Cumulating the Science of HCI: From S-R Compatibility to Transcription Typing[1]

Bonnie E. John
Department of Psychology
john@psy.cmu.edu

Allen Newell
Department of Computer Science
allen.newell@cs.cmu.edu

Carnegie Mellon University
Schenley Park
Pittsburgh, PA, 15213

ABSTRACT
In keeping with our claim that an applied psychology of HCI must be based on cumulative work within a unified framework, we present two extensions of the Model Human Processor. A model of immediate response behavior and stimulus-response (S-R) compatibility is presented and extended to a new domain: transcription typing. Parameters are estimated using one S-R compatibility experiment, used to make a priori predictions in four other S-R compatibility tasks, and then carried over into the area of typing. A model of expert transcription typing is described and its prediction of typing phenomena is demonstrated and summarized.

KEYWORDS: user models, cognitive models, GOMS, Model Human Processor

INTRODUCTION

We have maintained that engineering models of human performance are necessary for psychology to have an impact on the design of human-computer interfaces [1, 4, 9, 10]. Progress in constructing such models requires a cumulation of the science of HCI. This cumulation can proceed in several ways: experimental programs that build on previous work a new specific area, review projects that draw published literature together into one easily accessed place, and model-building that seeks to explain different HCI activities with one integrated framework. This paper reports on two steps forward using the latter approach.

Generally, the procedure begins by choosing a theoretical framework; in our case the Model Human Processor (MHP) [1]. A task is examined experimentally and analytically, and a model is constructed, within the theoretical framework, that quantitatively explains behavior on that task. Parameters of the task are identified and empirical data are used to make estimates of those parameters. Other similar tasks are then analyzed using the model and a priori predictions are made for performance on those tasks.

Empirical data is compared to the predictions to verify the model. Then tasks that share fewer common features of the original task are modelled using the same structure and parameter estimates, and predictions are made and verified against empirical data. Thus, the theory grows.

CONSTRUCTING THE MODEL AND ESTIMATING PARAMETERS: AN ABBREVIATION RECALL TASK

A task involving stimulus-response (S-R) compatibility was chosen for analysis. This domain was chosen because of its relevance to HCI and because Rosenbloom [11] had successfully modelled such tasks with an algorithmic, GOMS-like[2] approach.

The task was to generate the abbreviation of a computer command given the full name of the command.[3] Specifically, a set of commands and their abbreviations were taught to experimental participants. After learning these abbreviations to criterion, the participants had to type the correct abbreviation when presented with each command. The performance measures chosen for investigation were the time between presentation of the command and the typing of the first letter of the abbreviation (*initial-response time*) and the time between the typing of the first letter and the typing of the last letter (*execution time*).

Algorithms were written using elementary operations like "Get-command" (from the screen), "Is-the-letter-a-vowel?" (when generating an abbreviation consisting of only the consonants of the command name), and "Type-letter". Rosenbloom originally assumed a single duration for all such operators, but that proved to be too simple an

[1]This research was, in part, supported by the Office of Naval Research, Perceptual Science Programs, Contract Number N00014-87-K-0432. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Office of Naval Research or the U. S. Government.

[2]GOMS stands for Goals, Operators, Methods, and Selection rules, and is a type of analysis used with the MHP [1].

[3]Originally reported in [7] and subsequently re-analyzed in [6], producing slightly different parameter estimates, which are used here.

assumption for this more complex task. Therefore, we used separate estimates for three types of operators associated with the three MHP processor cycle times ("Get-command" would be a perceptual operator, "Is-the-letter-a-vowel?" would be a cognitive operator and "Type-letter" would be a motor operator). In addition, we estimated a fourth operator, a *retrieval* operator, that is a composite of several cognitive operators and serves to recall an arbitrary association.

The separation of operators based on the MHP theoretical framework produced an excellent explanation of the average human performance (r^2=0.98). The operator durations produced by a regression between the number of operators needed to accomplish the tasks and the initial-response and execution times observed were 340 msec for a perceptual operator that perceives a word (of about 6 letters), 50 msec for a cognitive operator that performs an elementary cognitive activity in one cognitive processor cycle time, 1200 msec for a retrieval operator that recalls a newly-learned arbitrary association, and 230 msec for a motor operator that types out a single letter at a rate of about 30 words per minute.

This analysis demonstrates the GOMS approach at a new level of detail. Goals and methods are set by the task (and therefore, selection rules are unnecessary). The operators are at the level of the MHP processor cycle times. This refinement of GOMS in itself is an extension of the MHP, and it forms a basis for extension to other tasks.

EXTRAPOLATION TO NEAR TASKS: ANOTHER ABBREVIATION RECALL TASK AND THREE CLASSIC S-R COMPATIBILITY TASKS

The next step in the process is to make zero-parameter, a priori predictions of performance on similar tasks to check out the predictive power of the model. This was done by examining an abbreviation recall task with two new abbreviation techniques and three classic S-R compatibility tasks from the experimental literature.[4]

The two abbreviation recall tasks were very similar. Physically, they were run on the same computer with the same CRT, keyboard, and timing devices. Conceptually, they employed the same perceptual-cognitive-motor activities: perceiving short, individually presented words, making a correspondence between computer commands and their abbreviations, and typing short letter strings.

On the other hand, the only common element between the parameter-setting task and the three S-R compatibility tasks [2, 3, 8] was that they all involved immediate response behavior. That is, each task involved a fast-as-possible response to a stimulus (on the order of 1 second). The response was well known, though not highly practiced; there was no problem solving involved.

Every other element differed from the abbreviation recall task. The stimuli were lamps in a two dimensional array [3], digits and symbols projected on a screen [8], and lines illuminated on an oscilloscope [2]. The responses were moving a stylus [3], naming a digit [8], and striking one of four keys [2]. The experiments were conducted by different experimenters with different goals, in different labs, in different decades.

Zero-parameter predictions were made of the response times for most of the conditions in these experiments.[5] The perceptual and motor components of the classic S-R compatibility tasks were sufficiently different from those in the parameter-setting experiment that the estimated durations were not used. Instead, we returned to the basic MHP for an estimate of generic perceptual and motor operators and used the typical perceptual and motor processor cycle times (100 msec for the perceptual processor cycle time and 70 msec for the motor processor cycle time).

Figure 1 plots the predicted response times versus the observed response times for all S-R compatibility zero-parameter predictions. The average absolute percent error is 19%. This degree of accuracy is acceptable in engineering design situations, especially since the order of performance times was preserved by the predictions in all but one situation where the observed performance differed by more than 20%. Thus, the theory predicts the big differences, and

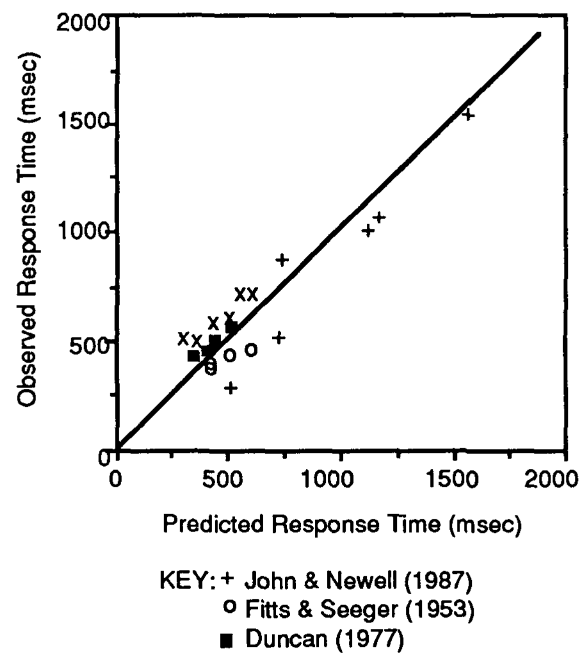

Figure 1. Predicted vs. observed response times for the four S-R compatibility experiments analyzed.

[4]Originally reported in [5], and subsequently re-analyzed in [6]. The new analysis was quite different from the previously reported work, using an information-theory approach to predicting the encoding of non-word stimuli.

[5]Only two of the three stimulus sets and two of the three response modes studied in the Fitts & Seeger experiment were analyzed because the remaining stimuli and responses involve the use of eye-movement and two-handed responses which were beyond the scope of the model.

identifies areas where human performance will be close.

EXTRAPOLATION TO A FARTHER TASK -- TRANSCRIPTION TYPING

The next step in cumulating the science is to pick another HCI activity and model it, carrying over as much of the preceding model as possible. The activity we chose was transcription typing. Typing is a major form of human-to-computer communication and is therefore an important skill to understand within HCI. It is also similar in many respects to the command abbreviation recall task. It involves perceiving and internally representing words, manipulating the representations of those words, and typing characters. This similarity makes transcription typing seem only a small step away from the previously successful work.

In other ways, the two domains are very different. While S-R compatibility involves a substantial proportion of time in the cognitive operations required to accomplish the task, transcription typing is primarily a perceptual-motor task. The S-R compatibility tasks were very short, discrete behaviors, whereas transcription typing is of longer duration, a flow of behavior. S-R compatibility tasks seem inherently sequential: see the stimulus, do the cognitive mapping, execute the response. Typing is parallel in nature: look ahead at what is coming while executing the motor responses for the current letters. Also, since specific tasks were again taken from the experimental literature, the experiments were run by different people, on different apparatus, with different goals in mind.

The Model of Expert Transcription Typing

A model of expert transcription typing (METT) [4] was derived from the MHP, experience in the S-R compatibility analyses, and some qualitative information about typing. The METT is defined by six assumptions.

1. Basic Algorithm. The basic algorithm is that a person perceives a *chunk* (word, syllable, letter) using a perceptual operator. If the chunk is a word or syllable, the spelling of that chunk is obtained with a cognitive operator, the first character is initiated with a cognitive operator, and then executed with a motor operator. The second character is then initiated and executed, and so on. If a letter is perceived alone, then the character is initiated immediately following the perception, without unpacking a chunk. (This is the same algorithm used for typing in the no-abbreviation condition of the S-R compatibility experiments.)

2. Serial/Parallel Processing. In the MHP, each processor works serially within itself, and concurrently with the other processors, with the following typing-specific exceptions.

2A. Perception/Cognition Interaction. Perception has to be complete before the spelling can be gotten or initiation of a character can begin.

2B. Same-Hand Constraint. A character on the same hand cannot be initiated with a cognitive operator until the motor processor execution of the previous character is complete. (This assumption was included because same-hand successive keystrokes are significantly longer than different-hand successive keystrokes, one of the most striking aspects of human typing behavior.)

2C. Perception/Working-Memory-Limitation Interaction. The perceptual processor cannot perceive the next piece of information unless there is enough space in working memory (see assumption 3 for the implication of this).

3. Working Memory Limitation. In normal transcription typing, the perceptual processor stays three chunks ahead of the cognitive processor. The chunk is usually a word, but it can be a syllable or a character if words are not available in the specific typing task. (This three-chunk limitation is the typical capacity of working memory in the MHP. Long-term memory may be accessed for the spelling of one chunk at a time, extending the effective capacity of working memory to, on average, seven chunks. Assuming an average word length of five letters, these chunks would be the five letters of the first word, plus the two next words).

4. Perceptual Chunks. The perceptual processor perceives at the most meaningful level available at or below the word level. For example, if words are present, they are perceived and encoded. If the view of whole words is restricted or if there are no words present (as when typing random letters), the perceptual processor perceives syllables. If syllables are not visible because of restricted view or random characters, then the perceptual processor perceives single characters.

5. Cognition/Motor Interaction. Once a character is initiated with a cognitive operator, the motor operator that executes that character cannot be stopped. (This assumption was included because typists cannot immediately stop typing when signalled to do so.)

6. Operator Similarity. Across all domains to which the MHP is applied, similar operations involving similar perceptual and motor operators take similar amounts of time. Thus, the operator times estimated in the S-R compatibility experiments can be used here because the perceptual, cognitive, and motor operators are assumed to be similar to those in typing tasks.

6A. Perceptual Operator Duration. The time to perform a perceptual operation is 340 msec. A simplifying assumption is that this time is constant even if the chunk to be perceived is a word, a syllable, or a character.

6B. Cognitive Operator Duration. The time to perform a cognitive operation is 50 msec.

6C. Motor Operator Duration and Interaction with Skill. As a simplifying assumption, practice in typing decreases the motor operator time only; the estimates of the perceptual and cognitive operators remain constant.[6] A critical path analysis was used to estimate the motor operator duration for a range of expert typists (Figure 2).

[6] There are undoubtably individual differences in the perceptual and cognitive processes [13], but given the amount of practice an adult typist has had perceiving words (as a part of reading) and in spelling words (as a part of writing), we assume that these operators remain constant relative to the more newly acquired motor operators of typing.

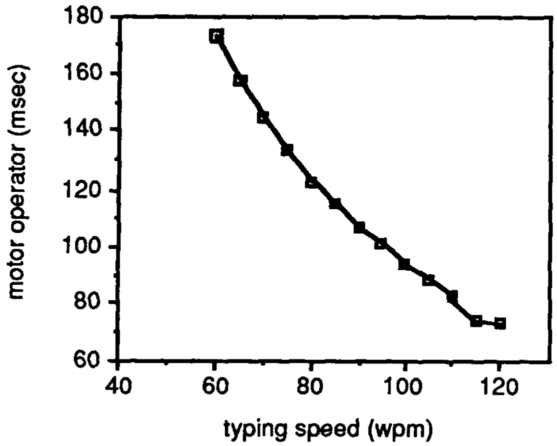

Figure 2. Estimated motor operator duration vs. typing speed.[7]

The use of the METT

Recently, Timothy Salthouse [14] reviewed the typing literature and laid out a list of 29 phenomenon that define "...the criteria by which alternative models...may be evaluated". Some of these phenomena are qualitative, expressing a relationship (e.g., #4: typing rate is independent of word order), and some are quantitative, giving a numerical aspect of typing (e.g.,#13: copy span for expert typists is 14.6 characters). The METT can be applied to the experimental tasks that demonstrate these 29 phenomenon. Just as the phenomena are expressed qualitatively or quantitatively, the METT can be used to reason qualitatively or make quantitative predictions. Two examples follow that demonstrate these different analyses.

Phenomenon 4: Typing rate is independent of word order

The METT works on the word level and has no comprehension or syntacticly high-level mechanisms, so random words will be treated no differently than meaningful text. Therefore, qualitatively, the METT predicts that the rate of typing would not be different for random, or scrambled, words than it is for meaningful text.

Phenomenon 13: Copy span for expert typists is 14.6 characters

The copy span, defined as the amount of material that can be typed accurately after a single inspection of the copy, has been measured in many different ways, different methods yielding vastly different results. Rothkopf [12] measured the copy span by asking the typists to glance at the copy, remembering as much as possible, and then type it out before glancing at the copy again. This is very different than normal transcription typing and it yielded the result that a typist can remember up to 40 characters at a time. Salthouse [13] measured the copy span in a way more appropriate to transcription typing. He instructed typists to type normally using a CRT to display the material to be typed. After the typists had gotten up to speed, he erased the display and the typist had to continue typing as much material as he or she was confident appeared on the display. An average copy span of 13.2 characters was found for all typists (speed range from 20 to 120 gwpm), and an average of 14.6 characters for expert typists (above 60 gwpm).

The METT can easily model this task. Predictions can be made at several different levels of detail. First, a quick and dirty analysis, then a more detailed analysis is presented

On average, a word is 5 letters long. With a 3-word look ahead, there is, on average, 15 letters in the perceptual buffer, or working memory. If the display is removed randomly, it will be removed, on average, 2.5 characters into a word. Hence, there will be a copy span of about 2.5 words, or 12.5 characters. This prediction is within 14% of the observed 14.6 characters.

For the more detailed analysis, the average typing speed of the expert typists must be known (81.6 nwpm). Salthouse used four different sentences of about 75 characters and erased the display after 15, 25, 35, or 45 characters were typed. Analytically, the same effect can be gotten by using one sentence, imposing a stop after every character, and averaging them all together. Assuming an 80 wpm typist, the copying span was predicted as if the display disappeared at each position between the first character and the 75th character in an 89-character sentence from a standard typing test (by the time the fingers have typed the 75th character, the METT predicts that the eyes have reached the end of this sentence). By the METT, there is a three-word look-ahead and as soon as the last character of a word (including the punctuation and space after it) is out of working memory, i.e., the cognitive processor has sent a signal to the motor processor to type the space, the next word can be perceived.

The disappearance of the copy is triggered by the typing of a character, which is the completion of a motor operator. Since the initiation of characters by the cognitive operator triggers the perception of the next word, the perception of the word takes a finite amount of time, and the cognitive and perceptual processors work in parallel with the motor processor, the relationship between the character just typed and the contents of working memory is not a straightforward one. The relationship is determined by the same- and alternate-handed history of the text being typed and the duration of the operators. Thus, the copy span is a combination of the letters that have already been initiated by the cognitive processor but not yet executed by the motor processor, and what is left in working memory that could be initiated and executed.[8] Table 1 shows the letters that can be typed for several different stopping points. The copy span for the average typist in the expert range (~80 gwpm), from this detailed calculation is 11.9 characters (19% from the observed 14.6 characters).

[7] Critical path analysis, used to make the estimates in Figure 2, is an engineering project management tool used to predict the total time to complete a series of parallel activities with information flow dependencies.

[8] This complex relationship is easily seen with a timeline display of a critical path analysis. The reader is referred to [4] for a complete description of this analysis.

Table 1: Copy span predictions for an 80 gwpm typist.

Character Typed	Available to Type	Copy span
O	ne_reason_	10
n	e_reason_	9
e	_reason_	8
_	reason_is_	10
r	eason_is_	9
e	ason_is_quite_	14
a	son_is_quite_	13
s	on_is_quite_	12
o	n_is_quite_	11
n	_is_quite_	10
_	is_quite_	9
i	s_quite_	8
s	_quite_obvious;_	16

The first, quick and dirty analysis gives a good prediction of the copy span, in fact, it is even slightly better than the prediction from the more detailed analysis. Why should more effort be spent to do the more detailed analysis? For HCI design purposes, there is no reason to do the more detailed analysis. For purposes of developing a model, it should be demonstrated that the mechanism of the model does not get in the way of good predictions. If situations occur where the entire mechanism of a model is not necessary for prediction, then the quicker predictions should be used, but more detailed analyses should not be substantially worse. The copy span task is one situation where considering only the model of working memory in the METT is sufficient for a good prediction. However, this simple analysis is not detailed enough to predict most of the 29 phenomena. For example, it would predict that the copy span is constant over skill, whereas the more detailed analysis procedure produces different predictions for different skill, reproducing the pattern observed in actual data.

Integrating the Models - Phenomenon 12: A concurrent task does not affect typing

Salthouse and Saults [16] gave typists a standard transcription typing task, a simple reaction-time task where a foot pedal was pressed in response to a tone, and a dual task where they had to type and perform the reaction-time task concurrently. The instructions and procedure were designed to emphasize the typing portion of the dual task. The typing task was not slowed by the concurrent reaction-time task (going from an interkeystroke mean of 181 msec to 185 msec), but the reaction-time task slowed down considerably (going from a mean of 269 msec to 431 msec), indicating that the reaction time task had not yet become automatic and that the devices used to emphasize the typing task were successful.

The first step in a GOMS analysis of this dual task situation is to estimate the duration of the motor operator that presses a foot pedal. One way to make this estimate, used in the S-R compatibility analyses, is to use the typical value of the motor processor cycle time given by Card, Moran and Newell [1]: 70 msec. In this case, however, we have the data available to make a more specific estimate of the parameter. We write an algorithm for the simple reaction-time task and fit the motor parameter to the observed performance. An algorithm for the simple reaction-time task is as follows.

```
BEGIN
Tone ← Get-Stimulus("Tone")             100 msec
IF-SUCCEEDED Right-Tone?(Tone)           50 msec
 THEN BEGIN
        Initiate-foot-press              50 msec
        Execute-foot-press               ?
        END
END
```

Subtracting the perceptual and cognitive operator times from the observed reaction-time, 269 msec, leaves an estimate of the motor operator to press a foot pedal of about 70 msec. This value happens to be the typical motor processor cycle time given by Card, Moran and Newell [1], providing converging evidence for the validity of that estimate.

With this new motor operator estimate, the algorithm for the simple reaction-time task is superimposed on top of the METT algorithm for a 60 gwpm typist typing a sentence from a standard typing test (Salthouse and Saults' typists averaged 63.1 nwpm). Since the reaction time task was not automatic and the typing task was emphasized, the operators that perform the reaction time task were woven in between those of the typing task, with the typing operators taking precedence. If the perceptual processor was not busy perceiving a word when the tone started, then the perception of the tone began at the onset of the tone, otherwise the perception of the tone began as soon as the perceptual processor completed the perception of the word. When the perception of the tone was complete, if the cognitive operator was not busy doing something for the typing task, the verification of the tone began, otherwise the verification waited until the typing cognitive operation was complete. Then the cognitive operators for the typing task and the reaction-time task were woven together, alternating between tasks if they were competing for cognitive processing time. The motor operator to press the foot pedal began after the foot-press was initiated by the cognitive processor and the motor processor was not busy typing a character.

The reaction time in the dual task was predicted by measuring the time between the onset of the tone and the completion of the foot-press motor operator. The average reaction time predicted by this analysis is 435 msec, 1% away from the 431 msec observed by Salthouse and Saults.

The effect on the average interkey interval for Salthouse and Sault's task is reported to be small. From Table 2, the mean of the interkey interval was 181 msec for the normal typing task and 185 msec when the concurrent reaction time task was added. The METT predicted an average interkey interval of 195 msec for the normal typing task, 8% above the observed value. The interkeystroke interval for those keystrokes that were interrupted by the reaction-time task was predicted to be 240 msec. However, this was only for those keystrokes that were directly involved in the concurrent task, not the average for the entire typing task. Salthouse and Saults presented 30 tones within a 1200 character passage [15]. With this ratio, the overall average interkey interval for typing with the concurrent task was predicted to be 196 msec, 6% above the observed time. This analysis supports the claim that a concurrent task has little or no effect on the typing speed of an expert typist.

This analysis of the dual task shows that the model of immediate response behavior and the METT are truly integrated. Because they spring from the same theoretical base, use similar algorithms, and share parameters, they can be woven together to predict the outcome of a complex task involving both types of behavior.

The METT was applied to the experimental tasks that demonstrate the rest of the 29 phenomenon and was shown to account for 72% of the 29 phenomena (21 phenomena) at least qualitatively. It accounts for 55% of the 29 phenomena (16 phenomena) quantitatively, making predictions to within 20% of the observed performance. If you remove from consideration those phenomena the METT does not attempt to predict -- phenomena associated with errors and separate fingers -- it accounts for all of the 21 remaining phenomena at least qualitatively and 76% quantitatively (John, 1988).

CONCLUSION

The model of immediate response behavior and the METT are now part of the MHP and the family of GOMS models; they share the architecture and several parameter estimates with each other and the earlier work. This process of expanding the MHP, or similarly motivated models of human performance, needs to continue until we can model all the activities present at the human-computer interface: reading, visual search, planning, problem solving, pointing, etc. These activities must be understood, not only in terms of performance time, but also with respect to errors and learning time. Expansion can occur by collecting available data about an activity and constructing a model that fits within an existing framework, by taking an existing, successful, stand-alone model and massaging it so that it is integrated with other models, and by pushing the current MHP into new domains (e.g. errors). Thus, work in this area of cumulating the science of HCI can have many starting points and go off in many directions, but the eventual goal is clear: an applied science of human performance useful for HCI design, based on an integrated, coherent model of human information processing.

REFERENCES

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

2. Duncan, J. (1977) Response selection rules in spatial choice reaction tasks. In S. Dornic (Ed.), *Attention and Performance VI.* Hillsdale, NJ: Lawrence Erlbaum Associates.

3. Fitts, P. M. & Seeger, C. M. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology, 46,* 199-210.

4. John, B. E. (1988). *Contributions to engineering models of human-computer interaction.* Doctoral dissertation, Department of Psychology, Carnegie Mellon University.

5. John, B. E. & Newell, A. (1987). Predicting the time to recall computer command abbreviations. In J. M. Carroll & P. P. Tanner (Eds.), *CHI+GI 1987 (Toronto, April 5-9).* New York: ACM, 33-40.

6. John, B. E. & Newell, A. (in press). Toward an engineering model of stimulus-response compatibility. In R. W. Proctor & T. G. Reeve (Eds.), *Stimulus-response compatibility: An integrated approach.* North-Holland.

7. John, B. E., Rosenbloom, P. S., & Newell, A. (1985). A theory of stimulus-response compatibility applied to human-computer interaction. In L. Borman & B. Curtis (Eds.), *CHI'85 Human Factors in Computing Systems (San Francisco, April 14-18).* New York: ACM, 213-219.

8. Morin, R. E. & Forrin, B. (1962). Mixing two types of S-R associations in a choice reaction time task. *Journal of Experimental Psychology, 64,* 137-141.

9. Newell, A. & Card, S. K. (1985). The prospects for psychological science in human-computer interaction. *Human Computer Interaction, 1,* 209-242.

10. Newell, A. & Card, S. K. (1986) Straightening out softening up: Response to Carroll and Campbell. *Human Computer Interaction, 2,* 251-267

11. Rosenbloom, P. S. (1983) *The chunking of goal hierarchies: A model of practice and stimulus-response compatibility.* Doctoral dissertation, Carnegie Mellon University.

12. Rothkopf, E. Z. (1980) Copying span as a measure of the information burden in written language. *Journal of Verbal Learning and Verbal Behavior, 19,* 562-572.

13. Salthouse, T. A. (1985) Anticipatory processes in transcription typing. *Journal of Applied Psychology, 70,* 264-271.

14. Salthouse, T. A. (1986) Perceptual, cognitive, and motoric aspects of transcription typing. *Psychological Bulletin, 99* (3), 303-319.

15. Salthouse, T. A. (1988) Private communication.

16. Salthouse, T. A. & Saults, J. S. (1987) Multiple spans in transcription typing. *Journal of Applied Psychology, 72,* 187-196.