Evolution of Response Time Distribution in Menu Search

Jake Costello (jcostel1@mail.depaul.edu)
Peter Hastings (peterh@cdm.depaul.edu)
College of Computing and Digital Media, DePaul University
243 South Wabash Avenue, Chicago IL 60604, USA

Abstract

This paper presents an ACT-R model of a realistic menu item selection task from a previous study. The model aims to explain observed trends in both the mean and dispersion of response time (RT) with respect to both menu length and amount of practice. Our model follows previous work in assuming that the primary strategies for determining the location of the target menu item are visual search and recall from memory. The model introduces a hypothesis about how the visual search is performed that yields a very good fit with no free parameters for the slope of the linear relationship between RT and menu length. We validated this hypothesis by testing the model's predictions for the dispersion of RT. The model was also used to generate predictions to test the hypothesis that the observed RT reduction with practice could be predicted as the sum of the time required for initiating the task and the lesser of the realized visual search time or realized recall time. The model generated predictions for both means and dispersion of RT as a function of practice.

Introduction

Menu item selection is an important practical topic in HCI. Locating a menu item can be accomplished by either visual search or by recalling the location from memory. Each of these strategies requires the careful coordination of cognitive, perceptual and motor mechanisms to be efficient. Previous models of practice effects in menu item selection have focused on the evolution of only the mean RT (Cockburn, Gutwin, & Greenberg, 2007; Das & Stuerzlinger, 2010). We investigated how the distribution of response time (RT) evolved across experimental conditions. Our aim was to explore the coordination of the underlying mechanisms within each of these strategies, as well as the interaction of the two strategies.

Response time variance has gained attention more recently as a useful window on the psychological processes underlying task performance. This approach has received significant attention in mathematical psychology accounts of speeded two-choice tasks (Wagenmakers & Brown, 2007). We extended that concern to menu item selection tasks by investigating the psychological processes involved and how they contribute not only to the evolution of the mean RT, as with previous work, but how they contribute to the evolution of the dispersion of the RT distribution.

The models from Cockburn et al. (2007) and Das and Stuerzlinger (2010) both explained the observed reduction in RT with practice as driven by a shift from a visual search strategy to a strategy based on recalling item locations. If this explanation is correct, then an accurate detailed description of these strategies and their interaction in terms of the underlying cognitive, perceptual, and motor processes involved should allow accurate predictions about the evolution of the RT distribution. We specifically aimed to make predictions about the dispersion of RT as well as the mean. Dispersion was quantified as median absolute deviation, or MAD, which is less sensitive to outliers than standard deviation.

There were two independent variables that we were interested in. Since visual search is necessarily used when there is no prior exposure to a menu, the distribution of RT without practice is determined entirely by the visual search strategy. To investigate this we modeled the increase in mean RT with menu length. Menu length was used instead of target position because prior studies have indicated that visual search is not strictly top-down (Byrne, 2001). Our second aim was to model the reduction in mean RT with practice. The model was constructed in ACT-R (Anderson, Matessa, & Lebiere, 1997), which has the requisite detailed perceptual-motor modeling capabilities, as well as a detailed model of declarative memory.

Cockburn et al. (2007) described an analytic model of menu performance. The model predicts response times to be a linear interpolation of visual search times and memory based *decision* times. As users gain experience with a menu layout, the interpolation shifts primarily to the decision strategy. The mean RT for visual search is assumed to be a linear function of the number of items in the menu. The decision strategy is modeled with Hick's law (Hick, 1952), which predicts RT for a decision task as an increasing logarithmic function of the number of alternatives.

Das and Stuerzlinger (2010) explored the memory recall strategy in more detail. Their model adds an additional term to the equation used by ACT-R to describe how the activations of declarative memory chunks change with time. This additional term is intended to account for the *proactive interference* with declarative memory learning caused by the presence of distractor menu items other than the target. In contrast to our model, the contributions to the total RT from elements of the task other than declarative memory latency are accounted for by a single multiplicative constant adjusted to fit to the data. It was their hypothesis that proactive interference explains why mean RT does not decrease as fast as would be expected if one assumes that the RT for the item recall strategy is determined primarily by the latency of declarative memory recall.

Our model is similar to both of these in that it describes the transition from a visual search strategy to a recall strategy. The visual search strategy in our model was motivated by the observation that the mean increase in RT per added menu item without prior practice is very low, approximately 82ms after factoring out movement time as described later. Our hypothesis to explain this is that pre-attentive search, requests to shift visual attention, and processing of encoded visual data occur in parallel. In this respect, the visual search component of our model is similar to the EPIC model from Hornof and Kieras (1997). Each of these operations corresponds to a different module in ACT-R. A key part of the ACT-R theory is that the top level modules represent psychological operations that can occur in parallel. To further validate the model, the model was also used to generate predictions about the how dispersion of RT changes with menu length.

Our hypothesis to explain the observed reduction in RT with practice was that RT could be predicted as the sum of the time required for initiating the task and the lesser of the realized visual search time or realized recall time. The model embodies this hypothesis by describing the detailed execution of each strategy in terms of the underlying cognitive, perceptual, declarative memory and motor operations. The lesser of the realized times from either strategy is used because our hypothesis supposes that the strategies operate in parallel. We used the standard ACT-R equations for the declarative memory latency calculations and standard values for most parameters. As described below, we compared our predictions for the reduction in RT with practice to empirical data and to Das and Stuerzlinger (2010).

The model was validated with data from a previous study described in Cockburn et al. (2007). That experiment was designed to test their analytic model for the evolution of mean RT with practice. Response time data was collected for multiple menu lengths across multiple blocks with the same menu. This makes the data useful for investigating trends in the RT distribution versus both menu length and amount of practice. In Cockburn et al.'s model (2007) for predicting mean RT as a function of menu length, a regression line is fit to a subset of the data to estimate visual search parameters. Likewise, a subset of the data was used to estimate the parameters for the decision process, which predicts response time as an increasing logarithmic function of the number of menu items.

In contrast, our model provides a good parameter-free fit for predicting mean RT for the same visual search data. There was an interesting discrepancy between our model and experimental data regarding the shape of the predicted relationship between menu length and MAD. Our model's predictions for how the RT mean and dispersion evolve with practice were not as close a match to the data, though the general trends were reproduced. A closer look at the data shows that there may be more than one effect responsible for the observed reduction in RT with practice.

Materials and Methods

The data used for validating our model came from the 'static + unfamiliar' condition in Cockburn et al. (2007). In that condition, country names were used for the items. For each block, the same set of country names was used throughout, so the menu was *static*. The positions for each item were randomly assigned, though constant throughout the block. Due this randomness, and because the item text was not as familiar as common menu items like 'File ' and 'New', the menus were *unfamiliar*.

The task on each trial was to select a cued menu item. The item to select was cued when the subject clicked a button marked 'Menu'. Movement times were measured from when the cue was displayed, which occurred immediately after clicking the button, to when the correct item was clicked. Trials that resulted in the subject clicking on an incorrect item were ignored.

In each condition, the subject proceeded through 4 menu lengths. There were 7 successive blocks with each menu length. The same set of country names was used to populate the menu for all trials with each length, in the same positions each time. For each block, each menu item was indicated as the target once. The items were indicated as the target in a different order for each block with that menu length. Each country name was used in at most one menu length for each subject.

Since movement times were not part of the dependent variable of interest in the original experiment, a procedure was devised to remove movement time from the mean RT. A separate experimental condition was used to determine average Fitts's Law (Fitts, 1954) coefficients across all participants. In this experimental condition, the menu presented to the subject had only one non-empty item. Response time was measured starting from when the mouse cursor exited the menu button. These coefficients were used to estimate the movement time for each trial, which was then subtracted from the total RT. The result is referred to as decision/search time (DST). DST was the dependent variable of interest in the original experiment, and is the dependent variable of interest in this study as well. As described later, we applied a similar procedure to the data generated from the model.

Cognitive Model

This section describes the complete cognitive model in more detail. The model executes the experimental task in its entirety, from clicking on the 'Menu' button and identifying the target, through moving the cursor, and finally clicking on the target item. For each trial, there are three steps that must be completed:

Search Initiation Clicking 'Menu' button, then shifting attention to and encoding the cued target

Search/Recall Locating the target item by visual search or recall

Search Completion Moving the cursor to the item and clicking on it once located

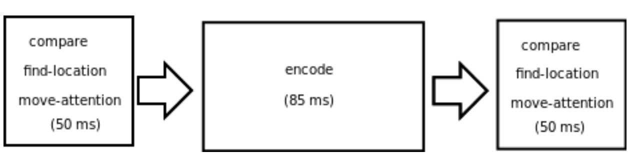
This breakdown provides a useful framework for describing our model. The time required for search initiation puts a floor under all response times. Search initiation and visual search actually overlap, as described shortly. The time required for locating the item is the lesser of the time required for visual search or for recalling the item location, since these strategies operate in parallel. The model executes the completion of the search through clicking on the correct item. This is important in order to accurately model the distribution of times between trials for a given target menu item, which in turn determines the strength of the memory for that item's location and how fast that location can be recalled. The description of the model focuses on the visual search and item recall strategies used for locating the correct item. The initiation and completion of each trial are discussed afterward.

Visual Search Strategy

Consistent with our hypothesis regarding visual search for the target menu item, the model implemented a *pipelined* visual search strategy. The term pipeline is used here, as in Hornof and Kieras (1997), because multiple sequences of serial cognitive and perceptual operations operate in parallel. The sequence of operations for finding and inspecting a possible target item location are: 1. Find a location meeting preliminary criteria using pre-attentive visual search (Treisman & Gelade, 1980). This is modeled with the visual-location module in ACT-R. The model uses a simple approach; it finds the first item below the current location. Thus, it is top-down search. 2. Shift visual attention to the location found using pre-attentive visual search and encode the information. This is modeled with the visual module in ACT-R. 3. Once the information is encoded, it can be processed to determine if the target item has been located. That is the final stage of the pipeline, modeled with the procedural module in ACT-R.

In our model, finding a location, initiating an attention shift, and processing previously encoded information occur in parallel, as shown in Figure 1. As seen from the diagram, the pipeline requires 135 ms per item searched. Noise in the time required for executing the cognitive processing in the pipeline is modeled by adding noise from a logistic distribution to the time required to execute a production, using the `:vpft` parameter in ACT-R (Bothell, 2004, p. 158).

Figure 1: Visual Search Pipeline



Item Location Recall Strategy

The item location recall strategy that runs in parallel with the visual search strategy is very simple. After the target cue has been encoded, an attempt to recall a matching target location from memory is made as soon as declarative memory is free. Item locations are committed to memory only when the correct item is successfully located. This is motivated by the observation that only the correct target item is likely to be fully encoded, and that it is also the most salient item in the visual field. When an item location is recalled, it is used immediately to direct a movement.

The simulated time elapsed between the presentation of a chunk to declarative memory and an attempted retrieval is an important component in determining the activation of the chunk, which in turn determines the latency for retrieving the chunk. In the current model, it was assumed that no time elapsed between trials or between blocks. That means that the time between preceding presentations of a chunk and an attempted recall is determined entirely by the time required to complete the intervening trials. Since the items occur in a random order with each block, the time between successive presentations varies. This in contrast to the model in Das and Stuerzlinger (2010), which assumed a constant time between succeeding trials with each item.

Standard parameters for the ACT-R declarative memory module were used, with two exceptions. The base activation `:blc` was set to 1.0 instead of the default 0.0. This value was chosen so that for menu length 12, the activations of the declarative memory chunks for the locations of each item were greater than zero when the second set of trials started. Without this adjustment, the activations of the chunks increased, but never significantly exceeded zero. Noise in declarative memory activations was modeled by setting `:ans` to 0.2. This adds noise from a logistic distribution with mean 0 and scale 0.2 to the current activation.

Search Initiation and Completion

Previous models have largely ignored the transition into the main visual search or item recall strategy. The time required to locate, attend to, and remember the target cue adds a constant amount of time on average to each trial in our model. Since several productions are required to model this phase, the simulated time required for this phase of the task is not constant, due to the simulated noise in production execution time. After encoding the target cue, both the visual search and recall strategies begin execution. The visual search strategy is initiated by requesting the location of the first menu item below the target cue display area. This initiates the visual search pipeline. Once the visual location of the first menu item is obtained, an attention shift to to that location is initiated in parallel with requesting the location of the next menu item. When attention has been shifted to the location of the first menu item, the text of that menu item can be compared to the target cue. In parallel with encoding the text of the first menu item, an attention shift to the previously determined location of the second menu item is initiated and the location of the next menu item is requested. At this point, the pipeline is full, as in Figure 1.

After the correct item is located, the model generates motor actions to move the cursor to the located item and click on it. As described above, both the experimental and model generated data are treated to remove the mean movement time from each trial to obtain the DST. Our model also addresses motor feature preparation for the mouse click by modeling this as occurring while the mouse movement occurs. Motor feature preparation for the mouse movement occurs when the search is initiated. Whether current models of motor feature preparation are accurate is an active research area (Kieras, 2009), but the model is designed so that changing the assumptions about the need for feature preparation would not significantly affect the results. In the current version of the model, there is no visual verification that the cursor is over the target after the mouse movement.

Using Model to Generate Predictions

In order to compare the model's predictions to those of Cockburn et al. (2007) and Das and Stuerzlinger (2010), we computed estimated decision/search times (DSTs) from the simulated task completion times. The procedure for doing so was as follows. The Fitts's law calibration data from the original experiment was used to fit the coefficient for the form of Fitts's Law used by ACT-R. Cockburn et al. (2007) used the traditional form of Fitts's Law, whereas ACT-R uses a modified form of Welford's formulation (Welford, 1968). This coefficient was used in the ACT-R model runs to generate movement times. To compute DST for the model trial completion times, the same form and coefficients that Cockburn et al. (2007) was used to calculate an MT to subtract from the total time. The reason for using the same form of Fitts's Law and coefficients for computing DST from total time for both the experimental data and the model is that the coefficients for the experimental data are based on times that include clicking on the target item, whereas ACT-R's simulated movement time calculation is strictly for movement. Note that only the movement time from the calibration data was used to to fit the coefficient for ACT-R.

Data from the first block was used to validate the model's predictions about visual search. There were two tests against the empirical data: 1. Mean DST vs menu length 2. Median absolute deviation (MAD) vs menu length

We provide results for both the square of the Pearson correlation coefficient and root mean squared deviation (RMSD) to emphasize how close the predictions are to the data without any scaling. 100 runs of the model were used.

Data from all blocks was used to test the model's predictions for the evolution of the RT distribution with practice. Block number is the dependent variable representing amount of practice. Again, there were two tests against the data: 1. Mean DST averaged across each menu length, with those results then averaged for each block. This specific measure is the one used in (Cockburn et al., 2007). 2. Since the medians for different menu lengths are not the same, we used the data for menu length 12 to look at the MAD of the DST versus block.

Results and Discussion

Results

The plots showing the comparison to experimental data for visual search are shown in Figure 2. The first plot shows mean vs menu length; the second shows MAD vs menu length. For the means, the square of the Pearson correlation coefficient between the model's predictions and the data is 0.99. RMSD is 39 ms. For the MADs, the square of the Pearson correlation coefficient between the model's predictions and the data is 0.96. RMSD is 75 ms.

The plots showing the comparison to experimental data for practice effects are shown in Figure 3. The first plot shows mean DST averaged across each menu length, with those results then averaged for each block. The square of the Pearson correlation coefficient between the model's predictions and the data is 0.75. The second plot shows the MAD vs block for menu length 12. The square of the Pearson correlation coefficient between the model's predictions and the data is 0.65.

Discussion

The model provides a good fit for the mean RT as a function of menu length, as shown by the low RMSD. That the model reproduces the slope is unsurprising since the model was constructed to achieve this result. The model's prediction for the intercept of this relationship is also fairly close, as the first plot in Figure 2 shows.

The comparison of the model's predictions for MAD versus menu length, shown in the second plot in Figure 2, provides additional support for the model. The expected upward trend in MAD with menu length is reproduced. This is expected because longer search require more underlying cognitive operations, each adding to the total noise. Given the way noise in the time to execute a production is modeled in ACT-R, the line for the model's predictions reflects a sum of logistic random variables. There is an interesting discrepancy between the model and empirical data, however. The model's predictions are best fit by a straight line, with the slope in the regression significant at a 0.1% level, but the data is better fit to a logarithmic function of menu length, with the slope again significant at a 0.1% level. In order to explain this effect we may need to account for dispersion in movement times. As described earlier, the mean movement time is removed from DST, but this procedure does not account for the dispersion in movement times.

The comparisons of the model's predictions for the reductions in the mean and MAD of RT with practice are shown in Figure 3. The model reproduces the general trends in the data, but does not provide as good a fit as the visual search model. Recalling that the values in the mean DST trend versus block plot are computed as an average of averages, it is useful to look at some of the values that go into the averages for the data and the model. For menu length 12, the drop in mean RT from block 1 to block 7 for the model is 480 ms versus 496 ms for the experimental data, so the overall reduction in RT matches quite well. The model doesn't reproduce the drop with the same shape, however. The data shows a much bigger drop from block 1 to block 2, 272 ms, than the model does, 66 ms. For menu length 2, on other hand, the data shows a reduction in mean RT of 245 ms from block 1 to block 7, while the model produces a reduction of only 36 ms. In fact,

Figure 2: Mean/MAD of DST vs Menu Length

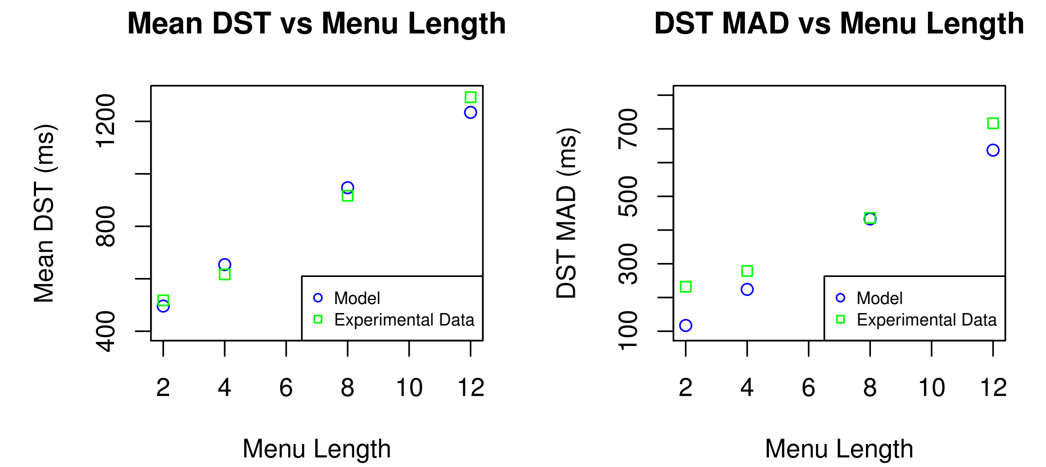
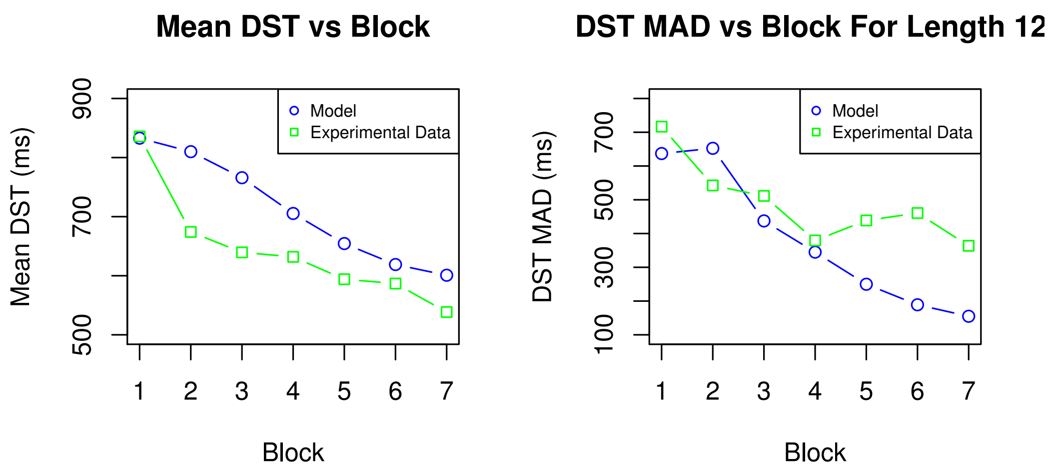


Figure 3: Mean/MAD of DST vs Block



the reduction in mean RT seen in the data is greater than the observed time required to visually search 2 additional items, as reflected in the slope of the line of DST vs menu length. This may reflect more efficient initiation of the search or decision process with practice, in addition to decreased latency for memory recall.

Our model does not fit the observed reduction in RT as well as the model in (Das & Stuerzlinger, 2010). However, our model uses only one non default value for ACT-R parameters and has no free parameters. Das et al. adjusted the activation decay factor to a small value to reflect minimal activation decay due the time between trials with the same target item (2010, p. 40). Our model also predicts a sustained reduction in MAD with increased practice, but the data shows continued reduction in MAD only through the first 4 blocks. This may reflect the fact that our model does not account for all significant sources of RT dispersion, especially between subjects differences.

Conclusions

The model's fit to the visual search data is encouraging. The parameter free fit for the dispersions also provides support for the model. The observed difference between the empirical data and the model in the shape of this relationship merits further investigation. The first step is a more careful accounting for movement times, so that the dispersion in movement times is not ignored. The challenge is that Fitts's Law predicts only mean movement times. Thus, further investigation may require a more detailed model of movement times, as in (Meyer, Abrams, Kornblum, Wright, & Keith Smith, 1988). In addition, the mouse movement may overlap with the cognitive and perceptual processes for finding the correct item, as Cockburn et al. (2007) and others have noted. If the amount of overlap is a function of one or more experimental conditions, such as menu length, this may yield spurious effects in the analysis of trends in RT dispersion.

The model's predictions for the reduction in both the mean and MAD of RT with practice reproduce the expected downward trends, but this part of the model requires more substantial investigation. One step is to determine whether there is more than one underlying effect is responsible for the reduction in RT.

The model provides an alternate explanation for the slower than expected reduction in RT, as compared to Das and Stuerzlinger (2010). Tests of the model's predictions for the evolution of the RT distribution confirm that the mean and dispersion for the RT distribution follow roughly the expected trends given the assumption that the reduction in RT is driven by a transition from visual search to item location recall.

We intentionally used standard ACT-R declarative memory parameters and equations for the initial model, but it may be that the most accurate model could constructed by accounting for not only the detailed execution of the strategy, as our model does, but for additional significant effects in declarative memory learning, as Das et al. proposed (2010).

Acknowledgments

We would like to thank Dr. Alexander Cockburn for sharing his experimental data.

References

Anderson, J., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, *12*(4), 439–462.

Bothell, D. (2004). ACT-R 6.0 reference manual. *Working Draft*.

Byrne, M. (2001). ACT-R/PM and menu selection: Applying a cognitive architecture to HCI. *International Journal of Human-Computer Studies*, *55*(1), 41–84.

Cockburn, A., Gutwin, C., & Greenberg, S. (2007). A predictive model of menu performance. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 627–636). New York, NY: ACM.

Das, A., & Stuerzlinger, W. (2010). Proactive interference in location learning: A new closed-form approximation. In *Proceedings of the 10th international conference on cognitive modeling* (pp. 37–42). Philadelphia, PA: Drexel University.

Fitts, P. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, *47*(6), 381.

Hick, W. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*(1), 11–26.

Hornof, A., & Kieras, D. (1997). Cognitive modeling reveals menu search in both random and systematic. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 107–114). New York, NY: ACM.

Kieras, D. (2009). Why EPIC was wrong about motor feature programming. In *Proceedings of the international conference on cognitive modeling 2009* (pp. 68–73). Manchester, UK: University of Manchester.

Meyer, D. E., Abrams, R. A., Kornblum, S., Wright, C. E., & Keith Smith, J. (1988). Optimality in human motor performance: ideal control of rapid aimed movements. *Psychological review*, *95*(3), 340.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97–136.

Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, *114*(3), 830.

Welford, A. T. (1968). *Fundamentals of skill* (Vol. 423). London, UK: Methuen London.