Jones, G., & Ritter, F. E. (2000). Over-estimating cognition time: The benefits of using a task simulation. In *Simulating Human Agents, American Association for Artificial Intelligence Fall 2000 Symposium Series*. 67-74. Menlo Park, CA: AAAI Press.

Over-estimating cognition time: The benefits of interacting like a human with a task simulation

Gary Jones (gaj@psychology.nottingham.ac.uk)

School of Psychology, University of Nottingham, Nottingham NG7 2RD (UK)

Frank E. Ritter (frank.ritter@ist.psu.edu)

School of Information Sciences & Technology, Penn-State University

Abstract

Tying a cognitive model to a task simulation and having it interact using a model eye and hand provides many benefits, such as accounting for both the physical constraints of the task and the time spent interacting with the task. A cognitive model and task simulation of a physical problem solving task are presented. The model is shown to match adult behaviour on the task very well. Analysing the interactions between the model and the task simulation shows that approximately 50% of the model's task time is spent on interaction, that is, eye movements, eye fixations, and hand movements. The breakdown shows that any cognitive model of a physical task, including all human-computer interaction tasks, that does not simulate task interactions is likely to over-estimate the time spent on cognition and therefore attribute too much emphasis to cognition and cognitive learning.

Introduction

Cognitive models have provided valuable insight into the possible ways in which humans may be solving particular tasks (e.g. Anderson & Lebiere, 1998; Newell & Simon, 1972). The models can provide a test of a particular verbal theory, or the model itself can be the theory of how the task is accomplished. Modelling has been applied to various task domains, including those involving physical interaction (i.e. eye movements, hand movements, or both).

Which aspects of human behavior are most worth capturing in a human modeling architectures? This paper presents a case for including models of perception to interact with an external task simulation. The central part of the paper illustrates several of these benefits by detailing a task for which a model and simulation have been developed, and shows that the influence of cognitive aspects of the task could easily have been over-estimated if the model was not linked to an external task simulation. In addition, the benefits of examining particular aspects of simulation use (e.g., the amount of time spent on visual search) are shown to provide insights into the task that would not have been available had an external task simulation been omitted.

Why have relatively few cognitive models used a task simulation?

There appear to be three simple reasons as to why task simulations have not been used extensively with cognitive models. First, the task particulars may not demand it. For example, in modelling children's performance on a balance beam task (e.g. McClelland & Jenkins, 1991), there are really only two observable states: the initial state of the beam, and the final state of the beam (i.e. whether it balances or tips to one side). The mental processes that the child may be using have to be inferred because there is so little visible, overt behaviour to aid the modeller. Such tasks increase the scope for criticism of the model because there is little support that the mechanisms by which the model performs the task map onto the mechanisms that the child may use.

Second, the model of the task may have been developed in an architecture or modelling environment for which there is little support for the development of a task simulation. This is true of most model development languages. For example, Lisp has graphical simulation aids (e.g. Garnet, Myers et al., 1990), but specifying how the model and simulation are linked together is left to the modeller. Only recently have cognitive architectures began incorporating simulation environments within the architecture (Byrne, 1994, for CAPS; Byrne & Anderson, 1997, for ACT-R; Ritter, Baxter, Jones & Young, 1999, for Soar) to create integrated cognitive architectures (Pew & Mavor, 1998).

Third, in the absence of an assortment of tools to help build a task simulation, the development of an adequate simulation requires additional time and effort which many researchers do not have. This is partly because the simulation often needs to be developed in a different language from the model, and in an environment where there is little support for building simulations. Although current cognitive architectures (in particular: ACT-R, Anderson & Lebiere, 1998; EPIC, Kieras & Meyer, 1997; Soar, Congdon & Laird, 1996) now include support for creating task simulations, they are recent developments. There is still comparatively little use of task simulations in models of tasks which involve interaction.

Benefits of including an external task simulation

An interactive task means that a cognitive model of the task is very likely to require a task simulation. There are two methods for linking the task simulation with the model: either include the task environment within the model's implementation language, or have the model interact with an external simulation of the task environment (usually a graphical simulation written in a fairly specialised language, but possibly a graphical environment attached to a cognitive architecture)[1]. Incorporating aspects of the simulation within the modelling environment (e.g. John, Vera & Newell, 1994; Peck & John, 1992) takes less time but is not ideal because there is every likelihood that some aspects of the task are taken for granted when they are actually difficult. Many existing models may perform tasks too quickly for this reason. Even external simulations will have idealisations (e.g. no slips of action) meaning they may also perform the task too quickly (Kieras, Wood & Meyer, 1997), but the magnitude of the under-prediction will be greatly reduced. Another problem for representing the simulation within the modelling environment is that the modelling environment is not ideal for representing a simulation (because it is not a language which is specialised for writing simulations) and so the task representation is often simplified.

The above problems highlight the need to complete the modelling task by including an *external* task simulation, which can be linked to any cognitive model of the task and not just the particular model which has been developed. The development of an external task simulation has clear advantages over including the task environment within the model (or not including a task environment at all):

1. The simulation can indicate how complex the task is and how great a role the eye and hands play, based on the number of times the model has to interact with the simulation, and for what length of time.
2. Modelling only the high-level processes involved in the task assumes that access to the external task information is effortless. Accessing the external task information may in fact influence speed and accuracy in the task (Anderson, Matessa & Lebiere, 1997). For example, the main source of the extra time required to complete a subitizing task (Jensen, Reese & Reese, 1950) is likely to be the extra fixations required when there is a larger number of objects.
3. An external task simulation enables the parameters associated with the eye and hands to be changed easily. If the representations of the eye and hands were within the model, the parameters can be difficult to modify, because the extent to which alterations can be made is restricted by the cognitive modelling environment. For example, altering the area that the eye covers would be easy for an external simulation because the simulation should be written in an appropriate specialised language. The language of the model (e.g. rule based or connectionist) could make these changes awkward.
4. The simulation may indicate possible aspects of the task which occur in parallel. Aligning the model/simulation behaviour against subject behaviour may reveal task processes where the model is too quick or too slow, based on matching specific behaviour to specific time points. Although a simulation is not strictly required to elicit this knowledge, in tasks which emphasise use of the eye and hands, it is likely that mismatches will be found with regard to eye/hand behaviour being done concurrently (indeed, the EPIC architecture assumes this).
5. Modelling only the high-level processes involved in the task, and not modelling how information is obtained, may mean modellers are “granting themselves unanalysed degrees of freedom in terms of choice of representation” (Anderson et al., 1997, p.442). The success of the model may simply be because of the chosen representation and not because of the high-level processes that have been modelled.

6. The behaviour of the model can be viewed on the graphical representation that the simulation provides.
7. A task simulation offers the opportunity to examine task behaviour that is difficult to obtain from subjects (i.e. providing further measures of behaviour). For example, the time spent performing visual search would require tracking eye movements which for some tasks are difficult to obtain (especially ones where the subjects are children).

Allowing a model to interact with a clear task simulation allows more fine grained comparisons to be performed. Models that have interacted with a task similar or identical to that seen by subjects have generally been able to profit from it. Work within a variety of architectures, for example, EPIC (Kieras, Wood & Meyer, 1997), EPIC-Soar (Chong & Laird, 1997), Soar (Nelson, Lehman & John, 1994), and ACT-R (Anderson, Matessa & Lebiere, 1997) have shown that it is possible and that the resulting models can be quite accurate and useful.

What has not been fully done, which is presented here, is a break down and summary of where the task time goes—what proportion of the behaviour in an example task is spent on cognition and what proportion is spent on interaction. The example task used will show that the time spent on interaction is about equal to the time spent on cognition.

An example task to illustrate some of the advantages of using an external task simulation is presented, together with the actual task model and task simulation used. The model's match to adult behaviour is summarised. Some of the benefits of using task simulations are then illustrated using the cognitive model and task simulation environment. The model predicts that for this task (which is similar to many tasks in HCI and cognitive psychology) over half of the task time is spent on interaction.

The task, the model, and its simulation

The task, the model, and the task simulation are only discussed briefly here. More detailed descriptions are available for the task (e.g. Wood, Bruner & Ross, 1976), and the model and task simulation (Jones, 1998; Jones, Ritter & Wood, 2000).

The Tower task is analogous in many ways to direct manipulation graphical user interfaces. Like them, the problem solver in this task has to choose objects to manipulate, pick up or select objects,, and arrange or manipulate objects. The difference here is the objects are modelled and represent three-dimensional objects, but in terms of timing and distribution of effort, it will be very similar to numerous tasks, such as drawing in MacDraw, circuit layout with a graphical user interface, and simulation games like Sim-City.

<u>The Tower task</u>

The Tower task (Wood & Middleton, 1975) is a problem solving puzzle in which a pyramid (shown in Figure 1) must be assembled from a set of 21 wooden blocks. There are six layers to the pyramid; the lower five consist of four blocks each, with a single block as the top layer. The blocks which comprise each layer are all of the same size, but the size of blocks changes uniformly across layers. The blocks in the lower layers all share the same characteristics (as shown in Figure 1), differing only in size.

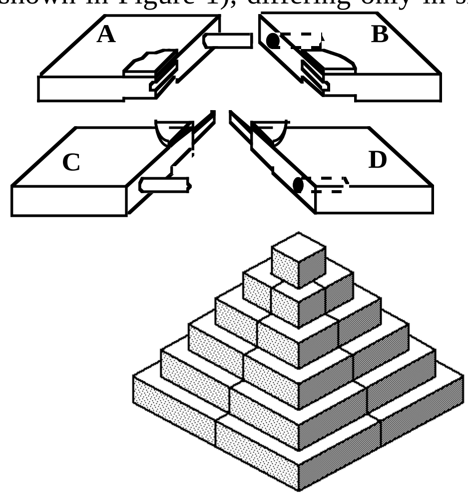

Figure 1. On the top are the four blocks that make up each of the lower five layers in the Tower task, together with (on the bottom) the final assembly of the Tower.

The interactive nature of the task enables a variety of measures of behaviour to be taken. Matching subject data on multiple measures provides more constraints on the model because it has to fit the subject data on more data points. Providing a good match to subject data also allows the processes of cognition and interaction to be examined to give indications as to how tasks are being completed, and where task learning is occurring. In addition, the task allows timing data to be recorded. Timing data is often neglected by cognitive models even though it provides a very important measure, partly because it indicates possible areas where learning takes place, and helps to indicate areas where the model is either too quick or too slow in accomplishing a component of the task. Where the model does not match the subject data can indicate where the model can be improved.

The task has been used extensively to examine the effects of instruction and tutoring in children, who show a wide range of behaviour across different ages. Three year old children are complete novices who can hardly be taught the task, whilst eight year old children are relative experts who can teach themselves. In general, older children accomplish more correct operations, produce less errors, and take less time than their younger counterparts (Murphy & Wood, 1981; Wood & Middleton, 1975).

The blocks that comprise the task have several different features. The single block which comprises the top layer has two salient features: its size, and a circular depression. Every other block in the Tower task has five salient features: size; a quarter circle depression (except for blocks in the bottom layer); a quarter circle elevation; a peg or a hole; a halfpeg or a halfhole. One or a combination of these features will be used when subjects select one or more blocks. For example, one of the largest blocks which has a peg may be required if one of the largest blocks having a hole is currently being held.

The block features allow the task to have several interesting characteristics. First, every peg and half-peg can fit into every hole. Second, the position of pegs and holes from the edge and bottom of each block is the same, such that placing the peg of one block into the hole of a different sized block can result in a construction which is “flush” on its outer edge. Third, each layer is formed by putting together correctly the four blocks comprising a layer (excluding the single block top layer), such that the quarter circle elevations on each block form a circular elevation, and the quarter circle depressions form a circular depression. The diameters of the circles are the same for every layer, permitting the stacking of layers in any order of size. Fourth, the six layers all differ in size by the same magnitude (the size6 blocks are the largest; the size1 block is the single top block). For example, the difference between the size two and size three layers is the same as that between the size five and size six layers (this is shown in Figure 1).

The features of the blocks allow various incorrect constructions to be made, such as placing halfpegs into holes, fitting different sized blocks together, and fitting blocks so that their outside edges are flush, but the blocks are not connected in any way (e.g. blocks B and D can be connected in this way if block B is rotated 180 degrees and placed alongside block D with their quarter circles aligned). How an incorrect construction is produced can help give insights as to what task knowledge is known. For example, if a subject always produces incorrect constructions which are flush on their outer edges, then this implies that some knowledge of the appearance of correct constructions is known. Error-free performance involves twenty correct constructions; three for the production of each of the five layers (for example, placing the peg of block A into the hole of block B to make a pair, the same process for blocks C and D, and then fitting the two pairs together), and five for stacking layers.

A model and simulation of the Tower task

Both a cognitive model and a task simulation have been developed for the task; the two interact in order to complete the pyramid. The cognitive model is based in the ACT-R cognitive architecture (Anderson, 1993) and consists of 317 rules. Learning involves altering the strength of rules based on the perceived success or failure of the rules in achieving construction goals. The model interacts with a simulation of the task which includes all blocks and block features, and an eye and two hands. The model directs the eye and hands in order to look at what objects are on the table, and to pick up, drop, assemble, and disassemble blocks and constructions.

For the simulated eye, three areas of decreasing visual quality are defined: fovea, parafovea, and periphery. To be certain of viewing blocks and features correctly, they must be seen in the fovea. New information concerning what the simulated eye sees is only given to the model when the model requests a fixation from the simulated eye. Both the decreasing visual quality and the model being forced to request information from the simulated eye mean that the model's view of the world is not the same as the external task simulation's. This causes occasions where blocks will be selected which do not have the particular features that the model was looking for. Subjects also show this type of behaviour.

The model also incorporates timing estimates, meaning timings for the complete task and sub-components of the task can be predicted by the model. The timings are based on combining cognition and interaction. Times for cognition are taken from ACT-R and are based on its default parameters. For interaction times, times for the eye are in accordance with a review of the vision literature (Baxter & Ritter, 1996), and the hand movement timing is based on an estimate from the adult subjects (Jones & Ritter, 1997).

A comparison of the model's behaviour and adult's behaviour on the Tower

The behaviour of ten runs of the model will be compared to the behaviour of five adult subjects completing the Tower task. The methods by which the subject and model data were obtained will be briefly described, followed by a comparison of the two across several measures of behaviour. The validated model can then be used to examine specific aspects of interaction, such as the specific allocation of task time to each different process of interaction.

Obtaining the adult and model data

Five adult subjects who had never encountered the task before were used. The subjects were asked to build the Tower unaided, whilst giving verbal protocols. All construction behaviour and verbalisations were transcribed. The results of the construction behaviour will be reported here (the verbalisations aided the development of the cognitive model). Ten runs of the model are used to compare the model's results to the subject's results, in order to minimise the effects of the random components that the model has (e.g. a random unseen block is selected when the model wishes to fixate on a block it does not yet know the features of). The construction behaviour of the model was automatically transcribed in exactly the same way as the transcriptions for the adult behaviour. In this way, exactly the same analyses can be carried out for both the adults and the model.

Comparison of adult and model behaviour

There are two types of measure that the model has been compared with: overall measures and the same measures taken over a series of layers (Jones, 1998). The overall measures describe general task behaviour in building the Tower, such as the time taken and the number of constructions made.

These same measures taken over a series of layers indicate if there is any learning taking place whilst building the Tower. This type of analysis is necessary because the task involves building five layers where the blocks in each layer share the same characteristics. Learning should occur because each subsequent layer should take less time to construct. The overall measures and layer-by-layer measures will be shown in turn so as to illustrate the general fit of the model to the adult data.

Overall measures

A variety of overall measures exist. The two most important measures, which are reported here, are the time taken to complete the Tower, and the number of constructions made in completing the Tower (these define the task and influence scores on other overall measures). Table 1 shows that the model provides a close match to the subjects for the two primary measures of overall behaviour. Jones (1998) shows that the model matches the subject data on a total of seven out of nine measures of overall behaviour.

Table 1. Time taken and number of constructions made in completing the Tower for adults and the model. Standard deviations are in parentheses.

	Adults (N=5)	Model (N=10)
Time taken	126.6 s (34.0)	129.0 s (31.5)
Construction attempts	22.8 (2.9)	23.1 (2.4)

Layer-by-layer measures

Learning is expected to occur throughout the task because the layers of the Tower share the same block characteristics. Subjects should be faster at constructing subsequent layers as they become more familiar with the block and construction characteristics. Figure 2 shows the mean time taken to construct each layer for the adults and the model. There is a good correlation between the adults and the model for the mean time taken to construct each layer (r=0.96), but not for the mean number of constructions made in producing each layer (r=0.40) because the curve for adults is flat. Comparisons are also favourable for the RMS error for each layer, which indicates the average percentage difference between the model scores and subject scores for each layer. The RMS error is good for both the time to construct each layer (4.1%) and the number of construction attempts made in producing each layer (5.7%).

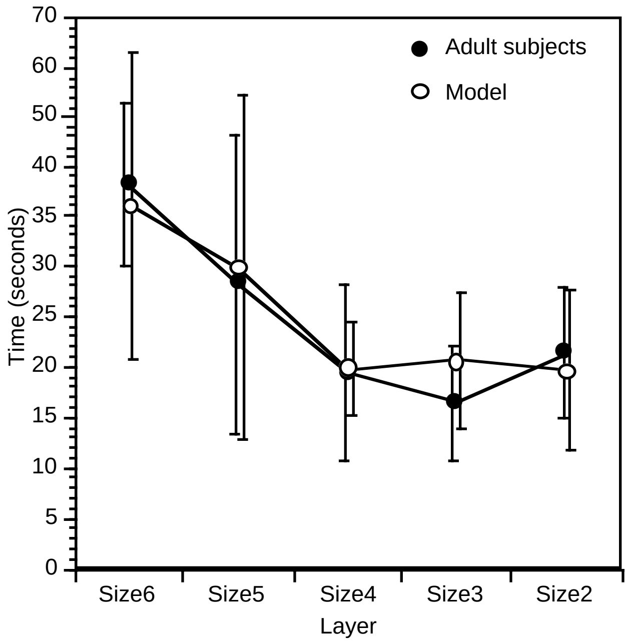

Figure 2. Time taken (in seconds) for adults and the model to complete each layer. Error bars are to the left for adults, right for the model.

Looking at the inner working of the model

Having a model which matches the data this well and which interacts with an external task simulation, allows several questions regarding interaction to be answered in a more theoretical way. The first is to examine whether the speed-up in constructing subsequent layers is because of the reduced visual search that is required (because there are less blocks on the table). The second is to examine to what extent eye and hand timings predict task behaviour. The third shows additional benefits of using a task simulation which is external to a cognitive model.

Does reduced visual search account for all of the speed-up in layer building?

With these adult subjects, the reduction in time to construct layers combined with a relatively fixed number of construction attempts per layer suggests that task learning by adults is not due to more efficient construction strategies or due to making less construction errors. The reduction in the time to produce subsequent layers is perhaps because there are less blocks to select from as the task progresses (i.e. the reduction is due to reduced visual search).

The task simulation enables the behaviour of the model to be analysed in more detail to find where the time is spent when constructing each layer. The interaction between the model and the task simulation allows the extraction of timings for moving and fixating the eye, manipulating the blocks, and cognizing. The time spent on each of these processes can be seen in Figure 3.

The model does not incorporate learning in its simulation eye and simulation hands, so the timings allocated to a single eye movement (50 ms), a single fixation (200 ms), and a single hand movement (550 ms) are fixed across the task. A decrease in interaction time thus represents less interactions, not faster interactions. The lack of learning in perception will slightly over-estimate eye movements and fixations, but the perceptual skills used in the task should be fairly well practised.

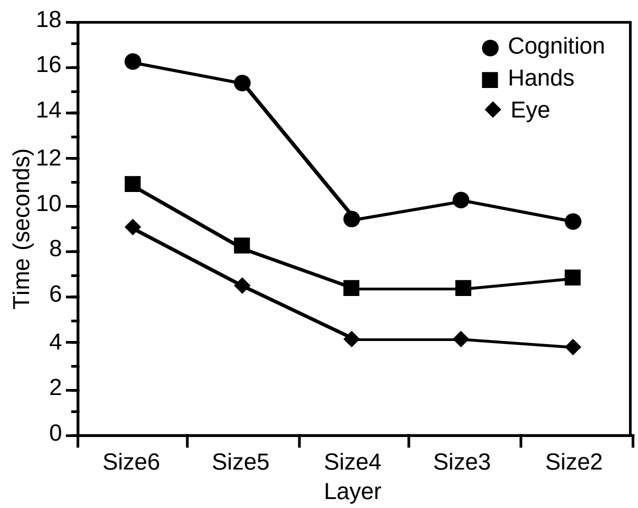

Figure 3. Contributions (in terms of time) of cognition, hand movements, and eye movements and fixations involved in completing each layer of the Tower. The final layer (the single pinnacle block) is omitted because it is trivial.

The reduction in the amount of eye and hand use accounts for 84% of the total reduction in time taken between constructing the first layer (size6) and the second layer (size5). The reduction in the amount of eye and hand use accounts for 42% of the reduction in time taken for the second and third layers constructed (size5 and size4). A reduction in cognitive effort therefore accounts for 16% and 58% of the reduction in time taken between the first and second, and second and third layers respectively. The reduction in time between the first, second, and third layers is due to cognitive learning as well as a reduction in visual search. This would suggest that adults are learning on the task even though it is not reflected in their construction attempts. The timings for the eye and hands remain constant after the size4 layer is produced, suggesting a minimum time for searching and constructing the blocks involved in the task. This helps to explain why adult performance does not improve much after completing the size4 layer.

The fact that eye movements, fixations, and hand movements are having a marked influence on the reduction in layer timings (they also account for 52% of the total time to complete the Tower) indicates that to ignore interactions with the environment will lead to cognitive models under-predicting task elements.

Simulations carried out within the EPIC architecture also show the influence of interaction on task times. For example, one simulation shows that behaviour in highly interactive short time-span tasks such as menu search cannot use a serial-search strategy. For serial-search, the time expenditure in eye movements and fixations would mean subjects had no time to process any of the information – therefore there has to be some parallel processing (Kieras & Meyer, 1997). Using EPIC's interaction constraints enabled some hypothesised task strategies to be ruled out because subjects could not have completed the task as quickly as they did if they had used the strategies.

There could be problems in placing so much emphasis on eye and hand timings, because the model and task simulation do not incorporate any learning in their perception and action. However, the findings should still be reliable for two reasons. First, a large amount of the eye and hand behaviour occurs during the building of the first two layers—no significant speed-up would be expected this early into the task (in fact, Zelinsky and Sheinberg, 1995, did not find any differences in fixation duration for a visual search task involving 17 items over the same task involving only 5 items – time differences were due to the time taken to initiate the first eye movement). Second, the eye and hand timings account for more than half of the overall task time; even if speed-up was included, the eye and hand timings would still account for a significant portion of task times.

To what extent do eye and hand timings predict task behaviour?

The extent to which the eye and hand timings influence the overall behaviour of the model can be examined by seeing how well the eye timings, and the hand timings, correlate with the model timings as a whole. For clarity, cognition timings will also be correlated. The time spent on each of the three individual processes (eye, hands, cognition) in the model correlate very well with the full model for timings per layer (minimum r=0.97). The eye and hand timings per layer also correlate very well with the model's construction attempts per layer (r=0.97 and r=1.00 respectively), although the correlation between the cognitive timings per layer and construction attempts per layer is not as high (r=0.87).

The eye and hand timings are better predictors of the number of construction attempts that the model makes than cognition timings are. This is because some fit attempts are aborted before a physical construction attempt is made. For example, features seen in the parafovea are subject to a certain amount of noise. This means one feature (e.g. a halfpeg) can be mistaken for another (e.g. a peg). Once the features are in the fovea the mistake will be caught and the fit attempt may be aborted. This represents cognitive effort without the observed behaviour of attempting to fit blocks. To some extent this may explain why adult timings and construction attempts do not correlate well, because the task time increases without there being a corresponding construction attempt.

Having eye and hand timings that are easy to extract from the model means that it is relatively straightforward to see the extent to which the eye and hand timings predict the behaviour of the adult subjects on the task. If they are a good predictor, then this suggests that the eye and hands play a significant role in task behaviour. The extent to which the eye and hand timings predict the behaviour of adult subjects can be examined by correlating the time spent on each process with the time the adult subjects take to complete each layer. This is shown in Table 2 (which includes cognition timings, for clarity). The correlations for the eye, hand, and full model timings are all similar in how well they correlate with the adult subject layer timings. This suggests that a good predictor of task time is the time spent looking at and manipulating blocks (this data is not available for the adult subjects).

Table 2. Correlations between adult subjects and the model when individual model processes are extracted out of the timing data.

Model process	Process time	Correlation with adult layer timings	Correlation with adult layer construction attempts
Eye movements and fixations	27.5 s (21.8%)	0.97	0.35
Hand movements	38.5 s (30.6%)	0.98	0.37
Cognition	60.0 s (47.6%)	0.91	0.32
Full model (excluding stacking final top block)	126.0 s (100.0%)	0.96	0.40

The time spent interacting with the simulation eye and hands should also correlate well with the number of construction attempts that adults produce, because any interaction is toward the goal of completing each layer. Further construction attempts should mean more fixations on blocks, and more blocks being manipulated by the hands. Table 2 shows that this is not the case, mainly because the curve for adult construction attempts is flat.

The correlations show that the eye and hand timings are a better predictor of task behaviour than cognition timings. The is an important finding because it again shows the importance of using a task simulation for interactive tasks: the time spent on interaction is the best predictor of task behaviour.

What benefits does an external task simulation give?

Tasks of a physical nature, such as the Tower task, gain clear advantages by having an external task simulation. One reason for this is that all of the important features of the task environment can be properly represented. In the Tower task, the simulation is able to precisely ascertain whenever any object prevents two other objects from being fit together (for example, if there is an obstructing block in between two blocks that the model is trying to fit together). When the model knows the hands are holding two blocks (or constructions), it performs a mental operation of fitting the blocks together. The model only proceeds with actually fitting the blocks if the result of this mental operation is positive. There is an average of 17.4 of these pseudo-construction attempts (in that they are not actually carried out) each time the model completes the Tower, suggesting that they account for a reasonable amount of the task time. Determining whether a block will obstruct the fitting together of two other blocks would be more difficult to accomplish if done within the modelling environment, because the modelling environment is not normally a specialised language for dealing with graphical objects.

There are other advantages to having an external task simulation, such as the ease with which simulation specific variables can be altered to test alternative theories. For example, the timings associated with eye movements and fixations can be altered easily to test developmental psychology hypotheses that propose that children's behaviour may be due to taking longer to look and act. When eye movement timings are increased to 100 ms (from 50 ms) and fixation timings increased to 400 ms (from 200 ms), the time for the model to complete the task increases from 129.0 s to 156.5 s. When hand movement timings are increased to 1050 ms (from 550 ms), the time for the model to complete the task rises to 164.0 s. These and other similar modifications have been used examine theories of behavioural differences between adults and children's performances on the Tower (Jones, 1998).

Conclusions

Having a model interact with an external task simulation provides several benefits in this example task. Most importantly, it has allowed a fine grained comparison between subject behaviour and the model's behaviour. An external task simulation provided the model with an environment where it could perform more of the actions that the subject performed. This allowed the model's behaviour to closely match subject behaviour on the task. The explicit perceptions and actions supported more detailed predictions of task times, which could be compared with subject times. These comparisons also hold across learning—the timing predictions from the model, on a layer-by-layer basis, matched the adult subjects.

The closeness of the match between the behaviour of the model and that of subjects' supported a detailed analysis of task time. Analysing the time spent on cognition and interaction showed that the interactive nature of the task accounted for just over half of the total task time. The model was able to predict this based on the number of interactions with the external task simulation that were necessary to complete the Tower. Although the time spent on interaction may be slightly high because no learning occurs in the model's perception and action components, and because interaction and cognition are performed serially, interaction still clearly represents a significant amount of total task time. The close fit between the model and subject data suggests that adults must also spend a significant amount of time interacting with the task.

The model shows that, for tasks of a highly interactive nature (such as many direct manipulation HCI tasks), any cognitive model which does not include a simulation of the task is likely to over-estimate the cognitive aspects of the task. This is likely to be one of the aspects of behaviour that Kieras (1985) has alluded to when models run too fast. The task used is comparable to larger scale HCI tasks such as CAD/CAM design where a large proportion of the task involves interaction. As such, any models of user behaviour on HCI tasks will be best served by having the model interact with an external task simulation.

The results have shown that a physical and highly interactive task can be successfully modelled using a cognitive model and an external task simulation. The variety of measures used meant that a detailed analysis of the task behaviour (such as how much time was spent interacting) was possible. Using various measures to match model-subject data also has a welcome side-effect: it helps to reduce the "black box" criticism of cognitive models (e.g. Searle, 1980/1997). This criticism suggests that there is little way of knowing that the processes carried out in the model reflect the processes carried out by subjects. By matching the subject behaviour on as many measures as possible, the scope for this criticism is reduced. This is a general advantage of modelling tasks which have many observable measures.

The findings that have been presented indicate that when modelling any task which involves interaction, it is important to also have a simulation of the task which the model is able to interact with. Two critical benefits arise from including a task simulation: there is less likelihood that cognition time will be over-estimated, and a more fine grained analysis of task behaviour can be carried out. In combination, these benefits allow attention to be focused more appropriately on how much time the user spends performing each task process.

References

Anderson, J. R. (1993). Rules of the mind. Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Lawrence Erlbaum Associates.

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction, 12*, 439-462.

Baxter, G. D., & Ritter, F. E. (1996). Designing abstract visual perceptual and motor action capabilities for use by cognitive models. (Technical Report 36). CREDIT, Psychology Department: University of Nottingham.

Byrne, M. D. (1994). Integrating, not debating, situated action and computational models: Taking the environment seriously. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society,* 118-123. Hillsdale, NJ: Lawrence Erlbaum Associates.

Byrne, M. D., & Anderson, J. R. (1997). Enhancing ACT-R's perceptual-motor abilities. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum Associates.

Chong, R. S., & Laird, J. E. (1997). Identifying dual-task executive process knowledge using EPIC-Soar. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society,* 107-112. Mahwah, NJ: Lawrence Erlbaum Associates.

Congdon, C. B., & Laird, J. E. (1996). The SOAR user's manual version 7.0.3. Ann Arbor, MI: University of Michigan.

Jensen, E. M., Reese, E. P., & Reese, T. W. (1950). The subitizing and counting of visually presenting fields of dots. *Journal of Psychology, 30,* 363-392.

John, B. E., Vera, A. H., & Newell, A. (1994). Towards real-time GOMS: A model of expert behavior in a highly interactive task. *Behavior and Information Technology, 13,* 255-267.

Jones, G. (1998). Testing mechanisms of development within a computational framework. Unpublished doctoral dissertation, University of Nottingham, Nottingham.

Jones, G., & Ritter, F. E. (1997). Modelling transitions in children's development by starting with adults. In *Proceedings of the European Conference on Cognitive Science (ECCS '97),* 62-67. Manchester, UK: AISB.

Jones, G., Ritter, F. E., & Wood, D. J. (2000). Using a cognitive architecture to examine what develops. *Psychological Science, 11*(2), 1-8.

Kieras, D. (1985). The role of cognitive simulation models in the development of advanced training and testing systems. In N. Frederiksen, R. Glaser, A. Lesgold, & M. G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition.* 365-394. Hillsdale, NJ: LEA.

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,* 391-438.

Kieras, D. E., Wood, S. D., & Meyer, D. E. (1997). Predictive engineering models based on the EPIC architecture for a multimodal high-performance human-computer interaction task. *Transactions on Computer-Human Interaction, 4,* 230-275.

McClelland, J. L., & Jenkins, E. (1991). Nature, nurture and connections: Implications of connectionist models for cognitive development. In K. VanLehn (Ed.), *Architectures for intelligence*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Murphy, C. M., & Wood, D. J. (1981). Learning from pictures: The use of pictorial information by young children. *Journal of Experimental Child Psychology, 32,* 279-297.

Myers, B. A., Guise, D. A., Dannenberg, R. B., Vander Zanden, V., Kosbie, D. S., Pervin, E., Mickish, A., & Marchal, P. (1990). Garnet: Comprehensive support for graphical, highly-interactive user interfaces. *IEEE Computer, 23,* 71-85.

Nelson, G., Lehman, J. F., & John, B. E. (1994). Integrating cognitive capabilities in a real-time task. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, 658-663. Lawrence Erlbaum Associates.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Peck, V. A., & John, B. E. (1992). Browser-Soar: A computational model of a highly interactive task. In *Proceedings of the CHI '92 Conference on Human Factors in Computing Systems*, 165-172. New York, NY: ACM.

Pew, R. W., & Mavor, A. S. (1998). *Modeling human and organizational behavior: Application to military situations*. Washington, DC: National Academy Press.

Ritter, F. E., Baxter, G. D., Jones, G., & Young, R. M. (in press). Exploiting user interface management systems in cognitive modelling. *ACM Transactions on Computer-Human Interaction.*

Ritter, F. E., & Major, N. P. (1995). Useful mechanisms for developing simulations for cognitive models. *AISB Quarterly, 91(Spring),* 7-18.

Searle, J. R. (1980/1997). Minds, brains, and programs. In J. Haugeland (Ed.), *Mind Design II*. Cambridge, MA: MIT Press.

Wood, D., & Middleton, D. (1975). A study of assisted problem solving. *British Journal of Psychology, 66,* 181-191.

Wood, D. J., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17,* 89-100.

Zelinsky, G., & Sheinberg, S. (1995). Why some search tasks take longer than others: Using eye movements to redefine reaction times. In J. M. Findlay, R. Walker, & R. W. Kentridge (Eds.), *Eye movement research: Mechanisms, processes and applications*, 325-336. Amsterdam: North-Holland.

---

[1] A further review of approaches is available in Ritter and Major (1995).