Psychological Review
2006, Vol. 113, No. 3, 461–482

0033-295X/06/$12.00 DOI: 10.1037/0033-295X.113.3.461

The Soft Constraints Hypothesis: A Rational Analysis Approach to Resource Allocation for Interactive Behavior

Wayne D. Gray and Chris R. Sims
Rensselaer Polytechnic Institute

Wai-Tat Fu
University of Illinois, Urbana-Champaign

Michael J. Schoelles
Rensselaer Polytechnic Institute

Soft constraints hypothesis (SCH) is a rational analysis approach that holds that the mixture of perceptual-motor and cognitive resources allocated for interactive behavior is adjusted based on temporal cost-benefit tradeoffs. Alternative approaches maintain that cognitive resources are in some sense protected or conserved in that greater amounts of perceptual-motor effort will be expended to conserve lesser amounts of cognitive effort. One alternative, the minimum memory hypothesis (MMH), holds that people favor strategies that minimize the use of memory. SCH is compared with MMH across 3 experiments and with predictions of an *Ideal Performer Model* that uses ACT-R's memory system in a reinforcement learning approach that maximizes expected utility by minimizing time. Model and data support the SCH view of resource allocation; at the under 1000-ms level of analysis, mixtures of cognitive and perceptual-motor resources are adjusted based on their cost-benefit tradeoffs for interactive behavior.

Keywords: reinforcement learning, rational analysis, embodied cognition, resource allocation, interactive behavior

The night before the birthday party you open the box and separate the assembly instructions from the parts for the child's new toy. Do you memorize all of the instructions, put them aside, and then assemble the toy from memory? Or, do you read the first line, put the instructions down, do the first step, pick up the instructions, read the next line, put the instructions down, do the next step, and so on until the toy is complete? Whatever you do, you are making tradeoffs between strategies that minimize the use of memory by making repeated interactions with the task environment versus strategies that minimize interactions by increasing their demands on the memory system.

At a second-by-second level of analysis, interactive behavior can be analyzed as a complex mixture of elementary cognitive, perceptual, and motor operations (e.g., Gray & Boehm-Davis, 2000). Although all three types of operations are required for any interactive behavior, as in the example of the assembly instructions for the new toy, frequent accesses of knowledge in-the-world (Norman, 1989, 1993) will be characterized as more interaction-intensive, whereas greater reliance on knowledge in-the-head will be characterized as more memory intensive.

Few people would be surprised by the observation that sometimes they take notes and sometimes they memorize things, or that they sometimes look at their notes and sometimes simply remember what they have written. However, although such interactions are commonplace, until recently the interleaving of cognition, perception, and action has been little noted and less studied by the cognitive community.

An important spur to the status quo came when researchers (Card, Moran, & Newell, 1980, 1983; Larkin, 1989; Larkin & Simon, 1987; Norman, 1982, 1989) began trying to apply cognitive theory to real world problems. These attempts at cognitive engineering (Norman, 1982, 1986), although productive (Gray, John, & Atwood, 1993), revealed the limits of cognitive theory (Gray, Schoelles, & Myers, 2004) and spurred many cognitive researchers to study how cognition, perception, and the motor system worked together when moderately complex laboratory (Freed, Matessa, Remington, & Vera, 2003; Gray & Boehm-Davis, 2000; Howes, Lewis, Vera, & Richardson, 2005; Kieras & Meyer, 1997; Ritter, Van Rooy, St. Amant, & Simpson, in press; Taatgen & Lee, 2003) or complex real-world tasks were performed (Byrne & Kirlik, 2005; Salvucci, in press).

Wayne D. Gray, Chris R. Sims, and Michael J. Schoelles, Cognitive Science Department, Rensselaer Polytechnic Institute, and Wai-Tat Fu, Human Factors Division and Beckman Institute, University of Illinois at Urbana-Champaign.

The work on this project was supported by a grant from the Office of Naval Research ONR# N000140310046, as well as the Air Force Office of Scientific Research AFOSR #F49620-03-1-0143. We thank Robert Sorkin for his enthusiastic support for this work and his many conversations about optimality analysis and the ideal observer analysis. We wish to thank Hans Neth, Maria Angelica Velazquez, and Brittney Oppermann for their close reading of an earlier version of this paper. The authors especially wish to thank John Anderson and Erik Reichle for their pointed and persistent prodding on earlier versions of the article.

Correspondence concerning this article should be addressed to Wayne D. Gray, Rensselaer Polytechnic Institute, Carnegie Building, 110 8th St., Troy, NY 12180. E-mail: grayw@rpi.edu

Initially, researchers were content to demonstrate that the task environment in which interactive behavior takes places could influence the higher-level strategies that people adopt for decision making (Lohse & Johnson, 1996), problem solving (O'Hara & Payne, 1998, 1999), or game playing (Kirsh & Maglio, 1994). Recently, attention has turned to studies that have shown systematic effects of the design of the task environment on the methods that people adopt for routine tasks such as simple mental arithmetic (Neth & Payne, 2001; Stevenson & Carlson, 2003). Although each of these studies implies a general sensitivity of the human control system to perceptual-motor costs, what is lacking is a functional mechanism that adjusts the mixture of low-level cognitive, perceptual, and motor resources to produce the observed higher-level changes in behavior.

Gray and Boehm-Davis (2000) noted that the procedural steps that implement low-level goals are selected as if milliseconds matter. Although other researchers tend to agree that the selected routines conserve milliseconds, they do not agree that temporal costs are the causal basis of selection as opposed to a correlated measure. In a series of studies, Carlson and associates (Carlson & Sohn, 2000; Cary & Carlson, 1999; Sohn & Carlson, 1998, 2003; Stevenson & Carlson, 2003) have shown that people adapt their interactive behavior to the tools they have available. Indeed, if left to their own devices, people spontaneously adopt methods for doing simple arithmetic that shave 200 ms off of alternative routines. However, rather than basing selection on time per se, Cary and Carlson (1999, p. 1067) concluded that, "Participants without memory aids tended to choose solution paths that minimized working memory demands."

Similarly, when the cost of accessing needed information was increased by milliseconds from an eye movement to a head movement, Ballard, Hayhoe, and Pelz (1995; Pelz, 1996) noted a small decrease in gaze frequency to an external display. However, like Carlson and associates, rather than concluding that the selection of interactive behaviors minimizes effort defined by time, they concluded that, "Observers prefer to acquire information just as it is needed, rather than holding an item in memory" (Hayhoe, 2000, p. 50). As elaborated later, this *minimum memory hypothesis* appears related to views that cognitive limitations (in this case, working memory) bias the control system to offload work onto the perceptual-motor system (Wilson, 2002). The minimum memory hypothesis is thus one candidate explanation for the functional mechanism that adjusts the mixture of low-level cognitive, perceptual, and motor resources.

Throughout this paper the implications of the soft constraints hypothesis for resource allocation will be contrasted with those of the minimum memory hypothesis. The next section introduces the soft constraints hypothesis as an alternative functional mechanism to the minimum memory hypothesis. The distinction between soft constraints and minimum memory hypotheses is elaborated, and the concept of an ideal performer analysis as a tool to study the implications of constraints on cognition is introduced. The Experiments section is an overview of three experiments that provide increasingly persuasive evidence in favor of soft constraints. Our *Ideal Performer Model*, based on our ideal performer analysis, is presented next. This model serves as an explicit test of the sufficiency of the soft constraints hypothesis as an explanation for the functional mechanism underlying the control of interactive behavior. As we will show in the model results section, the Ideal Performer Model provides a close fit to the human data. The last section summarizes the results and concludes that the human control system is not biased to conserve cognitive resources at the expense of other resources, but rather that the selection of interactive behaviors is driven by cost-benefit considerations. When the expected utility (i.e., the cost-benefit tradeoff) of alternative interactive behaviors can be quantified in terms of time, those that minimize milliseconds are selected over those that minimize cognitive resources.

Soft Constraints, Minimum Memory, and the Ideal Performer

The essence of soft constraints is a hypothesis about the functional basis for selecting one low-level interactive routine over another. Interactive routines are envisioned as dependency networks of low-level cognitive, perceptual, and motor operators that come together at a time span of about 1/3 to 3 seconds in the service of low-level interactive behavior (Gray & Boehm-Davis, 2000).[1] Interactive behavior proceeds by selecting one interactive routine after another or by selecting a stable sequence of interactive routines (i.e., a method) to accomplish a unit task (Card et al., 1983). Adopting Ballard's (Ballard, Hayhoe, Pook, & Rao, 1997) analysis of embodiment, we see these interactive routines as the basic elements of embodied cognition.

The Soft Constraints Hypothesis

The rational analysis perspective (Anderson, 1990, 1991; Oaksford & Chater, 1998) has shown that it is important to step back from the study of mechanisms to ask about the environments in which these mechanisms are applied (Gray, Neth, & Schoelles, in press). If we assume that the mechanisms responsible for goal-directed human behavior are adapted to the structure of their task environment, then finding an appropriate description of the environment may yield important constraints on the nature and behavior of functional mechanisms. Anderson and Schooler's classic work on the structure of the environment for human memory (Anderson & Schooler, 1991) is a prime example of this approach, as is the more recent work on the statistical properties of the perceptual environment (Geisler & Diehl, 2003; Purves, Lotto, & Nundy, 2002).

Interactive behavior is usually in the service of higher-level goals. Anything that increases its performance helps us achieve these goals faster. In the nonlaboratory world, besides decreasing costs in terms of time (and presumably, resources), efficient interactive behavior may make the difference between the success or

[1] In Gray and Boehm-Davis (2000) we used the term "basic activity" to describe these combinations of low level operators. Our current use of the phrase "interactive routine" is, in part, a homage to Hayhoe's (2000) and Ullman's (1984) use of the term "visual routines." However, in larger part, "interactive routine" better reflects the notion that certain combinations of low-level cognitive, perceptual, and action operations can be regarded as building blocks of interactive behavior as well as the notion that at this level of description all behavior is composed of cognitive, perceptual, and motor operations.

failure of higher-level tasks. Hence, in situations as diverse as playing computer games, tuning a radio while driving in busy traffic, searching for information amid the near-infinite space defined by the World Wide Web, and assembling a child's toy, the time required for interactive behavior may be a cost, whereas achieving the goals of the behavior may be a benefit.

Simply stated, the soft constraints hypothesis maintains that at the 1/3 to 3 sec level of analysis, the control system selects sequences of interactive routines that tend to minimize performance costs measured in time while achieving expected benefits. Cost-benefit considerations provide a soft constraint on selection as they may be overridden by factors such as training or by deliberately adopted top-down strategies.

Negotiating cost-benefit tradeoffs in the selection of interactive routines does not guarantee optimal performance in a task; that is, locally optimal interactive routines may not lead to globally optimal performance. Rather, the soft constraints hypothesis predicts optimal performance only in tasks where maximizing the expected gains and minimizing the expected costs of interactive routines (i.e., over 1/3 to 3 sec) is congruent with an optimal strategy at the global task level. In environments that violate this property, the soft constraint hypothesis predicts persistently suboptimal performance (Fu & Gray, 2004, in press). This focus on local optimization is consistent with the rational analysis position that "Specifying the computational constraints essentially amounts to defining the locality over which the optimization is defined" (Anderson, 1990, p. 247). The extent to which human goals can be achieved by optimizing at the level of interactive routines is the extent to which the soft constraints hypothesis represents a rational adaptation to the environment.

In summary, the soft constraints hypothesis applies the rational analysis (Anderson, 1990, 1991) approach to the allocation of cognitive, perceptual, and motor resources for interactive behavior. These resources are encapsulated in interactive routines that are described at the 1/3 to 3 sec level of analysis. To the extent that the elements going into the calculation of expected utility are variable, unstable, or overridden by deliberately adopted policy, then cost-benefit calculations provide a soft, not hard, constraint on the selection of interactive behavior. However, the soft constraints hypothesis assumes that the selection of interactive routines minimizes performance costs measured in the currency of time. The objective of minimizing time is a soft constraint, and it is the deviations from this policy that must be explained. In this paper we seek to strengthen the soft constraints hypothesis by showing that its predictions are supported by empirical data and that an Ideal Performer Model, which enforces a strict temporal cost-benefit accounting, fits the empirical results.

Soft Constraints Versus the Minimum Memory Hypothesis

In contrast to the soft constraints hypothesis, alternative views of embodied cognition suggest that cognitive resources are conserved by biases that favor the use of perceptual-motor resources (Wilson, 2002). The minimum memory hypothesis provides a specific instance of this view of embodiment which suggests that the control system is biased toward reducing memory costs even when the costs of information access (as measured by time) for perceptual-motor strategies are much greater than the costs for memory strategies (Ballard et al., 1997). An attraction of the minimum memory hypothesis is that it offers a simple heuristic for governing behavior, and unlike the soft constraints hypothesis, does not require an accounting of costs sensitive at the level of hundreds of milliseconds.

The minimum memory hypothesis seems to embrace a limited capacity view of memory in which capacity is defined either by the number of slots available in a short-term or working memory buffer (Miller, 1956) or a limit on the amount of activation available to that buffer (Just & Carpenter, 1992; Just, Carpenter, & Keller, 1996). (For more detailed and more recent discussions of limited capacity see, e.g., Cowan, 1997, 1999; Engle, Tuholski, Laughlin, & Conway, 1999.) If there is only "so much" memory available for use, then it is reasonable that this precious resource is conserved whenever possible either to avoid overloading the system or to have reserves available if needed for more important tasks.

All memory theories of which we are aware hold that encoding items into memory requires time and that once items enter memory they may be forgotten. The soft constraints hypothesis implies that on the memory side of the tradeoff between interaction-intensive and memory-intensive strategies, the only factors that matter are the time required to encode, the time required to retrieve an item from memory, and the probability that an encoded item can be retrieved (i.e., is not forgotten) when needed. An item that is forgotten represents time wasted in the original encoding, time wasted in the attempted retrieval, and additional time required to recode and reretrieve the item. Hence, the soft constraints view on use of memory as a resource is that only milliseconds matter; there is no particular premium on conserving memory and no inherent bias favoring perceptual-motor effort.

In a search of the literature we have found no tests that directly pit any form of the minimum memory hypothesis against any form of the soft constraints hypothesis. However, at least two studies have indirectly examined tradeoffs between memory utilization and perceptual-motor effort, one by Ballard (Ballard et al., 1995) and one by Gray and Fu (2004).

Ballard, Hayhoe, and Pelz (1995) used a Blocks World task (for our version of the Blocks World task see Figure 1) to study patterns of information access. The participant's task was to reproduce the pattern of blocks presented in the Target Window in the Workspace Window using blocks obtained from the Resource Window. In Ballard's study (and unlike ours) all windows were freely visible at all times. Information access required only an eye movement.

Ballard and colleagues report that participants preferred an interaction-intensive strategy in which they would look at the Target Window first to encode a block's color, get a block of that color from the Resource Window, look again at the Target Window to encode the block's location, then move to the Workspace Window to place the block. They report that the interaction-intensive strategy of looking twice took 3 s to execute, whereas the more memory-intensive strategy of encoding color and location at the same glance took 1.5 s to execute. They comment that "It is surprising that participants choose minimal memory strategies in view of their temporal cost" (Ballard, Hayhoe, & Pelz, 1995, p. 732).

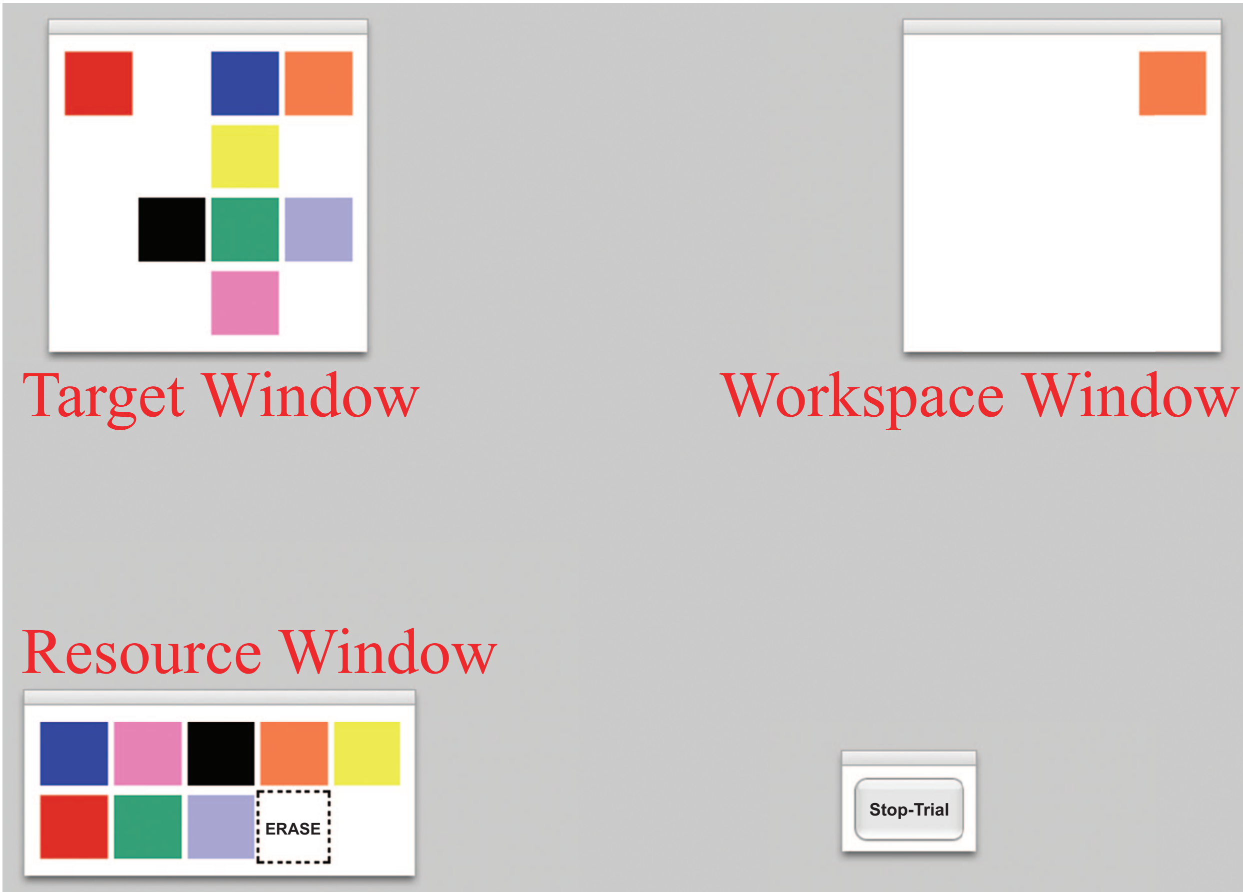

Figure 1. The Blocks World task. The figure shows a random arrangement of eight colored blocks in the Target Window (top left), eight colored blocks plus an eraser in the Resource Window (bottom left), and one block (correctly placed) in the Workspace Window (upper right). In the actual task all windows are covered by gray boxes, and at any time only one window can be uncovered. (Note that the window labels do not appear in the actual task.)

Although this dramatic bias toward perceptual-motor access costs seems to support the minimum memory hypothesis, the study that Ballard and colleagues report contains a potential confound. Participants used the interaction-intensive (i.e., mostly perceptual-motor) strategy at the beginning of the task and used the memory-intensive strategy "only at the end of the construction" (Ballard, Hayhoe, & Pelz, 1995, p. 732) of the 8-block trial. The differential use of the two strategies at different phases of construction raises the question of whether the cost of encoding required by the memory-intensive strategy was paid at the end of the trial, as Ballard seems to assume, or whether it was amortized over the entire trial. If memory for the pattern of blocks was strengthened throughout the trial (e.g., Chun & Nakayama, 2000; Ehret, 2002), by the time the last few blocks were placed, their color and position information could be retrieved from memory with little additional encoding. Hence, if encoding time is amortized over both early and late block placements, then end of trial events do not provide clean estimates of the time costs for encoding blocks in memory.

In a study involving programming a simulated VCR, Gray and Fu (2004) showed a progressive increase in errors and in trials-to-criterion as the cost of information access increased. We manipulated the cost of accessing the information required to program shows. For all groups, show information was located in a window 5 in. below the VCR window. For the Free-Access group, the show information was clearly visible at all times. For the Gray-Box and Memory-Test groups, field labels (such as Channel, Start Time, End Time, and Day-of-Week) were clearly visible, but the values of these fields (such as 32, 11:30, 12:30, and Sat) were covered by gray boxes. To access, for example, the current value of the Channel field, participants were required to move the mouse to and click on the gray box. Prior to programming a show, the Memory-Test group was required to memorize the show information (thus the term, Memory-Test).

For each group, Gray and Fu estimated the costs of accessing information in-the-head versus in-the-world. The retrieval latency for well-learned information was estimated as between 100 and 300 ms (Memory-Test group); whereas the latency for less well-learned information (the Free-Access and Gray-Box groups) was estimated as between 500 and 1,000 ms. Contrariwise, the cost of shifting visual attention and the eyes to freely accessible information in-the-world was estimated as 500 ms (Free-Access group), whereas the cost of moving the mouse, visual attention, and clicking on a gray box was estimated as 1,000–1,500 ms (Gray-Box and Memory-Test groups).

By informal standards it would seem that the Free-Access and Gray-Box groups (i.e., the two groups that were not forced to memorize show information) had easy access to perfect knowledge in-the-world; such access could easily compensate for their less than perfect knowledge in-the-head. Hence, it was somewhat surprising that the Memory-Test group made fewer errors and reached criterion in fewer trials than either of these groups. Indeed, for these two groups, performance was inversely correlated with the cost of external information access. The Free-Access group, which could obtain show information at any time by shifting their point-of-gaze by 5 in., performed better than the Gray-Box group, which had to move their mouse cursor 5 in. and click the mouse to uncover an information field.

These findings were interpreted as suggesting a race between the time costs for memory retrieval versus the time costs required either to move, click, and perceive, or to saccade and perceive. Rather than obtaining perfect information from in-the-world as they needed it, both the Free-Access and Gray-Box groups preferred to rely on knowledge in-the-head. Unfortunately, this knowledge was obtained in the course of programming a show and, as the data suggest, was not as well learned as that obtained by the Memory-Test group. Surprisingly, this increased reliance on imperfect knowledge in-the-head over perfect knowledge in-the-world was obtained even though it produced more errors and kept participants in the experiment longer. This surprise is consistent with our earlier observation that soft constraints work locally to select least-effort interactive routines. However, locally optimal interactive routines may not lead to globally optimal performance (Fu & Gray, 2004, in press).

Unfortunately, neither Ballard's study nor ours directly compared minimal memory with the soft constraints hypothesis. Neither study attempted to rule out attempts to conserve memory or to demonstrate a bias favoring perceptual-motor effort. In the work presented here, we attempt to show that differences of several hundreds of milliseconds are enough to shift the allocation of the resources used for interactive behavior from more interaction intensive to more memory intensive.

To summarize, although tradeoffs between interaction-intensive and memory-intensive strategies have been documented, it is less clear what the nature of these tradeoffs are. Gray and Fu argued (2004) that, when alternative means of performing a task exist, costs-benefit tradeoffs act as soft constraints in choosing one set of interactive routines (i.e., one pattern of cognitive, perceptual, and action operations) over another. Hence, in contrast to the minimum memory hypothesis, soft constraints posits that the control system is indifferent to the source of the resources it uses and is sensitive only to their expected utility as measured in time. Likewise, while the minimum memory hypothesis implies a bias to conserve a limited resource, soft constraints implies that the operative factor is not a limit in the number of slots or amount of activation available, but rather the time needed to encode items in memory, time required to retrieve items from memory, and the probability of retrieving an encoded item over time.

Ideal Performer Analysis

Both the minimum memory hypothesis and soft constraints hypothesis present theories for the functional mechanism underlying the selection of low-level, interactive routines. Although behavioral data will be extremely important in establishing the plausibility of the soft constraints account of resource allocation over that of the minimum memory hypothesis, it is not clear to us that behavioral data by themselves can be decisive. The minimum memory hypothesis does not deny that effort is an important factor in deciding the mix of resources brought to bear on interactive behavior. It merely asserts that, all else equal, the control system is biased to expend perceptual-motor resources to conserve memory resources. Unfortunately, it is difficult for an empirical approach to determine when "all else" is equal.

A stringent test of the two hypotheses requires behavioral data plus a modeling approach that combines two key components. In predicting human performance, Simon told us that it is vital to nail down the "side conditions" such as "visual acuity, strength, short-term memory, reaction times, and speed and limits of computation and reasoning" (Simon, 1992). Hence, the first component is a detailed and accurate estimate of the constraints or "side conditions" that bounded rationality places on human performance (Simon, 1996). In the Blocks World task, these side conditions include the time spent encoding an item; the time spent retrieving an item from memory; and the probability that retrieval will be successful given the amount of initial encoding and the retention interval. The second component is a computational or mathematical approach that is formally guaranteed to optimize temporal costs as opposed to any other metric. To conjoin these two key components (as well as several other necessary components) we combine elements of the ideal observer analysis approach from signal-detection theorists (Geisler, 2003; Macmillan & Creelman, 2004) with rational analysis (Anderson, 1990, 1991) to present an *Ideal Performer Model*.

In our case, the Ideal Performer Model will use a machine learning approach, reinforcement learning (Sutton & Barto, 1998), to optimize the tradeoff between time costs of the human perceptual-motor system and the time costs of the human memory system across the six conditions of our third Blocks World experiment. As discussed in a later section, the time of each interactive routine is derived from empirical or theoretical accounts of human cognition. Obtaining the optimal sequence of these interactive routines for each of the experimental conditions is left to a type of reinforcement learning that is formally guaranteed (Watkins & Dayan, 1992) to converge on the sequence of model components that minimizes time for each of our six conditions. Following other uses of reinforcement learning (e.g., Berthier, 1996), we make no claim that the process followed by the algorithm mimics any process followed by human cognition. We do claim, however, that the outcome of this approach approximates what would be expected if human cognition calculated costs as if milliseconds mattered. Hence, a good fit of the model to the data will be taken as support for the soft constraints hypothesis and as evidence against the minimum memory hypothesis.

The Experiments

Three experiments were conducted using the Blocks World task shown in Figure 1. As in Ballard's studies (e.g., Ballard et al., 1995, 1997) there are three windows: a Target Window containing a pattern of colored blocks, a Workspace Window where the participant must reproduce the pattern, and a Resource or parts Window containing blocks that may be picked up, carried to, and placed in the Workspace Window.

Unlike Ballard's studies, a gray window covered each of the three task windows. The Resource and Workspace Windows were uncovered as soon as the participant moved the cursor into one of the gray windows; however, the method and cost of uncovering the Target Window varied across the three studies. Experiment 1 combined an intuitive estimate of low versus medium perceptual-motor cost with a time consuming (but presumably low perceptual-motor effort) manipulation for medium versus high cost. Experiment 2 manipulated the perceptual-motor effort along with time by varying the Fitts Index of Difficulty (MacKenzie, 1992) (discussed in the following section). As the results from both of these studies suggested that the tradeoffs we observed were sensitive to time per se, and not perceptual-motor effort, Experiment 3 increased the range of access costs studied by varying lockout time of the target window across six between-subjects conditions from 0 to 3,200 milliseconds. As the three studies were very similar, we present and discuss them together.

Method

Participants

Across each of the three studies a minimum of 16 and a maximum of 18 participants were assigned to each condition. For each study undergraduates participated in the study for course credit and were randomly assigned to experimental conditions.

Equipment and Software

The experiments were conducted on Macintosh computers running versions 8.6 (Experiments 1 and 2) or 9 (Experiment 3) of the operating system. All experiments used a mouse for input and a 17-inch monitor set at 1024 × 768 resolution. Blocks World was written in Macintosh Common Lisp (MCL). All window events (e.g., mouseEnter and mouseLeave) and key presses were recorded and saved to a log file with 16.67 ms accuracy.

Design

For each 8-block pattern, each of the (48 × 48 pixel) blocks was chosen randomly with the constraint that no color be used more than twice. The blocks were placed at random in the Target Window's nonvisible 4 × 4 grid. The Workspace Window was the same size as the Target Window and contained the same 4 × 4 grid (see Figure 1).

Across all conditions of all experiments the Target, Resource, and Workspace windows were covered by gray boxes. Only one window was visible at any one time. In all three experiments, the Resource or Workspace windows opened as soon as the mouse cursor entered the window. Except for the low-access cost condition of Experiment 1 (e1-low, discussed below), all windows in all conditions stayed open for as long as the cursor remained inside of them and closed as soon as the cursor left. Across the three studies, the only difference in procedure was in the method and cost of opening the Target window. For all experiments, all manipulations were between subjects.

Experiment 1. Three levels of access cost were varied. In the low-cost condition (e1-low) the Target Window opened and stayed open when the control key on the keyboard was pressed and remained open for as long as the control key was held down or until the mouse cursor entered another window. In the medium-cost condition (e1-med) the Target window opened as soon as the cursor entered (same method and cost as to open the Resource and Workspace windows). In the high-cost condition (e1-high), a 1-s lockout was imposed between the time the cursor entered the Target window and before the window opened.

Experiment 2. To open the Target Window, all participants in Experiment 2 moved the cursor to a button located at the center of the Target window and clicked. In this experiment, the cost of accessing information was manipulated by changing the size of the button in the Target Window. For e2-low the button was as big as the window, 260 × 260 pixels. For e2-med the button was 60 × 60 pixels. For e2-high the button was 8 × 8 pixels.

Changing the button size manipulated perceptual-motor effort along with time by changing the mean Fitts Index of Difficulty (MacKenzie, 1992) for moving to the button from either the Resource or Workspace window from 1.7 (e2-low) to 2.8 (e2-med) to 6.2 (e2-high). The Fitts Index of Difficulty (ID) is a continuous scale defined as,

$$ID = \log_2\left(\frac{D}{W} + 1\right),$$

where D is the distance to the target and W is the width of the target. Fitts' law predicts movement time (MT) as, $MT = a + b \times ID$, where a is the intercept and b is the slope (these parameters are not used in computing the ID). Fitts' law is an approximation that has held up for over 50 years. Hence, although the reasons for why this equation usually works and an explanation of deviations from it continue to be researched (Meyer, Smith, Kornblum, Abrams, & Wright, 1990), the Index of Difficulty can be considered a standard and generally accepted measure of the type of information access costs varied in this study.

Experiment 3. For the third study, the buttons inside the Target Window were removed and the Blocks World display was restored to the look it had in Experiment 1 (see Figure 1). Six between-subjects conditions varied lockout time from 0 to 200 to 400 to 800 to 1,600 to 3,200 ms. Due to software errors, data from four participants were lost, one each from lockout Conditions 0, 200, 1,600, and 3,200.

Procedure

To select a block, participants moved the mouse cursor to the Resource Window and clicked on a colored block. The mouse cursor then changed to a small version (16 × 16 pixels) of the colored block. To place a block in the workspace, the cursor was moved into that window (which opened as soon as the cursor entered it), moved to the desired position, and the mouse clicked.

When the participants believed that the model pattern had been copied to the Workspace Window, they pressed the "Stop-Trial" button. The program notified the participants if the patterns differed and required them to revise or complete the pattern before they could move on to the next trial.

Misplaced blocks could be corrected at any time during the trial (i.e., before or after the Stop-Trial button was pressed). Wrong color placements could be corrected by selecting the correct color block from the Resource Window and placing it on top of the wrong color block. Wrong location placements could be corrected by selecting a white "erase" block from the Resource Window and placing this on top of the wrong location block.

For each experiment, all participants received instruction by being led by the experimenter through a PowerPoint™ demonstration. Within each experiment, the same slides with the same prerecorded narration were provided to each group. After this demonstration, the participants completed one practice trial while the experimenter watched and answered any questions the participant might have. As the participant typically had no problems with this practice trial, the experimenter typically said nothing. After the practice trial the experimenter left the room and the participants completed the remaining 39 trials in Experiment 1 and 47 trials in Experiments 2 and 3 by themselves. All experiments lasted approximately 45 minutes.

Results

For each experiment, we provide one general measure of the differences between conditions and then focus on two specific measures. The general measure is a count of the mean number of times during a trial that the Target Window was uncovered. The two specific measures look at events surrounding the first uncovering of the Target Window: median duration of the first uncovering and mean number of correct placements following the first uncovering. There are two rationales for focusing on events surrounding the first uncovering. First, for each trial, at the time of the first uncovering of the Target Window, there were eight not-yet-placed blocks. For all subsequent uncoverings, the mean number of not-yet-placed blocks varied between conditions. Comparing across conditions is easiest when the number not-yet-placed is equal for each condition. Second, focusing on events prior to the second and subsequent uncoverings avoids any potential confound with any cumulative memory trace for the block pattern. This ensures that the measures of duration and correct placements can be attributed to events surrounding the first uncovering and are not influenced by a cumulative memory trace for the block pattern.

As we are interested in the strategies that participants use after they adapt to the access costs in their condition, the first 10 trials were eliminated, and for each participant on each measure either the mean or median score (depending on the measure) across Trials 11–40 (Experiment 1) or 11–48 (Experiments 2 and 3) was used.

For each of the three experiments, an independent analysis of variance (ANOVA) was performed on each dependent variable. A summary of all ANOVAs performed on each dependent variable is provided in Table 1. The mean or median scores for Experiments 1–3 are reported in Tables 2-4, respectively.

Table 1
Analysis of Variance Table for All Dependent Measures for Each of the Three Experiments

Experiment	Degrees of freedom	F-value	Mean-square error	Significance level (p)
Number of target window accesses				
E-1	(2, 45)	7.53	34.50	.0015
E-2	(2, 51)	9.27	10.83	.0004
E-3	(5, 104)	11.60	16.99	.0001
Duration of first look				
E-1	(2, 45)	9.16	6,756,009	.0005
E-2	(2, 51)	6.01	8,055,996	.0045
E-3	(5, 104)	13.18	26,924,234	.0001
Blocks correctly placed following the first look				
E-1	(2, 45)	9.84	6.56	.0003
E-2	(2, 51)	8.85	3.72	.0005
E-3	(5, 104)	17.39	5.85	.0001

Table 2
Mean Results for Experiment 1 Over Trials 11–40

	Information access condition		
	(keypress) Low	(0-lock) Medium	(1000-lock) High
Number of target window accesses	6.8	6.4	4.1
Duration of first look (ms)	1179	1241	2334
Blocks correctly placed (first look)	1.7	1.9	2.9

Number of Target Window Accesses

Each study showed a main effect of access cost condition on the mean number of times the target window was accessed (see the top third of Table 1). For Experiment 1 (see Table 2), a series of three planned comparisons showed that accesses for e1-low and e1-med did not differ, but that each made more accesses than e1-high (low vs. high, $p = .0008$; med vs. high, $p = .0039$). For Experiment 2 (see Table 3), a series of three planned comparisons revealed e2-low > e2-med ($p = .016$) and e2-low > e2-high ($p < .0001$), but that e2-med did not significantly differ from e2-high. For Experiment 3 (see Table 4), the slope of the linear trend across conditions significantly ($p < .0001$) differed from zero and accounted for 98% of the variance for condition. The linear trend shows that the changes across the six conditions are all in the same direction.

Duration of First Look

Each study showed a main effect for condition on the median duration that the Target Window stayed open on its first access (see the middle rows of Table 1). For Experiment 1 (see Table 2), planned comparisons showed significant differences (p's $< .001$) between e1-high and each of the other two conditions. There were no differences between e1-low and e1-med. For Experiment 2 (see Table 3), a series of three planned comparisons revealed e2-low < e2-med ($p = .035$), e2-low < e2-high ($p = .0012$), but that e2-med did not significantly differ from e2-high. For Experiment 3 (see Table 4), the linear trend across conditions was significant ($p < .0001$) and accounted for 87% of the variance for condition.

Blocks Correctly Placed Following the First Look

This measure examined the mean number of blocks placed after the first look that correctly matched the color and location of a block in the Target Window. Across all three studies the differ-

Table 3
Mean Results for Experiment 2 Over Trials 11–48

	Information access condition		
	Low-ID	Med-ID	High-ID
Index of difficulty	1.7	2.8	6.2
Number of target window accesses	5.1	4.2	3.5
Duration of first look (ms)	1345	2182	2669
Blocks correctly placed (first look)	2.22	2.69	3.13

Table 4
Mean Results for Experiment 3 Over Trials 11–48

	Information access condition (lockout duration in ms)					
	0	200	400	800	1600	3200
Number of target window accesses	5.6	4.8	4.5	3.7	3.5	2.9
Duration of first look (ms)	1603	1702	1929	2392	3614	4634
Blocks correctly placed (first look)	2.00	2.39	2.49	2.94	3.11	3.58

ences across conditions were significant (see bottom third of Table 1). For Experiment 1 (see Table 2), a series of three planned comparisons revealed a significant difference between e1-high and each of the other two conditions (see Table 2, $p = .0015$). For Experiment 2 (see Table 3), planned comparisons revealed e2-low < e2-med < e2-high (e2-low vs. e2-med, $p = .034$; e2-low vs. e2-high, $p = .0001$; e2-med vs. e2-high, $p = .048$). For Experiment 3 (see Table 4), the linear trend across conditions was significant ($p < .0001$) and accounted for 97% of the variance for condition.

Discussion of the Experimental Data

Each of the three studies found a progressive switch from more interaction-intensive to more memory-intensive strategies as information access costs increased. The number of times the Target Window was opened decreased, while the duration that it was opened increased. Presumably, the increased duration that the Target Window was opened reflects increased time spent encoding its contents. This interpretation is supported by the increase in the number of blocks placed following the first look. As access costs increase, people minimize time per trial by accessing the Target Window less and using memory more.

Differences Between Methods of Information Access

Across the three studies we varied the method of accessing the Target Window. For Experiment 1 we were disappointed to find no significant differences between the e1-low and e1-med conditions on any of our three measures. Our intuitive notions of effort seem not to have produced the expected difference. Could these results be better understood by using access time to characterize the differences between conditions in access costs?

Unfortunately, access time for the Experiment 1 conditions is hard to compare since for e1-low the log file only collected the time at which the control key was pressed and for e1-med and e1-high the log file only reported the time at which the cursor entered the Target Window. However, in prior research (Gray & Boehm-Davis, 2000), we measured key down time as 100 ms. For the Blocks World paradigm, we estimated the time to move the cursor into the Target Window as 146 ms. This estimate is the average of the Fitts' law (MacKenzie, 1992) time to move the cursor to the Target Window from the Workspace and Resource Window. Hence, by these estimates the difference in expected time between e1-low and e1-med is 46 ms[2] (i.e., 146 ms for el-med minus 100 ms for e1-low), 1,000 ms between e1-med and e1-high (due to the 1,000 ms lockout for e1-high), and 1,046 ms between e1-low and e1-high.

If access costs are measured in time, then the Experiment 1 results are very regular. As access time increased, participants opened the Target Window less often, but the duration of the look increased, as did the number of correct and incorrect retrievals from memory. Although the e1-low versus e1-med difference in access time of 46 ms was not enough to produce significant differences, it was enough to produce the expected pattern across the three measures. All three measures found a significant difference between e1-high and each of the other two conditions.

Experiment 2 replicated the results of Experiment 1 using a manipulation that covaried difficulty of perceptual-motor activity with time. The Experiment 1 and 2 results suggested that, for the Blocks World task, time is the operative factor and it does not matter whether time for information access is manipulated by varying the Fitts Index of Difficulty or by lockout. We tested this suggestion in Experiment 3 by using six levels of lockout time as our independent variable. The use of lockout time in Experiment 3 also enabled us to more precisely control access time while also producing a wider range of access costs. Hence, Experiment 3 provides our best empirical test of the notion that access costs can be measured by access time.

Across three studies, the empirical data support the view that as access costs increased participants switched from more interaction-intensive to more memory-intensive strategies. This strategic switch was signaled by the decreasing number of openings of the Target Window across conditions as well as by the increasing duration that the Target Window was open. We argue that the increase in the duration that the Target Window is open reflects the greater amount of time that participants spent encoding the contents of the Target Window. This explanation is supported by the increase across conditions in the number of correct block placements following the initial uncovering of the Target Window.

[2] Alternative bases exist for estimating time difference in these two conditions. An alternative we tried was based on CPM-GOMS (Gray & Boehm-Davis, 2000; Gray et al., 1993). As the difference predicted by those models is 51 ms, we have elected to report and explain the simpler difference between keydown time and movement time (46 ms), rather than providing the level of detail required to understand the CPM-GOMS models.

Limits of the Experimental Data

The empirical data demonstrate that as access costs increase people adjust their strategies to be less interaction intensive and more memory intensive. However, although we view the steady increase in tradeoffs as persuasive evidence in support of the soft constraints hypothesis, the empirical data do not rule out weaker forms of the minimum memory hypothesis. For example, the soft constraints hypothesis argues that as information access costs increase, the use of interaction-intensive versus memory-intensive strategies is driven by their expected utility (i.e., cost-benefit tradeoff) as measured by time. The empirical data show a shift in strategies but, by themselves, do not relate the shift to expected utility. To make this argument, in the next section, we turn to a machine-learning algorithm, reinforcement learning, that is formally guaranteed to maximize expected utility (using time as its metric) if provided with sufficient training and adequate exploration of the problem space (Sutton & Barto, 1998). In fitting the model, the six between-subjects conditions of Experiment 3 will provide data on multiple measures against which to compare the predictions of the soft constraints hypothesis against the implications of the minimum memory one. As discussed in the next section, conformity to the reinforcement learning solution would support the soft constraints hypothesis. In contrast, deviations from the reinforcement learning solution would support the minimum memory hypothesis.

Ideal Performer Analysis: Ideal Observer Analysis Meets Rational Analysis[3]

Our ideal performer analysis combines elements of an ideal observer analysis (Geisler, 2003; Macmillan & Creelman, 2004) with those of rational analysis (Anderson, 1990, 1991). The ideal observer analysis (Geisler, 2003; Macmillan & Creelman, 2004) is used to "determine the optimal performance in a task, given the physical properties of the environment and stimuli" (Geisler, 2003). The ideal observer may be degraded in a systematic fashion by including side conditions, "for example, hypothesized sources of internal noise (Barlow, 1977), inefficiencies in central decision processes (Barlow, 1977; Green & Swets, 1966; Pelli, 1990), or known anatomical or physiological factors that would limit performance (Geisler, 1989)" (Geisler, 2003). In Simon's term (1992), the ideal performer analysis allows us to determine optimal performance given "side conditions" that represent the known limits of the performer.

Rational analysis "involves three kinds of assumptions: assumptions about the goals of a certain aspect of human cognition, assumptions about the structure of the environment relevant to achieving these goals, and assumptions about costs. Optimal behavior can be predicted by assuming that the system maximizes its goals while it minimizes its costs" (Anderson, 1990, p. 244).

Conjoining the ideal observer analysis with rational analysis yields four components of our ideal performer analysis: a description of the task environment; the systematic degradation of the ideal observer by adding in known human limits; defining sequences of interactive routines that allow us to characterize interactive behavior as more interaction intensive or memory intensive; and the optimal (ideal) sequencing of these interactive routines so as to minimize total time. Each of these aspects of the Ideal Performer Model is discussed in the sections that follow.

Hard Constraints: Defining the Task Environment

The goals of the human performer combined with the physical properties of the task environment act as hard constraints on how the task is performed. Given the task environment shown in Figure 1 and the goal to reproduce the pattern of Target Window blocks in the Workspace Window, then the task analysis breaks the task into a series of ENCODE-k strategies where k is the number of blocks (1–8) encoded on each round. Each ENCODE-k strategy consists of two unit tasks, an Encode Blocks unit task and a Get & Place unit task. As shown in the pseudocode provided as Table 5, the first unit task encodes some number of blocks from the Target Window pattern (lines 1–9) and the second gets blocks from the Resource Window and places them into the Workspace Window (lines 10–25).

This top level of description is completely objective in that it is based on the goals of the task and the task environment available for achieving these goals. For guidance on how to flesh out the interactive routines required by each unit task we turned to an ACT-R model that performed the task using the same experimental software as the human participants in Experiment 3 (Gray, Schoelles, & Sims, 2005). Although that model lacked a mechanism for optimizing time, it did provide a detailed cognitive task analysis that allows us to break each unit task down further. Each line with an entry in the cost column of Table 5 represents an interactive routine. If we further fleshed out the model, each interactive routine would be composed of an activity network of cognitive, perceptual, and motor operations (as illustrated and discussed in Gray & Boehm-Davis, 2000).

For the Encode Blocks unit task the performer must shift visual attention to and move the mouse into the Target Window (lines 2 and 3). Between conditions, hard constraints built into the task environment determine how long the performer must wait until the window opens (line 4). Once the Target Window is open, the performer encodes one or more blocks (lines 5–9). The number of blocks encoded in memory is not constrained by the task environment, and in our Ideal Performer Model the choice of number of blocks to encode corresponds to the selection of a particular ENCODE-k strategy. (The issue of selecting ENCODE-k strategies is discussed in the next section.) Functionally, the process of encoding a block in our model corresponds to creating a new declarative memory element (see Appendix A) and rehearsing the element by performing two retrievals before moving on to the next block.

The second unit task is Get & Place. In this unit task the performer must move visual attention and the mouse cursor into the Resource Window (lines 11–12), which then opens. The performer must then remember the color of an encoded, but not-yet-placed block, move to a block of that color, and click on the color.

[3] An annotated Common Lisp file of the model is available at the APA archive site for *Psychological Review* and is posted on our website http://www.rpi.edu/~grayw/pubs/papers/GSFS06_PsycRvw/GSFS06_PsycRvw.htm.

Table 5
Pseudo-code for the Ideal Performer Model

Line #	Cost (in ms)	Operation
00		Select strategy: ENCODE-k (where k = # of blocks to be encoded this round)
01		Unit Task: Encode Blocks
02	185	Shift visual attention to Target Window
03	217	Move mouse to Target Window
04	0-3200	Wait for lockout duration [Between-group independent variable]
		[System Event: Target Window opens]
05		Do Encode Blocks
06	185	Shift visual attention to a new block
07	50	Encode a new declarative memory element (DME)
08	Eqn. A-2	Rehearse the encoded DME (perform 2 retrievals)
09		Until k blocks have been encoded
10		Unit Task: Get & Place Encoded Blocks
11	185	Shift visual attention to Resource Window
12	249	Move mouse to Resource Window
		[System Event: Target Window closes and Resource Window opens]*
13		Do
14	Eqn. A-2	Attempt to retrieve the DME of an encoded, but not placed block
15		If a DME was retrieved
16	150	Move mouse to the block color (in the Resource Window)
17	150	Click on the block color
		[System Event: Cursor changes to 8 × 8 colored square]
18	185	Shift visual attention to Workspace Window
19	216	Move mouse to Workspace Window
		[System Event: Resource Window closes and Workspace Window opens]*
20	150	Move mouse to the block position in Workspace Window
21	150	Click on the position
		[System Event: Cursor changes to default arrow cursor]
22	185	Shift visual attention to Resource Window
23	249	Move mouse from Workspace Window to Resource Window
		[System Event: Workspace Window closes and Resource Window opens]*
24		End if
25		Until all encoded blocks are placed or a retrieval failure occurs
26		Until Workspace Window pattern matches the Target Window pattern
Apply Q-learning update rule using total time from the Encode + Get & Place unit tasks as penalty		

Note. Successful performance requires selecting a continual series of ENCODE-k strategies until the pattern in the Workspace Window matches that in the Target Window.
* Each window closes as soon as the cursor leaves it and before the cursor enters another window.

(At this point the cursor changes to a 16 × 16 pixel block the same color as the block that was selected.) The performer then moves the mouse and visual attention to the Workspace Window (which then opens), locates and moves the cursor to the position of the block, and clicks. (The cursor then changes back to the system default arrow cursor.) The performer then moves back to the Resource Window (which again opens) and attempts to retrieve another encoded, but not-yet-placed, block.

Adding Side Conditions to the Ideal Performer

Within the cognitive task analysis defined by the pseudocode of Table 5, the column "cost (in ms)" defines known human limits, or side conditions, to each step. The time to shift visual attention, 185 ms (lines 2, 6, 11, 18, 22), is taken from the estimate used by ACT-R (Anderson & Lebiere, 1998, pp. 150–151) for human attention to move to an object at a known location. All movement times (lines 3, 12, 16, 19, 20, 23) are based on the Fitts' law times (MacKenzie, 1992) to move a given distance to an object of a given size. We used the default ACT-R parameters for Fitts' law (a = 0.05; b = 0.10). These parameters are based on those established by Card, English, and Burr (1978) and have been shown to provide a good fit to moving a mouse cursor around a computer screen. Times to click on a block or position (lines 17, 21) are based on times from Gray and Boehm-Davis (2000) and includes an estimate of 50 ms to initiate the action and 100 ms to execute the click.

A key source of constraints imposed on the ideal performer is the memory limitations resulting from a fallible human memory (lines 8, 14 of Table 5). The estimates of retrieval times and probability of retrieval were based on the theory of memory incorporated into ACT-R (Anderson & Lebiere, 1998; Lovett, Reder, & Lebiere, 1999). According to Anderson's rational analysis of memory (Anderson, 1990; Anderson & Schooler, 1991), out of the multitude of memories that have been formed over a lifetime, any given memory should be made available to the performer according to the probability of its being needed as determined by its prior history of retrieval and relevance to the current environmental context. Implications of this approach have been validated across a wide range of tasks and task environments (Altmann, in press; Altmann & Gray, 2002; Lovett et al., 1999; Schooler & Hertwig, 2005; Todd & Schooler, in press). The functional consequence of this memory limitation is that if the model tries to encode, say, 5 blocks, it will have some probability of recalling and placing 5, 4, 3, 2, 1, or 0 blocks. (See Appendix A for a discussion of ACT-R's treatment of declarative memory). Encoding (line 7) and rehearsing (line 8) takes time as do attempts at retrieval (lines 8, 14). An item that is encoded but not retrieved adds cost but no benefit to task performance.

Defining Sequences of Interactive Routines: Generating Interaction-Intensive Versus Memory-Intensive Behavior

In the model of the Blocks World task, there are a maximum of eight possible ENCODE-k strategies. Each ENCODE-k strategy corresponds to encoding *k* blocks in memory, and then attempting to place those blocks in the Workspace Window. At the beginning of each trial eight strategies are available to the performer, ENCODE-1 through ENCODE-8, which correspond to actions available to the reinforcement learning agent. Along with the eight possible actions, there are eight possible states of the task. These states correspond to the number of blocks remaining to be placed into the Workspace Window. For example, if there are only 2 blocks left to place in the current trial, then only actions ENCODE-1 and ENCODE-2 are available to the performer. Across all task states there are $8 + 7 + 6 + 5 + 4 + 3 + 2 + 1$ or 36 possible state-action pairs. It is the sequence of state-action pairs that the performer chooses that enables us to characterize performance as more interaction-intensive or memory-intensive—consistently choosing the ENCODE-1 strategy corresponds to an extreme interaction-intensive strategy, while consistently choosing ENCODE-8 corresponds to an extreme memory-intensive strategy.

Defining an Objective Function to Optimize Sequencing of Interactive Routines

Unfortunately, we cannot predict the sequence of state-action pairs used across the six conditions of Experiment 3 simply from knowing the task structure and human performance limits. In addition to these constraints, a numerical *objective function* must be specified for an ideal performer to maximize its achievement according to this function. Although the constraints on human performance discussed above were based on hard constraints inherent in the task environment, previous research, or well-established theory, the selection of an objective function that would determine the sequence of state-action pairs is not so clearly defined.

One objective function might be provided by the minimum memory hypothesis. A strict, literal interpretation of this hypothesis suggests only that the ideal performer seeks to minimize the burden on its memory system. A direct way to maximize this objective function would be by choosing the ENCODE-1 strategy on each round, regardless of lockout cost. Although this extreme interaction-intensive strategy trivially fails to account for human performance in the Blocks World task, it is also a rather severe oversimplification of the minimum memory hypothesis. Other interpretations of the minimum memory hypothesis might only specify a penalty for exceeding a specified capacity limitation (e.g., by encoding more than 4 blocks at a time), or specify a bias toward interaction-intensive strategies in terms of a weight parameter. Unfortunately, as far as we know, there is no version of the minimum memory hypothesis specific enough to implement as a computational model.

In contrast, the soft constraints hypothesis makes a clear prediction regarding the objective function that should be maximized. If, as the soft constraints hypothesis assumes, the cognitive system is indifferent to the type of internal resources it exploits as well as to the location of the information it accesses (in-the-world vs. in-the-head) then it should simply maximize expected utility according to a cost-benefit tradeoff between competing interactive routines. The cost estimates defined in Table 5 can be used to maximize performance by selecting ENCODE-k strategies that minimize the total expected time to complete each trial for each of the six between-subjects conditions of Experiment 3.

Unfortunately, while specifying a suitable objective function is straightforward, maximizing achievement of the objective function to determine optimal performance is not an easy task. For example, if there remain 5 blocks to be placed, is the fastest strategy to ENCODE-5? Or, would the sequence ENCODE-3 and ENCODE-2 be faster, due to greater probability of successfully retrieving every block that was encoded? Further, how does the expected utility of each ENCODE-k strategy change across experimental conditions? Whatever the best solution, it is clear that given the probabilistic nature of memory, applying the soft constraints hypothesis to define the optimal strategy is not a simple matter.

To some degree, humans have some metacognitive sense regarding how likely they are to remember something, given how much effort they are willing to spend memorizing it, and given the length of time they need to remember it. For example, when looking up a telephone number in a directory, the time spent committing the number to memory reflects a tradeoff between the time it must be held in memory and the time required to relocate the number if it is forgotten while walking across the room to the telephone. In general, there seem to be many life events when information is temporarily needed and we make a tradeoff between encoding effort, retention interval, and the cost of reacquiring information if we forget it. Our ability to negotiate this tradeoff with our own memory limitations comes through experience remembering and forgetting things amortized over a lifetime of practice. However, given the varied nature of demands on memory, it does not seem likely that this metacognitive tuning would yield an immediate, optimal solution to each new memory challenge. In the case of the Blocks World task, we found that participants required on the order of 10 trials to fine tune their strategies to match the demands of the experimental condition.

A Reinforcement Learning Solution to the Objective Function

The final component of the ideal performer analysis is a formal mechanism for maximizing performance according to the objective function while simultaneously satisfying the constraints imposed by the human performer as well as the task itself. In our model we employed a reinforcement learning algorithm, Q-learning, that is formally guaranteed to converge on the optimal solution to this tradeoff if provided with sufficient training and adequate exploration of the problem space (Sutton & Barto, 1998; Watkins & Dayan, 1992). Reinforcement learning is a family of machine learning techniques in which agents learn directly from the outcomes of their actions. Reinforcement learning entails an unsupervised, trial-and-error exploration of the task environment, in which rewards can be defined in terms of minimizing solution time.

In recent years researchers in the neurocognitive community have examined reinforcement learning as a plausible model of how humans learn from their mistakes (Dayan & Abbott, 2001; Holroyd & Coles, 2002). The technique has also recently attracted the attention of the greater cognitive modeling community (Fu & Anderson, 2004, 2006; Nason & Laird, 2004; Phillips & Noelle, 2004; Wu & Liu, 2004). However, for the purpose of this research we are interested in reinforcement learning not as a theory of human cognitive functioning, but rather as a tool for determining optimal performance by maximizing expected utility under a set of explicit constraints. Reinforcement learning has similarly been used to approximate optimal motor control in reaching tasks and as a model of motor learning (Berthier, 1996; Berthier, Rosenstein, & Barto, 2005).

As discussed earlier, the Blocks World task has 36 state-action pairs defined by the number of states (i.e., not-yet-placed blocks can range from 1 to 8) and number of ENCODE-k strategies that can be applied to each state. The value function computed by reinforcement learning, $Q(s,a)$ (see Appendix B), ranges over these 36 state-action pairs. Each time the model completes an ENCODE-k strategy, it is penalized using the Q-learning update rule by the total time required to complete the strategy (the total duration for the Encode Blocks and Get & Place Encoded Blocks unit tasks, see Table 5). Over time, the value function learned by the Ideal Performer Model corresponds to its estimate of how long it will take to complete the entire trial given that a particular action is chosen in a particular state.

In introducing the soft constraints hypothesis, we wrote of maximizing expected utility in terms of a cost-benefit tradeoff. In implementing the soft constraints hypothesis in a reinforcement-learning approach, the outcomes of actions are defined only in terms of their local cost. Benefit in the model is implicitly defined as minimizing global costs—that is, the time required to complete an entire trial. Hence, a strategy that encoded 8 blocks, forgot 5, and placed 3 would not be as beneficial as a strategy that encoded and placed 3 blocks. The former strategy has wasted time encoding 5 blocks that it did not place. These 5 blocks require at least one other round of ENCODE-k strategy. Hence, in the reinforcement-learning model, just as costs are defined by time, benefits are defined as minimizing time. Optimizing benefits entails minimizing costs.

Summary of the Ideal Performer Analysis

The ideal performer analysis combined elements of a traditional ideal observer analysis (Geisler, 2003; Macmillan & Creelman, 2004) with a rational analysis (Anderson, 1990, 1991) to produce our Ideal Performer Model. At the top level of description, the requirements of the model were defined by the goals of the task and the task environment. We fleshed out the model with a cognitive task analysis that was based on an ACT-R model that performed the task using the same experimental software as the human participants in Experiment 3. The time required to perform each step in the model (see Table 5) was based on the known limits of the human performer. Most of the times for cognitive, perceptual, and motor operations reflected accepted estimates for performance. In our case, we took these times from the estimates used by ACT-R; however, the ACT-R estimate for these times is generally consistent with that of EPIC (Kieras & Meyer, 1997) as well as the much older Model Human Processor (Card et al., 1983; Newell, 1990). The most notable limit we discussed was the time required to encode an item into memory, the time required to later retrieve that item, and the probability that retrieval would be successful. Our estimate of these times and probabilities are directly derived from Anderson's rational analysis model of memory (Anderson, 1990; Anderson & Schooler, 1991).

Performing the Blocks World task was defined as a series of choices among ENCODE-k strategies for each state of a Blocks World trial. Optimizing this series of choices by an objective function that minimizes total time (according to the soft constraints hypothesis) is a hard problem in large part due to the probabilistic nature of human memory. As we lack a cognitively valid formal mechanism for maximizing achievement of this objective function, we turned to a reinforcement learning technique, Q-learning, that is formally guaranteed to find an optimal solution if certain assumptions are met. The training, testing, and performance of this Ideal Performer Model are reported in the next section.

Predictions From the Ideal Performer Model

In this section, we first walk through the training procedure as well as the utility estimates and memory estimates derived from the training phase. Next we compare model performance with human performance on each of the three dependent variables discussed in the experimental section: blocks correctly placed following the first look, duration of first look, and the per-trial number of target window accesses. From the measure of blocks placed following first look, we derive a fourth measure: the probability across lockout conditions that participants will place 0 to 8 blocks. This measure is also compared with model performance.

Training the Ideal Performer Model

For each of the six lockout conditions, the model was first trained for 100,000 trials. Although the model only had to explore 36 state-action pairs, in the Blocks World task completing a single trial requires a sequence of actions (i.e., multiple rounds of ENCODE-k strategies where each round is represented by the pseudocode in Table 5), and the outcomes of each action are probabilistic. If the model encodes 4 blocks (the ENCODE-4 strategy), there is some probability that it will actually place 4, 3, 2, 1, or 0 blocks.

For the case in which each ENCODE-k strategy results in the deterministic placement of a single block, there would be 8! or 40,320 different action sequences. As each action can result in as few as zero placements and one can result in as many as eight, the potential number of action sequences is very great. However, for

Table 6
The Utility Estimates Learned by Q-learning for the Initial State (8 to-be-placed blocks) of the Blocks World Task

Lockout (ms)	ENCODE-k strategy utilities (seconds) 1	2	3	4	5	6	7	8
0	−26.503	−26.421#	−26.551	−26.766	−26.994	−27.219	−27.574	−27.879*
200	−27.624	−27.439#	−27.488	−27.642	−27.814	−28.057	−28.301	−28.596*
400	−28.478	−28.271#	−28.275	−28.366	−28.529	−28.738	−28.943	−29.243*
800	−30.409	−30.047	−29.935#	−29.937	−30.020	−30.129	−30.291	−30.560*
1600	−33.629*	−33.052	−32.845	−32.726	−32.627#	−32.689	−32.769	−32.926
3200	−39.748*	−38.899	−38.379	−38.043	−37.786	−37.662	−37.609	−37.607#

Note. For each lockout condition # indicates the best Encode-k strategy and * indicates the worst.

the ACT-R memory equations (see Appendix A) and the memory parameters used in the study (see Appendix C) placements at the extremes (e.g., 0 or 8) will be very rare. Given these considerations and our experiences with Q-learning in the Blocks World paradigm, 100,000 training trials seem reasonable though somewhat conservative.

The challenge for the reinforcement-learning model is to extrapolate from local rewards following each ENCODE-k strategy to an estimate of the time required to complete an entire trial for each action and in each state. During the training, the model explored actions at random.[4] This ensured that it gained extensive experience with each combination of ENCODE-k strategy at every phase of a Blocks World trial.

The output of the Ideal Performer Model consists of two sets of information. The first is the table of utility estimates for each state-action pair. During training, the model was penalized by the negative time required for each ENCODE-k strategy. Under this approach, maximizing rewards corresponds to minimizing total time. Following training, the utility estimates correspond to the estimated minimum time required to complete the entire trial given that a specific action is chosen in the current state. Table 6 shows the utility estimates for the eight strategies available at the initial state (i.e., 8 to-be-placed blocks) of the trial. As the table shows, choosing a suboptimal action in the Blocks World task involves relatively little penalty—for each lockout condition, the difference between the best and worst ENCODE-k strategy for the first visit to the target window is on the order of 1 to 2 seconds. Given the small range of expected utilities, it is not obvious that participants in the task should be sensitive to these differences. As such, the ability of the Ideal Performer Model to fit the human data provides a strong test of the claim that time cost acts as a soft constraint in the Blocks World task.

The second piece of information produced by the Ideal Performer Model is the number of blocks successfully recalled and placed as a function of the number encoded in memory. The model's memory performance is jointly determined by the ACT-R memory equations and the retention interval imposed by the Blocks World task. The memory equations involve three parameters: a retrieval threshold, an activation noise parameter, and a latency scaling parameter, (see Appendix C). During training, the model's memory performance was recorded for each ENCODE-k strategy, producing the distribution of blocks placed that is shown in Figure 2.

From these two sets of information, the utility table (see Table 6) and memory performance (see Figure 2), it is possible to make a number of predictions for human performance in the Blocks World task. Although the utility table defines the optimal strategy for the first visit to the target window (deterministically choose the strategy with the highest utility), we have theorized that time is a soft, as opposed to hard, constraint in the task. Consequently, we expect that participants will not always select the optimal strategy, but rather will approximate the optimal policy to the extent that their behavior is influenced by time as a soft constraint. To transform a utility estimate into a selection probability, we used ACT-R's strategy selection equation, the "softmax" rule, which has also been widely used in other reinforcement learning models (Sutton & Barto, 1998). The probability of selecting strategy ENCODE-k at the start of a trial is related to its utility, U_k, as well as to the utility of all competing strategies:

$$p(k) = \frac{e^{U_k/t}}{\sum_{j=1}^{8} e^{U_j/t}}.$$

In this equation, t is a noise parameter controlling the probability that the model chooses a suboptimal strategy. As t approaches 0, the model will deterministically select the optimal strategy. Because of this property, the noise parameter reflects an estimate of the "softness" of time as a constraint on behavior.[5]

Given the probability of selecting each ENCODE-k strategy, $p(k)$, and the probability of placing a number n blocks given that strategy k has been selected, $p(n|k)$, it is possible to directly

[4] It might be objected that by exploring actions at random the model will only learn the utility of the random behavior policy. However, as *Q*-learning is an off-policy learning algorithm (Sutton & Barto, 1998), it is still able to learn the optimal policy through random exploration, and this approach produces the fastest learning by maximizing exploration of the full state space.

[5] The noise parameter t is related to the standard deviation of a logistic distribution according to $t = \sqrt{6}\sigma/\pi$. Since utility in our model is defined strictly in terms of time, this allows us to determine the probability that our model will discriminate between two strategies with a given difference in expected time cost. Using the value of t fit to our data ($t = 0.491$, see Appendix C), for a time difference of 1 second between competing strategies, the model will select the faster strategy on 88.5% of its choices.

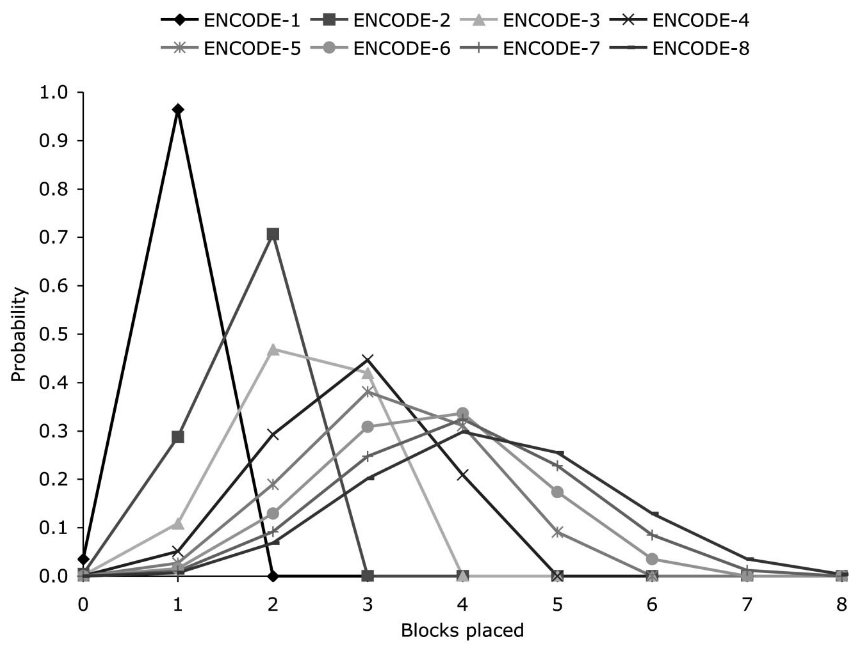

Figure 2. Probability of retrieving and placing n blocks given that k blocks have been encoded, $p(n|k)$, for each ENCODE-k strategy.

calculate the distribution of blocks placed following the first visit to the target window. If x is a random variable representing the number of blocks placed, then its distribution is given by:

$$p(x = n) = \sum_{k=1}^{8} p(n|k)p(k).$$

Likewise, the mean number of blocks placed is calculated as the expected value of x:

$$\bar{x} = E[x] = \sum_{n=1}^{8} p(x = n) \cdot n.$$

The ideal performer analysis also makes predictions about two other empirical measures reported for the human participants. The mean duration of the first look to the target window is jointly determined by the estimated costs from the task analysis in Table 5 and the probability of selecting each ENCODE-k strategy. Finally, the expected number of visits to the target window can be determined using Monte Carlo simulation of the Ideal Performer Model.[6] The next section presents the comparison of the model predictions to human performance for each of these measures.

Testing the Ideal Performer Model

The predictions of the Ideal Performer Model are dependent on four parameters (three parameters for the memory equations and one noise parameter for the strategy selection equation). The values for each parameter were fit to the human data on the key measure of number of blocks placed following the first look to the target window. The best-fitting parameters for the memory equations were determined using a grid search using a range of values based on previously published ACT-R models or established default values.[7] The noise parameter for the strategy selection equation was determined using least square error minimization. The best-fitting values for all the parameters, as well as estimates of perceptual-motor times used in the model are reported in Appendix C. The same parameter settings were used to produce all of the model predictions.

For the key measure of number of blocks placed following the first uncovering of the target window, the model has an RMSE of 0.092 and r^2 to the human data of 0.969 (see Figure 3). Although the standard error for the human data is quite low, the difference between the model's prediction and human performance is within 1 standard error for five of the six lockout conditions (for the 800-LOCK condition the model is within 1.15 standard errors).

Figure 4 compares the distribution of blocks placed following the first visit to the target window. The model showed an excellent fit to the human data, with an overall RMSE of 0.034, and r^2 =

[6] In theory, it may be possible to produce closed-form predictions for the number of visits rather than relying on Monte Carlo simulation. However, the number of visits is determined by the conditional probabilities of selecting each strategy on each visit, as well as the probabilistic outcome of each strategy, resulting in computations that quickly become unwieldy.

[7] Specifically, the latency parameter F was examined in the range 0.9–1.2 in increments of 0.1 units; the retrieval threshold was examined in the range 0.25–0.35 in increments of 0.025; and activation noise was examined in the range 0.28–0.32 in increments of 0.02. A grid search over a relatively small parameter space was necessary as changing any of the memory parameters requires re-training and running the Q-learning model, preventing more efficient gradient-based parameter fitting methods.

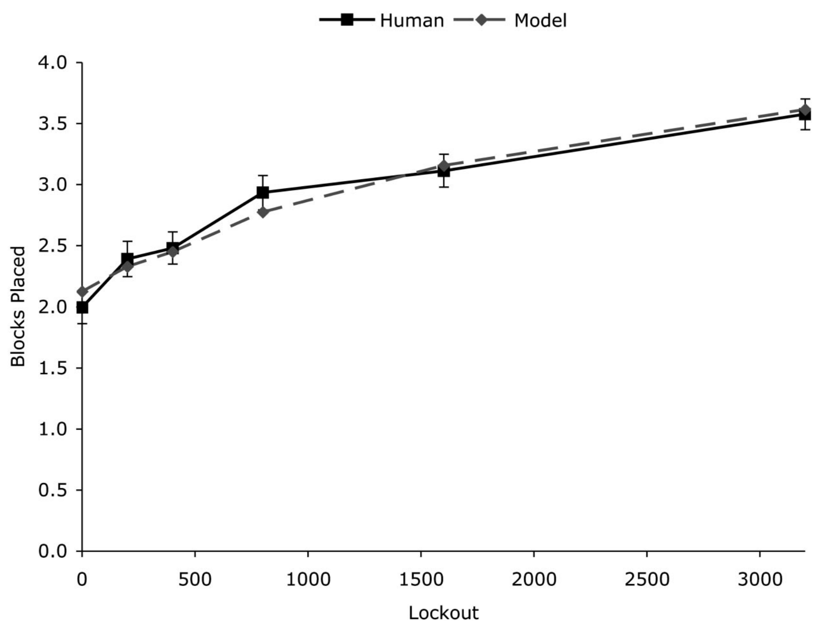

Figure 3. Number of blocks placed following first uncovering for human participants (Experiment 3 with +/− 1 standard error bars) and the Ideal Performer Model.

[0.892, 0.887, 0.947, 0.902, 0.958, 0.953] for the 0-LOCK through 3200-LOCK conditions respectively.

For the mean duration of the first look at the target window, the Ideal Performer Model also closely predicts the human data. The model prediction has an RMSE of 0.431 and r^2 of 0.980 to the human data, shown in Figure 5.

Finally, the model's prediction for the number of visits to the target window also closely matches human performance in the task, with an RMSE of 0.397 and $r^2 = 0.970$ (see Figure 6).

It is worth repeating that the model's predictions were fit to just one of the empirical measures (number of blocks placed, Figure 3), while the three remaining predictions—distribution of blocks placed (see Figure 4), number of visits to the target window (see Figure 6), and duration of first uncovering (see Figure 5)—all closely matched human performance using the same parameter settings.

Discussion of the Ideal Performer Model

As shown by the low RMSE and high r^2, the Ideal Performer Model predicts a number of blocks placed that is within the range of the standard error of the human data. Interesting enough, it does so by incorporating a rational analysis-based theory of forgetting that has accumulated a broad base of support across many diverse laboratory (Altmann & Gray, 2002; Anderson & Lebiere, 1998; Anderson & Milson, 1989; Lovett et al., 1999) and real-world tasks (Anderson & Schooler, 1991; Schooler & Hertwig, 2005).

The results of the Ideal Performer Model across four empirical measures suggest that human performance on the Blocks World task reflects a cost-benefit tradeoff between perceptual-motor and memory costs defined by time. Within the constraints of memory and perceptual-motor limits, the human control system adapts to the costs of information access in its task environment by making rational, cost-benefit tradeoffs among sets of more interaction-intensive and more memory-intensive strategies. The Ideal Performer Model is not biased to favor perceptual-motor effort over memory effort. Rather, it is sensitive only to costs and benefits defined by time. The noise parameter used to fit the human data suggests that humans in the Blocks World task adopt a close approximation to optimal behavior, and provides an estimate on the extent to which human performance in the task is driven by the soft constraint of time. Hence, the results support the soft constraint perspective on embodied cognition that views memory and perceptual-motor resources as allocated by a control system that attempts to optimize performance time. It seems improbable that a computational model employing the minimum memory hypothesis would be able to account for the same broad range of results.

Summary and Conclusions

The soft constraints hypothesis maintains that at the 1/3 to 3 second level of interactive routines, that is, the embodiment level (Ballard et al., 1997), tradeoffs among the use of cognitive, perceptual, and motor resources are made as if time is a resource that is to be preserved. In this paper we presented three experiments and an Ideal Performer Model that compared the predictions of the soft constraints hypothesis with that of the minimum memory hypothesis in a Blocks World task.

Human Performance

In all conditions, across each experiment, once the Target Window was uncovered the task environment was exactly the

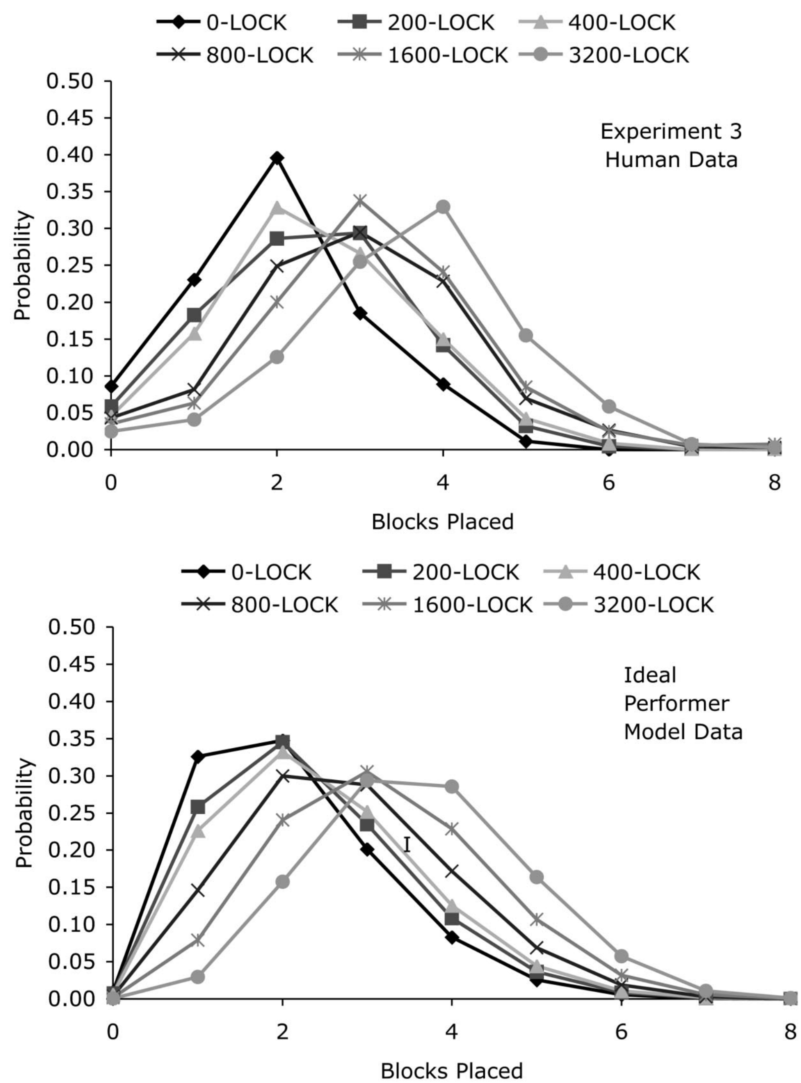

Figure 4. Comparison of the distribution of blocks placed following the first visit to the target window for humans (top) and the Ideal Performer Model (bottom).

same. The Target Window stayed open for as long as the mouse cursor remained inside it (in E1-low—for as long as the control key was held down). The Resource Window and Workspace Window worked exactly the same across all studies and conditions; both opened as soon as the mouse cursor entered and stayed open until the mouse cursor left. Another way of saying this is that once the Target Window opened, the task was exactly the same across all conditions and all studies, and no hard constraints existed that would account for why the task was not performed exactly the same. However, for the current studies, even when the comparisons between two conditions were not significant (e.g., as for e1-low vs. e1-med) an increase in the range of 50 ms to uncover the Target Window resulted in small, but consistent, increases in the duration for which the Target Window was uncovered and small, but consistent, increases in the number of blocks placed.

The Ideal Performer Model

Although the experimental studies documented a tradeoff between access costs and the use of more interaction-intensive or more memory-intensive strategies, the studies did not suffice to determine the nature of that tradeoff. To precisely predict what an optimal tradeoff would be between perceptual-motor and memory costs, we created an Ideal Performer Model that maximized performance in the Blocks World task by selecting ENCODE-k strategies that minimized the total expected time to complete each trial for each of the six between-subjects conditions of Experiment 3.

The Ideal Performer Model used realistic assumptions regarding the time required to execute each interactive routine. For memory operations it used a memory model, based on rational analysis, that yielded assumptions about encoding duration, retrieval latency,

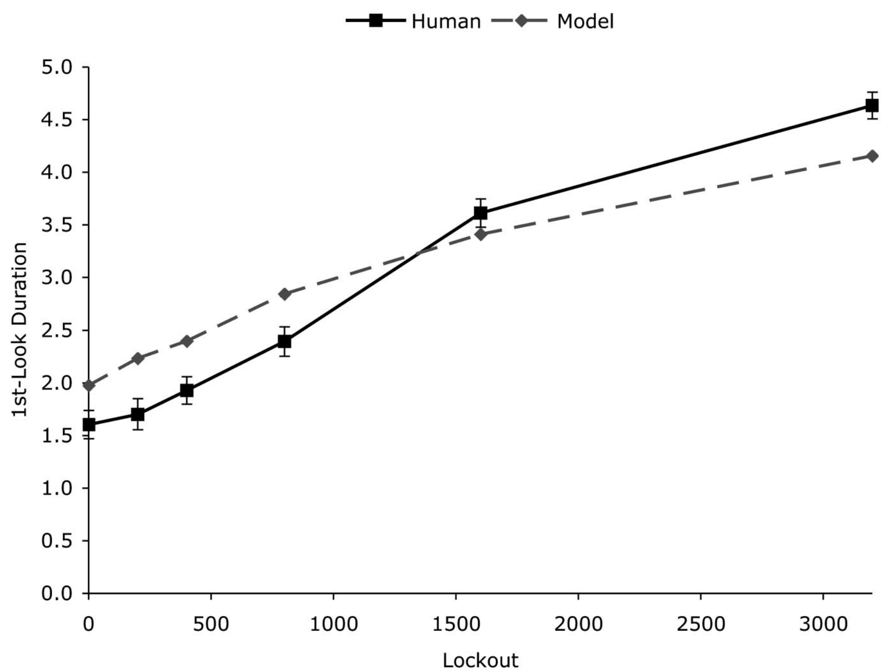

Figure 5. Duration of the first uncovering of the target window for the human participants (Experiment 3 with +/− 1 standard error bars) and Ideal Performer Model.

and the forgetting that would occur in the retention interval between encoding and placement. For the six conditions of Experiment 3, the performance of the model was nearly indistinguishable from human performance. We conclude that, subject to the limitations of the memory system, human performance is nearly identical to what would be expected if the allocation of cognitive, perceptual, and motor resources was based on their temporal costs and if overall benefit was defined by minimizing these costs. Cost-benefit tradeoffs among lockout time, perceptual-motor activity, and fallible memory act as soft constraints that select the interactive behaviors that are best adapted to the task environment.

Implications for Views of Memory and Metacognition

The success of the model has implication for theories of memory. First, it shows that a model based on a rational analysis of the demands the environment makes on memory can be successfully applied as a constraint on a rational analysis of interactive behavior. Given the vast differences between the nature of the memory tasks on which the model was derived (Anderson & Schooler, 1991) and the much more interaction intensive tasks required for performance in the Blocks World task, this success of the memory theory presents both a validation and important generalization of the theory.

Second, regardless of the ultimate validity of Anderson's model of memory, its use in the Ideal Performer Model provides a strong suggestion for the form in which theories of memory must take if they are to be usefully applied to interactive behavior. Rather than simply focusing on the number of slots or amount of activation, the Ideal Performer Model suggests that theories of memory must encompass three additional factors. First, is the time needed to raise the activation of an item so that it can be retrieved over the time period for which the item is needed. Second, is the time required to retrieve an item from memory. Third, is the probability that an encoded item will be retrieved due to decay and noise in the item's activation.

Additionally, the close fit of the human data to the predictions of the Ideal Performer Model suggests that people have implicit knowledge or metacognition of these three memory factors, and, with relatively little experience with a new task (within 10 trials in our studies), are able to near-optimally adapt their interactive behaviors to meet the demands of the task environment. (In a sense, it is this metacognitive knowledge that took the Ideal Performer Model 100,000 training trials to acquire.[8]) Although this extrapolation goes beyond the current study and model, imagining that human performance is adapted to experienced limits in cognition, perception, and action is congruent with recent results that show that human motor performance is exquisitely adapted to compensate for the effect of noise in the motor system (Maloney, Trommershäuser, & Landy, in press; Trommershäuser, Maloney, & Landy, 2003).

Embodied Cognition, Bounded Rationality, Rational Analysis, and the Ideal Performer Model

The soft constraints hypothesis is broadly compatible with many claims made for embodied cognition (Clark, 2003; Wilson, 2002)

[8] We thank Professor Ruth Maki (Texas Tech University) for pointing out that the training trails achieved in the model meta-cognitive knowledge regarding the limits of its memory system.

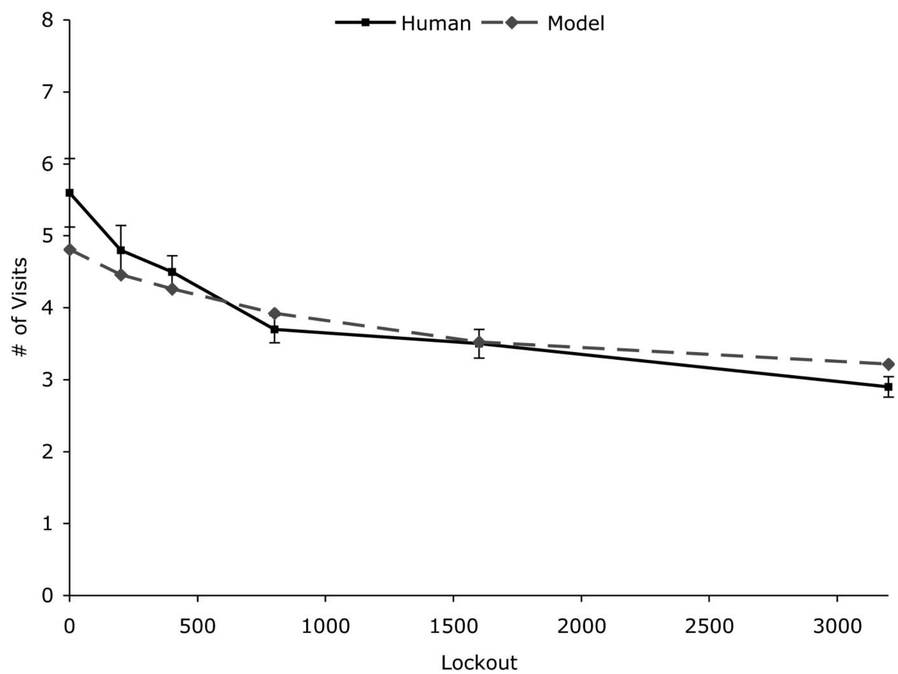

Figure 6. Mean number of visits to the target window to complete each trial for human participants (Experiment 3 with +/− 1 standard error bars) and the Ideal Performer Model.

but offers a more nuanced understanding of what these claims imply. For example, the soft constraints hypothesis addresses two claims in Wilson's (2002) taxonomy of embodied cognition. First, is the claim that we off-load cognitive work onto the environment. For this claim the soft constraints hypothesis implies that the control system is indifferent to information source; resources are allocated to knowledge in-the-world versus in-the-head not based on source, but based on the cost of accessing the source. Second, is the claim that the environment is part of the cognitive system. The soft constraints hypothesis offers the same comment on this claim as to the first—that the human information processing system is indifferent to the source of its information. The only bias imposed by biology is that of finding the most cost-effective means of using available cognitive, perceptual, and motor resources to accomplish a given task in a given task environment.

The power of the Ideal Performer Model flows directly from our combination of an ideal observer analysis with rational analysis. Perceptual-motor side conditions were derived from a variety of sources outside of the current study. The equations that described the side conditions for encoding time, retrieval latency, and probability of recall were themselves based on a rational analysis of human memory (Anderson, 1990, 1991; Anderson & Milson, 1989; Anderson & Schooler, 1991). As an approach, rational analysis is sometimes criticized for being the antithesis of the bounded rationality approach (Howes, Lewis, & Vera, in press). The Ideal Performer Model shows that a rational analysis of one side condition, in this case human memory, can provide an important bound that allows us to make progress on a rational analysis of another side condition, in this case, optimizing the use of internal resources by cost-benefit tradeoffs in the access of knowledge in-the-world versus in-the-head.

Conclusions

When you sit down the night before the birthday party to assemble the child's toy, you could force yourself to first memorize all of the instructions, or to memorize the first half, or to memorize every other line, or not. There are no hard constraints in the task environment that would prevent you from implementing any of these strategies. However, the work presented here suggests that you will treat time on task as a soft constraint that you will minimize by a cost-effective mixture of perceptual-motor and cognitive operations.

Our two sets of methods—experimental results and Ideal Performer Model—converge in their support for the soft constraints hypothesis. The control system is not biased to favor perceptual-motor over cognitive costs. Rather, at the 1/3 to 3 sec level of embodiment, the allocation of cognitive, perceptual, and motor resources is based on cost-benefit tradeoffs measured in time. The soft constraints view of embodiment suggests that many of the details of the cognitive system can be abstracted away and the function of the integrated cognitive-perceptual-motor system can be explained by expected utility measured in time. An information system that truly integrates cognition with perceptual-motor operations integrates the use of knowledge in-the-head with knowledge in-the-world so as to conserve the resource of time, not cognition.

References

Altmann, E. M. (in press). Control signals and goal-directed behavior. In W. D. Gray (Ed.), *Integrated models of cognitive systems.* New York: Oxford University Press.

Altmann, E. M., & Gray, W. D. (2002). Forgetting to remember: The functional relationship of decay and interference. *Psychological Science, 13*(1), 27–33.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences, 14*(3), 471–517.

Anderson, J. R., & Lebiere, C. (Eds.). (1998). *Atomic components of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review, 96*(4), 703–719.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2,* 396–408.

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7*(1), 66–80.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences, 20*(4), 723–742.

Berthier, N. E. (1996). Learning to reach: A mathematical model. *Developmental Psychology, 32*(5), 811–823.

Berthier, N. E., Rosenstein, M. T., & Barto, A. G. (2005). Approximate optimal control as a model for motor learning. *Psychological Review, 112*(2), 329–346.

Byrne, M. D., & Kirlik, A. (2005). Using computational cognitive modeling to diagnose possible sources of aviation error. *International Journal of Aviation Psychology, 15,* 135–155.

Card, S. K., English, W. K., & Burr, B. J. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys and text keys for text selection on a CRT. *Ergonomics, 21*(8), 601–613.

Card, S. K., Moran, T. P., & Newell, A. (1980). Computer text editing: An information processing analysis of a routine cognitive skill. *Cognitive Psychology, 12,* 32–74.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction.* Hillsdale, NJ: Erlbaum.

Carlson, R. A., & Sohn, M-H. (2000). Cognitive control of multiplestep routines: Information processing and conscious intentions. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 443–464). Cambridge, MA: MIT Press.

Cary, M., & Carlson, R. A. (1999). External support and the development of problem-solving routines. *Journal of Experimental Psychology-Learning Memory & Cognition, 25*(4), 1053–1070.

Chun, M. M., & Nakayama, K. (2000). On the functional role of implicit visual memory for the adaptive deployment of attention across scenes. *Visual Cognition, 7*(1–3), 65–81.

Clark, A. (2003). *Natural-born cyborgs: Minds, technologies, and the future of human intelligence.* New York: Oxford University Press.

Cowan, N. (1997). *Attention and memory.* New York: Oxford University Press.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance & executive control* (pp. 62–101). New York: Cambridge University Press.

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems.* Cambridge, MA: MIT Press.

Ehret, B. D. (2002). Learning where to look: Location learning in graphical user interfaces. *CHI Letters, 4*(1), 211–218.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*(3), 309–331.

Freed, M., Matessa, M., Remington, R. W., & Vera, A. (2003). *How Apex automates CPM-GOMS.* Paper presented at the Fifth International Conference on Cognitive Modeling.

Fu, W.-T., & Gray, W. D. (in press). Suboptimal tradeoffs in information seeking. *Cognitive Psychology.*

Fu, W-T., & Anderson, J. R. (2004). Extending the computational abilities of the procedural learning mechanism in ACT-R. In K. D. Forbus & D. Gentner & T. Regier (Eds.), *26th Annual Meeting of the Cognitive Science Society, CogSci2004* (pp. 416–421). Hillsdale, NJ: Erlbaum.

Fu, W-T., & Anderson, J. R. (2006). From recurrent choice to skill learning: A reinforcement-learning model. *Journal of Experimental Psychology: General, 135*(2).

Fu, W-T., & Gray, W. D. (2004). Resolving the paradox of the active user: Stable suboptimal performance in interactive tasks. *Cognitive Science, 28*(6), 901–935.

Geisler, W. S. (2003). Ideal Observer analysis. In L. Chalupa & J. Werner (Eds.), *The visual neurosciences* (pp. 825–837). Boston: MIT Press.

Geisler, W. S., & Diehl, R. L. (2003). A Bayesian approach to the evolution of perceptual and cognitive systems. *Cognitive Science, 27*(3), 379–402.

Gray, W. D., & Boehm-Davis, D. A. (2000). Milliseconds Matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior. *Journal of Experimental Psychology: Applied, 6*(4), 322–335.

Gray, W. D., & Fu, W.-T. (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science, 28*(3), 359–382.

Gray, W. D., John, B. E., & Atwood, M. E. (1993). Project Ernestine: Validating a GOMS analysis for predicting and explaining real-world performance. *Human-Computer Interaction, 8*(3), 237–309.

Gray, W. D., Neth, H., & Schoelles, M. J. (in press). The functional task environment. In A. Kramer & A. Kirlik & D. Wiegman (Eds.), *Applied attention.* Mahwah, NJ: Erlbaum.

Gray, W. D., Schoelles, M. J., & Myers, C. W. (2004). Meeting Newell's other challenge: Cognitive architectures as the basis for cognitive engineering. *Behavioral & Brain Sciences, 26*(5), 609–610.

Gray, W. D., Schoelles, M. J., & Sims, C. R. (2005). Adapting to the task environment: Explorations in expected value. *Cognitive Systems Research, 6*(1), 27–40.

Hayhoe, M. (2000). Vision using routines: A functional account of vision. *Visual Cognition, 7*(1–3), 43–64.

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis. of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109*(4), 679–709.

Howes, A., Lewis, R. L., & Vera, A. H. (in press). Bounding rational analysis: Constraints on the approach to optimality. In W. D. Gray (Ed.), *Integrated models of cognitive systems.* New York: Oxford University Press.

Howes, A., Lewis, R. L., Vera, A. H., & Richardson, J. (2005). Information-requirements grammar: A theory of the structure of competence for interaction. In B. G. Bara & L. Barsalou & M. Bucciarelli (Eds.), *27th Annual Meeting of the Cognitive Science Society, CogSci2005* (pp. 595–600). Austin, Tx: The Cognitive Science Society, Inc.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122–149.

Just, M. A., Carpenter, P. A., & Keller, T. A. (1996). The capacity theory of comprehension: New frontiers of evidence and arguments. *Psychological Review, 103*(4), 773–780.

Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction, 12*(4), 391–438.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science, 18*(4), 513–549.

Larkin, J. H. (1989). Display-based problem solving. In D. Klahr & K.

Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 319–341). Hillsdale, NJ: Erlbaum.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11,* 65–99.

Lohse, G. L., & Johnson, E. J. (1996). A comparison of two process tracing methods for choice tasks. *Organizational Behavior and Human Decision Processes, 68*(1), 28–43.

Lovett, M. C., Reder, L. M., & Lebiere, C. (1999). Modeling working memory in a unified architecture: An ACT-R perspective. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance & executive control* (pp. 135–182). New York: Cambridge University Press.

MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction, 7*(1), 91–139.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide.* Mahwah, NJ: Erlbaum.

Maloney, L. T., Trommershäuser, J., & Landy, M. S. (in press). Questions without words: A comparison between decision making under risk and movement planning under risk. In W. D. Gray (Ed.), *Integrated models of cognitive systems.* New York: Oxford University Press.

Meyer, D. E., Smith, J. E. K., Kornblum, S., Abrams, R. A., & Wright, C. E. (1990). Speed-accuracy tradeoffs in aimed movements: Toward a theory of rapid voluntary action. In M. Jeannerod (Ed.), *Attention and performance, XIII: Motor representation and control* (pp. 173–225). Hillsdale, NJ: Erlbaum.

Miller, G. A. (1956). The magical number 7, plus or minus 2: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81–97.

Nason, S., & Laird, J. E. (2004). Soar-RL: Integrating reinforcement learning with Soar. In M. C. Lovett & C. D. Schunn & C. Lebiere & P. Munro (Eds.), *6th International Conference on Cognitive Modeling* (pp. 208–213). Pittsburgh, PA.

Neth, H., & Payne, S. J. (2001). Addition as interactive problem solving. In J. D. Moore & K. Stenning (Eds.), *Twenty-Third Annual Conference of the Cognitive Science Society* (pp. 698–703). Hillsdale, NJ: Erlbaum.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Norman, D. A. (1982). *Steps toward a cognitive engineering: Design rules based on analyses of human error.* Paper presented at the 1982 conference on Human Factors in Computing Systems, Gaithersburg, MD.

Norman, D. A. (1986). Cognitive engineering. In D. A. Norman & S. W. Draper (Eds.), *User centered system design: New perspectives on human-computer interaction* (pp. 31–61). Hillsdale, NJ: Erlbaum.

Norman, D. A. (1989). *The design of everyday things.* New York: Double-Day.

Norman, D. A. (1993). Cognition in the head and in the world: An introduction to the special issue on situated action. *Cognitive Science, 17*(1), 1–7.

Oaksford, M., & Chater, N. (Eds.). (1998). *Rational models of cognition.* New York: Oxford University Press.

O'Hara, K. P., & Payne, S. J. (1998). The effects of operator implementation cost on planfulness of problem solving and learning. *Cognitive Psychology, 35,* 34–70.

O'Hara, K. P., & Payne, S. J. (1999). Planning and the user interface: The effects of lockout time and error recovery cost. *International Journal of Human-Computer Studies, 50*(1), 41–59.

Pelz, J. B. (1996). Visual representations in a natural visuo-motor task. *Dissertation Abstracts International: Section B: The Sciences & Engineering, 57*(2-B), 1480.

Phillips, J. L., & Noelle, D. C. (2004). Reinforcement learning of dimensional attention for categorization. In K. D. Forbus & D. Gentner & T. Regier (Eds.), *26th Annual Meeting of the Cognitive Science Society, CogSci2004.* Hillsdale, NJ: Erlbaum Publisher.

Purves, D., Lotto, R. B., & Nundy, S. (2002). Why we see what we do – A probabilistic strategy based on past experience explains the remarkable difference between what we see and physical reality. *American Scientist, 90*(3), 236–243.

Ritter, F. E., Van Rooy, D., St. Amant, R., & Simpson, K. (in press). Using a simulated user to explore human-robot interfaces. *IEEE Transactions on System, Man and Cybernetics, Part A: Systems and Humans.*

Salvucci, D. D. (in press). Modeling driver behavior in a cognitive architecture. *Human Factors.*

Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review, 112*(3), 610–628.

Simon, H. A. (1992). What is an "explanation" of behavior? *Psychological Science, 3*(3), 150–161.

Simon, H. A. (1996). *The sciences of the artificial* (3rd ed.). Cambridge, MA: The MIT Press.

Sims, C. R., & Gray, W. D. (2004). Episodic versus semantic memory: An exploration of models of memory decay in the serial attention paradigm. In M. C. Lovett & C. D. Schunn & C. Lebiere & P. Munro (Eds.), *6th International Conference on Cognitive Modeling* (pp. 279–284). Pittsburgh, PA.

Sohn, M. H., & Carlson, R. A. (1998). Procedural frameworks for simple arithmetic skills. *Journal of Experimental Psychology-Learning Memory and Cognition, 24*(4), 1052–1067.

Sohn, M. H., & Carlson, R. A. (2003). Implicit temporal tuning of working memory strategy during cognitive skill acquisition. *American Journal of Psychology, 116*(2), 239–256.

Stevenson, L. M., & Carlson, R. A. (2003). Information acquisition strategies and the cognitive structure of arithmetic. *Memory & Cognition, 31*(8), 1249–1259.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning.* Cambridge, MA: The MIT Press.

Taatgen, N. A., & Lee, F. J. (2003). Production composition: A simple mechanism to model complex skill acquisition. *Human Factors, 45*(1), 61–76.

Todd, P. M., & Schooler, L. J. (in press). Disintegrated architectures of cognition: The adaptive toolbox for decision making. In W. D. Gray (Ed.), *Integrated models of cognitive systems.* New York: Oxford University Press.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2003). Statistical decision theory and trade-offs in the control of motor response. *Spatial Vision, 16*(3–4), 255–275.

Ullman, S. (1984). Visual Routines. *Cognition, 18*(1–3), 97–159.

Watkins, C., & Dayan, P. (1992). Q-Learning. *Machine Learning, 8*(3–4), 279–292.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625–636.

Wu, C., & Liu, Y. (2004). Modeling psychological refractory period (PRP) and practice effect on PRP with queing networks and reinforcement learning algorithms. In M. C. Lovett & C. D. Schunn & C. Lebiere & P. Munro (Eds.), *6th International Conference on Cognitive Modeling* (pp. 320–325). Pittsburgh, PA.

Appendix A: Declarative memory in ACT-R

To implement human memory limitations in the reinforcement learning model, we used the memory theory incorporated into the ACT-R cognitive architecture (Anderson & Lebiere, 1998; Lovett et al., 1999). This theory has been widely tested, compares well to alternative approaches (Sims & Gray, 2004), and has been successful at capturing human performance on a wide range of memory tasks. At its core, the ACT-R memory model makes quantitative predictions regarding the probability of successfully recalling a previously encoded declarative memory element, or DME, as well as the retrieval latency for that DME. Both the probability of recall and retrieval latency are governed by activation, which increases with practice and successful retrieval of an item, and decays as a function of time. The equation below gives the formula for computing the base activation of a DME.

$$a_i = \ln\left(\sum_{j=1}^{n} t_j^{-d}\right) + \varepsilon \qquad \text{(Eqn. A-1)}$$

In this equation, a_i is the activation of DME i, t_j is the time since its jth retrieval, and d is a decay parameter governing how quickly each retrieval's influence on the activation decreases. The summation is over the entire history of retrievals of the DME. The last term is a noise component that is drawn from a logistic distribution and allows the activation of the DME to fluctuate from moment to moment. In the complete ACT-R memory model, environmental context and relevance to the current goal also influences the activation of a DME, however this component introduces additional complexity not relevant to the Blocks World model.

Retrieval probability is governed by adding a threshold parameter to the model. If retrieval of a DME is attempted and the DME's base activation is below the threshold, then a retrieval failure occurs, meaning that the item has effectively been forgotten. However, as the noise component of activation is dynamically generated, it is possible for a DME to be below threshold on one retrieval attempt but then above threshold on a second attempt.

The time it takes for a retrieval or a retrieval failure is governed by the activation of the DME such that more active DMEs are recalled faster than less active DMEs. The exact equation used by ACT-R is given below.

$$RT_{\mathrm{i}} = F \cdot e^{-a_i} \qquad \text{(Eqn. A-2)}$$

As before, a_i is the activation of DME i, while F is a latency scaling parameter, and RT_i is the retrieval time in seconds for that DME. In general, the DME with the highest level of activation is the one retrieved. If no DME is above the threshold at the time of retrieval, then a retrieval failure occurs. In this case, the retrieval threshold parameter is used in lieu of the DME activation (a_i) to compute the time taken by the failed retrieval, with the consequence that retrieval failures take longer than successful retrievals. Since retrieval time is based directly on activation, the moment-to-moment noise in activation also causes the retrieval time to fluctuate.

Appendix B: Q-Learning and ENCODE-k strategies

At its core, reinforcement learning is concerned with learning a value function *Q(s,a)* that transforms states of the environment and actions into a numerical expected reward outcome. This value function is followed by the agent according to a policy function that maps expected rewards into a particular sequence of actions. *Q*-learning, the particular reinforcement learning algorithm used here, has the additional property that it can learn an optimal behavioral policy while randomly exploring actions in the environment, so long as certain reasonable assumptions are met (for instance, sufficient training and exploration of the problem space). The exact *Q*-learning update rule is given below, though see Sutton and Barto (1998) for a more thorough treatment of the algorithm.

$$Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma \cdot \max_{a'} Q(s', a') - Q(s, a)] \qquad \text{(Eqn. B-1)}$$

In this equation the value of a particular action a is updated according to the local reward received, r, as well as the future expected rewards as a consequence of reaching the successor state, s'. Alpha is a parameter controlling how quickly the agent learns and can range from 0.0 to 1.0. At the lower end, the model stops learning completely, while at the upper end each new experience obliterates all previous learning by the agent. In training the Ideal Performer Model alpha was initially set to 1.0 and then decreased with increased experience according to $1/n$, where n is the number of experiences with a particular action. This scheme is equivalent to taking the arithmetic average of all rewards, and in the Q-learning algorithm is sufficient to guarantee that the optimal policy can be learned with sufficient practice. The parameter gamma controls whether the model discounts future compared to immediate rewards. In the task this parameter was set to 1.0, meaning that the algorithm should strive to maximize global performance rather than select actions locally greedily.

In the Blocks World task, optimal performance is defined as completing the overall task as quickly as possible. Therefore, after the Q-learning model selects each action, it is penalized according to how long that action took. As discussed in the text, in the model of the Blocks World task, there are a maximum of eight possible actions and 36 possible state-action pairs. Over time the value function *Q(s, a)* learned by the agent corresponds to its estimate of how long it will take to complete the entire task given that a particular action is chosen in a particular state. The costs used as rewards in the model are simply the total time needed to complete a particular ENCODE-*k* strategy.

(*Appendixes continue*)

Appendix C

Parameters Used by the Ideal Performer Model

Parameter	Value	Source
Motor parameters		
Mouse-target-to-resource	249 ms	Fitts' Law
Mouse-resource-to-workspace	216 ms	Fitts' Law
Mouse-workspace-to-resource	249 ms	Fitts' Law
Mouse-workspace-to-target	217 ms	Fitts' Law
Mouse-block-to-block	150 ms	Fitts' Law
Mouse-click	150 ms	(Gray & Boehm-Davis, 2000)
Shift of visual attention	185 ms	ACT-R default
Memory parameters (ACT-R equivalent)		
Activation decay (BLL)	0.5	ACT-R default
Activation noise (ANS)	0.28	Free parameter
Retrieval threshold (RT)	0.325	Free parameter
Latency scaling factor (F)	0.9	Free parameter
Q-learning parameters		
Utility noise (*t*)*	0.491	Free parameter
Alpha	1/*n*	ACT-R default; *n* is the number of experiences with a particular action
Gamma	1.0	Default value

* The noise parameter is also related to ACT-R's expected gain noise parameter (EGS) according to EGS $= t/\sqrt{2}$ (Anderson & Lebiere, 1998). Specifically, in our model EGS = 0.347.

Received November 15, 2004
Revision received January 11, 2006
Accepted January 25, 2006 ■