STRATEGY ADAPTIVITY AND INDIVIDUAL DIFFERENCES

Christian D. Schunn
Lynne M. Reder

I. Conceptions of Individual Differences

Two approaches have traditionally dominated the study of individual differences in cognitive performance, including studies of aging, brain damage, child development, expertise, giftedness, intelligence, schizophrenia, and adult individual differences. The first is the parameters approach. This approach assumes that individuals vary in performance because of differences in some fundamental aptitude or parameter of the brain or cognitive architecture. This parameter approach includes both psychometric approaches to individual differences (e.g., Ackerman, 1989; Snow, Kyllonen, & Marshalek, 1984; Spearman, 1904), and information-processing approaches (e.g., Hunt, Joslyn, & Sanquist, 1996; Just & Carpenter, 1992; Lovett, Reder, & Lebiere, in press; Sternberg, 1977).

Many different parameters have been proposed (e.g., g, gain, inductive reasoning, processing speed, working memory capacity, etc.). There are two especially popular parameters: processing speed and working memory capacity. For example, in terms of processing speed differences, researchers have argued that older children think faster than younger children (Fry & Hale, 1996; Kail, 1988), the elderly think slower than younger adults (Salthouse, 1994), the gifted think faster than average children (Saccuzzo, Johnson, & Guertin, 1994), and schizophrenics think slower than normals

0079-7421/98 $25.00

(Schooler, Neumann, Caplan, & Roberts, 1997). In terms of working memory capacity, researchers have argued that aphasics have reduced working memory capacity (Miyake, Carpenter, & Just, 1995), children have higher working memory capacity (Case, 1985; Fry & Hale, 1996), and general intelligence differences depend heavily on working memory capacity differences (Just & Carpenter, 1992).

The other traditional approach is the strategies approach that assumes there are many different ways in which a task can be attacked and that individuals vary in the strategies they use. For example, proposals have been made that older children use different strategies than younger children in a wide variety of domains (Siegler, 1983), the aged use different strategies than younger adults (Dunlosky & Connor, 1997; Reder, Wible, & Martin, 1986; Shapira & Kushnir, 1985), good students self-explain and poor students do not (Chi, Bassok, Lewis, Reimann, & Glaser, 1989), experts use different strategies than novices (Chi, Feltovich, & Glaser, 1981; Ericsson & Polson, 1988; Larkin, McDermott, Simon, & Simon, 1980), optimists use different strategies than pessimists (Carver & Scheier, 1992), and individuals from different cultures use different strategies (Greenfield & Lave, 1982; Wagner, 1978). A variant of the strategies approach is the styles approach, in which individuals are thought to vary in terms of their general cognitive styles, or typical modes of processing information (see Sternberg & Grigorenko, 1997 for a review).

Typically researchers tend to emphasize one approach over the other. Yet, the strategies approach and the parameters approach need not be mutually exclusive. One could argue that individuals select different strategies in order to compensate for parameter differences. For example, the aged have been argued to rely more heavily on plausible reasoning strategies because their exact memory retrievals are more effortful than for younger adults (e.g., Reder, Wible, & Martin, 1986).

More recently researchers have argued that the simple form of the strategies approach was incorrect. Specifically, they have argued that almost everyone uses multiple strategies, and the different groups of people shared many, if not most, strategies. This pattern was first found by Reder (1982, 1987) in a question answering domain, and subsequently has been found in every domain in which it has been studied (see Siegler, 1996 for a review). This observed ubiquity of multiple strategy use then led the strategies proponents to propose a general perspective on individual differences: groups vary in their distribution of use of strategies (i.e., when each strategy is used). For example, although older and younger children do simple addition problems using both counting and retrieval, the way they vary is that older children use retrieval more often, especially on more difficult problems.

A new approach to individual differences is the strategy adaptivity approach (Reder & Schunn, in press). This approach builds on the multiple-strategies approach, but assumes that people vary in how adaptive they are in their strategy selections. That is, although two individuals may have the same set of strategies, they may differ in their abilities to select the best strategy for a given situation. For example, some people may be less capable of flexibly changing their strategy use as the situation changes. This approach does not assume that some people suffer from lack of metacognitive knowledge of the particular strategies; rather, the approach assumes that people vary in their general ability to detect situational change and/or select strategies appropriate to the new situation. This chapter will explore this approach to individual differences.

The importance of this new approach is twofold. First, it provides a new view of individual differences that can be explored in many domains. Perhaps differences across individuals or groups in various domains that have been ascribed to parameter or strategy differences are in fact better characterized as differences in strategy adaptivity. Second, this approach can provide new information regarding the processes of strategy selection and strategy adaptivity. For example, correlations of individual differences with various ability tests can provide information regarding what types of underlying processes are important in strategy selection and strategy adaptivity.

This chapter has two goals, and the structure of the chapter reflects these goals. First, the chapter seeks to explicate the concept of strategy adaptivity and its components. Second, the chapter seeks to demonstrate the existence of individual differences in strategy adaptivity, as well as explore the stability and possible causes of such individual differences. These issues will be explored in several different task domains, with a section devoted to each task domain.

II. Strategy Adaptivity in Arithmetic

A. Retrieve versus Compute

A domain in which strategy selection has been studied quite frequently is arithmetic, which includes addition (Geary & Brown, 1991; Lebiere & Anderson, in press; LeFevre, Sadesky, & Bisanz, 1996; Siegler & Jenkins, 1988), subtraction (Siegler, 1987), multiplication (Lemaire & Siegler, 1995; Reder & Ritter, 1992; Schunn, Reder, Nhouyvanisvong, Richards, & Stroffolino, 1997; Siegler, 1988; Siegler & Lemaire, 1997), and artificial alpha-arithmetic (Logan, 1988; Zbrodoff, 1995). In this domain, the primary strategy decision is whether to retrieve the answer or compute the answer (e.g., count out the addition on one's fingers for addition or multiplying out the numbers in multiplication). Although other decisions do occur (e.g., which strategy to use if computation is selected), this particular strategy choice accounts for the largest percentage of the variance in accuracy and reaction times of responses (Siegler & Shipley, 1995). Computation is slow but accurate, and retrieval is fast but error-prone. Thus, in this domain, the primary issue of strategy adaptivity is defined in terms of how people select to retrieve when they know the answer well and use computation when they do not know the answer well.

Why is this a strategy choice? Under one commonly held view, people are assumed to always retrieve first and switch to computation once retrieval has failed (e.g., Lebiere & Anderson, in press; LeFevre, Greenham, & Waheed, 1993; Siegler & Shrager, 1984). Under this view, there is no strategic selection between retrieval and computation before a strategy is tried. Although this view is quite intuitive, several facts make it untenable. For example, people typically switch to computation faster than they would wait for retrieval to fail. Consider tip-of-the-tongue phenomena in which people will persevere in trying to retrieve for many seconds. Yet, for other problems, people either switch to computation or say "don't know" right away (e.g., what is 24 × 38?).

Another view is that people do retrieval and computation in parallel, and whichever finishes first is selected (Logan, 1988). Such a model can produce a very good account of the shape of learning curves. It also makes the selection of retrieve versus compute a non-issue. However, it also has fundamental problems. First, people don't always try computation right away (consider tip-of-the-tongue phenomena). Second, making the answer more available by priming it does not lead to a higher rate of selecting to retrieve (Reder, 1979), whereas priming parts of the problem statement (which does not make the answer more available) does affect the rate of selecting to retrieve (Reder, 1987).

The alternative view, which currently has the most empirical support, is that people make a decision to retrieve or compute *before* trying either (Reder, 1982, 1987, 1988; Reder & Ritter, 1992; Schunn et al., 1997; Siegler & Shipley, 1995). The key issue, then, is how people are able to make this strategy selection decision so rapidly.

B. Methods for Studying Strategy Selection

In studying the selection of retrieve versus compute strategies, the most common methodologies are to either look for overt strategy use (e.g., using pencil and paper to calculate, or fingers to count) with the assumption that no overt strategy use implies retrieval, or to ask subjects after the fact which strategy they used. The problem with these approaches is that they leave open the possibility that for trials in which the subject apparently used computation they might have tried to retrieve first, and on trials in which the subject apparently used retrieval they might have initially started computation.

One alternative approach is the game show or feeling-of-knowing paradigm developed by Reder (1987; Reder & Ritter, 1992). In this paradigm, subjects are shown a question or problem, and have to quickly decide whether the answer is known. For example in Reder and Ritter (1992) and Schunn et al. (1997), subjects had to decide in less than 850 ms whether they will be able to retrieve the answer to an arithmetic problem, or whether they will have to calculate the answer. This decision was reported with a key press. If they choose to retrieve, then the answer had to be produced in 2.4 s. If they choose to calculate, then they are given 30 s to calculate the answer. A payoff structure was provided to reward subjects for accurately predicting strategy use: 50 points if retrieve is chosen and the correct answer is produced within the 2.4 s deadline; 5 points if calculate is chosen and a correct answer is produced within 30 s; 1 point if a correct answer is produced too late; and 0 points for incorrect answers. The problems are selected such that the answers cannot be calculated within the 2.4 s (e.g., 26 × 37). Although none of the problems were initially familiar to the subjects (to control individual differences in preexisting knowledge), the problems were repeated many times such that the subjects learned the answer to many of the problems. In this paradigm, subjects quickly become quite adept at making their strategy selections within the 850 ms decision deadline, making retrieve–calculate strategy selections well before they are able to see whether retrieval will succeed.

C. An Example Study

Schunn, Reder, Nhouyvanisvong, Richards, and Stroffolino (1997) contains a simple study (Experiment 2) that employs the game show paradigm and has many useful properties for the purposes of this chapter. This study consisted of two experimental sessions on consecutive days. On the first day, subjects were given 16 problems that appeared over and over again with varying frequencies (either 27, 18, or 12 times over the course of the session). The problems were either multiplication or sharp (a novel operator similar in difficulty to multiplication), and the operands were all in the 12–38 range). On the second day, subjects were given 16 completely new training problems, also involving multiplication and sharp and operands in the 12–38 range, and also varying in presentation frequency (20, 10, or 5 times over the course of the session). On every trial during both sessions, subjects used the game show paradigm: first rapidly deciding whether to retrieve or calculate by selecting between two keys, and then providing the answer within the appropriate deadline.

On only the second day, there were a few operator swap problems intermingled with the training problems. These operator swap problems were copies of the training problems except that the operator was swapped (e.g., if 34 × 17 was a training problem, then 34 # 17 would be an operator swap problem). Some of the swap problems were swaps of training problems from Day 1, and some were swaps of training problems from Day 2. The purpose of these swap problems was to investigate whether subjects were using the familiarity of the problem (rather than the availability of the answer) to make their retrieve–calculate strategy selections. If the subjects choose to retrieve just as often for the operator-swapped problems (to which they did not know the answer) as for the original training problems, then they could not be using the retrievability of the answer to make their strategy decision.

The next sections reanalyzes the data from this study set to illustrate both the various components of strategy adaptivity at the global level and individual differences in adaptivity.

D. Overall Strategy Selections Are Adaptive

At the global level, subjects' strategy selections are quite adaptive. Adaptivity is defined as selecting to retrieve when the answer could be retrieved quickly (i.e., the answer time was less than 2.4 s and the answer was correct) and selecting to calculate when the answer could not be retrieved quickly (i.e., the answer time was greater than 2.4 s or the answer was incorrect). The majority of the trials in which they do know the answer, they choose retrieval (i.e., a high hit rate), and the majority of the trials in which they do not know the answer, they choose to compute (i.e., a low false alarm rate). Table 1 presents the mean hit rate, false alarm rate, and d' (Swets,

TABLE I

The Mean (and SE) Hit Rate, False Alarm Rate, and d' for the Two Days

	Day 1	Day 2
Hit Rate	73 (.05)	.64 (.05)
False Alarm Rate	.33 (.05)	.24 (.05)
d'	1.18 (.14)	1.21 (.12)

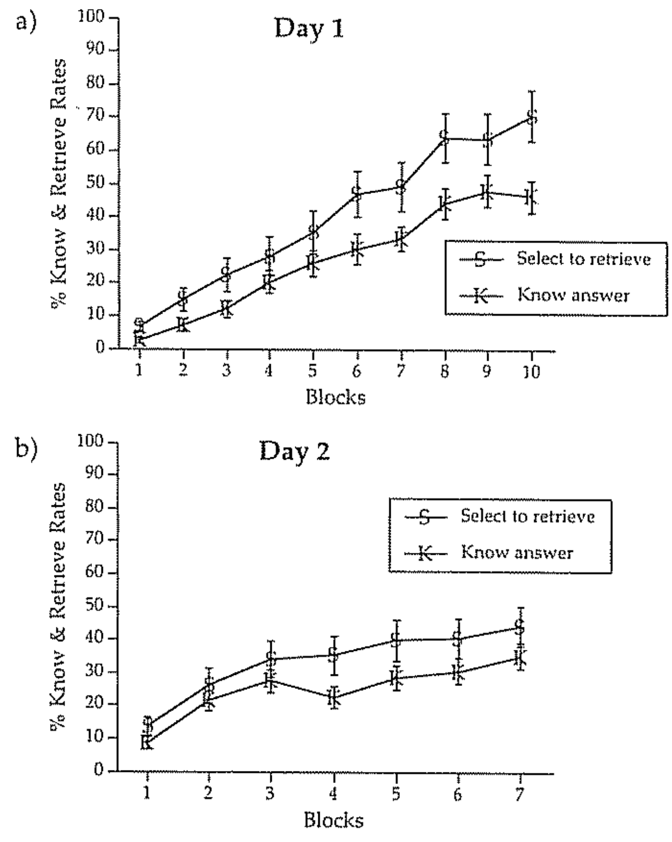

Fig. 1 The mean select to retrieve and know answer rates within each block (a) for Day 1. and (b) for Day 2

1986a, 1986b) for each of the two days in the experiment. These levels of d' are considered quite good when compared to those typically obtained from feeling-of-knowing judgments in other types of paradigms (e.g., Nelson & Narens, 1990; Schwartz & Metcalfe, 1992).[1] These levels of accuracy are especially impressive given that the decisions are being made in less than 850 ms (in fact, they are typically made in approximately 500 ms).

To illustrate their adaptiveness, we can examine how they select to retrieve in different blocks (of 30 trials) on each day. Fig. 1a presents the

[1] In this study, subjects made their retrieve–compute strategy selections and produced their answers using the keyboard. In previous studies (e.g., Reder & Ritter, 1992) in which subjects made the strategy selection using the keyboard and simply spoke their answers into a voice key, their d''s were even higher (e.g., above 2.0).

mean rates of selecting the retrieval strategy for each block of Day 1. Figure 1a also presents the mean proportion of problems for which the participants knew the answer (i.e., produced the correct answer within 2.4 s, regardless of which strategy decision they made). Fig. 1b presents the same information for Day 2. On both days, in the beginning of the experiment, the subjects rarely knew the answers and rarely selected to retrieve. Across blocks on both days, the subjects knew the answers to an increasingly larger proportion of problems and selected to retrieve for an increasingly larger proportion of problems. Although there is a slight bias to select to retrieve more often than when the answer is known (probably reflecting the large point value for on-time retrievals), there was a very high correlation between the mean strategy selection proportions for each block and the mean know rates for each block ($r^2 = .98$, $p < .0001$ and $r^2 = .93$, $p < .001$ for Day 1 and Day 2, respectively).

Not only was there a close correspondence at the aggregate level, but there was also a pretty close correspondence at the individual subject level. Subjects were not all equally adept at memorizing and quickly retrieving the answers. In general, the subjects were quite tuned to their own retrieval success levels. Using the same block sizes (30 trials), the subject strategy selection proprotions for each block were strongly correlated with the subject know rates for each block ($r^2 = .70$, N = 250, $p < .0001$ and $r^2 = .79$, N = 175, $p < .0001$ for Day 1 and Day 2, respectively). These subject-level correlations assume no separate biases for each individual, making the strong correlations even more impressive. Fig. 2 presents a few individual subject curves (the first four subjects on Day 1). Three of these four subjects were quite tuned in their strategy selections. A later section will explore the issue of whether subjects vary in how adaptive, or how tuned, they are.

E. Decomposing Adaptivity

Before turning to the issue of individual differences, the notion of strategy adaptivity must be unpacked further. Past research has identified two separate components of strategy selection, and both components are present in this data set. The first component has to do with learning the base rates of success of a strategy. If an individual were blinded during the strategy selection phase of each trial (i.e., if they had to make the decision without first seeing the problem), by wagering whether they were likely to know the answer they could exhibit strategy adaptivity by simply responding calculate on all early trials (since they did not know the answer to any problems in the beginning), responding retrieve on all trials at the end of a session (since they know the answer to most problems by then), and

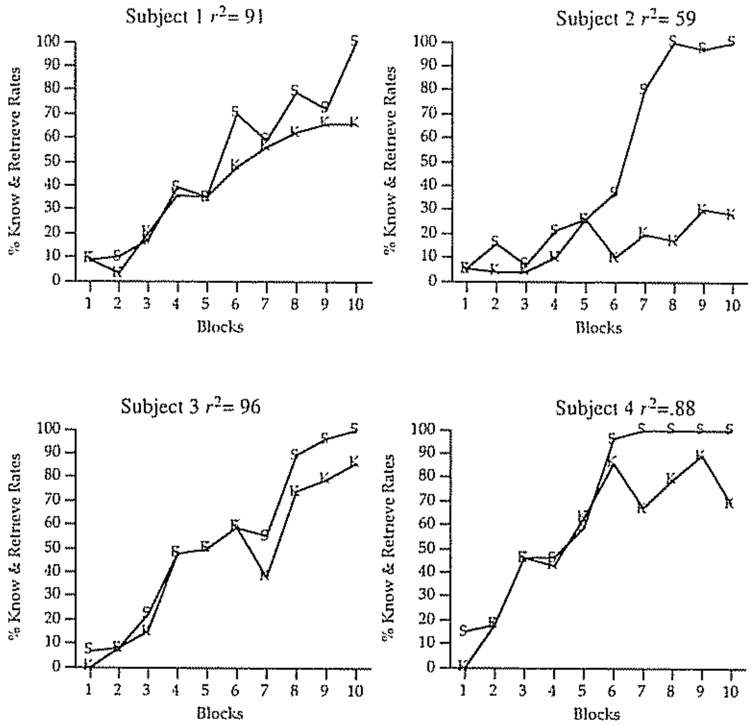

Fig 2. The mean select to retrieve and know answer rates within each block for the first 4 subjects on Day 1

responding with a mix of retrieve and calculate in the middle trials. The second component has to do with recognizing which particular problems are very familiar. Because problems were presented at different frequencies (high, medium, and low), not all problems were equally familiar at a given point during the experiment. Subjects could use the familiarity of a particular problem to decide to select to retrieve for the very familiar problems. Although the two components have been given different names by different researchers, this chapter will adopt the terminology developed by Reder (1987; 1988). The first kind of strategy adaptivity is called extrinsic because the adaptivity uses information external to the particular problem, and the second kind is called intrinsic because the adaptivity uses information internal to the particular problem. Variants of these two components can be found in the strategy choice models of Siegler and Shipley (1995) and Lovett and Anderson (1996).

The influence of intrinsic adaptivity can be found by examining the operator swap data—problems that look familiar but the answers are un-

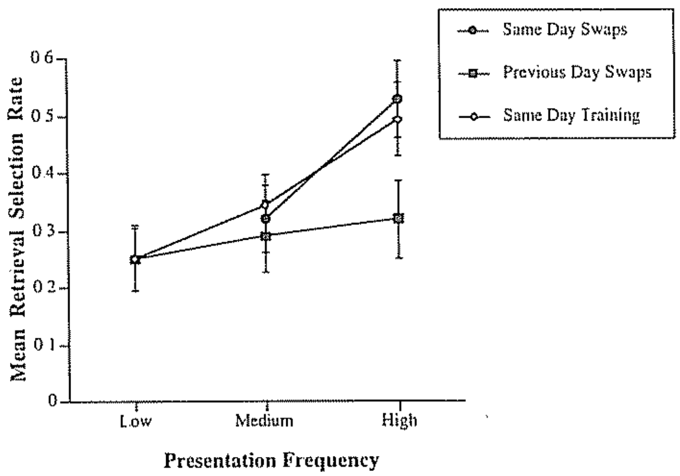

Fig. 3 The mean rate of selecting retrieval (and SE) at each frequency level for same and previous day operator-swap test problems and for comparable-time training trials (see text for details) for Day 2. Adapted from Schunn et al. (1997)

known. Fig. 3 presents the mean rates for selecting retrieval for the Day 2 training and operator-swap problems. To make the time periods comparable, training data are only taken from the interval during which swap problems were presented. As the figure shows, there was no effect of switching operators on the same day problems with respect to subjects' strategy selections—subjects were just as likely to select to retrieve for an operator-swap problem as they were for the original training problem. The effects of frequency of presentation on the previous day operator-switch test problems were in the same direction but weaker. This is to be expected, given that those problems would be much less familiar after a one day delay. These data illustrate how subjects use the familiarity of the problem statement per se to make their retrieve–calculate strategy decisions.

The presence of extrinsic adaptivity must be examined jointly with the presence of intrinsic adaptivity because the two covary in the stimuli. This combined analysis was done two ways for the data on both days; statistically and visually (the analyses for each day are presented separately). First, statistically, both frequency of presentation (how many times that problem had been seen before) and trial number (serial position of the trial on that day) were entered into a multiple regression predicting strategy choice on individual trials. For Day 1, both factors were independent predictors (partial r's of .14, .11 for trial and frequency respectively, N = 7500, p's < .0001). Second, visually, the 300 trials were divided into five blocks of 60

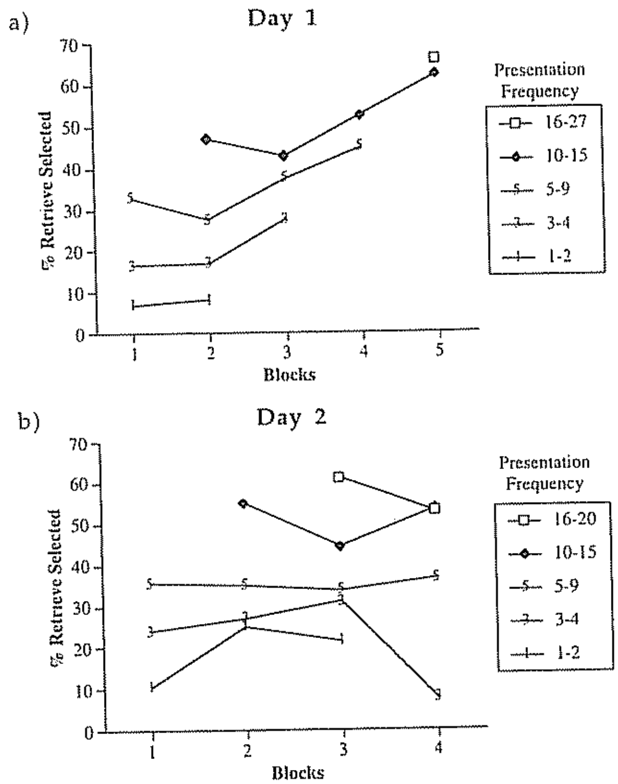

Fig. 4 The mean retrieval rate as a function of presentation frequency and block (a) for Day 1, and (b) for Day 2.

trials, and the presentation frequency range of 1–27 was divided into 1–2, 3–4, 5–9, 10–15, and 16–27 subranges. Fig. 4a presents the mean rate of selecting to retrieve within each block and frequency subrange for Day 1. There are clearly independent effects of both factors on strategy choice: the higher frequency lines are above the lower frequency lines illustrating the intrinsic effect, and the lines have a general positive slope illustrating the extrinsic effect.[2]

The data from Day 2 demonstrate that the increase in bias to select retrieval depends upon the increasing base-rates of knowing the answer. For Day 2, the overall frequency of presentation was lower than for Day

[2] To calculate the mean for a subrange, the mean of each presentation frequency mean was used—this removed any differential weighting bias across blocks. Correspondingly, for each subrange, only blocks for which all presentation frequencies in the subrange occurred could be included.

1 and there were many swap problems in the later trials. Thus, there was a much smaller increase in the base rate of knowing the answer (see Figs. 1 and 2). Correspondingly, there was a much smaller predictiveness of trial on strategy selections, whereas the predictiveness of problem frequency was just as large (partial r's of .04, .19 for trial and frequency, N = 4900, $p < .05$, $p < .0001$, respectively). Fig. 4b presents the subrange by block (of 50 trials) decomposition for the Day 2 strategy selection data. There is still an effect of presentation frequency but very little effect of trial. From block 1 to block 2 there is a small increase, but there are equal decreases from blocks 2 to 3 and 4. These changes in bias reflect the introduction of swap problems in blocks 3 and 4—the subjects adjusted their biases to compensate for spurious feeling-of-knowing. Note that the data plotted do not include the swap problems, but merely reflect their influence on the tendency to try retrieval.

F. Individual Differences in Adaptivity

Although subjects overall were quite adaptive in their strategy selections, did they vary systematically in how adaptive they are? To investigate individual differences in adaptivity, analyses were conducted on the subject d' scores (as defined earlier)—d' being a measure of how accurately the subjects chose to retrieve when they actually did know the answer and chose to compute when they did not actually know the answer. The individual difference analyses begin at this aggregate level, collapsing over the intrinsic and extrinsic components of adaptivity. Fig. 5 presents frequency histograms of subject d's on Day 1 and Day 2 (N = 25). There is a bimodality in the distribution of d's for both days. The rough dividing line is a d' of 1.00, with a cluster of subjects with d's below 1.00 and a cluster of subjects with d's above 1.00. Note the large range in d's. There are some subjects with adaptivity levels near chance, and there are other subjects with very high sensitivities (above 2.0).

How stable are these individual differences across days? Table 2 presents the correlations of d' and β across the two days. In fact, both d' and β are quite stable across the two days, although d' is more so than β. Returning to the d' threshold of 1.00 in Fig. 5, 84% of the subjects are on the same side of the threshold on both days. These stabilities indicate that the individual differences in d' cannot be attributed to random noise or insufficient Ns to properly estimate d'.

Is adaptivity in this task simply synonymous with skill in arithmetic? To assess this issue, the subject d's were regressed against the mean subject know rates (the percentage of trials for which the correct answer was given within 2.4 s), a measure of how quickly subjects were able to learn the

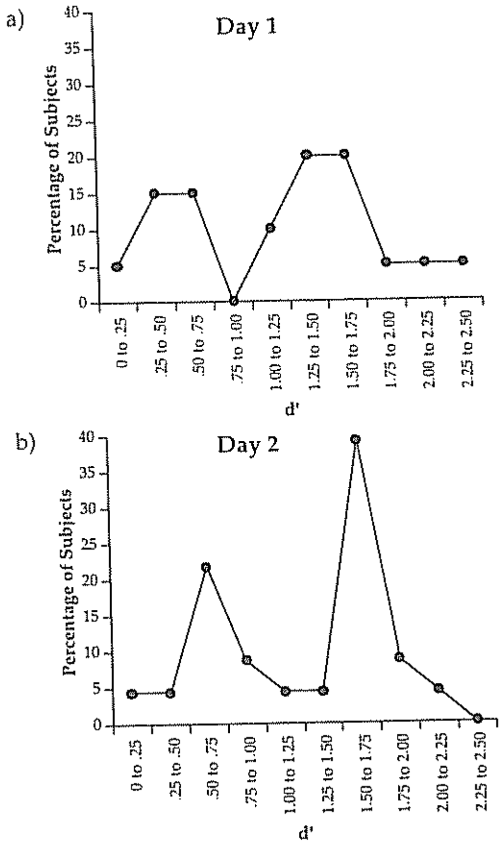

Fig. 5 Frequency histogram of subject d's (a) for Day 1, and (b) for Day 2

TABLE II

CORRELATIONS BETWEEN d' AND β FOR BOTH DAYS

	d' Day 1	d' Day 2	β Day 1	β Day 2
d' Day 1	1 00	76	.44	38
d' Day 2	.76	1 00	31	06
β Day 1	44	31	1 00	54
β Day 2	38	06	.54	1 00

arithmetic facts. In fact, d' was not correlated with mean know rates on either day ($r = -.05$ and $r = .23$ for Day 1 and Day 2, respectively, p's $> .3$), suggesting that adaptivity differences were independent of arithemetic or memorization skill differences.

Another worry is that the individual differences are simply due to floor or ceiling effects. That is, it would be hard to demonstrate strategy adaptivity if only one strategy was ever used. For example in the case of Subject 2 in Fig. 2, there appeared to be less adaptivity due to a large bias to select retrieve. To examine this issue, subject d's was regressed against the mean subject retrieval strategy selection rate (the percentage of trials for which the retrieve strategy was selected). Because both floor and ceiling effects were possible, a quadratic relationship was also assessed. In fact, neither the linear nor quadratic relationships were significant on either days (β's $< .01$, p's $> .5$), indicating that the individual differences could not be attributed to floor and ceiling effects—there were some subjects who used both strategies frequently but simply could not select adaptively when to use a particular strategy.

Another worry is that the more sensitive subjects were simply the ones who took more time to make their retrieve–compute decisions. However, there was no correlation between subject d' and mean strategy decision time on Day 1 ($r = .02$, $p < .5$), and the correlation was (weakly) in the wrong direction on Day 2 ($r = -.27$, $p < .25$).

G. Intrinsic versus Extrinsic Individual Differences

The preceding analyses did not establish whether the individual differences in strategy adaptivity were differences in intrinsic adaptivity, extrinsic adaptivity, or both. The following sections attempt to tease apart individual differences on the two components.

There are several ways in which individual differences on the two separate dimensions of adaptivity can be measured. The key to these measures is that they take into account the predictiveness of the extrinsic and intrinsic factors *for each individual*. For example, suppose that for individual A, there was a large difference in memorability of high-frequency problems and low-frequency problems, whereas for individual B, there was only a very small difference. Further suppose that individual A chose to retrieve much more often for the high-frequency problems than the low-frequency problems, whereas individual B showed only a small preference in choosing to retrieve for high-frequency than low-frequency problems. In this case, individual B should not be rated as being less intrinsically adaptive because the particular intrinsic predictor is simply less predictive for that individual. Instead, one might divide the degree of strategy change by the degree of retrievability change. This transformation also standardizes the range of scores across different intrinsic predictors. A similar transformation can be applied to extrinsic predictors—divide the degree of strategy change in response to an extrinsic variable by the degree of retrievability change in response to that variable. For both intrinsic and extrinsic measures, a score of one is exactly tuned adaptivity, a score of less than one is undersensitivity (zero being no adaptivity), and a score of more than one is oversensitivity.

Another important factor in measuring intrinsic and extrinsic adaptivity is to hold the potential influences of one of them constant while measuring the other. For example, if one examined intrinsic adaptivity in the first two blocks of either day, the influence of extrinsic adaptivity would be very small, as Fig. 4 revealed. This is exactly what was done. For blocks 1 and 2 combined (to increase power), frequencies 1–2 were compared with frequencies 5–9. The difference in retrieval strategy selection rates for the two frequency groups was divided by the difference in know answer rates for the two frequency groups. These ratios ranged from 0 to 1.6 and had a mean of .71 on Day 1, and ranged from 0 to 1.2 with a mean of .51 on Day 2. Over half (57% on Day 1 and 75% on Day 2) of the subjects showed undersensitivity (adaptivity $< .8$), with the remaining subjects mostly closely tuned ($.8 \leq$ adaptivity ≤ 1.2) and a very few oversensitive (adaptivity $>$ 1.2). The intrinsic adaptivity correlated $r = .51$, $p < .02$ across the two days, suggesting that there was a stable individual difference in intrinsic adaptivity.

Nonsignificant linear and quadratic regressions with mean subject retrieval selection rates ($\beta_1 = -.005$, $\beta_2 = .00003$, p's $> .5$ for Day 1, and $\beta_1 = -.003$, $\beta_2 = -.00006$, p's $> .5$ for Day 2) suggest that the individual differences were not due to floor and ceiling effects in strategy use. However, there were significant correlations with mean subject strategy decision time ($r = .51$, $p < .02$, and $r = .44$, $p < .05$ on each day, respectively). Because these correlations are approximately as large as the stability of the individual differences in intrinsic adaptivity, it is likely that those differences are conceived of as differences in how quickly decisions are made.

It was difficult to do a similar type of analysis of individual differences in extrinsic adaptivity because, as Figs. 5 and 6 revealed, the influences of intrinsic adaptivity were ubiquitous on both days. The alternative is to do multiple regressions for each subject to assess the independent effects of extrinsic and intrinsic factors in predicting strategy use within each subject. Thus, with this method, one can simultaneously assess individual differences in intrinsic and extrinsic adaptivity. As with the preceding analysis, one must divide by the degree of predictiveness on memorability. Specifically, the β-coefficients for problem frequency (intrinsic) and trial number (extrinsic) in the multiple regressions on strategy selection for a particular subject are divided by the corresponding β-coefficients for the multiple regressions on know answer rates (i.e., answering correctly within 2.4 s) for that subject. Subjects for whom the coefficients in the know answer regression are very small (absolute value $< .0005$) are excluded because adaptivity to an unpredictive feature is not meaningful. As one would expect, the impact of this was to exclude subjects for only the extrinsic sensitivity measure, the overall less reliable predictor of know rates in this task.

Table 3 presents the means and standard deviations for each measure on each day. There was considerably more variance on the extrinsic measure. Although there appears to be less extrinsic adaptivity on Day 2, this is an artifact of different subjects being excluded on different days. There was almost no difference in extrinsic adaptivity values across days for those subjects included on both days, .58 vs. .44, $F(1,10) < 1$, $MSE = 1.5$.

The measures of intrinsic and extrinsic adaptivity proved stable across the two days, $r = .73, p < .0001$, and $r = .67, p < .02$, respectively. Moreover, the two measures were independent of one another, $r = -.07$, $r = .12$ for each day, p's $> .5$, suggesting that the measures were tapping independent constructs. Regressions with mean subject know-answer rates established that intrinsic adaptivity may have been related to arithmetic skill or memorization rates, $r = .47, p < .02$ and $r = .08, p > .5$ for each day, respectively. However, extrinsic adaptivity appeared unrelated to know answer rates, $r = -.02, p > .9$, $r = -.23, p > .3$. Similarly, quadratic regressions with mean retrieval strategy selection rates revealed that intrinsic adaptivity may be related to floor and ceiling effects, $\beta_1 = .015, p < .2, \beta_2 = -.00011, p > .3$ for Day 1, and $\beta_1 = .026, p < .05, \beta_2 = -.00032, p < .05$ for Day 2. By contrast, extrinsic adaptivity appeared to be also unrelated to mean retrieval strategy selection rates, $\beta_1 = .056, \beta_2 = -.001$, p's $> .5$ for Day 1, and $\beta_1 = .016, \beta_2 = -.000067$, p's $> .5$ for Day 2. Finally, intrinsic

TABLE III

THE NUMBER OF SUBJECTS NOT EXCLUDED, THE MEAN STRATEGY ADAPTIVITY, STANDARD DEVIATION, AND PERCENTAGE OF SUBJECTS SHOWING UNDERSENSITIVITY.

	N	Mean	St Dev	% Undersensitive (< .8)
Intrinsic				
Day 1	25	.41	.40	84.0
Day 2	25	.46	.40	88.0
Extrinsic				
Day 1	15	.85	2.1	60.0
Day 2	19	.23	1.0	78.9

adaptivity was partially related to mean strategy selection decision time, $r = .20$, $p > .3$, and $r = .37$, $p < .1$ for each day respectively, whereas extrinsic adaptivity appeared unrelated to decision time, $r = .17$, $p > .5$, $r = -.22$, $p > .3$ for each day, respectively.

In sum, individual differences in intrinsic sensitivities may simply be attributable to differential arithmetic proficiency, large strategy selection biases, and–or speed–accuracy tradeoffs. By contrast, individual differences in extrinsic adaptivity seem NOT to be attributable to these sources, and may instead reflect a more fundamental individual difference factor.

III. Investigating Individual Differences in a More Controlled Environment

One of the problems of analyses of adaptivity individual differences in tasks like arithmetic is the subjects themselves determine the strategy success rates. That is, not only do the subjects vary in terms of which strategies they chose, but they also vary dramatically in terms of how successful are the strategies themselves, and, even worse, the relative successes under different circumstances. The preceding analyses attempted to equate for such differences mathematically by dividing by the predictiveness of the features, but some doubts remain. An alternative approach is to use a task in which the success of the strategies can be controlled entirely by the experimenter, and thus be equated precisely across all individuals. This section reports on a new experiment on individual differences in strategy adaptivity using just such a task, focusing on individual differences in extrinsic adaptivity. Another advantage of the new task is that strategy use is measured implicitly via behavior, and does not require that subjects make a separate, explicit self-report of their strategy use. Finally, the new task is a slower paced task, not requiring rapid strategy selections—it is possible that some of the individual differences in adaptivity for the arithmetic task were due to ineptitude in making rapid key presses.

A. The Building Sticks Task

The Building Sticks Task (BST) (Lovett & Anderson, 1996) is a problem-solving task similar to the classic water jars task (Luchins, 1942). For a given BST problem, subjects must add and subtract an unlimited supply of three different-sized building sticks to create a stick of the desired length (see Fig. 6, "initial state"). BST problems can be solved by one of two strategies. The *undershoot* strategy involves starting with a building stick that is shorter than the desired stick and then lengthening that stick by

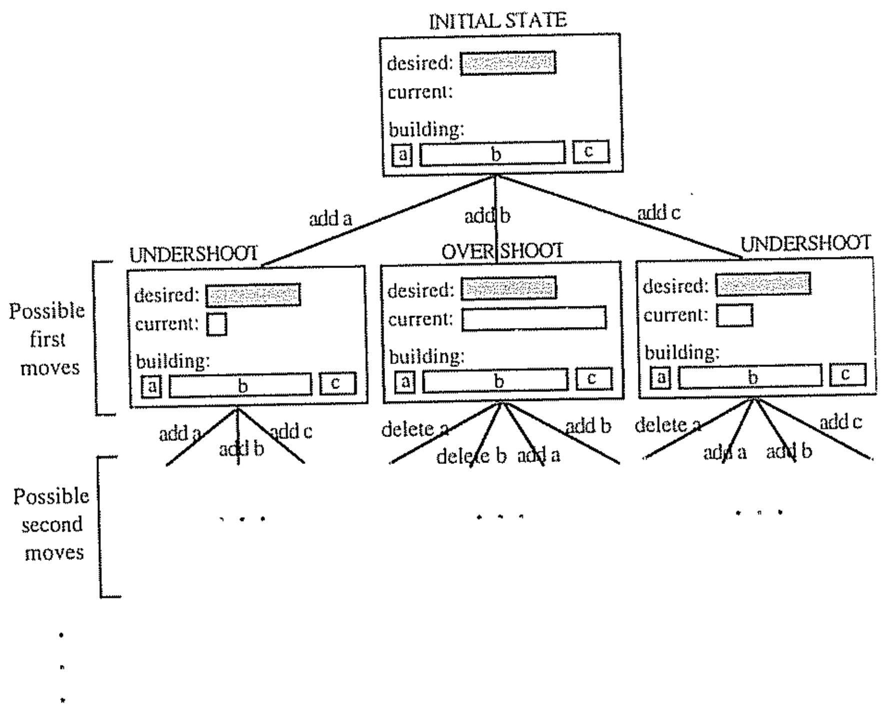

Fig. 6. Initial and subsequent states in a BST problem. The subjects' task is to build a current stick (initially length 0) that matches the desired stick in length by adding and subtracting various combinations for the building stick lengths. From Lovett and Anderson (1996).

additional building stick lengths until the desired stick's length is reached. In contrast, the *overshoot* strategy involves starting with the building stick that is longer than the desired stick and then shortening that stick by the other building stick lengths. As Fig. 6 shows, subjects implicitly choose between these two strategies in their first step.

For example, suppose the desired stick was of length 13 units, and the three sticks *a, b,* and *c,* were of lengths 3, 17, and 5, respectively. To obtain the desired stick length of 13 units, the subjects might start with stick *b* of 17 units and remove segments (the overshoot strategy), or the subjects might start with stock *c* of 5 units and add more segments (the undershoot strategy). In this example, a solution can only be obtained by the undershoot strategy ($c + c + a = 5 + 5 + 3 = 13$). The overshoot strategy will not work because subtracting lengths *a* and *c* from *b* will never lead exactly to a stick of 13 units length. Of course, in other problems the overshoot strategy may be the correct one to use.

It is very important to note that the subjects were never given the exact numerical lengths of the sticks—the example above was simply used for expository purposes. In the experiment, subjects had to visually estimate the length of each stick. This prevented subjects from being able to solve the task algebraically, and forced them to try a strategy (i.e., make their choice externally) to determine whether or not it would work.

Within this task, it is easy to manipulate the base-rates of success of the undershoot and overshoot strategies. Each problem is designed to be solvable by either undershoot or overshoot (but not both) and then the proportion of problems with each solution type is varied across blocks of time. In this way, it is possible to directly control the success rates of each strategy, and thereby more clearly measure individual differences in adaptivity to change success rates (i.e., extrinsic adaptivity).

B. Methods

1. *Subjects*

Fifty-six CMU undergraduates participated for course credit, and were randomly assigned to one of two complementary conditions.[3]

2. *Procedure*

Each subject was given 70 BST problems to solve. To measure extrinsic adaptivity, the base-rates of success of the undershoot and overshoot strategies were manipulated over time. For 10 trials, both strategies were equally successful (i.e., 5 overshoot and 5 undershoot). For the next 30 trials, one strategy was successful on 80% of trials. For the final 30 trials, the other strategy was successful on 80% of trials. Varied across two conditions was the strategy that was more successful first. That is, in one condition (50–80–20) the undershoot strategy was successful on 80% then 20% of trials, whereas in the other condition (50–20–80) undershoot was successful on 20% then 80% of trials.

The experiment was conducted on a Macintosh IIci, which ran the BST interface, provided initial instructions to subjects, and collected data, including timing data for each mouse click.

Another issue of interest is how individual differences in strategy adaptivity relate to explicit awareness of base-rate change. To address this issue, subjects were asked a series of questions at the end of experiment. Specifically, the overshoot and undershoot strategies were described to the subjects, and the subjects were asked, "Did there seem to be any pattern to which strategy worked?" and "Did you notice any changes in the effec-

[3] Because of a software bug, there were 42 subjects in the first condition and 14 in the other

tiveness of one strategy or the other as the experiment progressed?" The subjects' responses were coded for awareness of the base-rate manipulation: a 1 was given if the subject reported awareness of a change and correctly described the direction of the change, and 0 was given otherwise. Only a few subjects had intermediate awareness; apparent awareness of change, but inaccurate in reported direction. These subjects behaved like the unaware subjects and thus were collapsed into that group. Sixty percent of the data was recoded by a second coder, and the inter-rater reliability was 97%.

C. RESULTS

1. *Overview*

The results are broken into three sections. First, there is a presentation of overall extrinsic adaptivity to the manipulations along with a mathematical characterization of aggregate group adaptivity. Second, there is an analysis of individual differences in extrinsic adaptivity. Finally, there is analysis of the relationship between explicit awareness and extrinsic adaptivity.

2. *Overall Adaptivity*

The subjects' first choices on each trial were used to categorize strategy use into overshoot and undershoot strategies. The 70 trials were divided into blocks of 10 trials. Fig. 7 illustrates the mean undershoot use within each block within each condition, as well as the manipulated success rates of undershoot within each block for each condition. A Block × Condition ANOVA conducted on the mean undershoot use revealed that there was no overall effect of Block, but the interaction of Block by Condition was

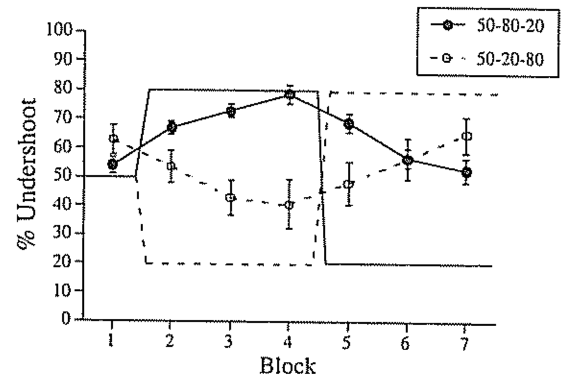

Fig. 7. Mean (and SE) use of the undershoot strategy within each block in each condition of the BST experiment, along with the manipulated base-rates.

quite strong, F(6,324) = 14 4, MSE = .03, $p < .0001$. The 50–80–20 condition showed the expected increase in use of the undershoot strategies from blocks 2 through 4, and then the expected decrease in use of the undershoot in blocks 5 through 7 (analysis of linear trends F(1,82) = 17.4, MSE = .02, $p < .0001$, and F(1,82) = 32.7, MSE = .02, $p < .0001$, respectively). The 50–20–80 showed the reverse pattern (analysis of linear trends F(1,26) = 4.0, MSE = .03, $p < .06$, and F(1,26) = 12.3, MSE = .02, $p < .005$, respectively). Thus, on average, the subjects adapted to the base-rate changes.

Because the two conditions behaved so similarly, the two conditions were collapsed (to gain greater power for the individual difference analyses) by reversing the values for the 50–20–80 condition (i.e., subtracting the proportions from 1). But before we turn to the issue of individual differences, how might be the group average performance be described? One good approximation is the average success rate of the strategy up until that point, called the global average. For example, by the end of block two, the global average is (50 + 80)/2 = 65. Fig. 8 shows the fit of the global average to the mean subject data. This model with zero free parameters accounts for 93% of the across block variance in the aggregate data.

3. *Individual Differences*

Most importantly, despite the very good fit of the global average model to the aggregate data, there are large individual differences in strategy adaptivity. Some subjects adapt not at all, some adapted at the level of the global average model, and others adapt much more quickly than the global average model. Using a Monte Carlo simulation (N = 1000), we can estab-

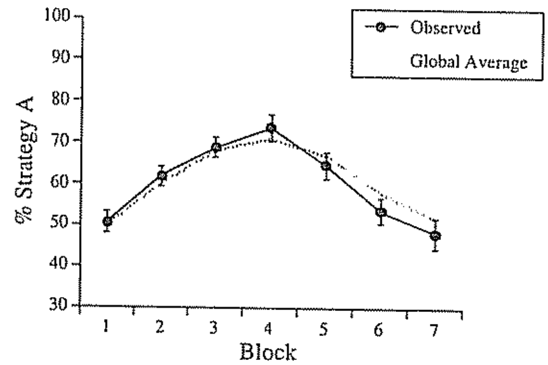

Fig 8. Mean strategy use (and SE) within each block of the BST experiment, and predictions of the global average model. Copyright 1998 by The American Psychological Association. Reprinted with permission.

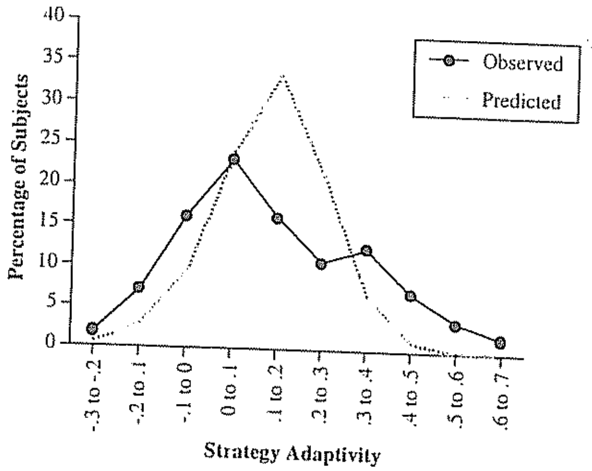

Fig 9 Observed and expected distribution of strategy adaptivity in the BST experiment.

lish the expected variation of the subjects assuming they all had true probabilities of selecting undershoot in a block that corresponded to the mean undershoot use across subjects in each block (i.e., followed the global average model closely). Because there were only 10 trials per block, and each trial produces a binary outcome (retrieve or compute), one would expect a certain amount of variability simply due to sampling noise. Fig. 9 plots the observed distribution of subjects' adaptivity (difference between the mean strategy use in blocks 2–4 and the mean strategy use in blocks 5–7)[4], and the expected distribution from the Monte Carlo simulation. The observed distribution is considerably flatter than the expected distribution, indicating more individual differences than could be attributed to sampling noise. Although we would expect 13% of the subjects to have zero or less adaptivity by chance, 25% of the subjects actually fell into this group. At the other extreme, we would expect only 8% of subjects to have greater than .3 adaptivity by chance, in fact 25% of the subjects displayed this high level of adaptivity. Using these cutoff points, the observed distribution of adaptivity did differ statistically from the expected distribution, χ^2 ($df = 2$; $N = 56$) $= 31.6$, $p < .001$. These comparisons establish that there are individual differences in extrinsic adaptivity that are not attributable to chance variation. In the subsequent analyses, the subjects whose adaptivity (as defined above) was at or below zero will be called the Nonadaptive subjects.

[4] Means for Blocks 2–4 and Blocks 5–7 were used to gain greater power per subject for the individual difference analyses. However, very similar results are obtained when strategy sensitivity is defined as the difference in strategy use between Block 4 and Block 7.

It is possible that these individual differences in adaptivity are due to floor or ceiling effects in strategy use (i.e., always using undershoot or never using undershoot). However, restricting the subjects to only those who start the first block in the range of 40% to 60% undershoot use, the observed adaptivity remains statistically different from the expected distribution, χ^2 ($df = 2$; N = 37) = 24.6, $p < .001$. Moreover, examinations of the distribution of mean percent undershoot across the 70 trials revealed that only 1 of the 16 Nonadaptive subjects was outside of the 20–80% range, and only 4 were outside of the 25–75% range. Therefore most of the Nonadaptive subjects did use both overshoot and undershoot strategies regularly, just not adaptively over time.

Another possible explanation for why some subjects may not have been adaptive is that they were simply doing the task too quickly—for lack of motivation or other reasons—to notice the change in base-rates. To investigate this possibility, a Block × Adaptivity (Adaptive–Nonadaptive) ANOVA was conducted on the mean time to make the first move (i.e., clock on the desired first stick and click on its target location). There was a speed-up over blocks, $F(6,324) = 7.0$, MSE = 1.5, $p < .0001$, as the mean time changed from 5.7 s in block 1 to 4.5 s in block 3. However, there was no effect of Adaptivity, or an interaction with Block, F's < 1. If anything, the adaptive subjects were slightly faster, 4.6 s versus 5.1 s. Therefore, the Nonadaptive subjects were not simply choosing too quickly—5.1 s is a long time to make two mouse clicks.

4. *Relationship of Explicit Awareness to Extrinsic Adaptivity*

How did explicit awareness of the base-rate changes (as measured by the debriefing questions) relate to the extrinsic adaptivity? As described earlier, the subjects were divided into those who explicitly noticed the direction of the shift (Aware, N = 18) and those who did not (Unaware, N = 37). An Awareness × Block ANOVA conducted on the mean strategy use in each block revealed a significant interaction, $F(6,318) = 3.6$, MSE = .02, $p < .005$. Although both groups showed the same transitions over time, the Aware subjects showed the transition to a greater degree (see Fig. 10).

To examine whether the Aware subjects were more likely to show any behavioral transition or whether they simply showed a larger transition, further analyses were conducted. First, a contingency analysis between Awareness (Aware–Unaware) and Adaptivity (Adaptive–Nonadaptive) revealed that the Aware subjects were no more likely to fall in the Adaptive group than the Unaware subjects, 78% versus 70%, respectively), χ^2 ($df = 1$; N = 55) < 1. Second, focusing on only the Adaptive subjects, an Awareness (Aware–Unaware) ANOVA conducted on the adaptivity measure (differ-

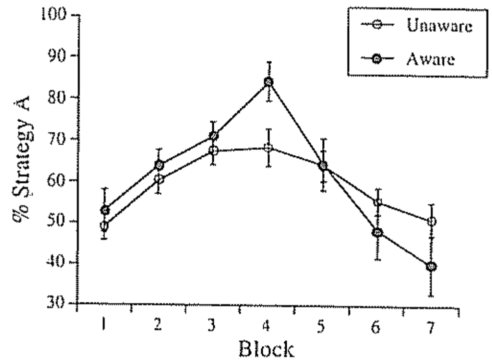

Fig 10 Mean (and SE) strategy use for Aware and Unaware subjects within each block of the BST experiment.

ence in mean strategy use between blocks 2–4 and blocks 5–7) found that the Aware subjects were more adaptive than the Unaware subjects (strategy adaptivity of .32 and .16, respectively), $F(1,38) = 10.9$, $MSE = .02$, $p < .005$. Thus, awareness was associated with larger transitions and not a greater likelihood of making any transition.

D. Discussion

As in the arithmetic domain, there were meaningful individual differences in extrinsic adaptivity in the BST domain that could not be attributed to various artifacts. Thus, even when strategy success rates are carefully controlled across individuals, individuals differ significantly in their abilities to change strategy use in response to shifting base-rates of success. The additional contribution of the BST results is that they establish that the individual differences are related to explicit awareness of base-rate change—awareness appears to led to faster or larger changes.

The relationship between awareness and strategy adaptivity deserves a cautionary note. Because the relationship found thus far is only correlational, the causality is (quite plausibly) ambiguous. All three logical interpretations are reasonable possibilities; (1) explicit awareness leads to greater shifts in strategy use; (2) greater shifts in strategy use are more likely to lead to explicit awareness after the fact; and (3) some third factor (e.g., working memory capacity or inductive reasoning skill) leads to both greater shifts in strategy use and a greater likelihood of noticing the base-rate change.

IV. Individual Differences in a Complex, Dynamic Task

A. Static versus Dynamic Tasks

Although there is now clear evidence that individuals can vary in their extrinsic adaptivity in a few particular tasks, there has been no discussion of whether there are real-world situations in which such individual differences will substantially effect performance. A particular hypothesis is proposed; individual differences in extrinsic adaptivity will be especially important predictors of overall performance in complex, dynamic tasks. Dynamic tasks (contrasted with static tasks) are those tasks in which the environment changes independently of the actions of the individual. For example, in Tower of Hanoi the task never changes unless the individual takes an action. By contrast, in driving a car, the weather, traffic, and roadway conditions can all change independently of the actions of the driver. Of course, the dynamic–static distinction is more of a continuum than a true dichotomy. Thus, the prediction is that the more dynamic the task, the more important differences in extrinsic adaptivity will be for predicting performance. Because most real-world tasks are dynamic, the range of tasks for which such differences will matter is predicted to be quite large.

Schunn and Reder (1997; Reder & Schunn, in press) examined individual differences in extrinsic adaptivity in a complex, dynamic task, and those results will be overviewed here. The task they used will be described briefly, followed by a presentation of the results of interest for the current chapter (see Reder & Schunn for the full details).

B. The Kanfer-Ackerman Air Traffic Controller Task

The Kanfer-Ackerman Air Traffic Controller Task (KA-ATC) (Ackerman & Kanfer, 1994; Kanfer & Ackerman, 1989) was designed to simulate dynamic aspects of real air traffic control (e.g., weather changes, consumption of fuel in real time). The object of the KA-ATC task is to accumulate as many points as possible. Points are earned by landing planes (+50 points) and are subtracted by rule violations (−10 points) or plane crashes (−100 points). In the task, subjects must monitor a variety of elements that are displayed on the screen (see Fig. 11): (a) 12 hold pattern positions that are divided into three altitude levels, (b) four runways—two short and two long, one of each running north–south and one of each running east–west, (c) a queue of planes waiting to enter into the hold positions, (d) messages indicating error feedback and changes in runway conditions, wind speed, and direction, and (e) the current score and penalty points. The time pressure in the task is quite severe; plane fuel levels decrease in real time;

```
     FLIN      TYPE      FUEL      POSN          Score       : 200
                                    3N           Landing Pts: 200     Penalty Pts  0
     233       PROP         5       3S           Runways     : DRY
     794       747          5       3E           Wind        : 25 - 35 knots from WEST
     892       DC10      *  4       3W
     888       DC10      *  3       2N             Flts in Queue
     420       747       *  3       2S                  <F1> to accept
                                    2E
     127       PROP      *  4       2W
     402       DC10      *  4       1N
 =>  822       PROP         5       1S
     878       747          5       1E
     429       DC10      *  4       1W

N    +++++++++++++++++++++++++    S    #1
N    ++++++++++++++++++           S    #2
W    ------PROP---------------    E    #3
W    ------------------           E    #4
```

Fig 11 The main screen of the Kanfer-Ackerman Air Traffic Controller Task

weather changes occur approximately every 25 seconds; and planes enter into the queue every 7 seconds.

There are six rules governing this task, four of which are particularly pertinent here. First, planes must land in the wind (e.g., use a north–south runway rather than an east–west runway if the wind is coming from the north or south). Second, planes can only land from hold level 1 (the lowest level). Third, the current weather conditions and wind speed determine the runway length required by different plane types (747s always require long runways; DC10s can use short runways if runways are not icy and the wind speed is less than 40 knots; 727s can use short runways only when the runways are dry or wind speed is 0–20 knots, and PROPs can always use short runways). Fourth, only one plane at a time can occupy a runway. Each violation of any of these rules produces a 10-point penalty.

The task consists of a sequence of 10-minute trials, with the total number of trials varying from study to study. Each trial begins with planes already in various hold positions and other planes in the queue. At the end of each trial, the subject is given a short, self-timed break. The next trial begins with a new screen display and the cursor at the top of the screen.

The task involves three primary subtasks: (1) accepting a plane from the queue into a hold position; (2) moving planes within the hold positions; and (3) landing planes. The analyses presented here focus on the third task, specifically on the strategy decision of landing a selected plane on the short or long runways when both are open. There are several advantages to selecting the long runway. First, the long runways are always legal for all plane types. Thus, the probability of making an error is lower. Second, the current wind speed and runway conditions need not be consulted before landing the plane (although wind direction must always be consulted). Third, the rules for landing a plane on the runways need not be retrieved. Fourth, the long runways are closer to the hold positions than the short runways and so require fewer keystrokes. The advantage of using the short runway (when it is legal) is that it keeps the long runway open for the planes that can only land on the long runway under the current wind speed and runway conditions. Because the planes require 15 s to land on a runway and only one plane can be landed on a runway at a time, subjects must maximize the use of both runways in order to maximize the total number of planes landed. In other words, the long runway is a scarce resource that should be used sparingly.

The particular operational measure of runway strategies was called OpShort, which is the proportion of times a subject opted to land a plane on the short runway of all the times a plane was landed and both runways were open. This ratio is computed separately for each plane type. Only ratios for DC10s and 727s are computed because 747s can never land on the short runway and PROPs can always land on the short runway.

C. A Base-Rate Manipulation in the KA-ATC Task

1. Methods

To measure extrinsic adaptivity, a base-rate manipulation was used in the KA-ATC task much like the base-rate manipulation in the BST domain. Specifically, OpShort success rates were manipulated by varying the proportion of 747s and PROPs in each 10-minute trial. Because 747s can only land on long runways and PROPs can always land on either runway, the proportion of 747s and PROPs drastically affect the importance of using the long runway effectively. When there are many 747s, then it is better to place the DC10s and 727s on the short runway whenever possible. By contrast, when there are few 747s, there is much less pressure to place the DC10s and 727s on the short runway. In this situation, it is better to place those planes on the long runway—landing on the short runway requires more keystrokes, knowing and accessing the rules for when the short runway is legal, and checking the current wind and weather conditions. Thus, when there are many 747s, OpShort rates should be high, and when there are few 747s, OpShort rates should be low.

Subjects were given nine 10-minute trials. These nine trials were divided into three blocks of three trials each. The proportion of 747s (and PROPs) were manipulated across blocks. The proportions of 747s over the three blocks were 25%–5%–50%.[5] Because the strategy adaptivity measure is defined by plane type, one plane type (DC10s) was set at a constant high level of 40% across all three blocks to insure sufficient Ns for each subject on at least one plane type. The frequency of PROPs was set to be 55% minus the frequency of 747s (i.e., 30%–50%–5%), thereby completing the manipulation of the scarcity of the long runways.

To contrast how well adaptivity differences predict overall performance in the task with the predictiveness of more traditional measures, and to understand what factors were related to adaptivity, the subjects were given an individual ability battery. The individual ability battery used was a subset of the CAM4 (Cognitive Abilities Measurement) (Kyllonen, 1993, 1994, 1995) battery. The CAM battery provides a broad range of tests that are plausibly related to adaptivity in strategy use, and has been used to predict learning and performance in a large number of training environments (e.g., Shebilske, Goettl, & Regian, in press; Shute, 1993). Eleven CAM tests were used, including multiple tests of associative learning, procedural learning, processing speed, working memory, and inductive reasoning.

2 *Results*

a. Overall Adaptivity To examine whether the subjects, overall, were able to adapt to the base-rate manipulation in the context of a complex task, an ANOVA was conducted on OpShort with Block as a within-subjects factor. The effect of Block was quite strong, $F(2,174) = 43.0$, $MSE = .04$, $p < .0001$, with subjects adapting in the predicted pattern of medium-low-high (see Fig. 12). Thus, people are able to show extrinsic adaptivity even in the context of a very complex, dynamic task.

b. Individual Differences in Adaptivity Although subjects on average adapted to the base-rate manipulation, the subjects did vary significantly in their adaptivity. Only 30% of the subjects followed the full pattern of adaptivity in OpShort use across the three blocks; medium, low, high (i.e., block 3 > block 1 and block 2 < block 1), and only 75% adapted at all to the largest transition from block 3 to block 2 (i.e., block 3 > block 2). Another measure of adaptivity is the extent of change from blocks 2 to 3

[5] A second control condition with a different order (25%–50%–5%) was used to insure that the results were not peculiar to one particular order, nor simply due to changes that would have occurred naturally as a function of practice with the task (i.e., independent of the manipulation). The control condition simply performed as expected and so will not be discussed here.

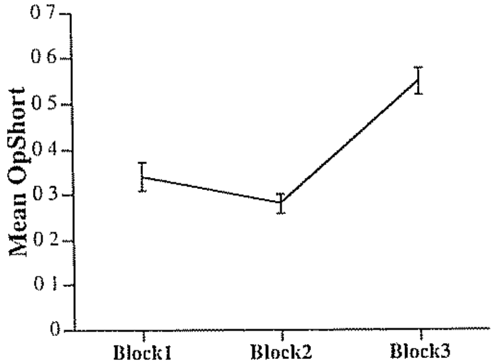

Fig 12 The mean proportion (and SE) OpShort within each block of three trials in the KA-ATC base-rate manipulation (25%–5%–50% 747s) Adapted from Schunn and Reder (1997)

(i e , the difference in mean OpShort use in those two blocks). Even in those who showed any adaptivity at all on the second transition, there was significant variability in OpShort use. A third of the adaptive participants adapted less than .2, and a third adapted .5 or more.

c. *Correlations with Performance and the CAM Battery* Do these differences in extent and rate of adaptivity relate to task performance? Because adapting strategy use may require cognitive resources, performance on other aspects of the task may suffer. Therefore, it is not necessarily true that differences in OpShort adaptivity will be related to differences in overall task performance. Moreover, the runway landing decision is only a very small part of the overall task. To examine this relationship, extent of adaptivity as measured by the rise in OpShort from blocks 2 to 3 was regressed against mean block score. Adaptivity proved to be quite strongly correlated to overall score, $r = .69$, $p < .0001$. For comparison, the largest correlation between the 11 CAM ability measures and score was only .56. Thus, the general trait of adaptivity, as measured on only one of the myriad of strategy decisions that must be made in this complex task, is a very important predictor of overall task performance.

What factors predict extent of adaptivity? To assess which measures were independent predictors of adaptivity, a stepwise regression was conducted including all 11 CAM test scores. Two measures entered into the stepwise regression (multiple $r^2 = .38$), in order of predictiveness: working memory ($r = .55$, $p < .005$), and the inductive reasoning ($r = .53$, $p < .01$). Fig. 13 illustrates the overall pattern of correlations among CAM, adaptivity, and overall task performance.

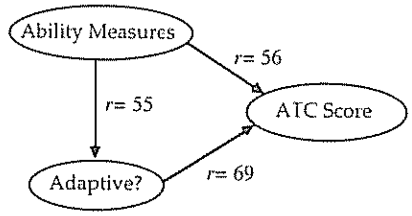

Fig 13 Best correlations in the KA-ATC base-rate manipulation between extent of adaptivity, ability measures, and overall task score

The relationship between the CAM measures and adaptivity are interesting because they provide information about what processes underlie strategy choice and strategy adaptivity. What possible mechanisms could explain these relationships? Inductive reasoning could play a role in adaptiveness in at least two ways. First, inductive reasoning skill may be related to being able to notice shifting patterns in the environment, as in the BST task. Second, inductive reasoning might be related to being able to quickly understand the relationship between a strategy and its effect, or diagnosing when a strategy is no longer appropriate. Relatedly, higher working memory capacity may involve an increased ability to keep information, specifically base-rate information, in mind while simultaneously performing the task. Under this account, ability to retain the recent set of outcomes would predict how quickly the pattern of change can be detected, and hence how fast one could adapt. Inductive reasoning ability, in contrast, would predict whether, given a pattern, the individual understood what strategy to adopt for best performance with the new pattern.

D. Micro-level Extrinsic Adaptivity

Another kind of measure of extrinsic adaptivity focuses at a more micro-level: adaptivity to the success of the strategy in the previous attempt. Even when base-rates of success remain roughly constant, the individual must use success and failure feedback to discover which strategies and what levels of strategy use are optimal in the current task. To investigate this kind of extrinsic adaptivity, and possible individual differences therein, Schunn and Reder also reanalyzed unpublished KA-ATC data collected by Phillip Ackerman.[6] In this data set, each 10-minute trial was roughly the same—what varied from trial to trial was the actions taken by the subjects and their knowledge and skills as they learned to perform the complex task. The data set also included individual ability scores from

[6] The data were taken from the Kanfer-Ackerman CD-ROM Database© (1994).

22 ability tests that were centered around six factors (Perceptual Speed, Movement Speed, Memory, Verbal, Reasoning, and Psychomotor factors).

1. *Results*

a. Overall Adaptivity To examine whether subjects were adaptive in their OpShort use at the micro-level, the OpShort data were analyzed as a function of whether the previous attempt to land that plane type on the short runway had been successful (i.e., had not resulted in an error). If subjects were adaptive, then they should reduce their tendency to use the short runway when that landing attempt resulted in an error previously and they should increase their tendency to use the short runway for that plane type when that action was previously successful. Only data from the first nine trials were used because error rates were very low after the ninth trial.

The analyses also distinguished situations in which the short runway was legal for the selected plane type and those in which it was illegal—the definition of OpShort required only that the two runways were open. Any observed adaptivity may reflect several types of strategy shifts. For example, it may lead subjects to learn the rules more completely following an unsuccessful attempt and then be less likely to use the short runway in illegal situations. Alternatively, it may reflect a simple shift in tendency to use the short runway that would have had equal impact in legal and illegal situations. Fig. 14 suggests that this second alternative is what occurred, and ANOVA results supported this view. There were main effects of legal-

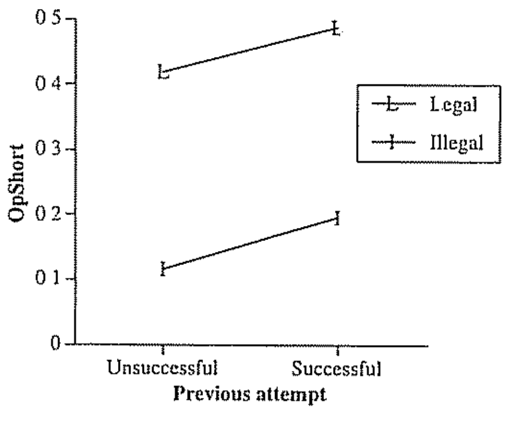

Fig. 14. Mean OpShort for legal and illegal short runway opportunities as a function of the success of the previous attempt to land that plane type. From Schunn and Reder (1997).

ity, F(1,42) = 76.2, MSE = .05, $p < .0001$, and success of the previous attempt, F(1,42) = 20.2, MSE = .01, $p < .0001$, but no hint of an interaction F(1,42) < 1. Thus, subjects do adapt at the micro-level, producing shifts in the bias to use a strategy as a result of success and failure feedback.

b. Individual Differences in Adaptivity Because evidence has been found for micro-level strategy adaptivity for OpShort, the issue of individual differences can be examined. In fact, there were significant individual differences in adaptivity, as defined by the difference in OpShort use following successes and OpShort use following failures (collapsing across legal and illegal situations). The modal adaptivity level, accounting for 23% of subjects, was −.05 to +.05 (i.e., no adaptivity). Moreover, this mode did not include the mean adaptivity level across all subjects (.13). Approximately 21% and 24% of the subjects showed adaptivity levels that were two and three or more times as high as the mean, respectively. Thus, a significant proportion of subjects showed no evidence of strategy adaptivity to success rates, whereas many subjects showed very strong strategy adaptivity to success rates.

c. Correlations with Performance and the CAM Battery How do these individual differences in strategy adaptivity relate to performance in the task? Because the primary goal (and primary determinant of score) in the KA-ATC task is landing planes, OpShort adaptivity was correlated against the mean number of planes landed. OpShort adaptivity correlated positively with planes landed, $r = .50$, $p < .02$. By contrast, of the 22 individual difference tests administered by Ackerman, the 4 best predictors of planes landed correlated in the .44 to .47 range. Using a stepwise regression on planes landed with the 22 individual difference tests and the Adaptivity measure as possible predictors, only two independent factors emerged (multiple $r^2 = .29$): OpShort Adaptivity $\beta = 11.4$, $p < .05$, and the Patterns test $\beta = 0.29$, $p < .05$. Thus, adaptivity appears to predict performance directly and not through indirect correlations with other determinants of performance. These results provide further evidence that the individual differences in adaptivity are a very important performance predictor and they demonstrate that the individual differences are systematic and cannot be attributed to noise.

Another issue of interest is whether adaptivity can be predicted from the psychometric ability tests. When the 22 psychometric scores were entered into a stepwise regression predicting OpShort adaptivity, only one factor entered: Letter Sets, $r = .45$, $p < .02$, a test that loads heavily on reasoning and general intelligence. The results are similar to those found in the KA-ATC base-rate manipulation study in that they also point to reasoning ability as an important correlate of strategy adaptivity. The overall patterns of correlations among adaptivity, ability scores, and task perfor-

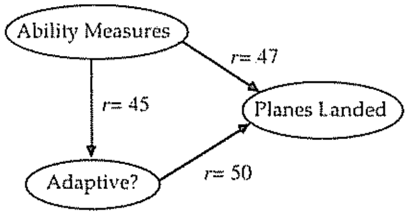

Fig. 15 Best correlations in the KA-ATC micro-level adaptivity study between extent of adaptivity, ability measures, and number of planes landed

mance are illustrated in Fig. 15. These correlations provide another piece of evidence that the individual differences in adaptivity were not simply due to chance variation.

E. Discussion

Overall, subjects were able to demonstrate extrinsic adaptivity, even in a complex dynamic task. However, across several KA-ATC studies, in both college (Ackerman data set) and noncollege populations (base-rate manipulation data set), using both macro- and micro-level measures, there were significant individual differences in extrinsic adaptivity, with many individuals showing no evidence of adaptivity.

By measuring adaptivity on one small strategy decision in a complex task, it was possible to predict overall task performance with greater accuracy than with all of the traditional individual difference measures that were tried. These results suggest that strategy adaptivity differences must have manifested themselves in many other strategic aspects of the overall task. The results also suggest that strategy adaptivity is a very important component of performance in complex, dynamic tasks.

These studies using the KA-ATC task also found that inductive reasoning and working memory capacity were correlated with individual differences in extrinsic adaptivity. These correlations provide new insights into the processes of strategy choice and strategy adaptivity. For example, they suggest that strategy adaptivity involves noticing patterns and keeping significant amounts of information in working memory—two features that are not part of current models of strategy choice (e.g., Logan, 1988; Lovett & Anderson, 1996; Reder, 1987; Schunn et al., 1997; Siegler & Shipley, 1995).

V. General Discussion

Previous accounts of individual differences in performance have tended to focus on either parametric differences in an underlying cognitive architecture (e.g., attentional capacity, working memory capacity, *g*, etc.) or strategy differences. This chapter has explored a possibility that unifies the previous two approaches: that there is an architectural parameter that controls strategy use, or more specifically, that individuals may have the same strategies but are differentially able to select the most appropriate strategy for the given situation. Reder and Schunn (in press; Schunn & Reder, 1997) explored this idea within the KA-ATC task. This chapter extends this work both in the sophistication of the analyses and the breadth of the tasks: examination of individual differences in three very different types of tasks (rapid arithmetic, simple problem solving, and complex problem solving) produced evidence of significant variability in strategy adaptivity in all three cases.

Although the individual differences were all investigated among normal, nonexpert, adult populations, these kinds of adaptivity differences might also be studied in all the other areas of individual differences (e.g., child development, aging, expertise, giftedness, schizophrenia, etc.). The prediction from the findings presented here is that populations that display working memory or inductive reasoning differences will also display adaptivity differences. In fact, many of these other areas of individual differences have already hypothesized group differences in working memory, and thus there are many places to test this prediction. Moreover, the differences in performance due to adaptivity differences is likely to be most pronounced in dynamic tasks.

Methodologically, this chapter described five different methods for establishing the meaningfulness of individual differences in strategy adaptivity. First, observed individual differences were compared to chance levels of variation due to sampling noise, as predicted by a Monte Carlo simulation. Second, the stability of individual differences was measured across consecutive days in the same task. Third, the individual differences were shown to be strongly correlated to overall task performance, even in situations in which base-rates of success are not directly manipulated. Fourth, the individual differences were compared to existing ability measures and found to be correlated to several of them. Finally, many correlational analyses were conducted to show that the individual differences were not due to simple artifacts (e.g., floor and ceiling effects in strategy use, overall skill in the task domain, and speed-accuracy tradeoffs).

Two kinds of strategy adaptivity were considered in the arithmetic domain: intrinsic and extrinsic. Although there appeared to be individual differences in both kinds of adaptivity, the intrinsic adaptivity differences appeared to be attributable to various artifacts—differential task proficiency, large strategy selection biases, and speed–accuracy tradeoffs. The extrinsic adaptivity differences did not have this problem, and were further established in the BST and KA-ATC tasks. Of course, individual differences in intrinsic adaptivity have not been ruled out altogether—it still might exist in other population comparisons or become more apparent in other tasks. Lovett and Schunn (1997) found that although people can learn not to pay attention to certain features (i.e., intrinsic adaptivity is malleable), they also found significant variation across individuals in intrinsic adaptivity under a given payoff schema. The distinction between intrinsic and extrinsic adaptivity is an important one, and future attempts to measure individual differences in adaptivity should continue to distinguish these two types of adaptivity.

There are many conceptions of adaptivity to found in the psychological literature. This chapter has focused on adaptivity in strategy use. It is an open question as to how strategy adaptiveness might relate to other kinds of adaptiveness. For example, it may be correlated with individual differences in the ability to adapt to instructions (Reder, 1987; Shebilske, Goettl, & Regian, in press), in the ability to select and change representations (Lovett & Schunn, 1997; Schunn & Klahr, 1996; Schunn & Lovett, 1996), or in the ability to adaptively control attention (Gopher, 1982, 1996; Gopher & Kahneman, 1973).

Another question is how strategy adaptivity relates to metacognition. Although the metacognitive view is related to the strategy adaptiveness view explored here, there are several important differences. The metacognitive view (e.g., Case, 1985; Flavell, 1979; Kuhn, 1988; Sternberg, 1985) postulates that some individuals use less sophisticated strategies because they have poor metacognitive knowledge of why and when different strategies are effective. Although this view seems quite plausible, empirical research has found that there is at best a weak relationship between metacognitive knowledge and the adaptiveness of strategy selections (e.g., Cavanaugh & Perlmutter, 1982; Schneider & Pressley, 1989).

Base-rate adaptivity and individual differences in base-rate adaptivity has also been studied in a much more explicit kind of task—text-based probability problems that explicitly present both base-rate and case-specific information and require subjects to predict the probability that one alternative will succeed. Studies using this paradigm have typically found base-rate neglect (e.g., Ginossar & Trope, 1987; Lyon & Slovic, 1976; Tversky & Kahneman, 1973). That is, people's predictions frequently do not adequately take into account the overall probabilities (or base-rates) of the various alternatives, and overweight the influence of case-specific information. However, various manipulations of the task wording can lead to greater use of base-rate information (cf. Koehler, 1996, or Lovett & Schunn, 1997 for a review). Most importantly for this chapter, Stanovich and West (1998) studied individual differences in base-rate neglect and other reasoning fallacies observed on text-based problems, and found that cognitive ability measures explain some of the variance of people's performance on these tasks. It may be interesting to explore how these individual differences in the ability to reason explicitly about base-rates may be related to individual differences in extrinsic strategy adaptivity.

A final remaining issue is the cross-task stability of individual differences in strategy adaptivity. More generally, the issue is whether the individual differences in strategy adaptivity are specific to particular tasks or types of tasks or whether they are general across many types of tasks. This chapter presented data showing stability within a task over time, similar kinds of individual differences in three very different tasks, and correlations with standard ability tests, all of which suggest that cross-task stability would be found if it were measured. However, no direct data were presented on stability across tasks. Although this issue is not yet resolved, it is important to note that the interestingness of individual differences in adaptivity does not hinge on cross-task stability: Even if the individual differences were specific to particular tasks, they nevertheless remain important features of task performance, and they provide insight into the mechanisms of strategy choice and strategy adaptivity.

REFERENCES

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology. General, 117*(3), 288–318.

Ackerman, P. L. (1989). Individual differences and skill acquisition. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences* (pp. 165–217). New York: Freeman and Company.

Ackerman, P. L., & Kanfer, R. (1994). *CD-ROM Database, Data Collection Program, and Playback Program.*

Carver, C. S., & Scheier, M. F. (1992). *Perspectives on personality* (2nd ed.). Needham Heights, MA: Simon & Schuster.

Case, R. (1985). *Intellectual development: A systematic reinterpretation.* New York: Academic Press.

Cavanaugh, J. C., & Permutter, M. (1982). Metamemory: A critical examination. *Child Development, 53,* 11–28.

Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145–182.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121–152.

Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory & Cognition, 25*(5), 691–700.

Ericsson, K. A., & Polson, P. G. (1988). A cognitive analysis of exceptional memory for restaurant orders. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise.* Hillsdale, NJ: Erlbaum.

Flavell, J H (1979) Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34,* 906–911

Fry, A F, & Hale, S (1996) Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade *Psychological Science, 7*(4), 237–241

Geary, D C, & Brown, S C (1991) Cognitive addition: Strategy choice and speed-of-processing differences in gifted, normal, and methematically disabled children *Developmental Psychology, 27*(3), 398–406

Ginossar, Z, & Trope, Y (1987) Problem solving in judgment under uncertainty *Journal of Personality and Social Psychology, 52,* 464–473.

Gopher, D (1982) A selective attention test as a predictor of success in flight training *Human Factors, 24*(2), 173–183

Gopher, D (1996) Attention control: Explorations of the work of an executive controller. *Cognitive Brain Research, 5,* 23–38

Gopher, D, & Kahneman, D (1973) Individual differences in attention and the prediction of flight criteria *Perceptual and Motor Skills, 33,* 1335–1342

Greenfield, P M, & Lave, J (1982) Cognitive aspects of informal education In D A Wagner & H W Stevenson (Eds), *Cultural perspectives on child development* San Francisco, CA: Freeman.

Hunt, E, Joslyn, S, & Sanquist, T (1996) *Evaluating individual differences in response to emergency situations* ONR Tech Report.

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

Kail, R (1988). Development functions for speeds of cognitive processes *Journal of Experimental Child Psychology, 45,* 339–364.

Kanfer, R, & Ackerman, P L (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology—Monograph, 74,* 657–690

Koehler, J J (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges *Behavioral and Brain Sciences, 19,* 1–53

Kuhn, D (1988) Cognitive development In M H Bornstein & M E Lamb (Eds), *Developmental psychology: An advanced textbook* (2nd ed). Hillsdale, NJ: Erlbaum

Kyllonen, P C (1993). Aptitude testing inspired by information processing: A test of the four-sources model *The Journal of General Psychology 120*(3), 375–405

Kyllonen, P C (1994). Cognitive abilities testing: An agenda for the 1990s In M G Rumsey, C B Walker, & J H Harris (Eds), *Personnel Selection and Classification* Hillsdale, NJ: Erlbaum.

Kyllonen, P C (1995) CAM: A theoretical framework for cognitive abilities measurement. In D Detterman (Ed), *Current topics in human intelligence Volume IV, Theories of intelligence* Norwood, NJ: Ablex

Larkin, J H, McDermott, J, Simon, D, & Simon, H (1980) Expert and novice performance in solving physics problems *Science, 208,* 140–156

LeFevre, J, Greenham, S, & Waheed, N (1993) The development of procedural and conceptual knowledge in computational estimation *Cognition & Instruction, 11*(2), 95–132.

LeFevre, J, Sadesky, G S, & Bisanz, J (1996) Selection of procedures in mental addition: Reassessing the problem-size effect in adults *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 216–230.

Lemaire, P, & Siegler, R S (1995) Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General, 124*(1), 83–97

Logan, G D (1988) Toward an instance theory of automatization *Psychological Review, 95*(4), 492–527

Lovett, M C , & Anderson, J R (1996) History of success and current context in problem solving: Combined influences on operator selection *Cognitive Psychology, 31* 168–217

Lovett, M C , & Schunn, C D (1997) Task representations, strategy variability and base-rate neglect Paper submitted for publication

Lovett, M C , Reder, L M , & Lebiere, C (1996) Modeling individual differences in a digit working memory task Paper presented at the *International Conference on Memory* Padova, Italy

Lovett, M C., Reder, L M , & Lebiere, C (in press) Modeling working memory in a unified architecture: The ACT-R perspective To appear in A Miyake & P Shah (Eds), *Models of Working Memory* Cambridge University Press

Luchins, A S (1942) Mechanization in problem solving *Psychological Monographs, 54*

Lyon, D , & Slovic, P (1976) Dominance of accuracy information and neglect of base rates in probability estimation *Acta Psychologica, 40,* 287–298

Miller, P H , & Seier, W L (1994) Strategy utilization deficiencies in children: When, where, and why In H W Reese (Ed), *Advances in child development and behavior* (Vol 25) New York: Academic Press

Miyake, A , Carpenter, P A , & Just, M A (1995) Reduced resources and specific impairments in normal and aphasic sentence comprehension *Cognitive Neuropsychology, 12,* 651–679

Nelson, T O , & Narens, L (1990) Metamemory: A theoretical framework and some new findings In G H Bower (Ed), *The psychology of learning and motivation* (pp 1–45) New York: Academic Press

Reder, L M (1989) The role of elaborations in memory for prose *Cognitive Psychology, 11,* 221–234

Reder, L M (1982) Plausibility judgments versus fact retrieval: Alternative strategies for sentence verification *Psychological Review, 89,* 250–280

Reder, L M (1987) Strategy selection in question answering *Cognitive Psychology, 19,* 90–137

Reder, L M , & Ritter, F E (1992) What determines initial feeling of knowing? Familiarity with question terms, not with the answer *Journal of Experimental Psychology Learning, Memory, & Cognition, 18,* 435–451

Reder, L M., & Schunn, C D (in press) Bringing together the psychometric and strategy worlds: Predicting adaptivity in a dynamic task To appear in D Gopher & A Koriat (Eds), *Cognitive regulation of performance Interaction of theory and application Attention and Performance XVII*

Reder, L M , Wible, C , & Martin, J (1986) Differential memory changes with age: Exact retrieval versus plausible inference *Journal of Experimental Psychology Learning, Memory, & Cognition, 12*(1), 72–81

Saccuzzo, D P , Johnson, N E , & Guertin, T L (1994) Information processing in gifted versus nongifted African American, Latino, Filipino, and White children: Speeded versus nonspeeded paradigms *Intelligence, 19*(2), 219–243

Salthouse, T A (1994) Aging associations: Influence of speed on adult age differences in associative learning *Journal of Experimental Psychology Learning, Memory, & Cognition, 20*(6), 1486–1503

Schneider, W , & Pressley, M (1989) *Memory development between 2 and 20* New York: Springer-Verlag

Schooler, C , Neumann, E , Caplan, L J , & Roberts, B R (1997) A time course analysis of Stroop interference and facilitation: Comparing normal individuals and individuals with schizophrenia *Journal of Experimental Psychology General, 126*(1), 19–36

Schunn, C D , & Klahr, D (1996) The problem of problem spaces: When and how to go beyond a 2-space model of scientific discovery Symposium talk presented at *The 18th Annual Conference of the Cognitive Science Society* (pp 25–26)

Schunn, C D., & Lovett, M L (1996) Representation and individual differences in base-rate neglect Poster presented at *37th Annual Meeting of the Psychonomic Society*, Chicago. IL (November, 1996)

Schunn. C D , & Reder, L M (1997) Another Source of Individual Differences: Strategy Adaptivity to Changing Rates of Success Paper submitted for publication

Schunn. C D., Reder, L M , Nhouyvanisvong. A . Richards. D R . & Stroffolino, P J (1997) To calculate or not calculate: A source activation confusion (SAC) model of problem-familiarity's role in strategy selection *Journal of Experimental Psychology Learning, Memory, & Cognition, 23*(1), 3–29

Schwartz, B L , & Metcalfe, J (1992) Cue familiarity but not target retrievability enhances feeling-of-knowing judgments *Joural of Experimental Psychology-Learning, Memory, & Cognition, 18*,1074–1083

Shapira, N , & Kushnir, T (1985) Aging and information seeking: Patterns in sampling of sucrose solutions *International Journal of Aging & Human Development, 20*(2) 103–112

Shebilske, W . Geottl, B , & Regian, J W (in press) Individual and group protocols for training complex skills in laboratory and applied settings To appear in D Gopher & A Koriat (Eds), *Cognitive regulation of performance Interaction of theory and application Attention and Performance XVII*

Shute, V J (1993) A macroadaptive approach to tutoring *Journal of Artifical Intelligence in Education, 3*(1), 61–93

Siegler, R S (1983) Five generalizations about cognitive development *American Psychologist, 38*, 263–277

Siegler, R S (1987) Strategy choices in subtraction In J Sloboda and D Rogers (Eds), *Cognitive processes in mathematics*, (pp 81–106) Oxford: Oxford Books

Siegler. R S (1988) Strategy choice procedures and the development of multiplication skill *Journal of Experimental Psychology General, 117*, 258–275

Siegler, R S (1996) *Emerging minds The process of change in children's thinking* New York: Oxford University Press

Siegler, R S., & Jenkins, E (1988) *How children discover new strategies* Hillsdale, NJ: Erlbaum

Siegler, R S , & Lemaire, P (1997) The choice/no-choice method: Applications to assessing the adaptiveness of older and younger adults strategy choices in multidigit multiplication *Journal of Experimental Psychology General, 126*, 71–92

Siegler, R S , & Shipley, C (1995) Variation, selection, and cognitive change In G Halford & T Simon (Eds), *Developing cognitive competence New approaches to process modeling*, (pp 31–76) New York: Academic Press

Siegler, R S , & Shrager, J (1984) Strategy choices in addition and subtraction: How do children know what to do In C Sophian (Ed), *Origins of cognitive skills*, (pp 229–293) Hillsdale, NJ: Erlbaum

Snow, R E., Kyllonen, P C , & Marshalek, B (1984) The topography of ability and learning correlations In R Sternberg (Ed), *Handbook of human intelligence* New York: Cambridge University Press

Spearman, C (1904) The proof and measurement of association between two things *American Journal of Psychology, 15*, 72–101

Stanovich, K E., & West, R F (1998) Who uses base rates and P(D/~H)? An analysis of individual differences *Memory & Cognition, 26*, 161–179

Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning.* Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* New York: Cambridge University Press.

Sternberg, R. J., & Grigorenko, E. L. (1997). Are cognitive styles still in style? *American Psychologist, 52*(7), 700–712.

Swets, J. A. (1986a). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin, 99,* 181–198.

Swets, J. A. (1986b). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin, 99,* 100–117.

Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky, (Eds.), *Judgment under uncertainty: Heuristics and biases.* Cambridge University Press.

Wagner, D. A. (1978). Memories of Morocco: The influence of age, schooling and environment on memory. *Cognitive Psychology, 10,* 1–28.

Zbrodoff, N. J. (1995). Why is 9 plus 7 harder than 2 plus 3? Strength and interference as explanations of the problem-size effect. *Memory & Cognition, 23*(6), 689–700.