Journal of Experimental Psychology: General
2008, Vol. 137, No. 1, 131–148

0096-3445/08/$12.00 DOI: 10.1037/0096-3445.137.1.131

Goal Neglect and Spearman's *g*: Competing Parts of a Complex Task

John Duncan, Alice Parr, Alexandra Woolgar, Russell Thompson, Peter Bright,
Sally Cox, Sonia Bishop, and Ian Nimmo-Smith
Medical Research Council Cognition and Brain Sciences Unit

In *g*oal neglect, a person ignores some task requirement though being able to describe it. Goal neglect is closely related to general intelligence or C. Spearman's (1904) *g* (J. Duncan, H. Emslie, P. Williams, R. Johnson, & C. Freer, 1996). The authors tested the role of task complexity in neglect and the hypothesis that different task components in some sense compete for attention. In contrast to many kinds of attentional limits, increasing the real-time demands of one task component does not promote neglect of another. Neither does neglect depend on preparation for different possible events in a block of trials. Instead, the key factor is complexity in the whole body of knowledge specified in task instructions. The authors suggest that as novel activity is constructed, relevant facts, rules, and requirements must be organized into a "task model." As this model increases in complexity, different task components compete for representation, and vulnerable components may be lost. Construction of effective task models is closely linked to *g*.

Keywords: goal neglect, intelligence, attention, working memory, frontal lobe

Supplemental materials: http://dx.doi.org/10.1037/0096-3445,137,1,131

In a previous article (Duncan, Emslie, Williams, Johnson, & Freer, 1996), we introduced the term *goal neglect* to describe a striking form of performance failure. In goal neglect, participants can say exactly what it is they should do, yet show no apparent attempt to do it. Such behavior is occasionally described in patients with major damage to the frontal lobes. For example, the patient might be asked to watch for a light and to raise one hand when it appears. When the light is switched on, the patient might say, "I should lift up my hand," yet make no attempt to do so (Luria, 1966). Similar failures have been described in performance of the Wisconsin Card-Sorting Test (Milner, 1963). The patient learns to categorize stimuli first by one rule (e.g., color), then by another (e.g., shape); sometimes, the patient continues with the first rule while explicitly stating that it is now wrong and should be abandoned. In our experiments (Duncan et al., 1996), goal neglect of this sort was also seen in people from the normal population. Notably, goal neglect was closely related to scores on a standard test of general intelligence or Spearman's *g* (Cattell, 1971; Spearman, 1904). The results suggested links among frontal lobe function, *g*, and the effective organization of goal-directed behavior.

The task used in our work is illustrated in Figure 1, Panel A. On each trial, a series of letter and number pairs was shown one after the other at the center of a computer screen. Each pair was presented for 200 ms, with a 200-ms blank interval between one pair and the next. The task was to watch just the characters on one side, left or right, and on that side, to ignore numbers but read letters aloud. Two cues told the participant which side to watch. At the start of the trial, there was a verbal instruction, WATCH LEFT or WATCH RIGHT, written in the center of the screen. In the display shown in Figure 1, Panel A, for example, the participant would see WATCH RIGHT and read the letters R, E, . . . aloud while ignoring A, B, ... Then, immediately before the last three pairs, a second side cue—this time it was either + or − symbol—also flashed for 200 ms in the center of the screen. Irrespective of the starting side, a + meant that for the final three pairs, the participant should watch right, whereas a − meant watch left. In the example, the participant would see +, stay on the right, and read the letters C and U aloud. In contrast, presentation of a − would have called for the participant to switch to the left and read the letters X and F aloud.

In this task, it was the +/− cue that was sometimes neglected. Usually, participants began each trial on the correct side but then continued to report letters from this same side no matter what cue was shown. Often, there was no pause or uncertainty in the response, as if cues had simply not occurred. Several things about this behavior are important. First, the rules were never explicitly forgotten; when asked again at the end of the task, participants were always able to repeat what the + and − called for them to do. Second, like the patients with frontal lobe damage described earlier, participants were actually quite capable of following the task instructions. If a failing participant was strongly cued to attend to the neglected task requirement—for example, by explicit error feedback from the experimenter after each trial—there was usually a rapid resolution to correct performance. The impression was that, as in the patients with frontal lobe damage, participants were perfectly capable of obeying the instruction but in some sense had not attempted to do so. This was confirmed by postexperiment questions, producing comments such as "I realize now that I

John Duncan, Alice Parr, Alexandra Woolgar, Russell Thompson, Peter Bright, Sally Cox, Sonia Bishop, and Ian Nimmo-Smith, Medical Research Council (MRC) Cognition and Brain Sciences Unit, Cambridge, England.

Correspondence concerning this article should be addressed to John Duncan, MRC Cognition and Brain Sciences Unit, 15 Chaucer Rd, Cambridge CB2 7EF, England. E-mail: john.duncan@mrc-cbu.cam.ac.uk

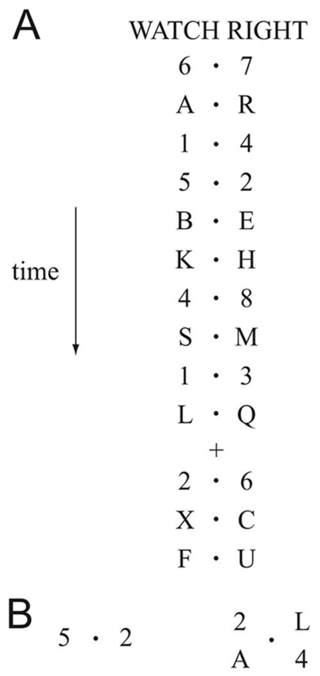

Figure 1. A. Sample stimuli from the work of Duncan et al. (1996). B. Sample stimulus frames for two-character (left) and four-character (right) tasks in Experiment 1. Stimuli are drawn approximately to scale.

haven't been looking out for the +s and −s," or "I've been letting those go over my head."

As noted above, a further important result was the close link between goal neglect in the letter-monitoring task and a standard measure of general intelligence. In psychometrics, the concept of general intelligence derives from the observation of ubiquitous positive correlations among different kinds of cognitive tests: to some extent, people who do well on one test also tend to do well on others (Spearman, 1904). To explain this result, Spearman (1904) proposed that some "general" or *g* factor contributes to success in all manner of different activities, both in the laboratory and everyday life. One possible interpretation is that *g* reflects the function of some particular cognitive system or process, important in effective behavior of many different kinds. If this interpretation is correct, it is easy to show which tests are the best measures of this common process. In factor analysis, they would be tests with the highest *g* loadings, reflected in the strongest average correlations with other tests, or strongest ability to predict overall success in different kinds of performance. Traditionally, the best tests involve novel problem solving based on spatial, verbal, or other materials. Even a single "fluid intelligence" test of this sort can give an excellent measure of *g*. Using one of these tests, Cattell's Culture Fair Test (Cattell, 1971; Institute for Personality and Ability Testing, 1973), we found that goal neglect was very well predicted by *g*. For people with *g* scores more than 1 standard deviation below the population mean, goal neglect was almost universal; for people scoring above the mean, it was almost absent. Such results suggest that the cognitive failure reflected in goal neglect may be something quite basic to Spearman's *g*—something closely related to the cognitive functions measured in standard tests of novel problem solving.

In this article, we investigate what this cognitive failure may be. Our work has several antecedents. It is frequently pointed out that, in some sense, more complex tasks have higher *g* correlations (e.g., Marshalek, Lohman, & Snow, 1983; Spilsbury, 1992; Stankov, 2000). Certainly the best tests of *g*, like the Culture Fair Test, have many different components that must be assembled for effective performance. Doubtless, complexity has many aspects, making it hard to operationalize and certainly hard to compare between tasks that are very different. One simple proposal, however, would be that different components of a complex task may compete for attention or representation. A related idea concerns working memory. Several working memory tasks—often involving some combination of processing and storage—show reasonably strong correlations with *g* (e.g., Kane et al., 2004; Kyllonen & Christal, 1990). An example is the reading span task (Daneman & Carpenter, 1980) in which the participant is required to read a series of sentences while at the same time storing the last word of each one for a subsequent memory test. Reasonably strong *g* correlations for such tasks have led to the proposal that *g* may reflect, at least in part, the capacity of working memory or its ability to focus on task-relevant information (Kane & Engel, 2002; Kyllonen & Christal, 1990). Again, a key idea is that *g* derives from limited capacity for concurrent processing. The same idea is reflected in studies of *g* and dual-task interference (e.g., Roberts, Beh, & Stankov, 1988; Spilsbury, 1992). Indeed, Spearman (1927) himself proposed that *g* might be seen as a limited mental energy driving different kinds of cognitive process.

In some sense, it seems obvious that competition must be a key factor in neglect. Surely, the +/− instruction would be followed if it were the only task requirement to be satisfied—that is, if the participant were not at the same time paying attention to other task requirements and aspects of performance. Indeed, in one experiment, Duncan et al. (1996) provided direct evidence for this idea while at the same time showing that a key factor in competition is the way that instructions are presented. In this experiment, a further requirement was added to the basic letter-monitoring task. Occasionally, while pairs of letters and digits flashed at the center of the screen, a dot was briefly presented on one side. Participants were to respond by pressing a key whenever a dot was detected. If this dot task was explained before the letter-monitoring task, omissions were rare and unrelated to *g*. If the dot task was explained after full instructions for letter monitoring, however, results were very different. In that case, in addition to neglect of the +/− cue, there was also frequent neglect of the dot. Such neglect took the form of a series of trials with no response, often followed by rapid resolution to good performance. Again, participants were capable of responding to the dot but often failed to do so for a series of trials; again, this neglect of one task component was closely related to *g*. The results suggest that task requirements specified early in instructions compete with those added later, with the result that late requirements, though explicitly remembered, can sometimes be disregarded in performance.

In the present experiments, we developed this idea of competition in goal neglect. Of course, there are many different kinds of dual-task interference (Allport, 1980), some very specific, such as difficulty in identifying two simultaneous stimuli in the same modality (Duncan, Martens, & Ward, 1997), some more general (Bourke, Duncan, & Nimmo-Smith, 1996; Kahneman, 1973). Similarly, the concept of working memory is used in many different ways in the psychological literature (e.g. Miller, Galanter, & Pribram, 1960). Commonly, working memory is taken to include a number of separate component systems, with each component contributing its own processing limitations (Baddeley, 1986). In our experiments, we asked what kind of attention or working memory limit underlies goal neglect.

Most conceptions of attentional or working memory limits are concerned with immediate processing demands at task execution. For example, experiments focus on difficulties in identifying multiple simultaneous stimuli (Broadbent, 1958), in making simultaneous choices of different responses (Pashler, 1994), or in holding more than a certain number of digits in short-term memory (Baddeley, 1986). In all these cases, processing difficulties resolve if stimuli are separated by more than a few 100 ms (Raymond, Shapiro, & Arnell, 1992), if responses can be selected in turn (Pashler, 1994), or if one set of digits can be reported before another set is stored. In Experiment 1, we tested the hypothesis that neglect derives from real-time task-processing demands during task execution. A manipulation of attentional demand for letter monitoring, however, left neglect of the +/− cue unaffected.

A second theme in the attentional literature is difficulty in keeping two different tasks adequately prepared (e.g. Pashler, 1994; Rogers & Monsell, 1995). For example, a block of trials requiring alternation between different tasks may be harder than a pure block of one task or the other. In Experiment 2, we tested and rejected the hypothesis that neglect derives from a requirement to keep different task components simultaneously prepared.

In Experiment 3, we turned to a different hypothesis. Often, a more complex task simply has a longer list of relevant facts, rules, or requirements. In Experiment 3, we showed that this aspect of complexity—the total complexity of task instructions, irrespective of complexity during actual performance—strongly influences goal neglect. We suggest that as new behavior is constructed, relevant facts, rules, and requirements must be organized into an effective control structure or mental program. We call this program a *task model*. As the complexity of the task model increases, there is increasing chance that one component will be lost. In Experiment 4, we confirmed this conclusion in a new goal neglect task.

As we have said, an original motivation for our interest in goal neglect was its similarity to clinical observations in patients with frontal lobe damage. In the present experiments, we were not directly concerned with the neural basis for goal neglect. We return to this subject in the General Discussion.

Experiment 1

In Experiment 1, we began with a simple idea. In the task used by Duncan et al. (1996), the participant watched a continuous stream of letters and digits, repeating every letter as it appeared. At the same time, he or she had to be ready for the + or − cue. Intuitively, it would seem that the demands of watching for letters might cause the cue to be missed. To test this hypothesis, we asked whether neglect of the side cue would be sensitive to the difficulty of search for letters.

Of course, this kind of idea has precedent in many dual-task experiments. Often, increasing the demands of one task interferes with others that are occurring simultaneously or at least close in time (e.g. Bourke, 1997; Kahneman, 1973; Kalsbeek & Sykes, 1967). We asked whether this kind of attentional competition is important in goal neglect.

Attentional demands of visual search are traditionally altered by varying the number of stimuli to be searched (Schneider & Shiffrin, 1977). Following this approach, we used two task variants. In the basic version (Figure 1, Panel B, left), each 200-ms frame contained just two characters, one to the left and one to the right. In the harder version (Figure 1, Panel B, right), each frame contained four characters, two (one above the other) on each side. Either all four characters were digits, or on each side, one was a digit and one was a letter. Again participants watched just for letters on one side, repeating them as they appeared. Now, however, the participants had twice as many characters to monitor for these letter targets.

Method

Participants. In the letter-monitoring task, goal neglect is largely restricted to people in the lower part of the fluid intelligence range. Thus, neglect experiments require a method for selecting participants that ensures good sampling from this range. In Experiment 1, we obtained such a sample by avoiding selection of younger participants. Fluid intelligence scores decline rapidly with age (Cattell, 1971), whereas the function relating goal neglect to fluid intelligence remains largely unchanged (Duncan et al., 1996).

We tested 56 (30 male, 26 female) participants between the ages of 40 and 69 years ($M = 58$, $SD = 8$). Participants were recruited from the volunteer panel of the Medical Research Council (MRC) Cognition and Brain Sciences Unit and were paid to participate.

Letter-monitoring task. The letter-monitoring task was presented on an Macintosh (Apple, Cupertino, CA) computer running Psyscope software (Cohen, MacWhinney, Flatt, & Provost, 1993). Participants sat comfortably, without precise control of viewing distance; calculations of visual angle were based on an approximate distance of 57 cm.

In the basic task version (two-character task), the events of each trial were modeled closely on the task used by Duncan et al. (1996; see Figure 1, Panel A). All letters, digits, and symbols were presented in Geneva Bold font (for letters, all upper case), at a character height of 0.50 degrees (deg). On a key press from the experimenter, the trial began with the instruction WATCH LEFT or WATCH RIGHT presented for 1 s at the center of the screen. There was a further 1-s delay, after which the stimulus sequence began. This sequence consisted of a series of stimulus frames, each presented for 200 ms, with a blank interval of 200 ms between one frame and the next. In each frame, there was a single pair of letters or digits, one centered 0.85 deg to the left of a dot marking the middle of the display, the other 0.85 deg to the right. After 10 such frames, the + or − cue was flashed at screen center, followed by three final stimulus frames. To maintain the flow of stimulus presentation, the +/− cue was also shown for 200 ms and was separated by 200-ms blank intervals from the stimulus frames immediately before and after.

The two characters in each stimulus frame were either both letters or both digits (Figure 1, Panel A). In the first part of each trial, there were five letter pairs and five digit pairs, randomly ordered. In the second part, following the +/− cue, there was always a single digit pair followed by two letter pairs. Digits were selected randomly with replacement from the set 1–8. Letters for each trial were selected randomly but without replacement from the set of all letters excluding D, I, O, V, and W.

The participants' task was to read aloud all letters from the attended side at their own speed. The experimenter wrote responses onto a prepared score sheet. For the first part of the trial, the attended side was indicated by the verbal instruction WATCH LEFT or WATCH RIGHT. For the final part of the trial, the attended side was indicated by the + or − cue, + indicating attend right and − indicating attend left. Thus, a perfectly correct report for one trial consisted of seven letters, all different—five from the first part of the trial and two from the final part.

In the alternative task version (four-character task), the only change concerned the number of characters in each stimulus frame. In this task, each frame contained four characters, two to the left and two to the right (Figure 1, Panel B, right). The two characters on each side appeared one above the other, separated by 1 deg center-to-center. Matching the easy task version, one pair of characters was centered 0.85 deg to the left of display center, the other 0.85 deg to the right. Now each frame consisted either of four digits or of one letter and one digit on each side. Letters on each side were presented randomly in the upper or lower location, independently for the two sides. As before, there were five all-digit frames and five letter-plus-digit frames in the first part of each trial and one all-digit frame followed by two letter-plus-digit frames in the final part. Again, accordingly, a perfect report consisted of five letters from the first part of the trial, followed by two from the final part.

Verbatim instructions for the task are given in the Supplemental Materials online. To encourage memory of the +/− rule, we placed two pieces of paper with + and − cues written on them on the table in front of the participant, + to the right and − to the left. Instructions were followed by one (33 participants) or two (23 participants) practice trials, the second trial being given if the participant failed to report more than one letter on the first trial. After this practice, the participant was asked to restate what should be done on seeing a + and a −. Only 1 participant answered incorrectly; after being given one further practice trial, this person described the rule correctly.

Data were then collected over a block of 12 trials, with the two-character task used for half the participants and the four-character task for the remainder. In each successive group of four trials, there was one with WATCH LEFT followed by −, one with WATCH LEFT followed by +, one with WATCH RIGHT followed by −, and one with WATCH RIGHT followed by +, in random order. When the first block of 12 trials was complete, each participant switched to the other task version for a second block. This design was intended to permit within-participant comparisons, but these are not reported here (though they would support the same conclusions). As documented later (see brief discussion of Block 2 data following Experiment 3), we had strong reason to suspect carryover from one part of the task to another, reducing the power of within-participant comparisons. In the following, accordingly, we present just Block 1 data and thus a between-participants comparison of task versions. Following Block 2, each participant was asked for a final description of the +/− rule. All such descriptions were correct.

Culture Fair Test. After finishing the letter-monitoring task, participants completed a standard test of *g*, Cattell's Culture Fair Test, Scale 2, Form A (Cattell, 1971; Institute for Personality and Ability Testing, 1973). The test assesses novel problem solving with geometrical figures. There are four timed subtests (series completions, odd-one-out, matrices, and topology). The manual was used to convert scores to IQs ($M = 100$, $SD = 16$, in a reference population of young adults). Note that Culture Fair IQs are absolute rather than age-corrected; a score of 84, for example, corresponds to 1 standard deviation below the young adult mean.

Results

To assess performance in the letter-monitoring task (Block 1 data only; as explained earlier), we used two scores. The first, intended as a basic measure of identifying and reporting letters, was based on the first part of each trial, before the +/− cue was shown. Five letters were presented on the cued side in this part of each trial; the score was simply the proportion of these letters that appeared (in any position) in the participant's report. (Note that attention to the wrong side, which in principle affected this score, was in practice extremely rare in the first part of a trial, occurring only three times in the whole experiment.) As assessed by this score, our manipulation of task difficulty was highly effective. In the two-character version of the task, the proportion of letters correctly reported was 0.97. In the four-character version, this proportion fell to 0.87. Analysis of variance (ANOVA) showed the difference to be strongly significant, $F(1, 54) = 23.4$, $p < .001$, $d = 1.29$.

The second score reflects the extent to which a participant's report was influenced by the +/− cue. Hence, it concerns the probability that letters reported from the final part of a trial (last two stimulus frames) came from the incorrect side. For each trial, we gave a score of 1 if more letters were reported from the incorrect than from the correct side, 0 if more letters were reported from the correct than from the incorrect side, and 0.5 if equal numbers were reported from the two sides. Averaging across the 12 trials produced a mean side error score ranging in principle from 1 (incorrect side attended on every trial) to 0 (correct side attended on every trial). In practice, any strategy that is insensitive to the +/− cue (most commonly, always remaining on the same side throughout each trial, though other incorrect strategies do occasionally appear) gives a score of 0.5, reflecting equal reporting from correct and incorrect sides. Scores thus range from approximately 0.5 (chance) down to 0 (perfect performance). Across the 12 trials, a closely similar score would be the number of letters reported from the incorrect side, divided by the total number of letters reported from either side. Our score, however, gives equal weight to all trials, even those on which participants managed to report no letters from either side.

Figure 2 shows scatterplots relating side error scores to Culture Fair IQs. The first important point concerns the distribution of side error scores. For most participants (36/56), the score was at or below .167 (the score for attention to the wrong side on 2/12 trials). For these participants, accordingly, there was good use of the +/− cue. For most of the remainder (11/56), however, the score was at or above .417 (the score for attention to the wrong side on 5/12 trials). For 10 of these 11 participants, there was the usual stereotyped strategy of continuing to report letters from the initially attended side, no matter what cue was presented. Again,

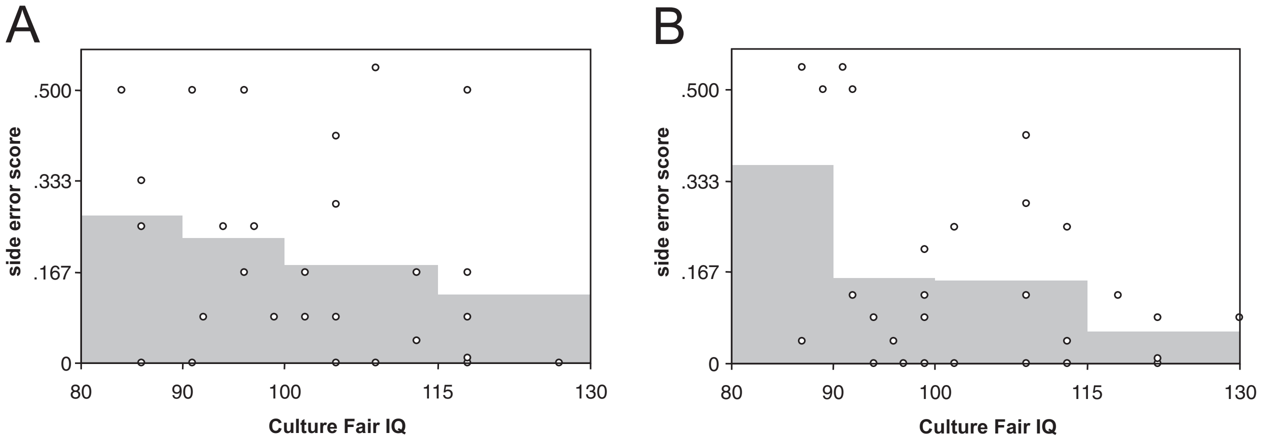

Figure 2. Experiment 1: scatterplots relating side error score to Culture Fair IQ, separately for (A) two-character and (B) four-character participants. Each point represents a single participant; participants with IQs $<$ 80 or $>$ 130 are plotted on the corresponding margin. Histograms (gray bars) show mean side error scores for participants with IQs $\leq$ 90, 91–100, 101–115, and $>$ 115.

therefore, many participants showed the typical pattern of goal neglect in this task.

The second point concerns relations between goal neglect and *g*. In the 28 participants given the two-character version of the task, the correlation (*r*) between side error score and Culture Fair IQ was −.28, lower than correlations in our previous work (Duncan et al., 1996). A similar, rather weak correlation was seen in the 28 participants given the four-character task, $r = -.35$. Later, we show substantially stronger correlations in more complex versions of the task.

In this experiment, however, the most important question concerned the relationship between goal neglect and letter-monitoring demand. Here, the result was clear-cut, with no evidence of stronger goal neglect in the four-character task. To compare the two participant groups, we used analysis of covariance (ANCOVA) to correct for the influence of IQ. Different slopes of the function relating dependent variable (side error score) to covariate (IQ) were allowed in the two groups. When slopes are allowed to differ, the significance of the group effect must be separately assessed at different values of the covariate (Maxwell & Delaney, 1990). The test suggested no difference between two-character and four-character tasks, separately assessed at IQs of 90, $t(52) = -0.4$, $d = 0$; 100, $t(52) = -0.6$, $d = -0.2$; and 110, $t(52) = -0.5$, $d = -0.2$ (negative sign indicating numerically greater error in the two-character task).

Discussion

In Experiment 1, we substantially altered the demand of the basic letter-monitoring task. In the first part of each trial, letter detection was close to ceiling in the two-character version of the task but substantially more difficult in the four-character version. This change of one task component, however, left another component—neglect of the +/− cue—unaffected. There was no suggestion that during task execution, side cues were missed because of the demands of letter monitoring.

Experiment 1 bears some resemblance to traditional studies of the attentional blink (Raymond, Shapiro, & Arnell, 1992). In attentional blink studies, processing of one visual event interferes with detection or identification of a second event, up to temporal separations of 500 ms or more. It might be supposed that in our task, processing of the +/− cue might have received blinklike interference from processing of the immediately preceding stimulus frame. In blink studies, however, the duration and depth of interference increase with the number of objects to be identified in the first display (Ward, Duncan, & Shapiro, 1996). In Experiment 1, equal rates of neglect for two- and four-character tasks suggest that the failure to respond to +/− cues did not derive from a typical blinklike limit.

Of course, limited conclusions are possible from a single negative result, and it is always possible that neglect would prove sensitive to some other manipulation of real-time letter-monitoring demand. Meanwhile, the suggestion is that neglect may not be sensitive to immediate processing demands during actual task execution. How else might task components compete? In the next two experiments, we consider competition over broader time scales. In Experiment 2, we addressed the number of task components to be borne in mind over a whole block of trials.

Experiment 2

Many dual-task experiments concern interference across a time scale of seconds or less. The expectation is that if events are well separated in time, interference should vanish as attention is switched from one task to the other. At the same time, it is often noted that cognitive activities can also compete in a different way. Even if participants must not actually perform two tasks at the same time, it may be hard to keep them both simultaneously prepared (e.g. Pashler, 1994; Rogers & Monsell, 1995). A block of trials requiring alternation between two different tasks may be harder than a pure block of one task or the other.

In Experiment 2, we asked whether this kind of competition might be important in goal neglect. Materials for the new task are illustrated in Figure 3. Again, pairs of characters were presented in the center of a screen at the same rate as before. This time, each trial was divided into three successive segments, separated by pairs

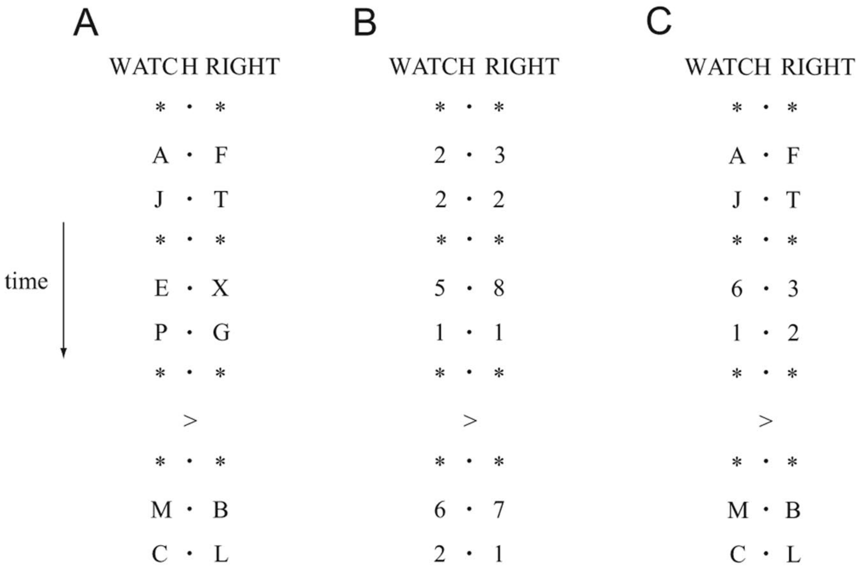

Figure 3. Experiment 2: sample stimuli for (A) pure letter trial, (B) pure number trial, and (C) mixed-task trial.

of asterisks. Within each segment, there were either two pairs of letters or two pairs of numbers. For letters, the task as before was to repeat those on the attended side. For numbers, the task was to add the two on the attended side and to state the result. After the first two segments and a further pair of asterisks, there was the second side cue, then one more segment to complete the trial. Thus for the mixed task trial in Figure 3, Panel C, the correct response would be "F, T, 5 B, L." Possible trial types were pure letter (Figure 3, Panel A), pure number (Figure 3, Panel B), and mixed task (Figure 3, Panel C).

To manipulate block complexity, we compared two groups of participants. For one group, there was a single block of 16 mixed-task trials. For the second group, trials were divided into two sets of 8, one a pure block of letter trials, the other a pure block of number trials. Extending the logic of Experiment 1, we asked whether this manipulation of block complexity would affect attention to side cues.

In previous experiments, neglected side cues have always been symbolic: + for right and − for left. It has often been suggested to us that neglect might somehow depend on this use of a novel and arbitrary stimulus–response mapping. In unpublished experiments, however, we have found that occurrence of neglect is very similar when the cue is a central arrow pointing directly to the side to be attended. To confirm this result, we used arrow cues in Experiments 2 and 3.

Method

Participants. In Experiment 2, participants were recruited from two sources: 32 (23 male and 9 female; age range, 16–25 years, $M = 18$, $SD = 2$) were recruited from an institute of higher education offering courses for a wide range of academic abilities and 24 (8 male and 16 female; age range, 40–68 years, $M = 56$, $SD = 8$) were recruited as before from the volunteer panel of the MRC Cognition and Brain Sciences Unit. Because participants from these two sources gave closely similar data, they were combined. All participants were paid to take part.

Letter and number task. In most ways, the letter and number task was very similar to the letter-monitoring task from Experiment 1. As described earlier, each trial was divided into three segments, separated by asterisks (Figure 3). For letter segments, the task was simply to repeat the two letters from the attended side. For number segments, the task was to add the two attended numbers together and state the result. Again, letters for each trial were drawn randomly and without replacement from the same set as before. Within each number segment, the first number on each side was drawn randomly from the set of 2–8, whereas the second was randomly either 1 or 2, minimizing the difficulty of addition. Again, each trial was initiated by a key press from the experimenter. This time, the initial instruction WATCH LEFT or WATCH RIGHT was shown for only 200 ms, with a 200-ms blank interval before the first pair of asterisks. In other respects, timing was as before, with each pair of asterisks, letters, or numbers, as well as the side instruction > or <, shown for 200 ms, with blank intervals of 200 ms between each display and the next. Arrows were drawn in the same font and size as other characters and were placed at the center of the display.

Verbatim instructions are given in the Supplemental Materials online. For all participants, the initial instructions were the same, describing the letter task, the number task, and the rules concerning which side to watch. There followed one (31 participants), two (21 participants), or three (4 participants) mixed-task practice trials (cf. Figure 3, Panel C), practice terminating after the first trial on which any response was given. After this, all participants were asked to repeat the rules describing what to do for letters, numbers, and arrows; all did so correctly.

For 28 participants, there followed a block of 16 mixed-task trials. For the other 28 participants, there were instead two sub-blocks of 8 pure letter trials and 8 pure number trials. For 14 of these participants, the letter trials were completed first. When these participants had completed their practice trial(s) and had correctly repeated the task rules, they were told that the first trial block would contain no numbers. They could temporarily disregard this task and concentrate only on letters. Then after the letter subblock was complete, they were told that the next set of trials would contain only numbers. For the other 14 pure-task participants, the order of letter and number tasks was reversed.

As in Experiment 1, mixed-task participants went on also to complete pure tasks, whereas pure-task participants went on to complete the mixed task, but again, these additional data added little and are only briefly discussed after the results of Experiment 3. At the end of the task, all participants again were able to describe the correct response to arrow cues.

Culture Fair Test. After completing the letter and number task, participants went on to complete the Culture Fair Scale 2, Form A.

Results

As before, we used separate scores for the first part of each trial, before the arrow cue, and for the final part. For the first part of the trial, we calculated the probability that a letter presented on the cued side would appear in the report and the probability that a number pair presented on the cued side would be correctly added and reported. The mean of these two probabilities gives a basic measure of letter and number task performance. ANOVA suggested that there was no difference between mixed-task (mean score = .90) and pure-task (mean score = .88) participants, $F(1, 54) = 0.1$, $d = -0.1$. Inspection of pure-task data showed that the letter task (mean score = .95) was somewhat easier than the number task (mean score = .82).

To assess the use of arrow cues, we calculated side error scores for the final part of each trial as before. Scatterplots relating side error scores to Culture Fair IQs are shown in Figure 4. As in Experiment 1, most participants (43/56) showed good use of arrow cues, with a side error score at or below .188 (the score for attention to the wrong side on 3/16 trials). For most of the remaining participants (10/56), the score was above .375 (the score for attention to the wrong side on 6/16 trials). Again, therefore, there was significant goal neglect, especially in the lower part of the *g* distribution. Correlations between side error score and Culture Fair IQ were substantially higher than those in Experiment 1: $r = -.50$ for pure-task participants and $r = -.66$ for mixed-task participants. This time, strategies in neglecting participants were mixed between continuing to respond to the initially attended side and simultaneous response to both sides.

In Experiment 2, the most important question concerned the effect of block complexity on goal neglect. The data in Figure 4 suggest little difference between pure and mixed tasks. ANCOVA as before showed no significant task difference at IQs of 90, $t(52) = 0.4$, $d = 0.1$; 100, $t(52) = 0$, $d = 0$; or 110, $t(52) = -0.6$, $d = -0.2$, (negative sign indicating numerically greater error in pure tasks).

Discussion

Once more, Experiment 2 showed frequent neglect of side cues among lower *g* participants. In this version of the task, correlations between side error score and Culture Fair IQ were substantially stronger than those obtained in Experiment 1. The results also showed that neglect can occur even when side cues do not require a novel, arbitrary stimulus–response association. Even an arrow directly pointing to the required side is often neglected under the conditions of these tasks.

The results give no support, however, to the idea that neglect might arise through a need to maintain active readiness or preparation for multiple task components. In Experiment 2, the two groups of participants had very different trial blocks, half the participants never knowing whether the next task segment would contain letters or numbers, the remainder having pure letter or number tasks for a whole block of eight trials. These two conditions, however, led to similar levels of side cue neglect. Again, we should be cautious in drawing strong conclusions from a negative result. As in Experiment 1, however, a standard manipulation of

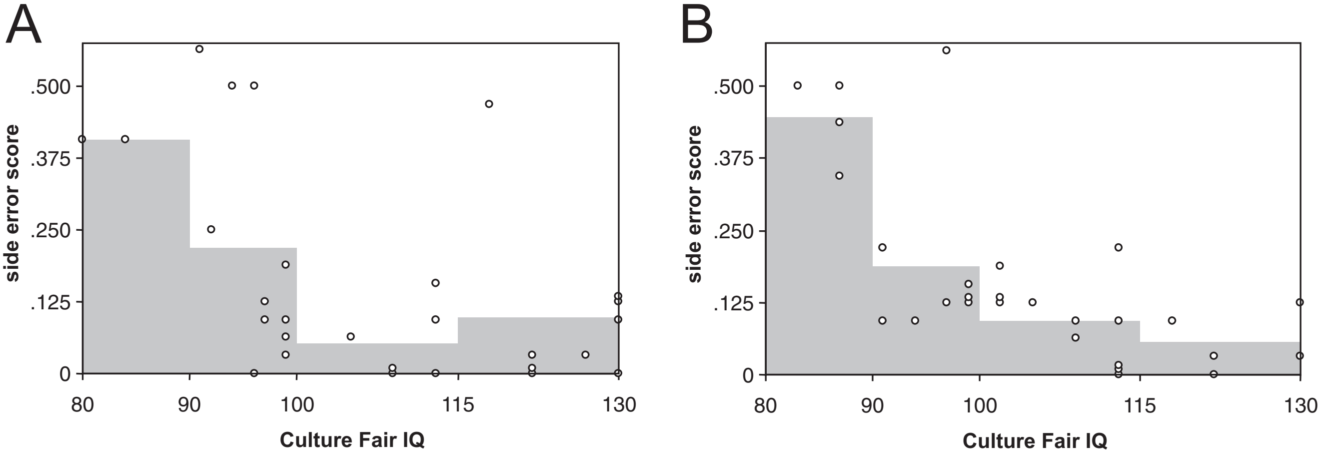

Figure 4. Experiment 2: scatterplots relating side error score to Culture Fair IQ, separately for (A) pure-task and (B) mixed-task participants. Each point represents a single participant; participants with IQs < 80 or > 130 are plotted on the corresponding margin. Histograms (gray bars) show mean side error scores for participants with IQs ≤ 90, 91–100, 101–115, and > 115.

task complexity—complexity of task preparation in a whole trial block—had little apparent bearing on the goal neglect phenomenon.

Experiment 3

In Experiment 3, we turned to a new perspective. If neither trial complexity nor block complexity determines the level of goal neglect, could the crucial events occur instead during instructions—that is, as the participant's body of knowledge bearing on the task is established? According to this hypothesis, the key aspect of task complexity does not concern what is actually done. Instead, it concerns the whole list of task-relevant facts, rules, and requirements specified at task onset.

Often, the complexity of task-relevant knowledge and the complexity of performance are confounded. A task with more components usually requires more complex instruction and then more complex execution. In Experiment 3, we designed a new kind of experimental comparison that unconfounds these factors. This time, all participants experienced pure blocks of letter and number trials, identical to those of pure-task participants in Experiment 2. These blocks, however, were preceded by different instructions. Half the participants ("full instructions") were instructed just like the participants of Experiment 2. They received a description of both letter and number tasks and had one or more mixed-task practice trials; then they received an additional instruction that tasks would be blocked so that either letters (half of this group) or numbers (the other half) could be temporarily disregarded. In the other half of the participants ("reduced instructions"), the task (letter or number) to be performed in the second trial block was never mentioned until the first block was completed. In terms of the tasks actually performed, full-instructions and reduced-instructions procedures were identical. Thus, we manipulated not the complexity of actual task performance, but the number of task components specified in the initial task description, in particular before mentioning the arrow cues and their meaning.

Method

Participants. Experiment 3 had 48 participants (11 male, 37 female; age range, 40–70 years, $M = 56$, $SD = 9$); 24 received full instructions and 24 received reduced instructions. Participants were recruited from the paid volunteer panel of the MRC Cognition and Brain Sciences Unit.

Letter and number task. Verbatim instructions for reduced-instructions participants are given in the Supplemental Materials online. For these participants, the first task to be performed (letter for 12 participants, number for the remainder) was described at the beginning of the experiment, with no mention of the other. After the eight trials of this task, the other task was described and performed. For full-instructions participants, the procedure was exactly the same except that as for the pure-task participants of Experiment 2, both tasks were explained at the beginning of the experiment.

Again, initial instructions were followed by one (28 participants), two (19 participants), or three (1 participant) practice trials, practice terminating as soon as any response was given. For reduced-instructions participants, practice trials used only the materials of the first task to be performed. For full-instructions participants, as in Experiment 2, practice trials were mixed task. All participants described task rules correctly after these practice trials and again at the end of the experiment. In all other respects, procedures were exactly as those for the pure-task participants in Experiment 2. As before, Culture Fair Scale 2, Form A, was administered.

Results

Scored as in Experiment 2, the proportion correct in the first part of each trial was .93 for full-instructions participants and .96 for reduced-instructions participants. The difference was not close to significance, $F(1, 46) = 0.7$, $d = 0.2$. This time, letter and number trials produced similar scores, .96 and .93, respectively, averaged over the two participant groups.

Scatterplots relating side error score to Culture Fair IQ are shown in Figure 5. Again, most participants (29/48) showed good use of arrow cues, with a side error score at or below .188 (the score for attention to the wrong side on 3/16 trials). For most of the remaining participants (13/48), neglect of arrows was indicated by a score at or above .375 (the score for attention to the wrong side on 6/16 trials). Twelve of these 13 neglecting participants showed the usual pattern of continuing to report from the same side, whichever arrow appeared.

The key result concerns the comparison of full- and reduced-instructions groups. For the full-instructions group, results were very similar to those obtained in Experiment 2 (Figure 4). In particular, neglect was common among participants with Culture Fair IQ < 100. For the reduced-instructions group, however, neglect was substantially less common. ANCOVA showed significant differences between full- and reduced-instructions groups, separately assessed at IQs of 90, $t(44) = 2.7$, $p < .005$, $d = 0.8$; 100, $t(44) = 3.1$, $p < .005$, $d = 0.9$; and 110, $t(44) = 1.8$, $p < .05$, $d = 0.5$ (all tests, one-tailed).

In the full-instructions group, the correlation between side error score and Culture Fair IQ was $-.54$, similar to results in Experiment 2. In the reduced-instructions group, even participants in the lower part of the IQ range often showed good use of the arrow cue. Accordingly, the correlation of side error score with IQ was reduced to $-.35$.

We conducted a further set of ANCOVAs to compare results from Experiment 3 with those from Experiments 1 and 2. At each IQ tested, the effect of instruction in Experiment 3 was significantly stronger than the effect of two versus four characters in Experiment 1: $t(96) = 2.7$, $p < .01$, $d = 0.5$ at IQ = 90; $t(96) = 3.1$, $p < .005$, $d = 0.6$ at IQ = 100; and $t(96) = 2.0$, $p < .05$, $d = 0.4$ at IQ = 110 (all tests, two-tailed). Similarly, the effect of instruction in Experiment 3 was significantly stronger than the effect of block complexity in Experiment 2: $t(96) = 2.4$, $p < .05$, $d = 0.5$ at IQ = 90; $t(96) = 2.9$, $p < .005$, $d = 0.6$ at IQ = 100; and $t(96) = 2.1$, $p < .05$, $d = 0.4$ at IQ = 110 (all tests, two-tailed). For tests at IQs 90 and 100, this latter difference remained significant ($p < .05$), even when we discarded data from the younger participants from Experiment 2. If we assume an effect size similar to that of Experiment 3, the power to detect a similar effect in Experiments 1 and 2 would have been approximately 0.9.

For all three experiments, we have confined ourselves to between-participants comparisons based on Block 1 data. In Block 2 of Experiments 1 and 2, each participant went on to the alternative task version (two- or four-character task in Experiment 1;

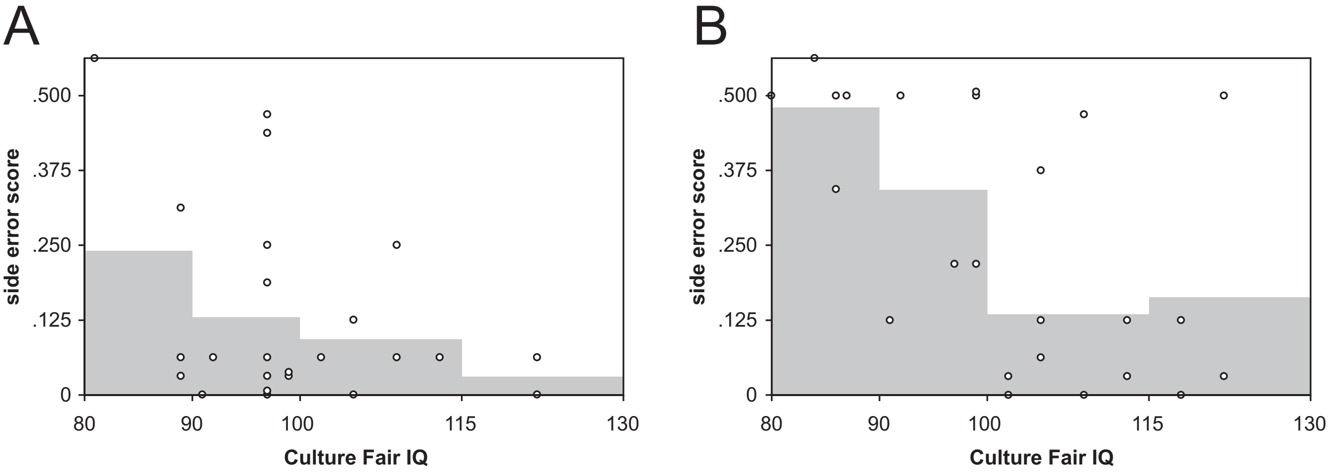

Figure 5. Experiment 3: scatterplots relating side error score to Culture Fair IQ, separately for (A) reduced-instructions and (B) full-instructions participants. Each point represents a single participant; participants with IQs < 80 or > 130 are plotted on the corresponding margin. Histograms (gray bars) show mean side error scores for participants with IQs ≤ 90, 91–100, 101–115, and > 115.

pure- or mixed-task in Experiment 2); in Experiment 3, all participants experienced pure tasks in Block 1, moving on to mixed tasks in Block 2. In each experiment, side error scores showed a small improvement in Block 2 (mean gain over Block 1 of .06 in Experiment 1, .03 in Experiment 2, and .04 in Experiment 3). Otherwise, however, each participant's performance in Block 2 was closely related to that in Block 1 (across participants, $r = .73$ in Experiment 1, .83 in Experiment 2, and .87 in Experiment 3). Across all experiments, only 4 participants changed from a neglect score (Experiment 1, error score ≥ 5/12; Experiments 2 and 3, error score ≥ 6/16) in Block 1 to good use of side cues (Experiment 1, error score ≤ 2/12; Experiments 2 and 3, error score ≤ 3/16) in Block 2. No participant showed the opposite transition, moving from good performance in Block 1 to neglect in Block 2. The results show that for the great majority of participants, the performance pattern established in Block 1 continued with only minor variation into Block 2. Again, this is consistent with key events occurring at the time of initial task specification.

In principle, our design would have allowed a within-participants comparison of two- and four-character tasks (Experiment 1) or of pure and mixed tasks (Experiments 2 and 3). In line with our between-participants comparisons, such analyses suggested no significant difference for either comparison ($p > .15$ in each case). Even had conditions truly differed, however, a strong tendency to carry the performance pattern from Block 1 over into Block 2 (e.g., improbability of neglecting side cues after first responding to them) could have obscured that difference. It is for this reason that we have rested our conclusions on Block 1 data alone.

Discussion

In Experiment 3, we finally confirmed the importance of task complexity. In line with our initial conjecture, the results show that neglect can become more frequent as the number of task components is increased. In some sense, accordingly, task components compete for attention, with a vulnerable component lost as the number of competitors increases. Together, however, the results of Experiments 1–3 suggest a kind of attentional limit that is rather different from many such limits in the literature.

Perhaps most surprisingly, the data suggest that neglect of side cues does not arise through immediate, real-time competition with monitoring for letters. Subjectively, it is not that participants are "busy" with the letter task and cannot at the same time look out for side cues. In Experiment 1, we found that neglect was insensitive to a strong manipulation of letter-monitoring demand. Specifically, doubling the number of characters to examine had no effect on neglect frequency, though it did substantially change the probability of finding letters themselves. Even more tellingly, Experiment 3 showed that neglect is changed by an added task instruction—even though this instruction has no impact on what participants must actually do at the time of performance.

Similarly, we find no evidence that neglect is sensitive to the number of things that a participant must be ready for at any given time. In Experiment 2, telling participants to disregard one task requirement for a whole block of trials had no impact on neglect frequency. In contrast, Experiment 3 showed very different levels of neglect in identical trial blocks, preceded by different descriptions of total task requirements.

Apparently, the level of neglect is not determined by processing demands during actual task execution. Instead, it depends on the total complexity of task-relevant knowledge and rules, as specified at the time of task instructions. Neglect is promoted by increased complexity in this body of knowledge, whether that knowledge is actually used or not. In the next section, we develop the implications of these results.

A competitive task model in novel behavior. The results we have described confirm that goal neglect reflects a kind of limit in attentional or working memory capacity. This limit, however, is something rather different from the real-time limits of traditional attention and working memory experiments. Specifically, the results suggest a limit in constructing and maintaining what we call a *task model*—a working-memory description of relevant facts, rules, and requirements used to control current behavior. It is this attentional limit, we propose, that underlies goal neglect and is closely related to *g*.

There are several antecedents to the idea that novel skills must be shaped by some explicit list of relevant facts and rules. Fitts and Posner (1967), for example, described the "cognitive phase" (p. 11) at the start of learning a new skill; in Anderson's (1983) work, this is called the declarative phase. More generally, in problem-solving programs like Anderson's (1983) ACT* (adaptive character of thought) or Newell's (1990) SOAR (state, operator, and result), a central working memory builds up an accumulating model of knowledge bearing on a current task—knowledge from the senses and from long-term memory. Typically, this model includes both the current state of relevant aspects of the world, goal states to be achieved, and candidate methods for achieving them. The contents of this model represent the active body of knowledge that bears on current decision making.

Our data suggest that goal neglect arises through competition in a novel task model of this sort. As a new task is attempted or new instructions are received, a new task model must be constructed. This model should incorporate all the knowledge that bears on current activity, organized into suitable chunks or subroutines of a mental program. These chunks must somehow be established in a form that can later be used when suitable trigger conditions occur. In our tasks, for example, letters must trigger the requirement to repeat them; numbers, the requirement to add them; and arrows, the requirement to direct attention to the correct side. A more everyday example is the process of reading the rules for a new card game. Our data suggest that as the task model is elaborated and more components are added, so the representation of each part is weakened. Again, this risk of overload is familiar when one is learning rules for a new game. In neglect, a vulnerable task component is lost and exerts no control over subsequent behavior. As participants sometimes report (Duncan et al., 1996), the importance of the side cue "slips their minds." For neglect, accordingly, the key aspect of task complexity is the complexity of all that information to be organized into a useful task model.

The task model must be distinguished from simple declarative memory for task requirements. As we have seen, participants always remember how side cues should be used and can describe this rule at the end of the task. In the representation that actually controls task performance, however, this information is lost. Subjectively, the relevant information does not come to mind when it is required.

Evidently, the task model is a form of working memory. How does it relate to other kinds of working memory limit? One popular conception, for example, combines a central executive with several lower level slave systems, each storing different kinds of information (Baddeley, 1986). Others equate working memory with immediate awareness for a few chunks of information (Cowan, 2000). In several respects, models such as these address processing limits that seem rather different from the one we are considering here.

One obvious difference concerns capacity. Traditional working memory experiments focus on the ability to remember just a few stimulus items, for example, typical spans of six or seven digits or five or six spatial locations. Correspondingly, traditional models of working memory have a limit of a few chunks of information. Although no method exists to compare chunks across different domains, it seems clear that the whole contents of more than a minute's task instructions—requiring more than a page to write down (see Supplemental Materials online)—have far greater storage demand than the few items of typical working memory lists. For our tasks, for example, the contents of the task model would include a description of relevant stimuli (characters on a computer screen, with their different locations), responses (spoken letters and numbers), different time periods of each trial (initial side instruction and the like), timing requirements on performance (speeded stimulus presentation but unspeeded response), as well as stimulus–response mapping rules (e.g., name letters, add numbers in pairs between asterisks, attend to side specified by arrows). As we have said, it is often proposed that the few chunks of working memory are closely related to the immediate contents of awareness. In contrast, nobody would suppose that a whole set of instructions is kept immediately in awareness until the time the instructions are used.

A second difference concerns dynamics. Traditional forms of working memory are highly labile—for example, emptying when a string of digits is reported and filling again with the digits for the next trial. In contrast, the task model must be a relatively stable control structure, surviving many intervening events to control performance when appropriate trigger conditions arise. Indeed, the data from Experiments 2 and 3 show a rather rigid structure, not easily changed once initially constructed. After full instructions describing rules for both letter and number tasks were given, there was little effect of a further instruction that one of these rules could be temporarily disregarded. There is little suggestion here that removing one part of the task model allows others to be better represented.

Our proposal is closer to Baddeley's (2000) "episodic buffer" (p. 417), a working memory component proposed to hold task-relevant facts over relatively extended periods of time. Like the capacity of our task model, the capacity of Baddeley's episodic buffer is substantially greater than the capacity of immediate processing systems; it underlies the ability, for example, to keep track of relevant material in a hand of bridge or other extended activity. Much the same is proposed by Ericsson and Kintsch (1995) in their concept of long-term working memory.

Limited capacity of the task model implies limited complexity in the set of rules that can shape task performance. Following on from the ideas of Fitts and Posner (1967) and Anderson (1983), however, we suppose that this limit only concerns a new or ill-learned structure of behavior. In Experiment 3, both groups of participants ultimately received the same total set of task requirements. Even reduced-instructions participants, after the first pure block of one task (letters or numbers), were given instructions and experience of the other task. When this happened, previously successful participants did not suddenly begin to neglect side cues. In fact, no participant showed such a change from success to failure. Having first mastered the use of side cues, they were able to add further requirements without harming performance. Such a result is reminiscent of the way that we very commonly give instructions for a complex task by serial addition of parts. It is natural to give instructions for one part of a task, allow practice on that part, and then add further parts until the whole structure is mastered.

The result is also reminiscent of the common observation that practice makes behavior more "automatic." Once behavior has been practiced, it is more strongly triggered by its eliciting conditions and less susceptible to interference from competing tasks or processes (Bryan & Harter, 1899; Fitts & Posner, 1967; James, 1890; Logan, 1988; Schneider & Shiffrin, 1977). A common proposal is that control is transferred from one processing mode to another (e.g. Anderson, 1983; Newell, 1990). In reduced-instructions participants, the proposal would be that once response to side cues had been practiced, it could resist impairment from further task requirements added after the first block of trials. Though performance may often be limited by the complexity of task-relevant knowledge, this limit can be bypassed with practice. A related observation is reduction of *g* correlations with practice (Ackerman, 1988).

In our laboratory tasks, we use a separate phase of task instructions to transmit the body of knowledge needed for effective behavior. In everyday performance, however, we should expect a much less clear separation between the phases of building and of using a task model. In typical behavior—and to some extent, in laboratory tasks too—experience itself suggests important facts about the world, requirements on behavior, or optimal methods of attack. Thus, task models are constantly revised or updated, corresponding subjectively to the active experience of reorganizing or reshaping task performance. Thus, we do not propose that building a task model is specifically a process of using task instructions. It simply happens that in typical experiments, this is the most active phase of new model development.

Later, we return to the relation between model complexity and *g*. Meanwhile, we are basing rather general conclusions on a very restricted database. In Experiment 3, the method of describing a task requirement and then telling the participant that it is no longer operative is rather different from methods used in typical studies of concurrent processing demands. In Experiment 4, we attempted a replication in a new task context.

Experiment 4

For Experiment 4, the task we chose was based on a series of unpublished experiments carried out in our laboratory (Bright, 1998). We chose this set of experiments for its complex, multicomponent structure, because our previous data showed it to have a strong correlation with *g* and because in other respects it was quite different from the letter-monitoring tasks of Experiments 1–3.

A typical sequence of trials is shown in Figure 6. On each trial, a pair of numbers appeared for 1 s on a computer screen. On most trials, numbers were surrounded by colored shapes. Full-instructions participants were given instructions for two tasks, one (vocal responses) for numbers without surrounds (no-surround trials), and the other (manual responses) for numbers with surrounds (surround trials). Although participants received instructions and practice for both tasks, the main experimental blocks involved only surround trials and manual responses. As before, these participants were explicitly told that no-surround trials would not occur during the main blocks. In contrast, reduced-instructions participants never saw or were instructed about no-surround trials. As in Experiment 3, accordingly, both groups ultimately carried out just the same tasks but in a context of more or less complex instructions.

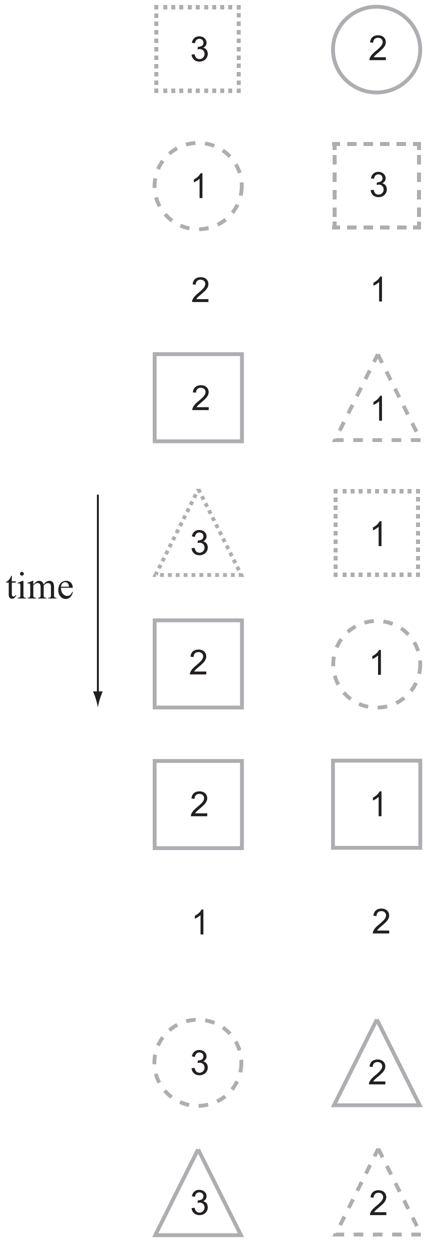

Figure 6. Sample stimuli for feature match task. In actual stimuli, outline shapes were colored rather than grayscale (solid shape = red, dashed shape = green, dotted shape = blue).

Method

Participants. We recruited 50 participants (11 male, 39 female; age range, 44-69 years, M = 59, SD = 7) from the paid volunteer panel of the MRC Cognition and Brain Sciences Unit.

Feature match task. The feature match task was performed on a PC (Dell Computer, Round Rock, TX) running Visual Basic (Microsoft, Redmond, WA). As in previous experiments, participants sat comfortably, without precise control of viewing distance; calculations of visual angle were based on an approximate distance of 57 cm. Screen background was white. On no-surround trials, stimuli were a pair of black digits, 0.7 deg in height and centered 2.3 deg left and right of a central fixation point. These two digits were always different, drawn from the set of 1–3. For surround trials, stimuli were similar except that each digit was contained within a colored outline shape. Shapes were squares, circles, or triangles 2.1 deg in height and presented in red, green, or blue. For both types of trials, stimuli were presented for 1 s; presentation was followed by a 1-s intertrial interval during which the screen was blank.

For no-surround trials, the task was to add the two numbers together and say the result aloud. Vocal responses were allowed at any time. For surround trials, the instruction was more complex. On the majority of trials (zero matches), surrounds were entirely different (matching in neither color nor shape; see first trial, Figure 6). For these trials, the participant was to do nothing. Less frequently, surrounds matched in a single feature (one match), either color or shape (see Figure 6, second, fifth, and final trials). For these trials, the task was to press a button on the side of the larger number. For the second trial in Figure 6, for example, the correct response would be to press the right key. Least frequently of all, surrounds were identical (matching in both color and shape; see Figure 6, seventh trial). For these trials (two match), response was to be withheld, as it was for zero-match trials. Manual responses were made with left and right index fingers on the *V* and *B* keys, respectively, of the computer keyboard. They were accepted through the whole 2-s period from stimulus onset on the current trial to stimulus onset on the next.

Verbatim instructions appear in the Supplemental Materials online. Full-instructions participants were first given the rule for the no-surround task and then the rule for the surround task. Reduced-instructions participants never were given a description of the no-surround task. There followed a block of 12 practice trials. For full-instructions participants, this block contained 2 no-surround trials and 10 surround trials: 7 zero-match, 2 one-match (1 matching in color and 1 in shape), and 1 two-match trials. For reduced-instructions participants, the 2 no-surround trials were replaced with 2 additional zero-match surround trials.

Following the practice block, participants were asked to describe the rules for both no-surround trials (full-instructions participants only) and surround trials. For surround trials, participants were asked when they should respond, on which side they should respond, and specifically what they should do for two-match trials. Any errors were corrected, and all questions were repeated until answers were correct. Among full-instructions participants, 13 described all rules correctly on the first attempt, 8 after one further repetition, and the remainder after between two and five repetitions. Among reduced-instructions participants, 16 described the rule correctly at the first attempt, 1 after one further repetition, and the remainder after between two and six repetitions. One participant who had special difficulty expressing the rules verbally was allowed instead to indicate what response was correct for sample stimuli of each sort.

After rules had been correctly described, reduced-instructions participants proceeded at once to the task. At this point, full-instructions participants were told that until further notice, only surround trials would appear. They were encouraged to disregard the no-surround task, which they would not be carrying out.

There followed four blocks of 30 trials each, separated by short pauses. For both groups, there were only surround trials: 18 per block zero-match, 9 per block one-match, and 3 per block two-match trials. The order of these trial types was pseudorandom, as was selection of digits, colors, and shapes for each trial.

After the fourth block, participants were again asked to describe the rules. The task then terminated for reduced-instructions participants, whereas full-instructions participants were given a final short block including some no-surround trials. Data from this final block were not scored.

Culture Fair Test. Again, *g* was measured with Scale 2, Form A, of the Culture Fair Test, which was administered immediately before the feature match task.

Results

For purposes of scoring, the task was regarded as having three separate rules, one for zero-match, one for one-match, and one for two-match trials. In each block, we scored participants' response for each rule as passed (>25% correct responses) or failed (<25% correct responses). For each participant, we counted the total number of failures across rules and blocks. For example, a participant who failed one rule in all four blocks and an additional rule in just one block would receive an overall rule failure score of 5.

Scatterplots relating number of rule failures to Culture Fair IQ are shown in Figure 7. In the reduced-instructions group, most participants performed well, with only scattered rule failures. Only 7 of the 25 participants had two failures or more. The correlation between number of failures and Culture Fair IQ was −.30. Results were very different in the full–instructions group. More than half the participants (17/25) in the group had two failures or more. The correlation between number of failures and Culture Fair IQ was −.55. ANCOVA, as before, showed significant differences between full- and reduced-instructions participants, separately at IQs of 90, $t(46) = 2.9$, $p < .005$, $d = 0.4$; 100, $t(46) = 3.1$, $p < .005$, $d = 0.4$; and 110, $t(46) = 2.0$, $p < .05$, $d = 0.3$ (all tests, one-tailed).

More detailed examination showed a consistent pattern of rule failures. When two or more failures occurred, in almost all cases they were either repeated block failures on the zero-match rule (6 participants with two or more failed blocks for zero-match) or repeated block failures on the two-match rule (16 participants with two or more failed blocks for two-match). Though occasional participants failed the one-match rule once, no participant did so more than once. Thus, failing participants tended to simplify the response rule; instead of responding for one-match and withholding response for zero- and two-match trials, they either responded for all shapes that were not identical (zero- and one-match) or for all shapes sharing any feature (one- and two-match).

In this experiment, 14 participants (9 full-instructions, 5 reduced-instructions) made some sort of error in their final (post-task) description of task rules. For only 6 people, however, was their erroneous description consistent with the actual pattern of rule failure. Specifically, of the 6 people showing repeated failure on the zero-match rule, 5 then described this rule incorrectly; of the 16 participants repeatedly failing the two-match rule, only 1 failed to describe the rule. Even after eliminating all 14 people with rule

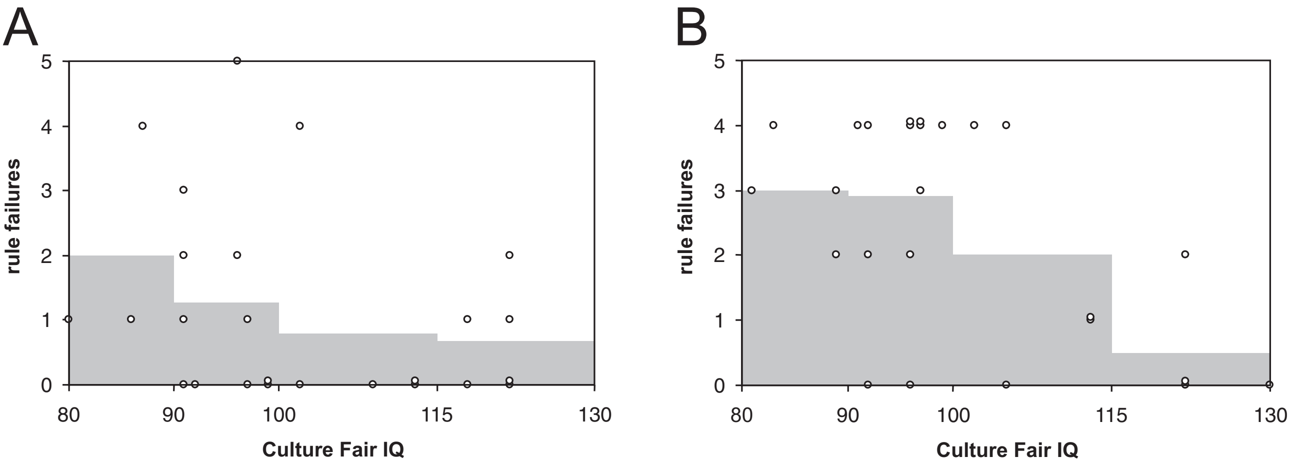

Figure 7. Experiment 4: scatterplots relating number of rule failures to Culture Fair IQ, separately for (A) reduced-instructions and (B) full-instructions participants. Each point represents a single participant; participants with IQs < 80 or > 130 are plotted on the corresponding margin. Histograms (gray bars) show mean rule failures for participants with IQs ≤ 90, 91–100, 101–115, and > 115.

recall failures, we found that major results were unchanged, with a significant difference between full- and reduced-instructions groups ($p < .05$ at each IQ tested) and correlations between number of rule failures and Culture Fair IQ of $-.40$ and $-.52$ in reduced- and full-instructions participants, respectively.

In a final analysis, we examined simple proportions of correct responses rather than rule failure scores. For each participant, proportion of correct responses was separately obtained for zero-, one-, and two-match trials; overall proportion of correct responses was calculated as the unweighted mean of these three. Results for proportion correct were very similar to those obtained with the rule failure score, with a significant difference between full- and reduced-instructions participants (ANCOVA, $p < .02$), and in each group, strong correlations with Culture Fair IQ ($r = -.50$ and $-.64$, respectively, for reduced- and full-instructions groups).

Discussion

In Experiment 4, we obtained strong confirmation that *g*-related errors depend on task complexity. Again, the key consideration is not complexity of the task actually performed—in both groups of participants, the task performed was the same. Rather it is complexity in the whole body of knowledge specified in initial instructions.

In the letter-monitoring tasks of Experiments 1–3, complexity affected just one vulnerable task component. As the number of task components increased, the result was increasing neglect of the requirement to watch for and respond to side cues. In Experiment 4, neglect was manifest in a tendency to simplify the response rule. Instead of making one response (no-go) for zero- and two-match cases and a different response (key press) for the one-match case, participants made key presses either for one- and one-matches (failure of two-match rule) or for zero- and one-matches (failure of zero-match rule). This simplification may be conceptualized in terms of mapping between a single stimulus dimension (number of matching features) and a single response dimension (go or no-go). In both simplified rules, one response was made for all events below a certain feature match threshold, the other for all events above that threshold.

In contrast to Experiments 1–3, explicit memory for the task rules was not perfect at the end of the experiment. In particular, participants who responded to zero-match trials—withholding response only for two-match trials—tended then to say that this was what the task rules required. Even when these participants were excluded, however, a clear difference remained between full- and reduced-instructions groups. Evidently, neglect in the feature match task, just as in Experiments 1–3, cannot be simply traced to lost explicit memory for task rules.

General Discussion

In these experiments, we sought the basis for goal neglect in competition between different parts of a complex task. Our hypothesis was that, in some form, neglect could be traced to an attentional limit.

Though our experiments confirmed this hypothesis, the attentional limit we found is quite different from anything we initially expected. Unlike most such limits in the literature, our attentional limit is not concerned with the real-time demands of actual task execution. Neither is it concerned with simultaneous preparation for different possible events. Instead, the key limit concerns the whole body of facts, rules, and requirements specified in what we have called a task model. As the complexity of the model increases, so it becomes increasingly likely that one task component will be lost. The task is accordingly simplified, losing important constraints on what should be done.

To show the importance of the task model, we have introduced a new kind of experimental comparison. In this comparison, one group of participants is given a relatively simple task model; the other is given a more complex model that is then simplified by an instruction that certain task events will not occur. Ultimately, the two groups carry out just the same task. Our experiments, nevertheless, showed there were different levels of neglect.

Our interest in neglect derives partly from its close relation to *g* and partly from its resemblance to mismatch of knowledge and behavior in patients with frontal lobe damage. In the following sections, we discuss the role of attentional limits in *g*, a possible neural basis for the task model, and possible relations to other psychological phenomena.

Goal Neglect, Task Complexity, and g

As in our previous work (Duncan et al., 1996), our results show a close link between goal neglect and *g*. The correlation was especially strong in more complex versions of the letter-monitoring task, with instructions for both letter and number subtasks and a higher overall rate of neglect. Combined data for all full-instructions participants (Experiment 2, both pure- and mixed-task participants; Experiment 3, full-instructions participants only) are shown in Figure 8. For this full set of 80 participants, the correlation between side error score and Culture Fair IQ is −.56. For full-instructions participants in the feature match task of Experiment 4, there was a comparable correlation between number of rule failures and IQ (Figure 7).

The strongest possible correlation between two variables is the product of the square roots of their two reliabilities (Nunnally, 1978). Even if the goal neglect score had a reliability of 1.0, the maximum possible correlation with the Culture Fair would be the square root of Culture Fair reliability, or .87 (according to the manual). If the reliability of goal neglect were .80, this maximum would be reduced to .78. Correlations are also limited by any difference in shape of the two correlated distributions (Nunnally, 1978). Whereas Culture Fair scores had an approximately normal distribution, neglect scores in the letter-monitoring task (Figure 8) tended to be bimodal, with most participants performing very well, and a second set showing close to complete neglect. Much the same applies to rule failures in the feature match task of Experiment 4 (Figure 7). With these limits in mind, the data suggest that goal neglect scores measure something very closely related to standard tests of *g*.

At one level, this conclusion confirms the common idea that *g* relates to task complexity or working memory demand. It suggests, however, a more specific version of this idea. Unlike some interpretations of complexity or demand, *g* is not closely related to the real-time processing limits of typical dual-task experiments. Instead, it concerns the ability to organize novel information into complex, effective task models.

A conflicting set of previous reports has compared *g* correlations for single and dual tasks. In some experiments, *g* correlations have been substantially higher for a dual task than for either of the single task components (e.g. Roberts et al., 1988; Spilsbury, 1992). These experiments have been interpreted as showing the importance of task complexity in *g*. Sometimes, on the other hand, a concurrent task is more harmful to high-*g* than to low-*g* participants, suggesting a reduced *g* correlation in the dual-task case (e.g. Rosen & Engle, 1997; Kane & Engle, 2000). In these cases, it has been suggested that a concurrent task interferes with processing strategies selectively used by high-*g* participants. Other experiments have shown mixed results, with little overall difference between single- and dual-task cases (e.g. Fogarty & Stankov, 1988).

Our experiments give some insight into why these previous results have been so variable. The strong suggestion is that although some aspects of processing competition are related to *g*, others are not. Very often, however, these different aspects of competition or complexity are confounded. Commonly, the body of knowledge bearing on decisions is more complex in a dual task. At the same time, the dual task may involve many different kinds of real-time processing conflict, arising in processes not closely related to *g*. With these different influences confounded, the overall effect on *g* correlations may be unpredictable.

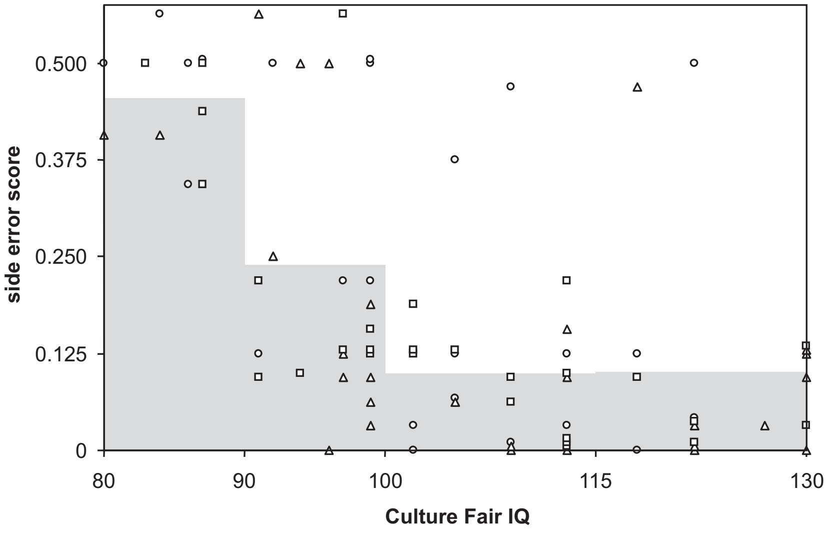

Figure 8. Combined data relating side error score to Culture Fair IQ for all full-instructions participants in Experiments 2 and 3. Triangles = Experiment 2, pure-task participants; squares = Experiment 2, mixed-task participants; circles = Experiment 3, full-instructions participants. Each point represents a single participant; participants with IQs < 80 or > 130 are plotted on the corresponding margin. Histograms (gray bars) show mean side error scores for participants with IQs ≤ 90, 91–100, 101–115, and > 115.

Indeed, the design of typical dual-task experiments may specifically minimize any limit on task-model construction. Commonly, for example, participants receive instructions and practice for each single task separately, with a second task only introduced once the requirements of the first have been mastered. As we have seen, this sort of training regime can directly bypass the kind of performance limit measured in our experiments. In Experiment 3, for example, participants who had practiced one pure task—either letter or number—could then add the other without performance cost. When participants receive instructions for and practice different parts of a complex or dual task in turn, limits on task model construction may be minimized, leaving other, real-time capacity limits to dominate behavior.

Similarly, we suggest that behavior can be influenced by many different kinds of working memory limitations (Baddeley, 1986). In a recent meta-analysis, Ackerman, Beier, and Boyle (2005) derived an average raw correlation of .17 between simple tests of short-term memory (such as digit span) and *g* as measured by Raven's Progressive Matrices (Raven, Court, & Raven, 1988). For more complex working memory tests, such as the reading span task of Daneman and Carpenter (1980), the average correlation was .39. To compare correlations across studies is difficult; for example, restriction of range in student samples inevitably leads to lower correlations than those we obtained with a broad spread of IQs. Nevertheless, raw correlations approaching .60 in our work suggest a form of working memory that is especially closely linked to *g*.

Our results also suggest an interpretation of complexity as manifest in a typical *g* test like the Culture Fair Test. Typically, each item in such a test requires participants to bear in mind a number of considerations in constructing a solution. Consider as an example the problem shown in Figure 9, Panel A, typical of matrix items in the Culture Fair Test. The task is to choose which of the four boxes in the lower row would correctly fill the empty box in the matrix. Choice of the correct solution (Box 4) requires separate consideration of object size, shape, and shading, all of which show relevant variation in the matrix, requiring appropriate computation in the solution.

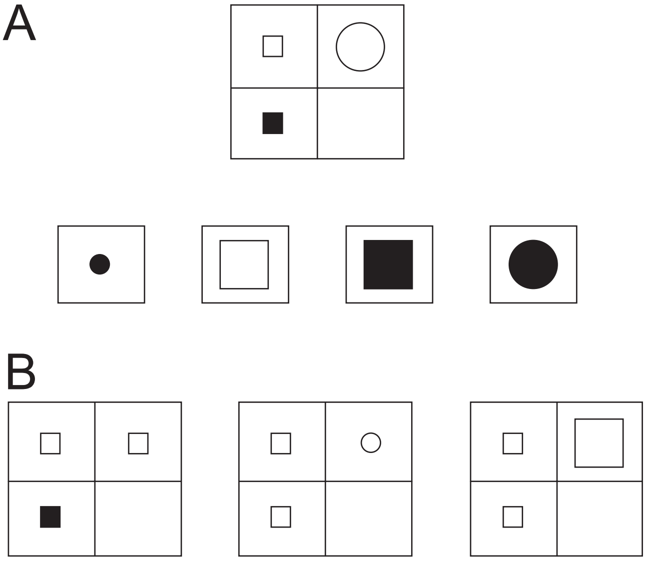

Figure 9. A. Typical matrix problem. The task is to decide which of the four boxes beneath the matrix would correctly complete its pattern. Correct solution requires recognition of relevant variations in object shading, shape, and size. B. Component single-feature problems.

On the one hand, complexity of some sort is certainly an important feature of this problem (Carpenter, Just, & Shell, 1990). When the three component problems are presented in isolation, each one appears trivial (Figure 9. Panel B). On the other hand, different aspects of complexity are confounded in the way that is typical of concurrent tasks. In a problem of this sort, a task model must be constructed as the problem itself is investigated. Presumably this model—a specification of relevant knowledge bearing on the solution—would focus on stimulus variations as these are detected in the matrix. Very much as must happen with verbal instructions, the information in the task materials must be split into useful segments, each with a bearing on the solution. This list of separate task components would certainly be more complex for the three-feature problem in Panel A of Figure 9 than for each of its one-feature components in Panel B. Equally, however, the three-feature problem leads to a number of more traditional concurrent task demands, for example, a demand to remember one aspect of the solution while working on others. According to our interpretation, it is model complexity that is the key consideration. With its increased model complexity, the three-feature problem makes it more likely that key task requirements will be missed. For example, having added to the model a requirement to address variations in shape, a low-*g* participant may have difficulty adding further task specifications.

These ideas also suggest why any task would have some significant *g* correlation. In even the simplest tasks, there are better and worse ways to bring relevant knowledge to bear, often described as better and worse strategies. As Rabbitt (1981) has argued, for example, even laboratory reaction-time tests include many components beyond the traditional stages of stimulus discrimination and response selection: hunting for the correct balance between speed and accuracy, optimal temporal and spatial preparation, and so on. Whenever a goal is to be achieved, there is a structure of task-relevant knowledge to be constructed and used in choice of behavior. If *g* reflects a core aspect of this process, it should contribute to success in all manner of different tasks.

Neural Basis

In some ways, goal neglect resembles knowledge–behavior mismatch in patients with frontal lobe damage. In one of our previous articles (Duncan et al., 1996), we suggested that goal neglect and *g* may be closely related to frontal lobe function.

On the basis of subsequent data, we would now propose a revised version of this hypothesis. In functional brain imaging studies, fluid intelligence tests produce a characteristic pattern of activity (Duncan, 2005; Duncan et al., 2000; Esposito, Kirkby, Van Horn, Ellmore, & Berman, 1999; Prabhakaran, Smith, Desmond, Glover, & Gabrieli, 1997). Frontal components include the cortex in and around the inferior frontal sulcus (IFS), the frontal operculum extending into the anterior insula (FO/AI), and, on the medial surface, the dorsal part of the anterior cingulate and adjacent presupplementary motor area (ACC/pre-SMA). This frontal activity is commonly accompanied by activity in the parietal lobe along the intraparietal sulcus (IPS). A very similar activity pattern—including the IFS, FO/AI, ACC/pre-SMA, and IPS—is part of the brain's response to many different kinds of cognitive demand (Duncan, 2006; Duncan & Owen, 2000). Together these results suggest that specific regions of both prefrontal and parietal cortex have broad involvement in many cognitive tasks, including responding to standard tests of fluid intelligence. Evidently, such regions are strong candidates for a broad task-modeling function—a distributed representation of facts and rules bearing on current behavior.

Supporting this idea, many experiments on single-cell recording have examined neural responses in the lateral frontal cortex of the behaving monkey. A striking result is that whatever task a monkey has been trained to carry out, many prefrontal cells seem to code the events of that task (e.g. Asaad, Rainer, & Miller, 2000). Responses are of many different kinds, coding stimuli, rewards, rules, and so on (Duncan, 2001; Miller & Cohen, 2001). Along with frequent coding of task-relevant information is reduced coding of information that is task irrelevant. As the monkey changes from one task context to another, so the prefrontal cortex changes in the information it codes (e.g. Everling, Tinsley, Gaffan, & Duncan, 2002; Freedman, Riesenhuber, Poggio, & Miller, 2001; Sakagami & Niki, 1994). A similar broad variety of task-relevant responses may also be seen in the ACC/pre-SMA (e.g. Niki & Watanabe, 1979; Procyk, Tanaka, & Joseph, 2000) and parts of parietal cortex (e.g. Caselli et al., 2004; Toth & Assad, 2002). In line with the task-model idea, these neurophysiological data are strongly suggestive of a flexible, distributed representation of information specifically relevant to current behavior.

That said, the only direct evidence concerning the neuropsychology of our goal neglect task comes from one experiment of Duncan et al. (1996). Though neglect was common in patients with frontal lobe lesions and apparently was not affected by lesions in posterior cortex, this evidence was weak because most patients in the frontal lobe group had closed head injuries. Typically, closed head injury is associated with diffuse brain damage in addition to any focal lesion (Richardson, 1990); and indeed, atrophy, rather than focal lesions, may be a better predictor of goal neglect in this group (Duncan, Johnson, Swales, & Freer, 1997). More research is needed to directly compare goal neglect following different kinds of focal brain lesion.

Of course, many questions remain concerning the link between psychological and physiological data. As we have emphasized, the task model that underlies goal neglect seems especially important in novel behavior. Once a task component has been learned, it is much less sensitive to increased model complexity. In the monkey, however, data are typically gathered after weeks or months of training. These data, accordingly, show broad prefrontal representation of task-relevant information even in well-learned behavior. Quite possibly, increasing learning simply produces multiple neural routes for generation of correct behavior. Parallel to the representation in the prefrontal cortex, complex task rules can be represented in the premotor cortex, the basal ganglia, and other locations (e.g., Muhammad, Wallis, & Miller, 2006). Parallel representation could underlie the increased robustness of well-learned, automatic behavior. As learning progresses, a prefrontal–parietal task model may not be lost but may simply become less crucial in ensuring that behavior is correct.

Relation to Other Phenomena

More work is also needed to address relations between goal neglect and a variety of other psychological phenomena. One obvious case is prospective memory. In prospective memory experiments, participants are given two kinds of task requirements. One requirement applies to the majority of trials; another applies more occasionally (e.g., when a particular stimulus is seen or at a specified later time). For example, participants might be asked to name a series of famous faces but also be asked to carry out some additional task for occasional faces with glasses (Maylor, 1998). Failures on the occasional task are the focus of prospective memory experiments.

There is certainly some surface resemblance between the experiments we have described and the typical prospective memory scenario. In both cases, the task model includes multiple different requirements, and the participant must bring each requirement into operation when the appropriate conditions apply. At the end of the experiment, prospective memory participants may be quite able to describe the requirements of the more occasional task, even after many failures to carry it out (Maylor, 1998). At the same time, there are potentially important differences between our experiments and the typical prospective memory experiment. In a typical prospective memory experiment, one salient factor is the long delay between specification of the requirement for performing the occasional task and receipt of the trigger stimulus indicating that the task should now be performed. Quite possibly, the forgotten task component loses salience through the passage of time, not just because of the total complexity of task instructions or rules.

Our experiments also have some resemblance to various kinds of interference or carryover from one task to the next. In full-instructions participants, neglect was strongly influenced by a task requirement previously specified but no longer in force. Might this be related to the proactive interference observed when participants must remember several successive lists of related material (e.g. Kane & Engle, 2000), or to task-set inertia when participants must change the rule mapping stimuli onto responses (e.g. Allport & Wylie, 1999; for an example especially close to our experiments, see Spieler, Mayr, & LaGrone, 2006)? Though again this is a matter for further work, again we may point to potentially important differences between these phenomena. In both proactive interference and task-set inertia, a key factor is specific similarity and thus specific interference between successive trials or tasks. In proactive interference, it is central that successive memory lists should contain related material. In task-set inertia, it is central that the same stimuli are remapped onto new responses. In neither case do we seem to be dealing with the core limitation that we have proposed for goal neglect—a simple limited capacity in coding rules and constraints that should shape novel behavior.

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., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131,* 30–60.

Allport, D. A. (1980). Attention and performance. In G. Claxton (Ed.), *Cognitive psychology: New directions* (pp. 112–153). London: Routledge and Kegan Paul.

Allport, D. A., & Wylie, G. (1999). Task-switching: Positive and negative priming of task-set. In G. W. Humphreys, J. Duncan, and A. Treisman (Eds.), *Attention, space, and action: Studies in cognitive neuroscience* (pp. 273–296). Oxford, England: Oxford University Press.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Asaad, W. F., Rainer, G., & Miller, E. K. (2000). Task-specific neural activity in the primate prefrontal cortex. *Journal of Neurophysiology, 84,* 451–459.

Baddeley, A. D. (1986). *Working memory.* Oxford, England: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423.

Bourke, P. A. (1997). Measuring attentional demand in continuous dual-task performance. *Quarterly Journal of Experimental Psychology, 50A,* 821–840.

Bourke, P. A., Duncan, J., & Nimmo-Smith, I. (1996). A general factor involved in dual task performance decrement. *Quarterly Journal of Experimental Psychology, 49A,* 525–545.

Bright, P. D. (1998). *The control of action: An exploration of Spearman's general factor.* Unpublished doctoral thesis, University of Cambridge, Cambridge, England

Broadbent, D. E. (1958). *Perception and communication.* London, England: Pergamon.

Bryan, W. L., & Harter, N. (1899). Studies on the telegraphic language. The acquisition of a hierarchy of habits. *Psychological Review, 6,* 345–375.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97,* 404–431.

Caselli, L., Fanini, A., Bertini, G., Tanaka, H. K., Mirabella, G., & Chelazzi, L. (2004). Posterior parietal cortex and the control of non-spatial attention. *Program No. 331.4. Abstract Viewer/Itinerary Planner.* Washington, DC: Society for Neuroscience.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action.* Boston: Houghton-Mifflin.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments and Computers, 25,* 257–271.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–185.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews Neuroscience, 2,* 820–829.

Duncan, J. (2005). Prefrontal cortex and Spearman's *g.* In J. Duncan, L. H. Phillips, & P. McLeod (Eds.), *Measuring the mind: Speed, control, and age* (pp. 249–272). Oxford, England: Oxford University Press.

Duncan, J. (2006). Brain mechanisms of attention. *Quarterly Journal of Experimental Psychology, 59,* 2–27.

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology, 30,* 257–303.

Duncan, J., Johnson, R., Swales, M., & Freer, C. (1997). Frontal lobe deficits after head injury: Unity and diversity of function. *Cognitive Neuropsychology, 14,* 713–741.

Duncan, J., Martens, S., & Ward, R. (1997, June 19). Restricted attentional capacity within but not between sensory modalities. *Nature, 387,* 808–810.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences, 23,* 475–483.

Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., et al. (2000, July 21). A neural basis for general intelligence. *Science, 289,* 457–460.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102,* 211–245.

Esposito, G., Kirkby, B. S., Van Horn, J. D., Ellmore, T. M., & Berman, K. F. (1999). Context-dependent, neural system-specific neurophysiological concomitants of ageing: Mapping PET correlates during cognitive activation. *Brain, 122,* 963–979.

Everling, S., Tinsley, C. J., Gaffan, D., & Duncan, J. (2002). Filtering of neural signals by focused attention in the monkey prefrontal cortex. *Nature Neuroscience, 5,* 671–676.

Fitts, P. M., & Posner, M. I. (1967). *Human performance.* Belmont, CA: Brooks/Cole.

Fogarty, G., & Stankov, L. (1988). Abilities involved in performance on competing tasks. *Personality and Individual Differences, 9,* 35–49.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001, January 12). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science, 291,* 312–316.

Institute for Personality and Ability Testing. (1973). *Measuring intelligence with the Culture Fair tests.* Champaign, Illinois: Institute for Personality and Ability Testing.

James, W. (1890). *The principles of psychology.* New York: Holt.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Kalsbeek, J. W. H., & Sykes, R. N. (1967). Objective measurements of mental load. In A. F. Sanders (Ed.), *Attention and performance I* (pp. 253–261). Amsterdam: North-Holland.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 336–358.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin and Review, 9,* 637–671.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133,* 189–217.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14,* 389–433.

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

Luria, A. R. (1966). *Higher cortical functions in man.* London: Tavistock.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence, 7,* 107–127.

Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data.* Belmont, CA: Wadsworth.

Maylor, E. A. (1998). Changes in event-based prospective memory across adulthood. *Aging, Neuropsychology, and Cognition, 5,* 107–128.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal function. *Annual Review of Neuroscience, 24,* 167–202.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior.* New York: Holt, Rinehart and Winston.

Milner, B. (1963). Effects of different brain lesions on card sorting. *Archives of Neurology, 9,* 90–100.

Muhammad, R., Wallis, J. D., & Miller, E. K. (2006). A comparison of abstract rules in the prefrontal cortex, premotor cortex, inferior temporal cortex, and striatum. *Journal of Cognitive Neuroscience, 18,* 974–989.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Niki, H., & Watanabe, M. (1979). Prefrontal and cingulate unit activity during timing behavior in the monkey. *Brain Research, 171,* 213–224.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116,* 220–244.

Prabhakaran, V., Smith, J. A. L., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1997). Neural substrates of fluid reasoning: An fMRI study of neocortical activation during performance of the Raven's Progressive Matrices Test. *Cognitive Psychology, 33,* 43–63.

Procyk, E., Tanaka, Y. L., & Joseph, J. P. (2000). Anterior cingulate activity during routine and non-routine sequential behaviors in macaques. *Nature Neuroscience, 3,* 502–508.

Rabbitt, P. M. A. (1981). Cognitive psychology needs models for changes in performance with old age. In J. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 555–573). Hillsdale, NJ: Erlbaum.

Raven, J. C., Court, J. H., & Raven, J. (1988). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* London: H. K. Lewis.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18,* 849–860.

Richardson, J. T. E. (1990). *Clinical and neuropsychological aspects of closed head injury.* London: Taylor & Francis.

Roberts, R. D., Beh, H. C., & Stankov, L. (1988). Hick's law, competing-task performance, and intelligence. *Intelligence, 12,* 111–130.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124,* 207–231.

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General, 126,* 211–227.

Sakagami, M., & Niki, H. (1994). Encoding of behavioral significance of visual stimuli by primate prefrontal neurons: Relation to relevant task conditions. *Experimental Brain Research, 97,* 423–436.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84,* 1–66.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology, 15,* 201–293.

Spearman, C. (1927). *The abilities of man.* New York: Macmillan.

Spieler, D. H., Mayr, U., & LaGrone, S. (2006). Outsourcing cognitive control to the environment: Adult age differences in the use of task cues. *Psychonomic Bulletin and Review, 13,* 787–793.

Spilsbury, G. (1992). Complexity as a reflection of the dimensionality of a task. *Intelligence, 16,* 31–45.

Stankov, L. (2000). Complexity, metacognition, and fluid intelligence. *Intelligence, 28,* 121–143.

Toth, L. J., & Assad, J. A. (2002). Dynamic coding of behaviourally relevant stimuli in parietal cortex. *Nature, 415,* 165–168.

Ward, R., Duncan, J., & Shapiro, K. (1996). The slow time-course of visual attention. *Cognitive Psychology, 30,* 79–109.

Received September 2, 2005
Revision received April 26, 2007
Accepted April 28, 2007 ■