer interaction.

Designing for a mix of initiatives

Harnessing probabilistic inference to provide an awareness of users and expected utility to guide actions offers an overall perspective that can guide the development of mixed-initiative architectures. However, the basic principles do not provide detailed blueprints for creating specific, valuable interleavings of direct manipulation and automation. Designs for mixed-initiative systems benefit greatly from careful consideration—from the earliest phases of the design process—of the detailed interactions between potential automated services and options for user manipulation and display.
The large space of design opportunities for mixed-initiative interaction includes

- developing automated services that are performed *in line* with a user's activity, allowing users to take advantage of contributions provided by a system while they work in a natural manner,
- identifying elegant metaphors that promote efficient grounding by providing efficient means for users and computers to communicate information *about* intended or ongoing contributions to a solution, and
- developing automated services that can provide solutions at varying levels of precision or completeness, giving mixed-initiative systems the flexibility to scope the precision of contributions in accordance with the uncertainty about a user's goals or the competency of an analysis.

The latter class of design opportunities is motivated by the notion that, as uncertainty grows about a user's intentions or about the quality of the result, a system should gracefully degrade its contribution so as to "do less, but do it well." That is, we prefer that a system provide users with a clear, valuable advance towards a solution—an advance that minimizes the need for the user to perform costly undoing or backtracking. We can enhance the ability of systems to make decisions about the most appropriate contribution by endowing those systems with the ability to decompose prototypical tasks into sets of subtasks that span a spectrum of precision or completeness.

Lookout

A prototype system named Lookout provides concrete instantiations of several key concepts that highlight the role of decision making under uncertainty in mixed-initiative interaction.[8] Lookout assists with the task of calendar review and appointment creation. A group of interested people scattered throughout Microsoft have employed the system since it was released for internal testing in early 1998.

Lookout monitors a user's interactions with the Microsoft Outlook messaging and calendar system. The system recognizes when users open and attend to new e-mail messages. Lookout decides whether, when, and how to best assist users with the tasks of accessing the appropriate view of their calendar and scheduling appointments associated with the messages.

For each message being reviewed, Lookout infers the probability that a user has the goal of invoking Outlook's calendar and scheduling subsystem. This is done by considering information in the header and patterns of text in the body of e-mail messages. Given this probability, and the costs and benefits of providing automation, the system performs an expected utility analysis and decides whether to simply do nothing (letting users continue to perform direct manipulation), or to interact with the user. The system considers the expected utility of pausing to ask the user if he or she might like assistance and of simply going ahead and performing the most appropriate calendar view or scheduling action without requesting the user's intput. When the system decides that automated calendar access or appointment generation would be a valuable contribution, it displays results in a manner that makes it easy for the user to further refine or undo the analysis.

In providing its service, Lookout uses knowledge about the typical ways people describe meetings and times. It understands the temporal implications of such phrases appearing in e-mail as "Fri afternoon," "tomorrow at 3," "next week," "in December," "get breakfast," "grab lunch," and so forth. In preparing its analysis, Lookout considers a spectrum of contributions at differing levels of precision. The system seeks to provide a user with a valuable step forward by displaying an automatically generated appointment or a calendar view with the most appropriate scope.

If the system can identify a single date and time with confidence, it will construct a proposed appointment by filling out the day, time, and subject fields of an Outlook appointment form and present it to the user for confirmation or modification. If the target appointment conflicts with another meeting on the calendar, the system will search to find an alternative time for the event before composing and presenting the appointment. If the system cannot identify a specific day and time with confidence, it will opt to introduce a less precise contribution. For example, the system will open the calendar to the most likely day, or the most likely week, and pass control back to the user for refinement.

Lookout relies on reasoning, learning, and communication to provide services in line with the flow of a user's work. Lookout employs a model for gauging the status of a user's attention in making decisions about when to jump in and query the user or to perform its service. The system infers the amount of time a user wishes to dwell on an e-mail message at hand by considering attributes of the message and the user's activity. Specifically, Lookout considers the length of the message and the time since the last paging or scroll event to decide on the ideal time to step in. Early versions of Lookout that did not employ such a model of attention had a very different feel; the appropriate timing of services dramatically improves the experience and relays a remarkable sense that an intuitive assistant is attempting to work with the user.

Lookout can be instructed to run in a hands-free, social-agent modality, employing an explicit animated assistant coupled with speech synthesis and recognition. When operating in the social-agent mode, Lookout establishes a separate audio channel for communicating with users about contributions, minimizing the potential conflict with ongoing keyboard and mouse activity. Figure 2 shows a sample interaction with Lookout as an embodied agent.

Integrating an explicit social presence has let us explore the use of gestures and utterances that might be expected in natural mixed-initiative interaction among people. For example, the agent selects a behavior from a set of gestures and utterances that communicates its confidence about taking an action. Also, the agent displays signs of confusion when the speech recognition subsystem has difficulty interpreting the audio signal. If the system does not receive a response when offering the user assistance, it uses gestures to communicate, in a noninvasive manner, the notion that it understands that the user is too busy to respond before disappearing (for example, the agent will shrug, relaying with visual cues that "I was just trying to be of service—no problem..."). At such times, the agent will wait patiently on the sidelines for a period of time that is computed dynamically as a function of the inferred belief that the system could have provided a valuable service.

Lookout continually attempts to improve its ability to provide valuable contributions by performing background learning. The system's models for inferring the goals and attention of users are updated over time through implicit observation of a user's behavior, using a learning process that collects evidence about such variables as the content of messages associated with a user's scheduling activity and the period of time between a message being opened and a user's direct execution of calendar and scheduling tasks.

Beyond implicit learning, Lookout allows users to directly indicate their preferences about the system's decision-making behavior. Preferences input during configuration are used in Lookout's cost-benefit analyses and timing decisions. Additionally, users can take the initiative to invoke Lookout's services at any time by simply clicking on the Lookout icon that is always available on the System Tray of the Windows shell.

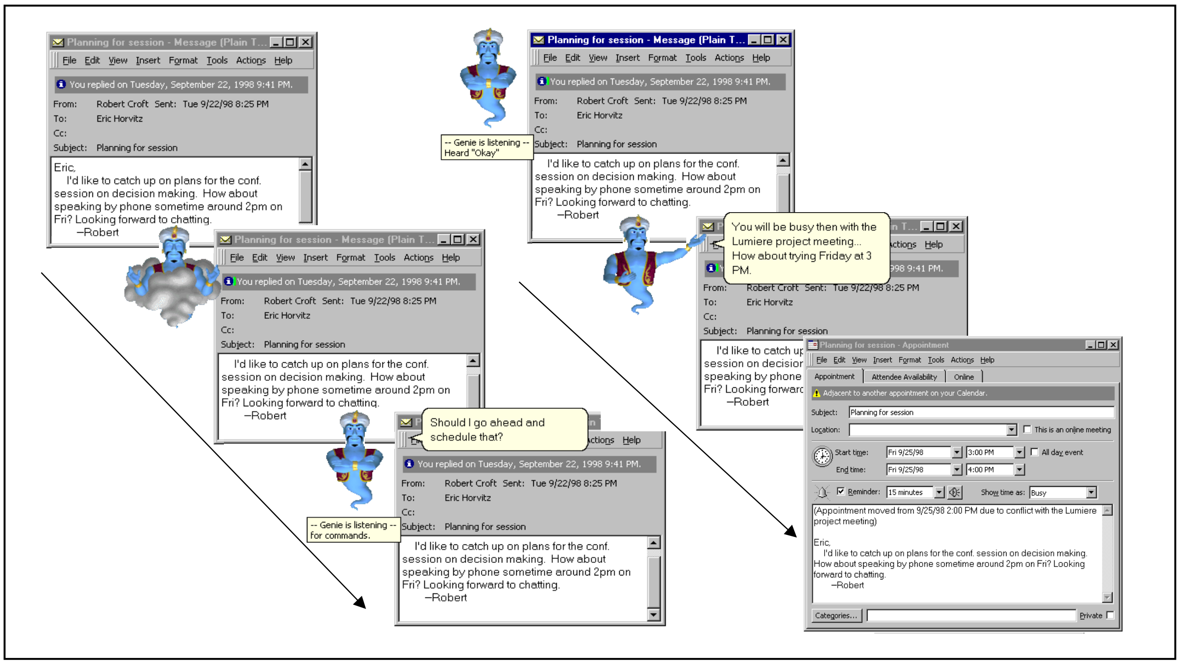

Figure 2. Lookout in action. Procedures that harness probability and utility guide Lookout's actions to assist users with accessing their calendar and composing appointments, based on a background analysis of e-mail being reviewed.[8]

We have explored designs for more deeply integrating Lookout's automated services with direct manipulation and display. The current version of Lookout's was designed to work with a legacy software application, rather than built as part of a more global design process taking a more comprehensive approach to interweaving direct manipulation and automation. As such, the behavior and value of the Lookout prototype hinges on the design of the direct-manipulation capabilities provided by Outlook. Without Lookout, users typically navigate to the appropriate graphical button or menu item to access their calendar, search for the appropriate day, input the appropriate times and fill in the subject of the meeting. Changes in the details of how Outlook operates would likely entail modifications of the actions and cost-benefit considerations employed by Lookout.

Frontiers of mixed-initiative interaction

The Lookout system has provided a testbed for utility-directed mixed-initiative interaction on relatively short-run sequences of interaction. Work is underway on leveraging Bayesian inference and expected-utility decision making in richer mixed-initiative systems that work with users on longer, more sophisticated communciation and problem-solving sessions. For example, work on the Bayesian Receptionist focuses on methods for supporting joint activity and grounding in conversation about goals that are typically handled by receptionists at the Microsoft corporate campus.[14,15] The Bayesian Receptionist decomposes goals into a hierarchical set of subgoals and employs sets of Bayesian networks and expected-utility decision making to navigate through a subgoal hierarchy in pursuit of common ground.

Lookout and the Bayesian Receptionist have highlighted the necessity and promise of endowing agents with beliefs and of employing probability and expected utility to mesh automated services with direct manipulation. Opportunities abound for harnessing probabilistic methods to weave automation more tightly together with methods that enable users to control, inspect, and guide computing. Although great challenges lie ahead, we believe these early prototypes, and others being developed by colleagues pursuing principles and machinery for mixed-initiative interaction, provide glimmers of the future of human-computer interaction.

References

1. P. Maes and B. Shneiderman, "Direct Manipulation vs. Interface Agents: A Debate," *Interactions*, Vol. IV Number 6, ACM Press, 1997.
2. P. Maes, "Agents that Reduce Work and Information Overload," *Comm. ACM*, Vol. 37, No. 7, pp. 31–40.
3. L. Birnbaum et al., "Compelling Intelligent User Interfaces: How Much AI?," *Proc. 1997 ACM Int'l Conf. Intelligent Interfaces*, ACM Press, New York, 1996; http://www.merl.com/reports/TR96-28/index.html.
4. E. Horvitz et al., "The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users," *Proc. 14th Conf. Uncertainty in AI,* Morgan Kaufmann, San Francisco, 1998, pp. 256–265; http://research.microsoft.com/~horvitz/lumiere.htm.
5. B. Schneiderman, *Designing the User Interface: Strategies for Effective Human-Computer Interaction,* ACM Press. 1992.
6. M.A. Walker and S. Whittaker, "Mixed Initiative in Dialogue: An Investigation into Discourse," *Proc. 28th Ann. Meeting Assoc. Computational Linguistics,* 1990.
7. G. Ferguson, J. Allen, and B. Miller, "TRAINS-95: Towards a Mixed-Initiative Planning Assistant," *Proc. Third Conf. AI Planning Systems,* AAAI Press, Menlo Park, Calif., 1996, pp. 70–77.
8. E. Horvitz, "Principles of Mixed-Initiative User Interfaces," *Proc. ACM SIGCHI Conf. Human Factors in Computing Systems*, ACM press, New York, 1999, pp. 159–166; http://research.microsoft.com/~horvitz/uiact.htm.
9. H.H. Clark, *Using Language*, Cambridge Univ. Press, New York, 1996.
10. P.R. Cohen and H.J. Levesque, "Preliminaries to a Collaborative Model of Dialogue*,"* *Speech Communication*, Vol. 15, 1994, pp. 265–274.
11. B.J. Grosz and C.L. Sidner, "Plans for Discourse," *Intentions in Communication*, MIT Press, Cambridge, Mass., 1990, pp. 417–444.
12. H.H. Clark and S.A. Brennan, "Grounding in Communication," *Perspectives on Socially Shared Cognition*, APA Books, Washington D.C., 1991, pp. 127–149.
13. E. Horvitz, A. Jacobs, and D. Hovel, "Attention Sensitive Alerting," *Proc. Conf. Uncertainty and Artificial Intelligence*, Morgan Kaufmann, San Francisco, 1998, pp. 305–313; http://research.microsoft.com/~horvitz/attend.htm.
14. E. Horvitz and T. Paek, "A Computational Architecture for Conversation," *Proc. Seventh Int'l Conf. User Modeling*, Springer Wien, New York, 1999, pp. 201–210; http://research.microsoft.com/~horvitz/converse.htm.
15. T. Paek and E. Horvitz, "Uncertainty, Utility, and Misunderstanding: A Decision-Theoretic Perspective on Grounding in Conversational Systems," *AAAI Fall Symp. Psychological Models of Communication in Collaborative Systems*, AAAI Press, Menlo Park, Calif., 1999; http://research.microsoft.com/~horvitz/grounding.htm.

Intelligent Systems, September / October 1999, IEEE Computer Society, pp. 17-20.