

## Another Source of Individual Differences: Strategy Adaptivity to Changing Rates of Success

Christian D. Schunn  
George Mason University

Lynne M. Reder  
Carnegie Mellon University

This article explores an alternative approach to the study of individual differences of cognitive function—that people may have the same strategies but differential ability to adaptively select among them in response to success and failure feedback from the environment. Three studies involving the complex and dynamic Kanfer–Ackerman Air Traffic Control Task (P. L. Ackerman & R. Kanfer, 1994) demonstrate (a) that individuals do differ systematically along this strategy adaptivity dimension, (b) that those differences have important consequences for overall task performance, and (c) that the differences are primarily associated with reasoning ability and working-memory capacity.

From an information-processing perspective, there are two popular approaches to the study of individual differences of cognitive function construed broadly (i.e., including studies of child development, giftedness, aging, intelligence, brain damage, schizophrenia, expertise, and adult individual differences). The first is the parameter approach. This approach assumes that individuals vary in performance because of differences in some fundamental aptitude or parameter of the cognitive architecture. Although many different parameters have been proposed, 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), that older adults think slower than younger adults (Salthouse, 1994), that gifted children think faster than average children (Saccuzzo, Johnson, & Guertin, 1994), and that people with schizophrenia think slower than people without schizophrenia (Schooler, Neumann, Caplan, & Roberts, 1997). In terms of working-memory capacity, researchers have argued that people with aphasia have reduced working memory capacity (Miyake, Carpenter, & Just, 1995), that children have higher working memory capacity (Case, 1985; Fry & Hale, 1996),

and that general intelligence differences depend heavily on working-memory capacity differences (Just & Carpenter, 1992). Although this parameter approach most naturally describes the analyses of information-processing psychologists (e.g., Daily, Lovett, & Reder, in press; Hunt, Joslyn, & Sanquist, 1996; Just & Carpenter, 1992; Lovett, Reder, & Lebiere, 1996, 1999; Sternberg, 1977), it can also be viewed as the information-processing take on the psychometric approach to individual differences (e.g., Ackerman, 1989; Snow, Kyllonen, & Marshalek, 1984; Spearman, 1904).

The second general approach is the strategies approach. A strategy is a method used for solving a problem. This approach assumes that there are many different strategies that can be used to solve any given problem and that individuals vary in the strategies they use. For example, researchers have argued that older children use different strategies than younger children in a wide variety of domains (Siegler, 1983), that older adults use different strategies than younger adults (Reder, Wible, & Martin, 1986; Shapira & Kushnir, 1985), that good students self-explain and poor students do not (Chi, Bassok, Lewis, Reimann, & Glaser, 1989), that experts use different strategies than novices (Chi, Feltovich, & Glaser, 1981; Ericsson & Polson, 1988; Larkin, McDermott, Simon, & Simon, 1980), that optimists use different strategies than pessimists (Carver & Scheier, 1992), and that 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 their typical modes of processing information (see Sternberg & Grigorenko, 1997, for a review).

Although many researchers tend to emphasize one approach over the other—in fact, they frequently design their experiments to minimize the influence of the other factor—the strategies approach and the parameters approach need not be mutually exclusive. It has been argued, moreover, that groups select different strategies to compensate for parameter differences. For example, it has been argued that older adults rely more heavily on plausible-reasoning strategies because their exact memory retrievals are more effortful than those of younger adults (e.g., Reder, Wible, & Martin, 1986).

---

Christian D. Schunn, Department of Psychology, George Mason University; Lynne M. Reder, Department of Psychology, Carnegie Mellon University.

This work was supported by Office of Naval Research Grant N00014-95-1-0223 through Helen Gigley and Susan Chipman. A preliminary form of Study 3 was reported in Reder and Schunn (1999).

Data collection was made possible with the generous assistance of the TRAIN laboratory of Air Force Office of Scientific Research (AFOSR) and in particular Barry Goettl and J. Wesley Regian, Jr., of Armstrong Laboratory. We would like to thank Frank Lee for making available the code for his variant of the KA–ATC task and Marsha Lovett for comments on earlier versions of this article.

Correspondence concerning this article should be addressed either to Christian D. Schunn, Department of Psychology, MSN 3F5, George Mason University, Fairfax, Virginia 22030-4444, or to Lynne M. Reder, Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213. Electronic mail may be sent to either schunn@gmu.edu or reder@cmu.edu.

Later research has shown that the simple form of the strategies approach was incorrect: Almost everyone uses multiple strategies, and the different groups of people shared many if not most strategies. This result was first described by Reder (1982, 1987, 1988) and has been found to be true of almost every domain in which it has been studied (see Siegler, 1996, for a review). The newer form of the strategies approach is that groups vary in their distribution of use of strategies (i.e., when each strategy is used). For example, both older and younger children do simple addition problems using retrieval and counting. The way they vary is that older children use retrieval more often, especially on difficult problems.

A third, less popular approach to individual differences is the strategy adaptivity approach. This strategy builds on the multiple-strategies approach and assumes that people vary in how adaptive their strategy selections are. 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. One reason for the lack of popularity of this approach is the currently popular assumption of the multiple-strategies approach that everyone is adaptive in their strategy selections (e.g., Anderson, 1990; Lovett & Anderson, 1996; Reder, 1987; Siegler & Shipley, 1995). That is, it is assumed that each individual selects the best strategy for them on the particular problem and that individuals may vary because of learning in the domain and, perhaps, process parameter differences that change with strategies are most adaptive for them. For example, young children choose to count for difficult addition problems because they have not learned answers sufficiently well and are likely to make an error if they try to retrieve. By contrast, older children know the answer and use retrieval because it is less effortful.

When various groups were compared on the adaptivity of their strategy selections, the groups were typically found to be equally adaptive. That is, all groups were equally able to select the most appropriate strategy for them. For example, Siegler and Lemaire (1997) found that young adults and the aged made equally adaptive strategy selections. Similarly, Kerkman and Siegler (1993) found that middle-income and lower-income children were equally adaptive. In only one case did people vary in their adaptiveness: adaptiveness improved with practice in a domain (Adolph, 1995; Lemaire & Siegler, 1995).<sup>1</sup> However, even at the beginning of practice, strategy selections were still quite adaptive.

This is in contrast to the metacognitive view (e.g., Case, 1985; Flavell, 1979; Kuhn, 1988; Sternberg, 1985), which 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). Instead, it has been argued that the people who use less sophisticated strategies are selecting appropriately because those strategies are best ones for them—either because the more sophisticated strategies are too effortful or because those strategies are too error prone because of lack of practice.

In this article, we explore a variant of the third approach—that individuals vary in their strategy adaptivity. In particular, we explore the proposal made by Reder and Schunn (1999) that individuals vary in their ability to adapt strategies to changing rates of success. That is, some people may be slow to change their strategy selections as the relative successfulness of the strategies change over time. Reder and Schunn's proposal is not that these

participants suffer from lack of metacognitive knowledge of the particular strategies; rather, the proposal is that the participants vary in their general ability to detect and make use of changing rates of strategy success.

To understand this proposal, one must first understand how adaptivity in strategy choice relates to strategy rates of success. Many factors influence strategy choice (cf. Lovett & Anderson, 1996; Reder, 1987; Siegler & Shipley, 1995). One of these factors is the strategy's rate of success. For example, Reder (1987) found that participants' strategy preferences were influenced by the proportion of trials for which a given strategy had been working in the recent past. In Reder's experiments, participants were to answer true-false questions based on short stories they had read. Two strategies were especially common: directly retrieving the answer (a close match to the query) from memory and judging the plausibility of the statement. Participants adjusted their tendency to adopt the plausibility strategy over the direct retrieval strategy as a function of the proportion of questions in the experiment for which a given strategy would work (i.e., proportion of trials in which the statement to be judged or its exact contradiction had been explicitly stated as part of the story). This proportion was varied across blocks of the experiment, and participants adjusted their use of the two strategies accordingly. This basic finding of participants adapting to shifting proportions of features of the experiment has now been seen in a number of other contexts, such as arithmetic verification (Lemaire & Reder, 1999) and in a problem-solving task (Lovett & Anderson, 1996).

There are several reasons why sensitivity to success rates is an important skill. First, many tasks in the world change dramatically independently of the actions of the individual, and in such dynamic tasks, the ability to shift strategies is crucial. To provide a few examples, one must change driving habits when it begins to rain, change hitting strategies as the tennis ball becomes soft, change walking strategies when the sidewalk is icy or when one is wearing high-heeled shoes, and change negotiating strategies when the opponent becomes irate. Second, even in tasks in which the structure of the task stays constant, the individual must shift strategy use as he or she becomes increasingly expert at the task (Adolph, 1995). For example, skiers begin with a snowplow strategy and only much later attempt parallel turns. Similarly, an algebra student shifts from implementing all the steps in an algorithm to skipping or combining simple steps (Blessing, 1996; Blessing & Anderson, 1996).

Are all individuals roughly equally sensitive to success rates, and do they all change strategies equally quickly? This article seeks to investigate whether there are systematic differences among adults in their ability to adapt strategies using success-rate information. Two related questions that will also be addressed are the following: (a) Is adapting or shifting strategies based on the success of the strategies indeed optimal (i.e., does increased strategy shifting produce higher levels of task performance)? and (b) What cognitive abilities underlie this strategy adaptivity?

Sensitivity to success rates can be measured at two levels: micro and global. At the microlevel, there is sensitivity to the success-

<sup>1</sup> An interesting related phenomenon is the case in which children start using a new strategy that will become the more effective strategy but temporarily is more effortful and more error prone (e.g., Miller & Seier, 1994).

fulness of a strategy on the immediately preceding attempt at using that strategy. For example, if one attempts to run on an icy surface and falls, one should be less likely to attempt to run on such a surface the next time. If the attempt is successful, one should be more likely to attempt to run the next time. This is similar to simple operant conditioning. Sensitivity to rates of success at the global level refers to changing strategy use in response to changes in the frequency of success defined over many of the past strategy attempts. For example, if a student first experiences 50% failure rates with a strategy for solving algebra problems and then later experiences 5% failure rates once some level of expertise has been reached, then that student should begin to use the strategy more often.

What is the relationship between micro and global sensitivity to success rates? Logically, if an individual is consistently sensitive at the microlevel, then the individual must be sensitive at the global level—the incremental sensitivity sums to at least some level of global sensitivity (although even greater levels of global sensitivity are possible). However, the reverse is not necessarily true. An individual may be sensitive at the global level but insensitive at the microlevel. For example, it may be that only large changes in success base rates measured over many trials result in changes in strategy use. Thus, this article will measure sensitivity at both the microlevel and global level to assess whether there are individual differences at both levels.

In this article, data from three studies are presented; the first two studies focused on microlevel sensitivity and the third study focused on global sensitivity. All three studies involved a particular dynamic task, the Kanfer–Ackerman Air Traffic Controller Task (KA–ATC; Ackerman & Kanfer, 1994), chosen because dynamic tasks bring to the forefront the importance of ability to adapt to changing success rates. Moreover, individual differences in this task have been studied before, both from a parameter-differences approach and from a strategy-differences approach. Coming from the parameter-differences approach, Ackerman (1988, 1989) found that what predicts individual differences in performance in this task is moderated by time in training: Early in training, reasoning ability best predicts performance; later in training, perceptual-speed ability best predicts performance; and by the end of training, simple reaction-time ability best predicts performance. By contrast, coming from the strategy-differences approach, Lee, Anderson, and Matessa (1995; Lee & Anderson, 1997) found that strategy differences predict individual differences in performance in the KA–ATC task at all points of training. Thus, the KA–ATC task is a good task for contrasting the strategy-adaptivity approach with the strategy-differences and parameter approaches. Before presenting the studies, the next section presents an overview of the task.

### The Kanfer–Ackerman Air Traffic Controller Task

The KA–ATC (Ackerman & Kanfer, 1994; Kanfer & Ackerman, 1989) was designed to simulate dynamic aspects of real air traffic control (e.g., weather conditions change, planes lose fuel in real time, certain types of planes require longer runways than others).<sup>2</sup>

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 lost by rule violations (–10 points) or plane crashes (–100 points). Crashes occur when a plane is allowed to run out of fuel before it is landed. In the KA–ATC task, participants must monitor a variety of elements that are displayed on the screen (see Figure 1): (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 the other running east–west; (c) a queue of planes waiting to enter into the hold positions (each queued plane is a dot); (d) two message windows (not shown), one indicating changes in runway conditions (dry, wet, or icy), wind speed (0–20, 20–40, 40–60 knots) and direction (N, S, E, or W) and one providing error feedback; and (e) the current score and penalty points. A weather change occurs approximately every 25 s; planes enter the queue every 7 s.

There are six rules governing this task. First, planes must land into 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, planes can move down only one hold level at a time and only into an unoccupied position. Fourth, the current weather conditions and wind speed determine the runway length required by different plane types (747s always require long runways, DC–10s 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 propeller planes (PROPs) can always use short runways). Fifth, planes with less than 3 min of fuel remaining must be landed immediately—points are subtracted even if the plane does not crash. Sixth, 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-min 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 (as in Figure 1). The number of minutes of fuel left for each plane is indicated at all times and decreases in real time. At the end of each trial, the participant 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 primary data for the KA–ATC task are taken from the keystroke protocol produced by the computer interface. The set of possible keystrokes on the computer keyboard includes up-arrow, down-arrow, F1, enter, and the number keys 1–6. The up-arrow and down-arrow keys move the cursor up and down (respectively) between the different hold positions and runways to indicate where planes are to be moved or landed. The F1 key accepts the planes from the queue into a holding pattern. The Enter key serves one of three functions, which is determined by the context: (a) it selects the plane in the hold corresponding to the current location of the cursor; (b) it moves a selected plane (either from the queue or from another hold position) into an empty hold position, indicated by the current location of the cursor; or (c) it lands a plane on the runway, if the cursor is next to one of the runways. The six number keys each display one of the six rules described earlier. These rules can be displayed at any time.

At the beginning of the task, before the first 10-min trial, the participants are given step-by-step instructions for the task. Once the task begins, the participants cannot refer back to the instructions and must rely on their memory or use the rule keys to display the six rules.

The KA–ATC task involves three primary subtasks: (a) accepting a plane from the queue into a hold position; (b) moving planes within the hold positions; and (c) landing planes. Although previ-

<sup>2</sup> The following description of the KA–ATC task is an abbreviated paraphrasing of the task description found in Ackerman & Kanfer (1994).

FLT#	TYPE	FUEL	POS.
481	PROP	5	3 n
295	DC10	x 4	3 s
738	747	x 4	3 e
			3 w
499	PROP	6	2 n
687	PROP	x 4	2 s
833	DC10	6	2 e
			2 w
-> 594	DC10	x 4	1 n
844	747	x 4	1 s
938	747	5	1 e
644	DC10	5	1 w

n	=====	s	#1
n	=====	s	#2
w		e	#3
w		e	#4

Score	:	388
Landing Pts:	488	Penalty Pts: 20
Runways	:	DRY
Wind	:	0 - 20 knots from NORTH

Flts in Queue...
<F1> to accept

Figure 1. The main screen of the Kanfer-Ackerman Air Traffic Controller Task.

ous researchers studying strategy use in the KA-ATC task have focused on the first subtask (e.g., John & Lallement, 1997; Lee & Anderson, 1997; Lee, Anderson, & Matessa, 1995), this article focuses on the third subtask. In particular, the focus is on the strategy decision of landing a selected plane on the short or long runways when both are open.

This runway-allocation decision is a strategic decision that involves a general tradeoff of physical and cognitive resources. 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, participants must maximize the use of both runways to maximize the total number of planes landed. In other words, the long runway is a scarce resource that should be used sparingly.

We analyzed data from three studies involving the KA-ATC task with the general goal of investigating individual differences in strategy adaptivity in a complex task. The first two studies were conducted by Ackerman, and the data were taken from the Kanfer-Ackerman CD-ROM Database (Ackerman & Kanfer, 1994). The first study (Study 1, PA-ATC on the CD) contains previously unreported data, and the second study (Study 2, ATC-SPR on the CD) contains data reported in Ackerman (1988). Because these two studies did not manipulate strategy success rates, the analyses of these studies focused on microlevel strategy adaptivity. The third study reports data from a new experiment that manipulated strategy success rates and thus focused on global-level strategy adaptivity. The analysis of all three studies addresses four basic questions: (a) Are people generally adaptive in their strategy use in the complex KA-ATC task?; (b) If they are, do people vary systematically in their adaptivity?; (c) Do those systematic differences relate to overall task performance?; and (d) What cognitive abilities are associated with strategy adaptivity?

## Studies

### Study 1 (Ackerman, 1994): Strategy Adaptivity While Learning a Complex Task

#### Method

**Participants.** The participants of Ackerman's Study 1 were 57 University of Minnesota undergraduates taking part for course credit and money.

**Procedure.** Participants were given a total of 27 10-min trials. After every three trials, they were given several ability tests. Nine trials were completed in a day.

**Dependent measures.** The data on the CD-ROM consisted of the following: (a) KA-ATC task computer protocols that were used to infer keystroke rates, strategy use, number of planes landed, and number of planes crashed and (b) scores from 22 ability tests that were centered around six factors (perceptual speed, movement speed, memory, verbal, reasoning, and psychomotor factors). The tests are listed in Appendix A. See Ackerman (1988) for a more complete description of these ability tests.

The primary strategy measure of interest that we used in our reanalysis of Ackerman's data is OpShort, which is the proportion of times that a participant opted to land a plane on the short runway of all the times that a plane was landed and both runways were open. This ratio is computed only for DC-10s-747s can never land on the short runway, PROPs can always land on the short runway, and 727s provide too few failure opportunities to evaluate individual differences in adaptivity.<sup>3</sup>

#### Results and Discussion

**Overall adaptivity.** We focused on the strategy decision of deciding whether to use the short or long runway when both are open, as measured by OpShort. To examine whether participants were adaptive in their OpShort use at the microlevel, we analyzed the OpShort data 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 participants were adaptive, then they should have reduced their tendency to use the short runway when that landing attempt had resulted in an error previously, and they should have increased their tendency to use the short runway for that plane type when that action was previously successful.

<sup>3</sup> In a few cases, there were sufficient numbers to evaluate OpShort adaptivity with 727s, and the results were always very similar to those found with DC-10s.

Only data from the first nine trials were used because error rates were very low after the ninth trial. Moreover, some participants only completed nine trials. Because different participants have different OpShort rates and different error rates, it was important to do this analysis by participant to avoid participant–strategy confounds. However, some participants landed few planes, and some participants made very few errors. This resulted in few opportunities to evaluate OpShort adaptiveness for these participants. To reduce noise levels due to low numbers, we removed all participants with minimum  $N$ s of less than 3, where the minimum  $N$  for each participant was defined as the minimum of two numbers: (a) the total number of OpShort opportunities after a success and (b) the total number of OpShort opportunities after an error. As expected, participants were more likely to use OpShort when the previous attempt was successful than when it was unsuccessful (mean OpShort of .32 vs. .24,  $F(1, 46) = 22.4$ ,  $MSE = 0.008$ ,  $p < .0001$ ). Thus, people were generally sensitive to the successfulness of their previous attempts.

Because error rates decreased over time and OpShort use increased over time, the preceding analyses may have confounded adaptivity with time-based change—OpShort increases and errors decrease simply as participants learn the rules, and the OpShort increases may be unrelated to the decrease in error rates. Moreover, strategy use may have become less flexible with increased experience in the task, and strategy successes and failures may have had weaker or no effect in later trials. To investigate these issues, we reanalyzed the data separating the first four trials from the later five trials—in the first four trials, error rates are above 50% for short landings; in the later five trials, error rates are well below 50%. As Figure 2 shows, participants appeared just as adaptive in the later trials as in the early trials. The analysis of variance (ANOVA) confirmed this assessment—the effect of success was significant,  $F(1, 38) = 25.8$ ,  $MSE = 0.015$ ,  $p < .0001$ , and effects of early or late and the interaction were nonexistent  $F(1, 38) < 1$ . Thus, the observed adaptivity was not a result of a time-based confound, and participants continued to show strong strategy adaptivity to success information in later trials (i.e., there was no evidence for a reduction in strategy choice adaptivity over time).

The previous analyses collapsed across 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

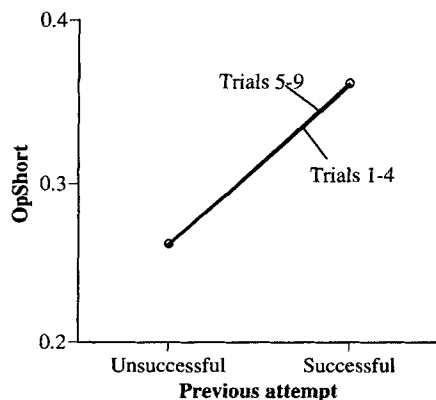


Figure 2. Ackerman Study 1, mean OpShort for early trials (1–4) and later trials (5–9) as a function of the success of the previous attempt to land that plane type.

the two runways were open. Thus, the observed adaptivity may have reflected several types of strategy shifts. It may have led participants to learn the rules more completely after an unsuccessful attempt and then be less likely to use the short runway in illegal situations. Alternatively, it may have reflected a simple shift in tendency to use the short runway that would have had equal impact in legal and illegal situations. Figure 3 suggests that this second alternative is what occurred, and ANOVA results supported this view: There were main effects of legality,  $F(1, 42) = 76.2$ ,  $MSE = 0.05$ ,  $p < .0001$ , and success of the previous attempt,  $F(1, 42) = 20.2$ ,  $MSE = 0.01$ ,  $p < .0001$ , but no hint of an interaction,  $F(1, 42) < 1$ .

In sum, we have found evidence for a general adaptivity to the success of previous attempts that biases participants in their use of the short runway. These results are in agreement with those found in previous research on strategy selection in simple problem-solving tasks (e.g., Lovett & Anderson, 1996; Reder, 1987). Of interest here is that participants were able to demonstrate this strategy adaptivity in the context of performing in a complex, dynamic task.

*Individual differences in adaptivity.* Now that we have established that there is strategy adaptivity for OpShort, we can examine evidence for individual differences in strategy adaptivity. Because individuals varied in how quickly their failures dropped over time, averaging performance over many trials results in differential weighting of early versus later trials across individuals for strategy use after successes and failures. Change in strategy use over trials also varied across participants. Therefore, averaging over many trials to produce estimates of individual strategy sensitivity to successes and failures may produce spurious individual differences. To reduce such artifactual individual differences, we focused on adaptivity in the first four trials. To reduce differences due to noise, participants were required to have a minimum  $N$  of greater than 5 to be included in this analysis. Using this threshold, 34 of the 57 participants were included. Figure 4 presents a histogram of the OpShort adaptivity measure—the difference between OpShort use after successes and OpShort use after failures. The histogram reveals that there was indeed a large range of sensitivity in adaptivity. The modal adaptivity level ( $-0.05$  to  $+0.05$ ) did not include the mean adaptivity level across all participants (0.13). One quarter of the participants showed adaptivity levels that were three or more times as high as the mean, and almost 40% showed no adaptivity at all. Each individual's adaptivity was tested against zero using a  $z$ -score approximation. Using a strict  $p < .05$  criterion, 3 of the 34 participants showed strategy adaptivity. Using a very lax  $p < .5$  criterion, only 17 of the 34 participants showed significant strategy adaptivity. Thus, some participants showed very strong evidence of strategy adaptivity to success rates, whereas many participants showed no evidence of strategy adaptivity to success rates.

How do these individual differences in strategy adaptivity relate to performance in the task? Because the primary goal of the KA–ATC task is to land planes, we correlated DC10 OpShort adaptivity against the number of planes landed. Because individual differences in OpShort adaptivity were measured using Trials 1–4, we used the mean number of planes landed in Trials 1–4. Over this range, OpShort adaptivity correlated positively with planes landed,  $r = .40$ ,  $p < .02$ . By contrast, of the 22 individual difference tests administered by Ackerman, the four best predictors of planes

landed correlated in the .44 to .47 range.<sup>4</sup> OpShort adaptivity continued to correlate significantly with performance even when each of the best individual test predictors was included first in a hierarchical multiple regression with OpShort ( $ps < .05$ ). The predictiveness of the psychometric battery was not curtailed because of restricted-range issues from removing so many participants from the analyses—correlations between performance and the top-predicting psychometric tests were lower when all participants were included. Moreover, when a factor analysis (principal-factor extraction using varimax rotation, extracting orthogonal factors with roots greater than 1) was conducted on the 22 individual difference tests, the six resulting factors were poorer predictors of performance than OpShort ( $rs < .2$ ,  $ns$ ). Taken together, these analyses suggest that adaptivity correlates with performance directly and not through indirect correlations with other determinants of performance. Thus, we have the first piece of evidence that individual differences in adaptivity may be an important factor in performance.

However, it is possible that these individual differences are due to floor and ceiling effect artifacts—in other words, perhaps some participants did not adapt because they either never used OpShort or always used OpShort. To investigate this possibility, strategy adaptivity was regressed against OpShort use after failures. If there was such an artifact, then there should be a strong negative quadratic relationship, such that there should be much weaker sensitivity at the extreme levels of OpShort after failure. Although the quadratic is negative, the relationship is not significant,  $t(31) = -1.5$ ,  $p < .2$ . Moreover, as Figure 5 reveals, the quadratic is driven primarily by two outliers—one with sensitivity more than 0.5 and the other with OpShort after failure more than 0.5. When these two outliers are removed, the quadratic relationship disappears entirely, and yet the remaining participants show the full continuum of sensitivity.

Another issue is the high attrition rate from all the preceding analyses. Of the original 57 participants, only 34 were evaluated for individual differences on OpShort adaptivity. The remaining 23 participants were removed because there were too few opportunities to evaluate whether they were adaptive or not—either because they selected the short runway too few times ( $N = 19$ ) or because they made too few errors ( $N = 4$ ). Did these excluded participants systematically differ from the included participants? To investigate

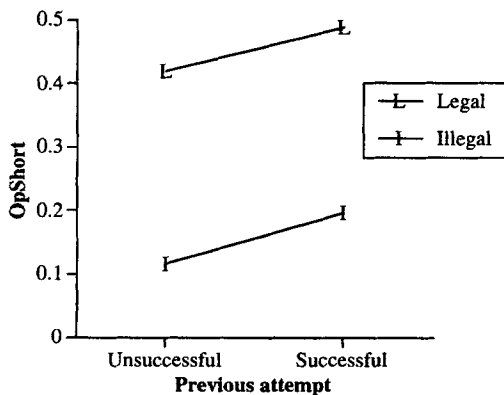


Figure 3. Ackerman Study 1, mean OpShort for legal and illegal short runway opportunities as a function of the success of the previous attempt to land that plane type.

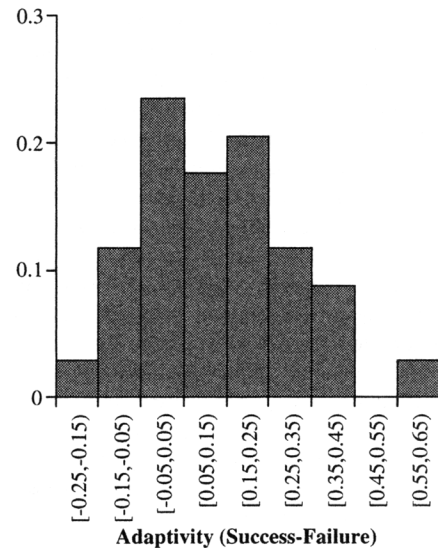


Figure 4. Ackerman Study 1, histogram of individual OpShort adaptivity to the success of the previous short runway landing attempt (OpShort after successes – OpShort after failures). Each range runs from the lower value (inclusively) to the upper value (exclusively).

this issue and to further investigate the issue of floor and ceiling effects, the participants were divided into five groups: excluded (OpShort minimum  $Ns \leq 5$ ), adaptive (OpShort adaptiveness  $> 0.05$ ),<sup>5</sup> unadaptive high (OpShort after failure  $> 0.5$ ), unadaptive low (OpShort after success  $< 0.1$ ), and unadaptive other (remaining unadaptive participants). The unadaptive high group ( $N = 1$ ) represented participants who chose a uniformly high level of OpShort. The unadaptive low group ( $N = 4$ ) represented participants who chose a uniformly low level of OpShort. These two groups may have been unadaptive because of floor and ceiling effects in OpShort use. The unadaptive other group ( $N = 8$ ) represented participants who had intermediate OpShort levels yet were unadaptive nonetheless.

Table 1 presents group means along several performance measures. The groups differed marginally in terms of number of planes landed,  $F(4, 52) = 2.1$ ,  $MSE = 37.1$ ,  $p < .1$ , with the adaptive participants landing more planes than those in all other groups. The groups did not differ significantly in the number of errors made,  $F(4, 52) = 1.3$ ,  $MSE = 42.2$ ,  $p > .3$ , the number of planes crashed,  $F(4, 52) < 1$ ,  $MSE = 1.6$ , or the number of keypresses  $F(4, 52) < 1$ ,  $MSE = 59,900$ . There were differences in terms of the number of calls to Rule IV—the rule for when each plane type may be landed—in that the unadaptive other group called Rule IV more often than did the other groups,  $F(4, 52) = 3.0$ ,  $MSE = 4.4$ ,  $p < .05$ , but this difference was driven entirely by one outlier in the unadaptive other group. There were also marginal differences in the Ravens progressive matrices scores—the best measure of  $g$  in

<sup>4</sup> Because 4 of the participants were missing some of the psychometric test scores, the number for these correlations was 30. Using only these participants, OpShort Adaptivity correlated at  $r = .50$  with planes landed.

<sup>5</sup> The .05 threshold for being included in the adaptive group was simply taken from the histogram analysis of adaptivity. Also, adaptivity levels below .05 were well within noise levels of zero adaptivity.

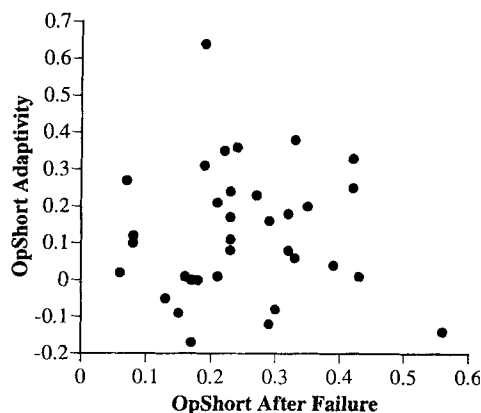


Figure 5. Ackerman Study 1, scatterplot of OpShort adaptivity (success - failure) against OpShort use after failure.

the individual differences battery—in that unadaptive low participants had the lowest scores,  $F(4, 52) = 1.9$ ,  $MSE = 44.0$ ,  $p < .15$ . Overall, the three unadaptive groups were quite similar to one another, and the excluded participants were more similar to the unadaptive participants (especially the unadapt other participants) than to the adaptive participants. These trends are consistent with the view that excluded participants were unadaptive in their behavior and made extreme behavioral choices relating to OpShort, which resulted in their being excluded. Thus, these measures of strategy adaptivity may have underestimated the proportion of unadaptive participants.

*What predicts adaptivity?* Another issue of interest is whether adaptivity can be predicted from the psychometric ability tests. Only 3 of the 22 psychometric scores were significantly correlated with OpShort adaptivity (Trials 1–4, minimum  $N > 5$ ): Letter Sets,  $r = .45$ ,  $p < .02$ , Simple RT,  $r = .42$ ,  $p < .02$ , and 2-Choice RT,  $r = .41$ ,  $p < .02$ . The first test is a measure of reasoning ability, and the other tests are measures of psychomotor ability. The  $N$  (30) is too low in this analysis for us to be sure which factors are really the best predictors, especially given the collinearity among several of the psychometric tests. Across the three studies reported here, it should become clearer which factors consistently predict adaptivity. However, these correlations do provide another piece of evidence that the individual differences in adaptivity were not simply due to chance variation.

#### Study 2 (Ackerman, 1988): Adaptivity During Environment Change

Study 1 found that people could adapt their strategy use to microlevel success and failure feedback, that people varied in how much they adapted, and that these individual differences were predictive of performance in the task. The analyses of Study 2 attempt to replicate these findings in a slightly different situation. Study 1 focused on adaptivity to success and failure feedback as people were just learning a complex task. By contrast, Study 2 focuses on adaptivity to success and failure feedback as people encounter a change in the task after having mastered many aspects of the task already. It may be that once people have settled on a particular strategy, they will be less adaptive in response to success and failure feedback. Also, if people are adaptive in this case, the

individual differences in adaptivity may disappear because the working-memory demands of the task are lower in later stages of training.

#### Method

*Participants.* The participants of Ackerman's Study 2 (also reported in Ackerman, 1988) were 63 University of Minnesota undergraduates who participated for course credit and \$25.<sup>6</sup>

*Procedure.* As in Study 1, Ackerman's participants were given 27 10-min trials. The participants were also given the same 22 ability tests as in Study 1. The primary difference from Study 1 was that the participants experienced only good weather (low wind speed and dry runways) for 18 trials, followed by 9 trials using the full weather conditions as in Study 1—a mixture of good and bad weather. The participants were not told in advance that the first trials would only involve good weather nor that this would change on Trial 19.

#### Results and Discussion

*Overall adaptivity.* As with Ackerman's Study 1, we analyzed the OpShort data 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). However, because the first 18 trials did not involve bad weather (and thus there could be no errors in landing on the short runway for the DC-10s), we used data from the first four foul-weather trials (i.e., Trials 19–22). As in overall adaptivity analyses of Study 1, we removed all participants with minimum numbers of less than 3 in any condition.<sup>7</sup> The mean OpShort rates after a successful landing attempt were significantly higher than those after an unsuccessful landing attempt, mean OpShort of .41 vs. .28,  $F(1, 46) = 18.4$ ,  $MSE = 0.022$ ,  $p < .0001$ . Thus, people were sensitive to the successfulness of their previous attempts even when they were well practiced with other aspects of the task.

As with Study 1, the effects of success of the previous attempt on OpShort rates were divided into legal and illegal cases. There was a main effect of success,  $F(1, 39) = 6.07$ ,  $MSE = 0.040$ ,  $p < .02$ , a main effect of legality,  $F(1, 39) = 174.1$ ,  $MSE = 0.034$ ,  $p < .0001$ , and no hint of an interaction,  $F(1, 39) < 1$ ,  $MSE = .026$ , with approximately 8% more OpShort selections when the prior landing attempt was successful, regardless of whether the current situation was legal or not. Thus, once again, the impact of previous successes and failures appears to be a bias in strategy use rather than learning the rules for runway applicability.

*Individual differences in adaptivity.* After considerable training with the task, were participants just as variable in their adaptivity? As with Study 1, to reduce individual differences due to noise, participants had to have a minimum number of greater than 5 to be included in the individual differences analyses. Using this threshold, 45 of the 54 participants were included. Surprisingly, there was approximately the same level of individual differences in both Study 1 and Study 2 in the adaptivity measure of the difference between OpShort use after successes and OpShort use after failures. The standard deviation in individual adaptivity was the same in both studies (0.18), and the range in values was

<sup>6</sup> The original Ackerman study had 65 participants. However, data from 2 of the participants could not be extracted from the CD-ROM.

<sup>7</sup> The number of participants removed varied by analysis—the  $N$ s can be inferred from the degrees of freedom in the ANOVAs.

Table 1

Means and Standard Errors for Each Group Over the First Four Trials for Measures in Study 1 (Ackerman, 1994)

Measure	Group									
	Excluded ( <i>n</i> = 23)		Adaptive ( <i>n</i> = 21)		Unadaptive high ( <i>n</i> = 1)		Unadaptive low ( <i>n</i> = 4)		Unadaptive other ( <i>n</i> = 8)	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
OpShort adaptivity	—	—	.23	.03	-.14	.00	-.07	.04	-.02	.02
Planes landed	25.0	1.2	30.5	1.2	24.3	0.0	24.6	0.7	25.7	3.3
Errors	10.1	1.6	12.9	1.4	4.8	0.0	13.3	3.0	8.3	3.1
Crashes	1.4	0.3	1.2	0.2	0.3	0.0	1.1	0.3	1.9	0.7
Key presses	993	49	1054	60	929	0	991	103	914	71
Calls of Rule IV	0.2	0.1	0.4	0.3	0.0	0.0	0.4	0.2	3.0	1.8
Ravens Progressive Matrices score	33.0	1.6	34.6	1.3	39.0	0.0	25.2	5.2	32.0	2.2

quite similar (0.81 for Study 1 and 0.88 for Study 2). Figure 6 presents a histogram of OpShort adaptivity in Study 2. In this case, the modal group (0.05 to 0.15) did include the mean adaptivity level across all participants (0.14). However, one third of the participants showed adaptivity levels that were two or more times as high as the mean, and over 30% of the participants showed no adaptivity at all.

Again, we can ask how well these individual differences in strategy adaptivity correlate with performance in the task. We used the mean number of planes landed per trial over Trials 19–22 (the range over which adaptivity was assessed) as the measure of performance. OpShort adaptivity correlated positively with planes landed ( $r = .52, p < .001$ )<sup>8</sup> at even higher levels than in Study 1. As in Study 1, adaptivity is a somewhat weaker predictor of performance than the best individual difference battery measures—the four best predictors correlated in the .62 to .65 range. The top eight predictors (all with  $r_s > .44, p_s < .01$ ) involve a mix of perceptual speed, psychomotor ability, and reasoning ability. If one enters OpShort adaptivity and these eight psychomotor tests

into a hierarchical multiple regression (with the eight psychomotor tests entered first) predicting planes landed, OpShort continues to have a significant contribution (partial  $r = .41, p < .01$ ). As with Study 1, there were no restriction of range issues: The individual difference battery measures correlated no better with performance in the full data set than in the reduced data set for which there were minimum numbers of 5. Thus, both in early training and later in training after changes in the environment, adaptivity appears to correlate with performance directly and not through indirect correlations with other determinants of performance—because the direct correlations were larger than correlations along the indirect path.

*What predicts adaptivity?* To examine whether adaptivity was again associated with performance on the psychometric ability tests, the 22 psychometric scores were correlated with OpShort adaptivity (minimum  $N > 5$ ). Only two tests produced significant correlations: Patterns ( $r = .39, p < .03$ ) and Letter–Number Substitution ( $r = .35, p < .05$ ). Although both of these tests are supposed to be measures of perceptual speed, when a factor analysis (principal-factor extraction with varimax orthogonal rotation) is conducted, they both load either primarily or heavily on reasoning and memory factors. Once again, the numbers are too low to be sure that these factors and not others are really the best predictors. Yet across Studies 1 and 2, the evidence suggests that some aspect of reasoning ability and possibly psychomotor ability or perceptual speed may be associated with adaptivity. This issue, among others, is explored further in Study 3.

### Study 3: Strategy Adaptivity to Manipulated Base Rates

The analyses of Studies 1 and 2 suggest that people are generally adaptive in their strategy use and that there are important individual differences in this strategy adaptivity to success rates, both early in training and later on. However, the measure of strategy adaptivity used in those studies had one very important flaw: It was defined relative to the participants' self-created successes and failures (i.e., it was a correlational measure). It is possible that the participants differed in the kinds

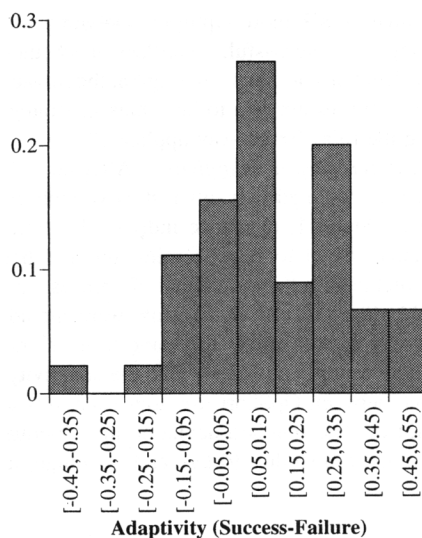


Figure 6. Ackerman Study 2, histogram of individual OpShort adaptivity to the success of the previous short-runway landing attempt. Each range runs from the lower value (inclusively) to the upper value (exclusively).

<sup>8</sup> Because 5 of the participants were missing some of the psychometric test scores, the number for these correlations was 40. Using only these participants, OpShort Adaptivity correlated at  $r = .58$  with planes landed.



of success and failures they produced and not in their strategy adaptivity. Study 3 provides a more controlled measure of adaptivity by manipulating the success rates of OpShort and observing the participant's adaptivity to these manipulated success rates. An additional advantage of manipulating success rates is that it becomes possible to study separately individual differences in amount of adaptivity (e.g., high vs. low amounts of strategy change) and individual differences in rate of adaptivity (e.g., fast vs. slow rates of strategy change). Because global success rates are manipulated, Study 3 also measures sensitivity to success rates at the global level.

Study 3 manipulated OpShort success rates by varying the proportion of 747s and PROPs in each 10-min 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, it should be adaptive to place the DC-10s and 727s on the short runway whenever possible. By contrast, when there are few 747s, there is much less pressure to place the DC-10s and 727s on the short runway. In this situation, it should be more adaptive to place those planes on the long runway—landing on the short runway requires making 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.

Another advantage of Study 3 is that it uses participants who vary more widely in their cognitive abilities than do the typical university undergraduates. This provides advantages both in terms of the greater external validity of the findings and greater power for correlational analyses. Finally, Study 3 also uses a different battery of individual ability measures. This battery has clearer information processing underpinnings than the battery used in Studies 1 and 2 and thus is more likely to provide useful information about the cognitive correlates of strategy adaptivity. For example, we may discover whether working memory is an important component of strategy adaptivity.

Finally, in Study 3, we are able to examine both macro- and microlevel strategy adaptivity, as well as the correlation between the two. Macrolevel adaptivity will be measured as the changes in OpShort to the base-rate manipulation. Microlevel adaptivity will be defined, as in Studies 1 and 2, as the difference between OpShort use after successful landings and OpShort use after unsuccessful landings.

## Method

**Participants.** There were 148 participants, ranging in age from 18 to 31 years ( $M = 23.1$ ). They were recruited from a temporary-employment agency and paid for their participation. The study was conducted at the Brooks Air Force Base TRAIN lab as part of a larger study on individual differences. There were 123 participants in Condition A and 25 participants in Condition B. Fewer participants were assigned to Condition B because the focus was on predicting individual differences, and Condition B was simply a control condition. Approximately 65% of the participants were men, 5% of participants did not have a high school diploma, and 21% had at least some college experience.

**Procedure.** Study 3 used a version of the KA-ATC task that was reimplemented for the IBM Windows environment. Also, in contrast to Studies 1 and 2, participants were given only nine 10-min trials. These nine trials were divided into three blocks of three trials each.

There were two between-subject conditions that manipulated the proportion of 747s (and PROPs) across blocks in different orders. In Condition A, the proportions of 747s over the three blocks were 25%, 5%, and 50%. A second condition, Condition B, with a different order was used to ensure 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 proportions of 747s for Condition B were 25%, 50%, and 5% across the three blocks. Because the strategy adaptivity measure is defined by plane type, one plane type (DC-10s) was set at a constant high level of 40% across all three blocks to ensure sufficient numbers for each participant on at least one plane type. The frequency of PROPs was set to be 55% minus the frequency of 747s (i.e., 30%, 50%, and 5% in Condition A and 30%, 5%, and 50% in Condition B), thereby completing the manipulation of the scarcity of the long runways. The proportion of 727s was held constant at a low level of 5% across the three blocks.

Because the focus of the study was on predicting individual differences, the remaining structure of the task was held as constant as possible across blocks and participants while still maintaining the overall dynamic structure of the task.<sup>9</sup>

**Dependent Measures.** As in Studies 1 and 2, the data consisted of the keystroke information (e.g., strategy use, planes landed, and planes crashed), and scores from the ability tests. The most important dependent measure is again OpShort. Because the DC-10 was the only plane type that occurred with a constant, high frequency in all blocks, the proportion of OpShort was only calculated for DC-10s.

The individual-ability battery was a subset of the Cognitive Abilities Measurement (CAM, version 4, 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, 1998; Shute, 1993). The reason for selecting the CAM is that it is structured around information-processing concepts (e.g., working memory, procedural learning, processing speed, etc.).

Because the full CAM is quite large and requires several days to complete, only 11 CAM tests were used. These 11 tests covered the main information processing constructs—all with plausible possible connections to adaptivity. The tests included measures of fact (or associative memory) learning, procedural learning, processing speed, working memory, and inductive reasoning. For each skill type, there was one test in the verbal domain and an isomorphic test in the spatial domain (e.g., word recognition and figure recognition). The single exception was inductive reasoning, for which only spatial reasoning tests were available, and so the three available spatial tests were included, without the complementary verbal tests. Table 2 presents each of the tests that were used (see Appendix B for more detailed descriptions of the tasks).

Each of the selected measures is a plausible correlate of strategy adaptivity. Fact learning may predict OpShort adaptivity because those participants who can memorize the rules of the task more readily may be better able to increase their use of the short runway. Similarly, participants with better procedural learning ability may learn the details of the task more quickly and then been more able to increase their use of the short runway. Participants with faster processing speed may have more free time to notice changes in base-rates and then be able to react to them. Participants with greater working-memory capacity may be better able to retain base-rate information while performing the task. Moreover, participants with greater working-memory capacity may have been better able to maintain the rules

<sup>9</sup> This was implemented by using different scripts for weather changes that repeated every three trials.

Table 2  
*The 11 Cognitive Abilities Measurement Tests Used in Study 3 Classified  
 by Skill Type and Content Domain*

Skill type	Content domain	
	Verbal	Spatial
Fact learning	Word Recognition	Figure Recognition
Procedure learning	Reduction: Future-Past-Present	Reduction: Circles
Processing speed	2-Term Ordering: Furniture/Animals	2-Term Ordering: Blocks
Working memory	4-Term Ordering: Furniture/Animals	4-Term Ordering: Blocks
Inductive reasoning		Figure Sets
Inductive reasoning		Figure Series
Inductive reasoning		Figure Matrices

*Note.* Verbal reasoning tests were not available for the tests of inductive reasoning.

in memory while doing the task.<sup>10</sup> Finally, participants with strong inductive-reasoning skills may be more likely to notice a change in the pattern of plane types—indeed, the results from Studies 1 and 2 suggest that reasoning ability may be related to adaptivity.

Each test produced one score per participant: overall percentage correct on that measure. The exception to this is Processing Speed, which involved a median reaction time. Thus, there were 11 ability measures per participant. Because these ability measures were selected because they map directly onto information processing concepts, we will use the measures themselves as predictors rather than factors derived from a factor analysis. Second, differences in correlations between the spatial and verbal component of each ability provide information about what aspect of the task is being tapped as well as suggestions about possible spuriousness of the correlations. However, including a large number of possible predictors has a nontrivial probability of finding spurious correlations. To reduce the number of variables (and yet retain factors that are meaningful information-processing constructs), the following procedure was used. First, a factor analysis was conducted (principal-factors extraction with varimax orthogonal rotation). This produced five factors, only three with eigenvalues greater than one. The measures that loaded only on the two factors with eigenvalues less than one (Fact Learning Verbal and Processing Speed Verbal) were removed, under the argument that these measures are essentially linear combinations of the other predictors. Then, using only the three inductive spatial tests, another factor analysis was conducted using an oblique rotation to extract the general factor underlying the inductive spatial tests. Thus, predictors were reduced to seven: fact learning spatial, processing speed spatial, skill learning spatial, skill learning verbal, working memory spatial, working memory verbal, and inductive reasoning spatial.

Another obvious source of differences in task performance is the degree to which participants pay attention to the instructions. To assess and partial out such differences, reading times were gathered for each of the 61 instruction pages. These timing data were compressed into four variables: the time spent reading each of the three rules pages (which displayed Rules I–III, IV, and V–VI, respectively) and the median time spent reading each of the remaining pages.

### Results and Discussion

Of the 148 participants, 25 participants (20 in Condition A and 5 in Condition B) were excluded from the analyses. Twenty-four of these participants performed so badly throughout the task (negative scores in every block; mean total score per block of  $-3,419$ ) and with little sign of improvement) that it is highly likely that they were not taking the task seriously. Eight of the 25 excluded participants had too few opportunities (less than three) to exhibit a preference in one or more of the blocks because they landed almost

no planes. Although one could assume that these 25 participants were not adaptive, these participants were simply excluded from the analyses because we wanted to exclude motivational issues from our analyses of the role that ability plays in performance.

**Manipulation check.** The goal of the plane-type manipulation over blocks was to change the adaptiveness of OpShort. That is, it should be adaptive to use low OpShort when there are few 747s and high OpShort when there are many 747s. To test whether the manipulation was successful in this regard, regression analyses were conducted on the relationship between OpShort use and task performance within each block. In particular, each participant's score for each block was regressed against the proportion of OpShort for that block, and this was done separately for Blocks 2 and 3, within each condition. Blocks in which the participant had fewer than three opportunities to select among two available runways were excluded. Within Condition A, Block 2 (5% 747s), OpShort was negatively correlated with score ( $r = -.35, p < .001$ ). Within Block 3 of condition A (50% 747s), OpShort was positively correlated with score ( $r = .35, p < .001$ ). The regression results for Condition B were similar, although not as strong because there were fewer data points. Within Block 3 of Condition B (5% 747s), OpShort was uncorrelated with score ( $r = -.12, p > .5$ ). By contrast, within Block 2 of Condition B (50% 747s), OpShort was positively correlated with score ( $r = .55, p < .02$ ). Thus, the regression analyses indicate that the manipulation was successful: High OpShort rates were adaptive when there were many 747s, and low OpShort rates were adaptive when there were few 747s.

**Overall adaptivity.** Although the preceding analyses suggest that participants should have changed their OpShort use across the blocks, it may have been that participants were too busy with the demands of the task to be able to adapt their strategy use. To examine whether participants did adapt, a Condition  $\times$  Block mixed ANOVA was conducted on OpShort. The predicted interaction was quite strong,  $F(2, 242) = 16.4, MSE = 0.022, p < .0001$ . For Condition A, the participants adapted in the predicted pattern of medium-low-high (see Figure 7). By contrast, the participants in condition B followed the predicted pattern for only

<sup>10</sup> In fact, participants with higher verbal working-memory capacity spent less time reading the instructions ( $r = -.25, p < .01$ ) and asked to see the task rules less often during the task ( $r = -.20, p < .02$ ).

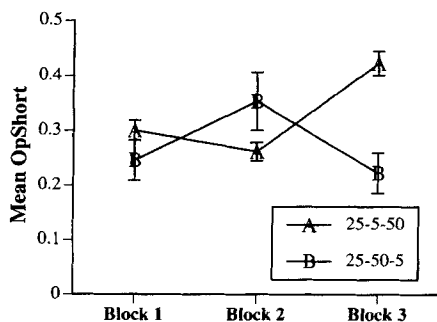


Figure 7. The mean proportion OpShort with each condition within each block of three trials in Study 3.

Blocks 2 and 3 (high, then low)—their Block 1 was somewhat lower than expected, perhaps reflecting a lack of knowledge of the rules at the beginning of experiment.<sup>11</sup> However, participants in general did adapt in response to the manipulation in the expected directions. Thus, as in Studies 1 and 2, participants overall were able to adapt their strategy use to changing rates of success in the context of a complex, dynamic task.

**Individual differences in adaptivity.** The remaining analyses focus on whether participants differed in their adaptivity. To address this issue, the participants were classified into adaptive and unadaptive groups using two different criteria: a strict criterion using all three blocks and a lax criterion using only the last two blocks. These analyses involve participants from both conditions, but the definitions of adaptivity will be defined in terms of Condition A (i.e., a medium, low, high pattern) to keep things simple (i.e., the reader can assume that the measures were appropriately reversed or otherwise modified for participants in Condition B). To be classified as adaptive using the strict criterion, the participant's OpShort for the second block had to be lower than in the first block and their OpShort had to be higher in the third block than in the first block (i.e., a medium, low, high pattern). To be classified as adaptive using the lax criterion, the OpShort for the third block had to be higher than that of the second block. According to the lax criterion, 69% (85 of 123) of the participants were adaptive. According to the strict criterion, only 29% (36 of 123) of the participants were adaptive. Thus, only a minority of the participants fully adapted their OpShort use throughout the task, and 31% of the participants did not meet even the very lax criterion of adaptivity.

Was it optimal to follow the expected strategy adaptation patterns? Deciding where to land a plane is only one small component of this complex task. For example, research by others with this task has shown that the very effective (and legal) strategy of moving planes from the queue directly to Hold 1 accounts for much of the performance variance throughout the task (John & Lallement, 1997; Lee et al., 1995). Thus, participants who pay close attention to where to land their DC-10s may be doing so at the cost of paying less attention to other important decisions. In other words, it is possible that OpShort adaptive participants are "penny wise but pound foolish" (i.e., locally optimal but globally suboptimal).

To examine this issue, adaptivity using the strict criterion was regressed against mean block score. Those participants classified as adaptive did indeed have much higher mean block scores than those classified as unadaptive, 3,480 vs. 2,102,  $F(1, 121) = 17.1$ ,

$MSE = 2.83 \times 10^6$ ,  $p < .0001$ . However, it is possible that this correlation is mediated by some other ability differences. That is, adaptive participants may have been generally more intelligent and may have performed at higher levels independent of their strategy adaptivity.

When adaptivity was placed in competition with the seven CAM ability measures (or any subset of these seven measures) in a hierarchical multiple regression (with the ability measures entered first) predicting score, OpShort adaptivity continued to be a significant correlate of score ( $ps < .05$ ).<sup>12</sup> Thus, it appears that the correlation between adaptivity and performance is not mediated through indirect correlations with psychometric ability.

**Amount and rate of adaptivity.** The binary classification of participants into adaptive and unadaptive does not distinguish between participants who adapted only slightly and those who shifted their strategy use a great deal in response to the base-rate manipulations. To examine whether the adaptive participants differed in how much they adapted their OpShort use, a measure of extent of adaptivity was developed: the difference in proportion of OpShort use between the second and third blocks, between which was the largest transition in 747 base rates.<sup>13</sup> This measure, as a difference in proportions, could range from 0 to 1. Only the 85 participants who showed an increase in OpShort use from Block 2 to Block 3 were included. Figure 8 shows that these adaptive participants varied widely in their extent of adaptivity—40% of the adaptive participants adapted less than .2, and 20% adapted .4 or more.

Another measure of adaptivity is how fast people adapt. This was measured as the proportion of an eventual adaptation made immediately, using the following method, again focusing on the transition from the second to third blocks for which there was the biggest base rate change. Eventual adaptation amount was first calculated—the difference between OpShort on the trial immediately before the transition (i.e., Trial 6) and the largest OpShort value on the three trials after the transition (i.e., maximum of Trials 7, 8, and 9). The amount of immediate adaptation was calculated as the difference between the OpShort for the last trial before a transition and the first trial after that transition (i.e., Trials 6 to 7). Then the amount of immediate adaptation was divided by the eventual transition amount, giving a proportion. A score of 1 reflected adaptation completed entirely immediately, zero reflected no immediate adaptation, and values in between reflected intermediate adaptation rates. Participants who adapted in the wrong direction on the first transition block were assigned a zero on this measure. As with extent of adaptivity, participants who did not adapt from the second to third blocks (i.e., were unadaptive according to the lax criterion) were excluded. Figure 9 shows very large differences in rate of adaptivity ( $N = 81$ )—20%

<sup>11</sup> The difference between conditions in Block 1 was only marginally significant,  $F(1, 121) = 1.5$ ,  $MSE = 0.033$ ,  $p < .25$ .

<sup>12</sup> The correlations between performance and the psychometric measures were approximately the same whether or not the excluded participants were included—thus, there were no problems of restricted range in these analyses.

<sup>13</sup> The transition between Blocks 1 and 2 involved a fairly small change in 747 base rates and therefore presents restricted range problems for examining individual differences.

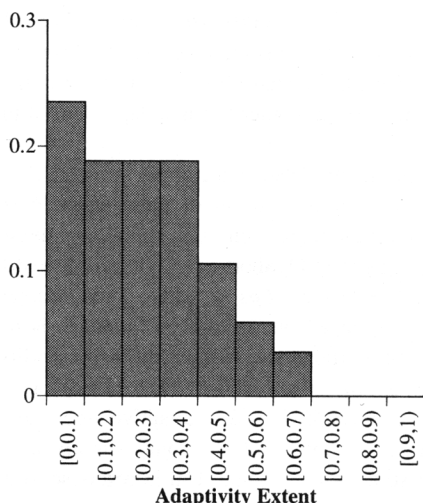


Figure 8. Histogram of adaptive participants' extent of adaptivity in Study 3. Each range runs from the lower value (inclusively) to the upper value (exclusively).

of the adaptive participants made almost no change immediately, and over 40% adapted almost entirely immediately.

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. As was done with the binary adaptivity measure, both of these measures were regressed against mean block score. Extent of adaptivity as measured by the rise in OpShort from Blocks 2 to 3 (including only those participants who showed an increase) was significantly correlated with overall score ( $r = .62$ ,  $p < .0001$ ). Rate of adaptivity as measured by the percentage of immediate change from Blocks 2 to 3 (again including only those participants who showed an increase) was not significantly correlated with overall score ( $r = .10$ ,  $p > .3$ ), but it did correlate with score in Block 3 ( $r = .23$ ,  $p < .05$ ). It may be that the low correlations of rate with performance are due to higher noise levels from using trial level rather than block level data in its calculation.

*Individual differences among the unadaptive.* Why did the unadaptive participants not adapt their strategy use? One potential explanation is that the participants may have always or never used the short runway. In other words, the unadaptive participants did not appear to adapt because of floor or ceiling effects in strategy use. Alternatively, these participants may have used both the short and long runways frequently but were simply unable to adapt their strategy use to the changing base rates. The unadaptive participants (using the strict criterion) were classified into three groups corresponding to these alternative explanations: unadapt high, with OpShort levels above 45% in all three blocks; unadapt low, with OpShort levels below 15% in all three blocks, and unadapt other, the remaining unadaptive participants. Table 3 shows that although there were participants who fell into the unadapt high and unadapt low groups, the great majority of the unadaptive participants fell into the unadaptive other group. In other words, the majority of the unadaptive participants did use the long and short runways frequently but could not adapt their use of them in response to changing success base rates.

What characterized the various participant groups? The five groups (excluded, adaptive, unadaptive high, unadaptive low, unadaptive other) differed in terms of score,  $F(4, 143) = 63.2$ ,  $MSE = 3.09 \times 10^6$ ,  $p < .0001$ ; errors,  $F(4, 143) = 10.9$ ,  $MSE = 5,639$ ,  $p < .0001$ ; crashes,  $F(4, 143) = 100.4$ ,  $MSE = 12.4$ ,  $p < .0001$ ; keypresses,  $F(4, 143) = 4.6$ ,  $MSE = 9.23 \times 10^5$ ,  $p < .001$ ; time spent studying Rule IV,  $F(4, 143) = 2.7$ ,  $MSE = 629$ ,  $p < .05$ ; and overall CAM scores,  $F(4, 143) = 54.8$ ,  $MSE = 165$ ,  $p < .0001$  (see Table 3). The excluded group had negative scores, a very large number of errors, very many plane crashes, a low keypress rate, little time spent studying the rule for when planes could land, and the lowest overall CAM scores—their performance was truly abysmal. By contrast, the adaptive group had the highest scores, had low error rates, had low crash rates, had a high keypress rate, spent a significant amount of time studying rule IV, and had the highest CAM scores—their overall performance was quite strong. How did the unadaptive subgroups compare? The unadaptive high (high use of the short runway) participants spent the most amount of time studying Rule IV and thus were able to use the short runway frequently without incurring more errors. The unadaptive low participants spent little time reading Rule IV and thus did well to land the planes infrequently on the short runway and incurred very few errors (and this difference in errors was entirely due to having fewer illegal runway selection errors). However, their scores were still much lower than those of the adaptive participants. The unadaptive other participants were not remarkable in any respect—neither strategy floor effects nor ceiling effects nor lack of time spent reading the rules nor motivational differences could explain their lack of strategy adaptivity.

*Predictors of adaptivity.* In Studies 1 and 2, reasoning ability tests correlated with strategy adaptivity, but somewhat inconsistently. In Study 3, the participants were more heterogeneous in their abilities, and a different set of ability tests was used. Were these ability tests predictive of adaptivity differences? If one compares ability test scores of the adaptive and unadaptive participants (using the strict criterion), the two groups differ on all but one (processing speed) of the seven psychometric measures. How-

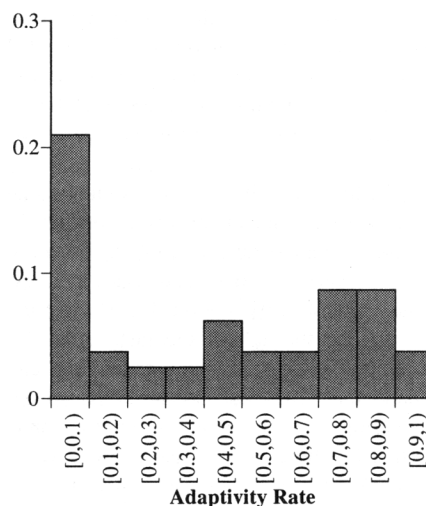


Figure 9. Histogram of adaptive participants' rate of adaptivity in Study 3. Each range runs from the lower value (inclusively) to the upper value (exclusively).

Table 3  
Means and Standard Errors for Measures in Study 3

Measure	Group									
	Excluded (n = 25)		Adaptive (n = 36)		Unadaptive high (n = 6)		Unadaptive low (n = 8)		Unadaptive other (n = 73)	
	M	SE	M	SE	M	SE	M	SE	M	SE
Score	-3,419	413	3,480	249	2,148	198	2,700	498	2,032	217
Errors	187.3	23.8	69.9	9.9	90.2	17.1	45.6	17.3	98.7	7.8
Crashes	17.7	1.0	1.1	0.3	1.8	0.4	1.7	1.0	3.1	0.4
Key presses	1,856	153	2,902	162	2,235	344	2,350	358	2,578	119
Study Rule IV <sup>a</sup>	28.3	3.5	39.2	3.6	64.9	9.4	38.1	4.7	37.1	3.4
CAM % correct	53.6	2.1	78.2	1.9	65.8	3.3	70.9	5.8	66.8	1.7

Note. CAM = Cognitive Abilities Measurement.

<sup>a</sup> Mean time in seconds spent reading Rule IV during the instruction phase.

ever, because all of the CAM scores are positively correlated, it is unclear which of these measures are actual independent predictors of adaptivity.

To assess which measures were independent predictors of adaptivity (as a binary outcome using the strict criterion), an all-possible-regressions procedure (with  $C_p$  as the selection criterion) was used including all seven CAM scores and the four measures of instruction-reading time. The overall best model involved only two predictors ( $R^2 = .30$ ): inductive reasoning (partial  $r = .31$ ,  $p < .001$ ) and skill learning spatial (partial  $r = .30$ ,  $p < .001$ ). All the best models involved those two factors. Once again, reasoning ability appears to be related to adaptivity.

Was this relationship between reasoning and adaptivity derivative of some other relationship between inductive reasoning and overall OpShort use? For example, perhaps inductive reasoning was predictive of the ability to learn and use the landing rules and not specifically adaptivity in runway use. In other words, perhaps the adaptive participants were simply better able to learn the rules. If this were true, then one would expect that predictors of OpShort adaptivity would also predict overall OpShort use (i.e., the overall absolute levels). However, none of the seven CAM scores were significantly related to overall OpShort use (mean across all blocks). The best predictor of overall OpShort use was the time spent reading Rule IV during the instructions ( $r = .16$ ,  $p < .07$ ), such that participants who spent more time reading that rule used the short runway more often. Inductive reasoning was not at all associated with overall OpShort use ( $r = .01$ ). Thus, raw OpShort use was a function of degree of rule learning, whereas adaptivity (differential use of the short runway in response to the manipulation) was related to inductive reasoning and not simple rule learning.

A further possible confound is that the CAM tests may have differed in their reliabilities, and the tests that were more predictive were simply the ones that were most reliable. Table 4 presents the even-odd reliabilities for each of the tests. As we see, the most predictive were not the ones with the highest even-odd reliability. Because even-odd reliability is a mixture of both internal consistency and item homogeneity, one might argue that the inductive spatial tests have the lowest item homogeneity and for this reason have the highest predictive power. However, an examination of the items in these particular inductive tests makes this argument unlikely: the items are all fairly similar, and the low even-odd

reliability scores are more likely attributable to the relatively small number of items in those tests. Thus, it is unlikely that the high predictiveness of the inductive reasoning tasks is due to relatively higher internal consistency or to relatively lower item homogeneity.

As discussed earlier, many of the tested participants were excluded from the adaptivity analyses. Although there were important reasons for excluding these participants, it is possible that excluding them has biased or misrepresented what factors are associated more generally with adaptivity. The analyses were reconducted including all the participants for whom adaptivity could be defined (i.e., no longer excluding those with extremely poor performance levels). With these 139 participants, the same two predictors were associated with adaptivity (multiple  $r^2 = .39$ ): inductive reasoning (partial  $r = .35$ ,  $p < .001$ ) and skill learning spatial (partial  $r = .34$ ,  $p < .001$ ). Thus, the conclusions regarding the associates of adaptivity did not appear to be influenced by problems of restricted range.

*Predicting extent and rate of adaptivity.* What factors predict extent of adaptivity? An all-possible-regressions procedure was conducted again, focusing on the OpShort change from Block 2 to Block 3 and including only those participants who showed an increase. The best model included the top three correlates (multiple  $r^2 = .35$ ) in order of predictiveness: working memory spatial ( $r = .49$ ,  $p < .0001$ ), inductive reasoning ( $r = .49$ ,  $p < .0001$ ), and fact

Table 4  
The Even-Odd Reliability for Each Cognitive Abilities Measurement (CAM) Ability Measure in Study 3

CAM measure	Spatial	Verbal
Fact learning	.77	.67
Processing speed (% correct)	.77	.89
Processing speed (median reaction time)	.91	.87
Procedural learning	.88	.96
Working memory	.74	.86
Inductive reasoning: Figure Sets	.15	
Inductive reasoning: Figure Series	.43	
Inductive reasoning: Figure Matrices	.43	

Note. Verbal reasoning tests were not available for the tests of inductive reasoning.

learning spatial ( $r = .47, p < .0001$ ). All the best models included these three predictors. Thus, although inductive reasoning does enter into extent of adaptivity, working memory and fact learning also seem to play an important role.

Do the same factors predict rate of adaptivity? A similar all-possible-regressions procedure was conducted using as a dependent measure the amount of immediate adaptation described earlier, again including only those participants who showed a positive increase in OpShort use. The best model included the top two correlates (multiple  $r^2 = .19$ ): skill learning spatial ( $r = .37, p < .001$ ) and processing speed spatial RT ( $r = -.31, p < .01$ ). For this measure, inductive reasoning was a very poor predictor ( $r = .09, p > .4$ ). Thus, rate of adaptivity appears to depend on different factors. Of course, it is important to note that none of the psychometric measures were particularly good predictors of rate of adaptivity. Once again, these weak correlations with rate of adaptivity may be due to higher noise levels from using trial level rather than block level data in its calculation.

*Micro- versus macroadaptivity.* Up to this point, the analyses of Study 3 have focused on the macrolevel adaptivity to the base-rate manipulation. What about the microlevel adaptivity that was examined in Studies 1 and 2? As in Studies 1 and 2, participants were also sensitive to local feedback: They were more likely to select the short runway after a successful landing attempt than after an unsuccessful landing attempt (mean OpShort of .42 vs. .33,  $F(1, 113) = 32.4, MSE = 0.015, p < .0001$ , for DC-10s, for Trials 1-9). To examine individual differences, micro-OpShort adaptivity was calculated as in Studies 1 and 2: the difference between the proportion of OpShort after successful landing attempts and the proportion of OpShort after unsuccessful landing attempts. This measure had a mean of .09 and a standard deviation of .17. Twenty-six percent of the participants showed no adaptivity at all on this measure.

There was a modest but significant correlation between this microlevel adaptivity and macrolevel adaptivity, as defined as adapting OpShort at all ( $r = .19, p < .05$ ) or as defined as the extent of adaptivity from the second to the third block ( $r = .17, p < .06$ ). There was also a small but significant correlation between microlevel adaptivity and performance in the task ( $r = .23, p < .02$ ). This correlation of microlevel adaptivity and performance is much lower than that found in Studies 1 and 2. This finding may indicate either that (a) microlevel adaptivity is less important over the course of shifting base-rates manipulations or (b) that we have not adequately measured microlevel adaptivity in this shifting base-rate situation. Because this measure of microlevel adaptivity has an even lower correlation with performance in the first block in which no base-rate manipulation had yet occurred ( $r = .17, p < .1$ ), the first alternative seems unlikely. Thus, the weak correlations with microlevel adaptivity are more likely to reflect some bias or extra noise in its measurement in this particular study.

### General Discussion

This article has further explored a new conception of individual differences: differences in strategy adaptivity, specifically adaptivity to changing success base rates (see also Reder & Schunn, 1999; Schunn & Reder, 1998). The three studies found evidence for significant individual differences in sensitivity to success rates, both at a microlevel (all three studies) and at a global level (Study

3). These individual differences were not attributable to chance variation because they were strongly associated with performance and because they could be predicted using cognitive-ability test batteries—most commonly predicted by reasoning ability. Study 3 also found evidence that individuals differ in terms of whether they adapt, the extent to which they adapt, and the rate at which they adapt. The differences did not seem attributable to differential knowledge of the task or to general intelligence differences.

How do our findings compare with other investigations of individual differences in the KA-ATC task? Lee, Anderson, and Matessa (1995; Lee & Anderson, 1997) found that differences in overall strategy use accounted for a large proportion of performance differences in the task.<sup>14</sup> This raises the question of whether the adaptive participants had more complex strategies rather than selecting among the same set of strategies more adaptively. For example, the adaptive participants may have had different explicit strategies for the different plane ratios (e.g., if the ratio of 747s is high, then use the short runway whenever possible). This alternative interpretation would make the observed differences consistent with the strategy-differences view of individual differences. However, it is important to note that the adaptive participants never shifted their OpShort levels in a binary fashion (e.g., from always using the short runway to never using the short runway). Instead, the participants merely changed the degree of short runway use. This kind of continuous shift is much more consistent with changing ratios of strategy use than is shifting from one strategy to another. Moreover, the participants also differed in terms of extent of adaptivity and rate of adaptivity, which is difficult to explain using just different strategies. One might ask whether the strategy differences observed by Lee et al. were in fact due to adaptivity differences. Perhaps more adaptive participants were better able to select appropriate strategies. Whatever the answer, the adaptivity- and strategy-differences approaches are complementary in that both emphasize the importance of strategies in the analysis of performance (Reder, 1982, 1987, 1988).

Coming from a parameter-differences approach, Ackerman (1988, 1989, 1990) focused on the relationship between predictors of performance and extent of training within the KA-ATC task. He found that different factors predicted performance at different phases of training: first reasoning ability, then perceptual speed, and finally reaction-time ability. Our studies also found that reasoning ability and processing speed were important components of performance. One potential contribution of this article is to provide an explanation for the relationships between the cognitive-skills assessment battery and task performance: The relationship is strongly mediated by differences in strategy adaptivity. Thus, the adaptivity-differences approach provides a link between the parameter- and strategy-differences approaches: Different strategies are selected because of different strategy adaptivity, which is related to parameter differences. Strategy selection is a process and as such may be affected by parameter differences in the cognitive architecture—differences in reasoning ability, working memory capacity, and processing speed.

In addition to the theoretical importance of understanding the nature of individual differences and the mechanisms underlying

<sup>14</sup> They focused on a strategy relating to moving planes from the queue to the hold patterns rather than on the strategies in this paper—runway use strategies.

strategy selection, there is practical importance to the findings of this paper. In particular, because individuals appear to differ in how well they adapt their strategy use to changes in the task environment, it might be advantageous to select adaptive individuals for tasks in which the task environment frequently and rapidly changes (e.g., air traffic controllers). Performance in the KA-ATC task was strongly related to strategy adaptiveness—adaptive participants had much higher scores. These studies also suggest which factors may be good predictors of adaptiveness: Inductive reasoning and skill learning predicted whether people adapted; working memory, inductive reasoning, and fact learning predicted how much people adapted; and skill learning and processing speed predicted how quickly people adapted. Future studies should be directed at determining exactly which factors prove to be the best predictors across a variety of tasks and situations.

Assuming, however, that the correlates of adaptivity found in these studies prove to be the best predictor of adaptiveness, 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. In the case of Study 3, this possibility seems unlikely because the base-rate manipulation was so heavy handed that it seems unlikely that any participants were unaware of the manipulation (at least among the nonexcluded participants). Moreover, Schunn and Reder (1998) found that noticing base-rate manipulations did not influence whether people adapted; instead, it was related to how much people adapted, assuming they adapted.

Second, inductive reasoning might be related to being able to quickly understand the relationship between a strategy and its effect or to being able to diagnose when a strategy is no longer appropriate. In a series of computational simulations of participant performance in these studies using the Adaptive Control of Thought-Rational framework (ACT-R, Anderson & Lebiere, 1998), Best, Schunn, and Reder (1998) found that developing an appropriate representation of the goal hierarchy for the landing task was key to being adaptive to the base-rate manipulation. In particular, it was necessary to set a goal to fill both runways (rather than simply a goal to land planes). If the participants did not induce this goal hierarchy, the model would predict that they would be insensitive to base-rate manipulations. Skill-learning ability might play a similar role here as well: How likely are people to adopt the correct skill decomposition for a task?

The role of working-memory capacity in extent of adaptation also has several possible causal chains. Presumably it involves an increased ability to keep information in mind while simultaneously performing the task, specifically base-rate information. In a related fashion, processing speed could be related to adaptation rate by allowing dual tasking: keeping track of outcomes while working at the basic 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. Reasoning ability, in contrast, would predict whether given a pattern, the individual understood what strategy to adopt for best performance with the new pattern.

How generalizable are the findings from these studies? Analyses of individual differences in question-answering and simple problem-solving tasks (Schunn & Reder, 1998) suggest that existence of individual differences in strategy adaptivity is a quite general phenomenon. An open question is whether the strategy

adaptivity approach might be applied fruitfully to other areas of individual differences (e.g., child development, aging, expertise, etc.). If the groups differ in reasoning ability, working-memory capacity, or speed of processing, then the results from this paper suggest there should be adaptivity differences across those other groups as well. Yet Siegler and Lemaire (1997) found no group differences in strategy adaptivity in their older-younger adult comparisons for which previous research suggests that there are processing speed or working-memory capacity differences.

One possible explanation of the lack of effect in the Siegler and Lemaire study is that Siegler and Lemaire looked at adaptivity in a static domain, mathematics. Base rates were not manipulated, and significant learning did not occur over the course of the experiment. Thus, although participants may have exhibited a crystallized strategy adaptivity in which they chose strategies appropriately based on features of the problem, they did not need to exhibit a fluid strategy adaptivity in which they chose strategies using feedback from (possibly changing) base rates of success. This analysis suggests that if Siegler and Lemaire had included a base-rate manipulation, they might have found group differences.

There are many conceptions of adaptivity. The preceding paragraph raised the possibility of static versus fluid adaptivity. The bulk of this article has focused on adaptivity in strategy use, in particular adaptivity to changing base rates of strategy success. It is an open question as to how base-rate-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, 1999), in the ability to select and change representations (Lovett & Schunn, 1999; Schunn & Klahr, 1996; Schunn & Lovett, 1996), or in the ability to adaptively control attention (Gopher, 1982, 1996; Gopher & Kahneman, 1973).

In conclusion, this article has provided evidence for a new kind of individual difference: differences in strategy adaptivity. The article has also described new methods for assessing, validating, and predicting such individual differences. There are several advantages to the strategy-adaptivity approach to individual differences. First, this approach builds on the strengths of the parameter-difference and strategy-difference approaches to individual differences. Because it analyzes the strategies underlying tasks as the strategy approach does, the strategy-adaptivity approach can provide a detailed account of performance on any particular task. Because it searches for predictive features relating to the individual outside of the details of the particular domain, as the parameter approach does, the strategy-adaptivity approach should be able to account for correlated performance differences across many tasks and domains. Finally, the strategy-adaptivity approach promises to provide new insights into the mechanisms underlying strategy selection and the nature of individual differences.

## References

- Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, *117*, 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.
- Ackerman, P. L. (1990). A correlational analysis of skill specificity: Learning, abilities, and individual differences. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *16*, 883-901.



- Ackerman, P. L., & Kanfer, R. (1994). *CD-ROM database, data collection program, and playback program*. Retrieved from Kanfer-Ackerman Air Traffic Controller Task CD-ROM database, 1994, Study 1, PA-ATC.
- Adolph, K. E. (1995). A psychophysical assessment of toddler's ability to cope with slopes. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 734-750.
- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R., & Lebiere, C. (1998). *Atomic components of thought*. Mahwah, NJ: Erlbaum.
- Best, B. J., Schunn, C. D., & Reder, L. M. (1998). Modeling adaptivity in a dynamic task. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the 20th annual conference of the Cognitive Science Society* (pp. 144-149). Mahwah, NJ: Erlbaum.
- Blessing, S. B. (1996). *The use of prior knowledge in learning from examples*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.
- Blessing, S. B., & Ross, B. H. (1996). Content effects in problem categorization and problem solving. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *22*, 792-810.
- 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., & Perlmutter, 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.
- Daily, L. Z., Lovett, M. C., & Reder, L. M. (in press). Modeling individual differences in working memory performance: A source activation account. *Cognitive Science*.
- 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.
- 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: Freeman.
- Gopher, D. (1982). A selective attention test as a predictor of success in flight training. *Human Factors*, *24*, 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.
- Hunt, E., Joslyn, S., & Sanquist, T. (1996). *Evaluating individual differences in response to emergency situations* (ONR Tech Report). Arlington, VA: Office of Naval Research.
- John, B., & Lallement, Y. (1997). Strategy use while learning to perform the Kanfer-Ackerman Air Traffic Controller Task. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the 19th annual conference of the Cognitive Science Society* (pp. 337-342). Mahwah, NJ: Erlbaum.
- 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.
- Kerkman, D. D., & Siegler, R. S. (1993). Individual differences and adaptive flexibility in lower-income children's strategy choices. *Learning & Individual Differences*, *5*, 113-136.
- 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: Vol. 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.
- Lee, F. J., & Anderson, J. R. (1997). Learning to act: Acquisition and optimization of procedural skill. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the 19th annual conference of the Cognitive Science Society* (pp. 418-423). Mahwah, NJ: Erlbaum.
- Lee, F. J., Anderson, J. R., & Matessa, M. P. (1995). Components of dynamic skill acquisition. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the 17th annual conference of the Cognitive Science Society* (pp. 506-511). Hillsdale, NJ: Erlbaum.
- Lemaire, P., & Reder, L. M. (1999). What affects strategy selection in arithmetic? The example of parity and five effects on product verification. *Memory & Cognition*, *27*(2), 364-382.
- Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, *124*, 83-97.
- 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., Reder, L. M., & Lebiere, C. (1996, June). *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. (1999). Modeling working memory in a unified architecture: An ACT-R perspective. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 135-182). New York: Cambridge University Press.
- Lovett, M. C., & Schunn, C. D. (1999). Task representations, strategy variability and base-rate neglect. *Journal of Experimental Psychology: General*, *128*, 107-130.
- 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.
- 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. (1988). Strategic control of retrieval strategies. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 227–259). San Diego, CA: Academic Press.
- Reder, L. M., & Schunn, C. D. (1999). Bringing together the psychometric and strategy worlds: Predicting adaptivity in a dynamic task. In D. Gopher & A. Koriat (Eds.), *Cognitive regulation of performance: Interaction of theory and application. Attention and Performance XVII* (pp. 315–342). Cambridge, MA: MIT Press.
- 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*, 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*, 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*, 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*, 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. In G. D. Cottrell (Ed.), *Proceedings of the 18th Annual Conference of the Cognitive Science Society* (pp. 25–26). Mahwah, NJ: Erlbaum.
- Schunn, C. D., & Lovett, M. L. (1996, November). *Representation and individual differences in base-rate neglect*. Poster presented at the 37th annual meeting of the Psychonomic Society, Chicago, IL.
- Schunn, C. D., & Reder, L. M. (1998). Individual differences in strategy adaptivity. In D. L. Medin (Ed.), *The psychology of learning and motivation* (pp. 115–154). York, PA: Academic Press.
- Shapira, N., & Kushnir, T. (1985). Aging and information seeking: Patterns in sampling of sucrose solutions. *International Journal of Aging & Human Development*, *20*, 103–112.
- Shebilske, W., Goettl, B., & Regian, J. W. (1999). Individual and group protocols for training complex skills in laboratory and applied settings. In D. Gopher & A. Koriat (Eds.), *Cognitive regulation of performance: Interaction of theory and application. Attention and Performance XVII* (pp. 401–426). Cambridge, MA: MIT Press.
- Shute, V. J. (1993). A macroadaptive approach to tutoring. *Journal of Artificial Intelligence in Education*, *4*, 61–93.
- Siegler, R. S. (1983). Five generalizations about cognitive development. *American Psychologist*, *38*, 263–277.
- Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.
- Siegler, R. S., & Lemaire, P. (1997). Older and younger adults' strategy choices in multiplication: Testing predictions of ASCM using the choice/no-choice method. *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.
- 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.
- 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.
- Wagner, D. A. (1978). Memories of Morocco: The influence of age, schooling and environment on memory. *Cognitive Psychology*, *10*, 1–28.

## Appendix A

### Ability Tests Used in Ackerman Studies 1 and 2

Test	Factor
Patterns	Perceptual Speed
Number Series	Reasoning
Circle Tapping	Movement Speed/Psychomotor
Name Comparison	Perception Speed
Object-Number	Memory
Analogies	Verbal, Reasoning
Scattered x's	Perceptual Speed
CA-2	Perceptual Speed
Letter Sets	Reasoning
Picture Number	Memory
Vocabulary	Verbal
Raven's Progressive Matrices	Verbal Reasoning
Word Beginnings	Verbal
Square Marking	Movement Speed/Psychomotor
Perceptual Speed	Perceptual Speed
First & Last Names	Memory
Figure Classification	Reasoning
Pursuit Aiming	Movement Speed/Psychomotor
Letter/Number Substitution II	Perceptual Speed
Simple RT	Psychomotor
2-Choice RT	Psychomotor
4-Choice RT	Psychomotor

## Appendix B

## Description of Ability Measures Used in Study 3

The verbal and spatial test of each factor were isomorphic, with differences only in the items presented on the screen (i.e., words vs. stick figures). Therefore, only descriptions of either the spatial or verbal test are included here.

**Working-Memory Capacity Verbal: Four-Term Ordering**

This test requires participants to produce the order of four items from three statements presenting partial information regarding the order of the items. The four items are divided into two categories of two items each. One statement describes the relative ordering of the two categories. The second and third statements describe the order of the items with each category. For example,

The ANIMALS come after the FURNITURE.  
The cow does not come before the bird.  
The chair does not come after the lamp.

After the three statements are presented, participants are given 15 s to select the correct order from eight numbered alternatives appearing on the screen. The correct response for this example is *chair lamp bird cow*. Participants are given feedback for incorrect responses and are allowed to review the three sentences and alternative answers. This test contains 24 items.

**Processing Speed Verbal: Two-Term Ordering**

Participants must decide as quickly as possible whether two presented words conform to the order specified in a sentence at the top of the screen. The sentences are presented first and either state that Word A will be before Word B or that Word A will be after word B (e.g., The bird comes before the cow). The two words are then presented in the middle of the screen. Participants are to respond as quickly as possible by typing *L* if the word order matches the sentence and *D* if it does not. This test contains 12 items.

**Fact Learning Spatial: Figures Recognition**

Participants are required to memorize 12 geometrical figures in a  $3 \times 4$  matrix and then to determine whether individually presented figures were in that matrix. Participants are given 60 s to study the figures. During a practice test, participants are given the hint to try to make associations with what the figures may resemble (e.g., a letter, a flag). Immediately after study, participants are shown individual figures and asked whether each was one studied. There are two sets of figure matrices, with 26 recognition test items per set.

**Procedural Learning Spatial: Reduction of Circles**

Participants are presented two circles that must be combined to form one circle, using the following rules:

**Rule 1**

If both circles are solid (i.e., each are either entirely white or entirely black), then the combined circle will keep the black parts of both circles. For example, if one circle is solid black and the other is solid white, then the result is a solid black circle.

**Rule 2**

If either circle is a mix, then the combined circle will keep the white parts of both circles. For example, if one circle is solid black and the other is black on the left half and white on the right half, the result is a circle with the right half white. Participants choose the answer from four numbered alternatives presented at the bottom of the screen. This contains 96 items.

**Inductive Reasoning Spatial****Figure Sets**

Participants are presented with three sets of figures. Two of the sets will be related according to various themes. Participants must determine which set is the odd set. Some of the various patterns include the following: figures formed with straight lines as opposed to curved lines, internal shading versus no shading, and so on. There are 10 items that must be solved within a 5-min period.

**Figure Series**

Participants are shown a series of shapes at the top of the screen and must choose the next shape occurring in the series from three numbered alternatives. For example, if the series was *\*/\_/\_*, the answer would be *\*\**. There are 10 problems that must be solved within a 5-min period.

**Figure Matrices**

Participants are shown a  $3 \times 3$  matrix in which a figure is contained in all but one of the cells. There are patterns or rules that apply across and down the figures that must be induced to decide what figure belongs in the empty cell. The matrix and eight alternatives responses are shown on the screen simultaneously. Some of the rules and patterns used are as follows: gradual shading of figures, successive additions or deletions to figures, rotation of figures, and so on. There are nine problems that must be solved within a 10-min period.

Received August 12, 1997

Revision received January 12, 2000

Accepted February 5, 2000 ■