Goal neglect and knowledge chunking in the construction of novel behaviour

Apoorva Bhandari [a,*], John Duncan [a,b]

[a] MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, UK
[b] Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK

ABSTRACT

Task complexity is critical in cognitive efficiency and fluid intelligence. To examine functional limits in task complexity, we examine the phenomenon of goal neglect, where participants with low fluid intelligence fail to follow task rules that they otherwise understand. Though neglect is known to increase with task complexity, here we show that – in contrast to previous accounts – the critical factor is not the total complexity of all task rules. Instead, when the space of task requirements can be divided into separate sub-parts, neglect is controlled by the complexity of each component part. The data also show that neglect develops and stabilizes over the first few performance trials, i.e. as instructions are first used to generate behaviour. In all complex behaviour, a critical process is combination of task events with retrieved task requirements to create focused attentional episodes dealing with each decision in turn. In large part, we suggest, fluid intelligence may reflect this process of converting complex requirements into effective attentional episodes.

© 2013 The Authors. Published by Elsevier B.V. Open access under CC BY-NC-ND license.

1. Introduction

Standard tests of "general intelligence" are important for their broad ability to predict success in all kinds of cognitive activities, from simple laboratory tasks to educational and other achievements (Spearman, 1904, 1927). Well-known examples are tests of novel problem-solving or "fluid intelligence", such as Raven's Progressive Matrices (Raven, Court, & Raven, 1988) or Cattell's Culture Fair (Cattell, 1971; Cattell & Cattell, 1973). It has long been recognised that task complexity is a critical aspect of such tests. The best tests of general intelligence – those best able to predict success in many kinds of activity – are complex tasks with many different parts, while simple tasks, often involving many trials of the same, small set of cognitive operations, correlate only weakly with others (Marshalek, Lohman, & Snow, 1983; Stankov, 2000). In this paper we ask why task complexity is so important, and what cognitive operations limit success as task complexity increases (Frye, Zelazo, & Burack, 1998; Halford, Cowan, & Andrews, 2007).

Our work builds on the phenomenon of *goal neglect* described by Duncan and colleagues (Duncan, Emslie, Williams, Johnson, & Freer, 1996; Duncan et al., 2008). Goal neglect is manifest when a participant is able to correctly remember and state a task requirement but fails to fulfil it during performance. This behaviour has been reported in patients with major damage to the frontal lobe (Luria, 1966; Milner, 1963) but also in people in the normal population (Altamirano, Miyake, & Whitmer, 2010; Duncan et al., 1996, 2008; Piek et al., 2004; Towse, Lewis, & Knowles, 2007). In some cases, a whole task may be completed incorrectly, followed by correct recall of the required rules (Duncan et al., 1996), though this may be the extreme case

⁎ Corresponding author. Tel.: +44 1223 355 294; fax: +44 1223 359 062.

E-mail address: apoorva.bhandari@mrc-cbu.cam.ac.uk (A. Bhandari).

of a more general difficulty in obeying novel rules (e.g. Duncan, Schramm, Thompson, & Dumontheil, 2012). Importantly, goal neglect is closely related to standard measures of fluid intelligence, supporting the proposal that, at least in large part, fluid intelligence concerns the cognitive control functions of the frontal lobe (for related ideas see Duncan, Burgess, & Emslie, 1995; Kane & Engle, 2003; Marshalek et al., 1983; Oberauer, Süß, Wilhelm, & Wittman, 2003). Goal neglect also shows intriguing and unexpected effects of task complexity, suggesting it may be a valuable test bed for cognitive analysis of complexity effects.

The phenomenon of goal neglect, and the critical effects of task complexity, can be illustrated by the *feature match task* used by Duncan et al. (2008) (Fig. 1). On each trial, a pair of digits, usually surrounded by a pair of coloured shapes, was shown on a computer screen. Participants were divided into two groups. One group were given instructions for two tasks (full-instructions condition). For digits without surrounding shapes, the task was to add them together and state the result. For digits with surrounding shapes, the task had three rules: if shapes are completely different (different colour and shape), do nothing; if shapes match in a single feature, press a key on the side of the larger digit; if shapes match in both features, again do nothing. Although the participants received instructions and practice for both tasks, the main experi-
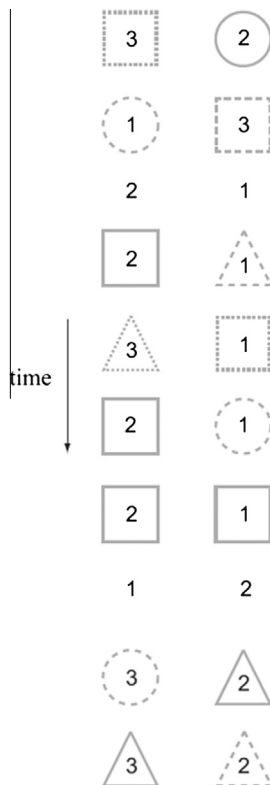

Fig. 1. Sample stimuli for a series of trials (top to bottom) from the Duncan et al. (2008) feature match task. Actual stimuli were in colour (solid shape = red, dashed shape = green, dotted shape = blue).

mental blocks included only trials with surrounding shapes, and participants were explicitly told that they would not see any digits without surrounding shapes. The other group of participants were never instructed about the trials with no surrounding shapes (reduced-instructions condition), and never received such trials. In both groups, neglect was observed as a tendency to simplify the response set and make consistent errors throughout the task. Commonly, for example, the double-match trials (match in both colour and shape) were treated as though they were single matches. As usual in goal neglect, errors were not explained by the simple forgetting of task rules, as shown by a post-experiment recall test. Instead, often, rules were correctly remembered but still ignored in actual behaviour. Also as usual, neglect was strongly correlated with fluid intelligence. Critically, full-instructions participants were much more likely to show this neglect pattern.

Note that the only difference between the two groups was that the full-instructions group had received a more complex set of task rules during instruction; the actual task performed by both groups was exactly the same. Along with similar results in a second task (Duncan et al., 2008), the findings show that neglect is not modulated by task complexity at actual task execution; for instance, by attentional load or the number of behavioural alternatives to be considered during a single trial or even trial block. In this sense it appears very different from standard effects of attentional or dual task load, which usually are tightly time-locked to actual task performance (e.g. Pashler, 1994). Instead, what seems to matter is the complexity of the rules as specified in the initial task instructions. In this paper we pursue the cause of such a complexity effect, and its relevance to fluid intelligence.

Task complexity effects have frequently been understood in terms of a limited capacity working memory which has also been linked to general intelligence (Kane & Engle, 2002; Kyllonen & Christal, 1990; Oberauer, Süß, Wilhelm, & Wittmann, 2008; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). In general, models of working memory concern combined maintenance and use of task-relevant information (Baddeley, 1986; Cowan, 2001; Daneman & Carpenter, 1980; Kane & Engle, 2002; Kane & Engle, 2003; Salthouse, 1991). Following up this suggestion, Duncan et al. (2008) proposed an account based on a form of limited working memory capacity. As instructions are presented, they proposed, task knowledge is assembled into a control structure they called the *task model*. The task model is a limited capacity working memory for the representation of task rules and facts. As more information is entered into the task model and this capacity is filled, multiple task components compete for representation and individual components may be weakly represented and lost, leading to neglect of that requirement during performance. On this account, neglect is driven by the total complexity of task instructions, i.e. the total amount of information to be memorised or entered into the task model as instructions are received. As Duncan et al. (2008) point out, such a task model would need somewhat different characteristics from more standard forms of working memory. Storage capacity would need

to be sufficient to hold a whole set of task rules, presented over an extended period of task instruction. The task model, furthermore, would need to be durable, maintaining the full set of task rules even when the participant knew that some rules would be unnecessary for a current block of trials. In these respects, Duncan et al. (2008) suggested that the task model might resemble the concept of 'long-term working memory' proposed by Ericsson and Kintsch (1995). Recent data suggest that goal neglect may show stronger correlations with fluid intelligence than several more standard measures of working memory capacity, including visual short-term memory and conventional simple and complex spans (Duncan et al., 2012).

In the current work we develop this proposal in light of the broader literature on task complexity. The challenges of complex tasks have been addressed from several perspectives, including artificial intelligence (e.g. Newell, Shaw, & Simon, 1958; Sacerdoti, 1974), functional brain imaging (e.g. Badre & D'Esposito, 2009), the role of complexity in cognitive development (e.g. Frye et al., 1998), and the link to working memory capacity (e.g. Halford et al., 2007). In such accounts, complexity is a matter not simply of the total amount of information in a task description, e.g. the total number of task rules, but also of structural relations between rules, e.g. relations between the components of an argument (e.g. Badre & D'Esposito, 2009; Frye et al., 1998; Halford, Wilson, & Phillips, 1998; Oberauer et al., 2008). In complex tasks, in particular, a critical factor is chunking into separate task parts (e.g. Halford et al., 2007). This requirement for chunking has been recognised in foundational work in artificial intelligence (Sacerdoti, 1974); here we develop it for the link of fluid intelligence to goal neglect.

The fundamental need for complex tasks to be divided into parts is straightforward. Complex tasks involve choosing appropriate, goal-directed actions by deploying task-relevant knowledge (Anderson, 1983; Newell, 1990). Correct actions must be chosen by searching through a space of alternatives which grows with problem complexity, making the search increasingly demanding. Without further structure, behaviour becomes chaotic, with too few constraints to shape effective action selection at any one point in the task (Sacerdoti, 1974). The search problem can be simplified by dividing it into a set of separate parts or chunks, delimiting the set of alternatives considered within any particular chunk. Often, this division of a complex problem into separate, more solvable parts is regarded as division of complex goals into simpler sub-goals (e.g. Anderson, 1983; Newell, 1990). In everyday behaviour, this chunking is obvious as a task such as driving to work is separated into independent parts such as finding car keys, leaving the house and locking the door, approaching the car etc. Thus complex tasks are divided into a set of cognitive or attentional episodes, each focusing on just one sub-part. Within each episode, relevant knowledge must be selectively retrieved from memory, and combined with current sensory input to guide appropriate behaviour.

This line of thought suggests a potentially critical role of chunking in goal neglect. Once a task model has been established, the knowledge within it must be used to shape correct behaviour, much as actions, productions, etc. are selected in classic cognitive architectures (Anderson, 1983; Newell, 1990). As task complexity increases, it is increasingly critical that each stage of the task is controlled by a focused attentional episode, excluding aspects of the model not relevant to the current decision. Without such focus, task rules are not effectively used, in extreme cases leading to goal neglect. On this account, neglect should depend, not just on total task complexity, but on how easily the information in a task model is divided into separate chunks. Obeying a task rule may be impaired by additional task knowledge that is closely linked to this rule, making clear focus on just the most relevant information hard to attain. Obeying the same rule may be much less influenced by knowledge that is easily separated into a distinct task chunk.

The role of chunking and task organisation relates to several themes in the study of fluid intelligence and frontal lobe function. In memory tasks, for example, frontal patients may be impaired in spontaneous organisation of a memory list (e.g. Gershberg & Shimamura, 1995; see also Incisa della Rocchetta, 1986) and helped by explicit grouping of the list into semantic categories (e.g. Kopelman & Stanhope, 1998). In young children, a form of goal neglect is sometimes linked to task switching, for example when cards must be sorted first by one feature and then another. Commonly, the child knows the new rule but continues to sort according to the old one (e.g. Zelazo, Frye, & Rapus, 1996); this form of goal neglect can be ameliorated by encouraging cognitive separation of the two rules, for example by spatial separation of the two stimulus features (Diamond, Carlson, & Beck, 2005). As for fluid intelligence, all tasks, certainly, require a control structure like that proposed in the task model, and variable ability to assemble and use such a structure could explain broad positive correlations across many types of task. Each new problem in a fluid intelligence test, for example, requires assembly of novel information into a complex set of internal operations. Plausibly, fluid intelligence could reflect the ability to manage complex task models, either through capacity to represent larger task chunks (e.g. Halford et al., 2007), or through more effective parsing into focused, functionally separate sub-parts (Duncan, 2010).

In the present experiments we extend the task model account to address the role of chunking in goal neglect. As a chunking account predicts, we find that neglect depends on how a total body of task rules is organised into component chunks or sub-tasks. Neglect of a task rule increases with added complexity within one chunk or sub-task; it is independent of complexity in a different sub-task. In a further analysis, we ask when the critical failures of action selection take place, and thus, how the incorrect behaviour of goal neglect develops. One possibility, presumed by Duncan et al. (2008), is that components of a task model are lost as instructions are received. The present account, in contrast, emphasises processes of knowledge use, as the information in instructions shapes subsequent behaviour. Here we show that, when neglect occurs, it is developed and stabilized over the first few performance trials, i.e. as the participant attempts to construct novel behaviour by recalling task instructions. It is during these early performance trials, we suggest, that complexity

affects use of a task model to shape behaviour, with poorer search of task-relevant knowledge as its complexity increases, and a strong effect of knowledge chunking.

2. Experiment 1

In Experiment 1, each task consisted of two independent sub-tasks. Only one of the two sub-tasks had to be carried out on any particular trial and each sub-task used largely different stimulus elements. Within each sub-task, one critical aspect – the response decision – was constant. Complexity was manipulated by adding additional rules, determining which element of a display should be chosen before the response decision was made. Though response decision and element selection rules were independent, affecting different aspects of behaviour, we anticipated that added element-selection rules might increase neglect of response-decision rules. For response decisions in each sub-task, we examined the effect of added complexity either within the same sub-task or in the other sub-task.

One simple possibility is that the critical factor determining response-decision neglect should be just the total number of rules specified in task instructions. In this case, neglect should increase with additional rules in either sub-task. A more complex pattern is predicted by the chunking account. At the time of response decision, behaviour could ideally be controlled by a focused attentional episode, combining relevant stimulus input with selected response-decision rules retrieved from the task model. Reflecting the usual role of task complexity, however, creating such a focused episode might be harder when additional, element-selection rules are also present in the task model. In particular, we reasoned that rules from different sub-tasks might be more easily separated into distinct episodes than rules from the same sub-task. Accordingly, response-decision neglect should be especially sensitive to added rules within the same sub-task.

In our design, the aim was to manipulate the complexity of the task model. A complex sub-task, however, differs from a simple sub-task not just in terms of requiring a more complex task model, but also presents a greater demand during performance. In order to disentangle the effects of task model complexity from those of real-time performance demand, we also included a set of blocks for each task where performance demand was manipulated on a trial by trial basis, independent of task instructions. This was done by including some trials requiring only the response decision, without additional trial complexity.

The term goal neglect has been used for errors in various kinds of tasks (De Jong, Berendsen, & Cools, 1999; Kane & Engle, 2003; West, 2001), and it is unclear how closely these are related. Here we link our results to the prior findings of Duncan et al. (1996, 2008) using three criteria. First is gross failure to follow task rules, strongly correlated with fluid intelligence. As in previous studies (Duncan et al., 2008), we score both overall task accuracy, and the frequency of gross performance failure. Second is performance limited by complexity at task instruction rather than complexity on an individual trial. Third is performance error not well explained by explicit rule recall, as tested at the end of each experiment.

2.1. Methods

2.1.1. Participants

We recruited 32 right-handed adult participants (12 males, 20 females; age-range: 40–67, $M = 59.1$, SD = 5.4) with no history of neurological disorder from the paid volunteer panel of the MRC Cognition and Brain Sciences Unit. Participants took part after informed consent was obtained. Our sampling from an older population, following Duncan et al. (2008) was motivated by the need to capture a wide range of fluid intelligence scores and ensure adequate sampling in the low range. Once fluid intelligence test performance is matched, goal neglect is independent of age (Duncan et al., 2012).

2.1.2. Apparatus

All experiments were conducted on a standard desktop computer running the Windows XP operating system and connected to a Higgstec 5-wire resistive touch screen display. The stimulus delivery program was written in Matlab v6, using the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner et al., 2007). Responses were collected via the touch screen. Analyses were conducted using Matlab, Microsoft Excel and SPSS.

2.1.3. Tasks and stimuli

Each participant completed 4 tasks, each composed of 2 sub-tasks (Fig. 2). There were accordingly 8 different sub-tasks in total, with the same basic structure but involving 8 different kinds of stimulus materials. We begin with an explanation of sub-task structure, and then consider how the combination of sub-tasks created the different experimental conditions.

As illustrated in Fig. 3, using the example of Vehicles, each sub-task came in two forms, simple and complex. In the simple form (Fig. 3A), only the response decision was required. The display consisted of a single panel, presented in one of four possible screen locations, two each to left and right of a central grey dot. This panel contained a picture of a motorbike, beneath it two labelled response boxes, and a number of other surrounding symbols. There were two types of trials, the majority *regular* and the remainder *critical*. On regular trials ($p = 0.75$), participants were asked to touch the box containing a lower case letter. On critical trials ($p = 0.25$), a different response was indicated by the symbol (*context cue*) appearing beneath the chosen response box. The context cue beneath one response box was a triangle, and beneath the other a dot. When the context cue beneath the chosen response box was a dot, the task was to touch not the response box itself, but the dot instead.

We designed critical trials to be especially sensitive to goal neglect. First, they were less frequent than regular trials. Second, the two symbols used to signal regular and critical trials were visually similar. Therefore, participants had to closely monitor the symbols to detect changes in context. Accordingly performance on critical trials was our primary measure, with neglect on regular trials anticipated to be less frequent.

In the sub-task's complex form (Fig. 3B), the response decision was supplemented by two additional decisions
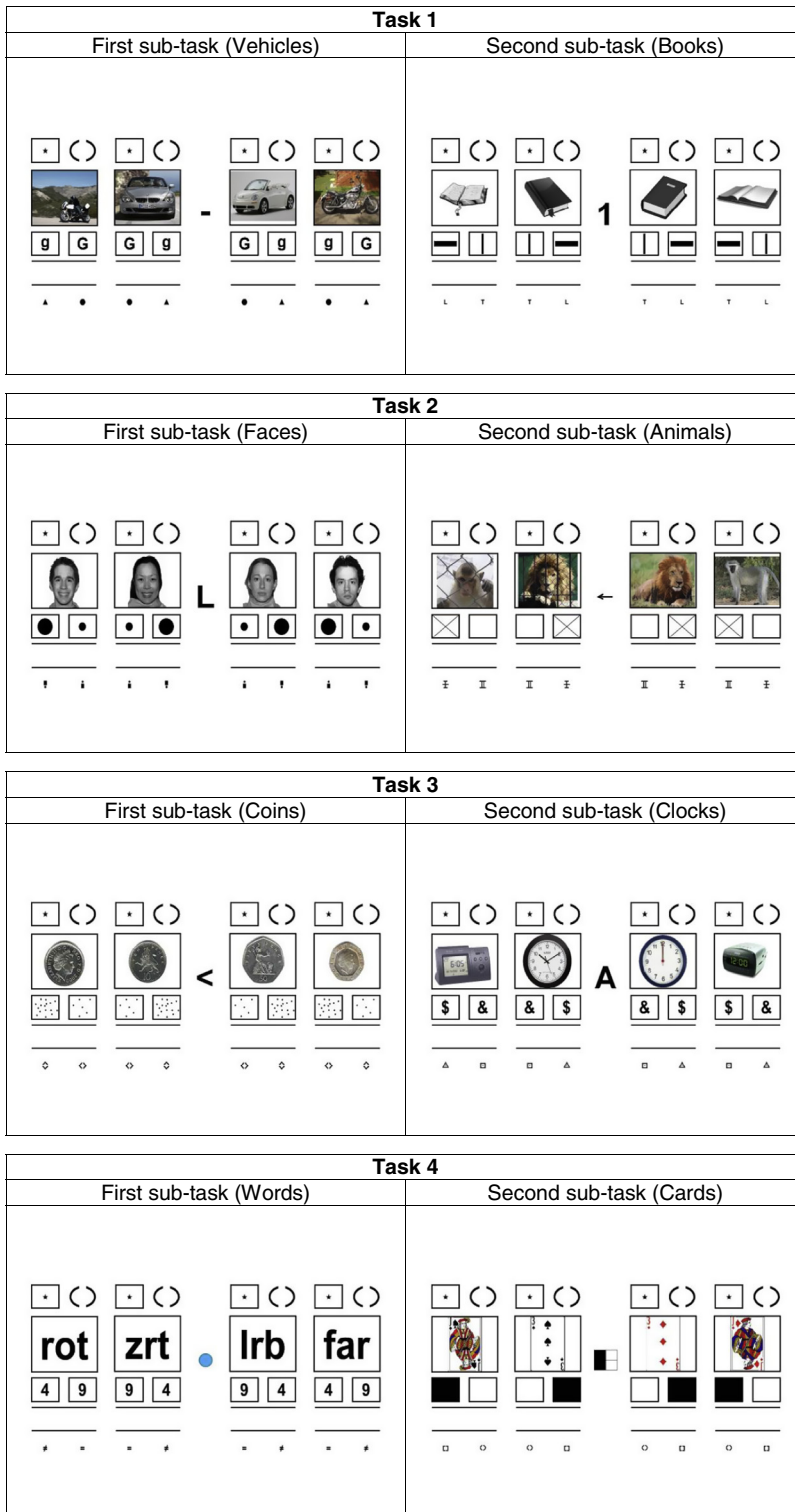
Fig. 2. Sample trial stimuli from all tasks used in Experiments 1 and 2. All stimuli are shown in complex form and represent a regular trial. Actual stimuli were in colour.
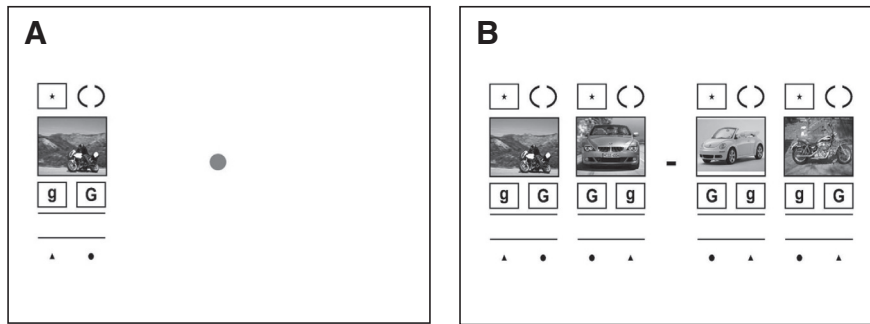
Fig. 3. Sample stimuli from one sub-task in Experiment 1. Stimuli are shown in both simple (A) and complex (B) form. Actual stimuli were in colour.

(*panel decisions*) determining which stimulus should be used for the response decision. There were now 4 different panels, two to either side of fixation. The correct panel to use was determined by two successive decisions, the first determining side and the second panel within that side. For the first decision the cue was a symbol at screen centre. When this cue was a −, the relevant side was left, while if the cue was +, the relevant side was right. The second decision was based on vehicle identity. On each side of the screen, one panel showed a motorbike and the other a car; the panel to choose was the motorbike (Fig. 3B, extreme left). Once the correct panel was chosen, the response decision proceeded as for the sub-task's simple form.

Since decisions could be scored independently (see below), our interest focused just on response decision accuracy, for sub-tasks in which this response decision was the only requirement (simple form), or in which it was preceded by two additional decisions (complex form). To avoid time limitations, performance was self-paced and participants were given up to 20s to make each response. The inter-trial delay was 500 ms.

Display structure and layout was similar for remaining sub-tasks (Fig. 2), all with simple forms in which only a response decision was required, and complex forms with additional panel decisions. The complete set of rules for all sub-tasks is listed in Table 1. For all tasks, displays were 18.75 deg in height and 36.02 deg in width. Viewing distance was not precisely controlled and all visual angle calculations are based on an approximate viewing distance of 50 cm.

Fig. 2 shows the pair of sub-tasks contributing to each task. Within each task, trials of the two sub-tasks occurred in random order. The four different experimental conditions were created by varying simple/complex form within each sub-task. For one task (simple–simple condition), both sub-tasks were in the simple form; for one (simple–complex), the first sub-task was simple and the second complex; for one (complex–simple), the first sub-task was complex and the second simple; for one (complex–complex), both sub-tasks were complex. (Note that, in these descriptions, the words "first" and "second" refer to the order of sub-tasks in Fig. 2, and the order in which they were explained in task instructions. As noted above, during actual performance, trials of the two

sub-tasks appeared in random order.) The allocation of the 4 tasks or sets of materials (Fig. 2) to conditions was counterbalanced across participants, along with both order of tasks and order of conditions within the session.

2.1.4. Protocol

Each participant served in a single experimental session lasting approximately 2 h. Session structure is illustrated with an example in Table 2.

The session began with the participant getting general instructions and training on using the touch screen display. This training involved successfully performing single and double taps at different locations on the screen. They were then instructed on a sample task (with a single sub-task in complex form) and asked to respond to 5 sample trials to familiarise them with the structure of the tasks. The materials presented in this sample task were not used in the actual experiment.

Each new task began with an instruction period during which the rules of the two sub-tasks were given to the participant one after the other. For each sub-task, rules were explained using an example stimulus display. The experimenter read out the rules verbatim (see Table 1 for text) while pointing to the various stimulus elements on the screen. Participants were told that the tasks were self-timed but that they should respond as quickly as they could while being accurate.

After the rules for both sub-tasks were read out, the sample screen was cleared and the participants were asked to repeat all the rules of the task. To aid recall, participants were cued on each rule by being asked a series of questions. To illustrate again using Vehicles (Fig. 3), the questions would be: "How do you decide whether to focus on the left or the right? Which of the two vehicles do you pick? Which of the two letters do you pick? How do you usually respond? What do you do when you see a dot below the lower case letter?" (For simple forms, the first two questions were omitted.) If any error was made during the cued repetition, the experimenter would correct the participant, and the entire repetition sequence would be repeated until the participant recalled all components of both sub-tasks correctly.

Participants were then told that they could not be reminded of the rules again. The total number of cued repe-

Table 1
Experiment 1. Verbatim instructions for all tasks (complex form).

First sub-task (vehicles)	Second sub-task (books)
Task 1	
First look at the centre of the screen. If there is a minus (−) sign, focus on the left half of the screen and if it is a plus (+) sign, focus on the right half. You will see a bike and a car. Look for the bike. Below the bike, there will be a pair of letters. Touch the lower case letter, unless you see a dot below the lower case letter, in which case touch the dot instead	First look at the centre of the screen. If the digit is a 1, focus on the left half of the screen and if it is a 2, focus on the right half. You will see a pair of books. Look out for the open book. Below the open book, there will be two bars. Touch the horizontal bar, unless you see the letter T below the horizontal bar, in which case touch the both the bars one after the other (the horizontal first)
First sub-task (faces)	Second sub-task (animals)
Task 2	
First look at the centre of the screen. If the letter is L, focus on the left half of the screen and if it is R, focus on the right. You will see a pair of faces. Look out for the male face. Below the male face, there will be a pair of shapes. Touch the larger shape, unless you see a symbol that looks like an i, just below the larger shape in which case touches the male face instead	First look at the centre of the screen. If the arrow is pointing left, focus on the left half of the screen and if it pointing right, focus on the right. You will see a pair of animals. Look out for the monkey. Below the monkey, there will be a pair of boxes. Touch the box that is crossed out, unless you see a II symbol below the crossed box, in which case touch the crossed box twice
First sub-task (coins)	Second sub-task (clocks)
Task 3	
First look at the centre of the screen. If the symbol is pointing left, focus on the left half of the screen and if is pointing right, focus on the right. You will see a pair of coins. Look out for the coin showing heads. Below the heads coin, there will be a pair of boxes with dots in them. Touch the box with more dots, unless you see a <> symbol below the box with the most dots, in which case touch the symbol below the box with the fewer dots	First look at the centre of the screen. If the letter is an A, focus on the left half of the screen and if it is the letter B, focus on the right. You will see a pair of clocks. Look out for the digital clock. Below the digital clock, there will be a pair of symbols. Touch the $ symbol, unless you see a box with a * in it just below the $ symbol in which case touch the star just above the clock
First sub-task (words)	Second sub-task (cards)
Task 4	
First look at the centre of the screen. If the circle is blue, focus on the left and if it is red, focus on the right. You will see a pair of words. Look out for the real/proper word. Below the real word, there will be a pair of numbers. Touch the lower number, unless you see an = sign below the lower number, in which case touch between the two parallel lines just below the numbers instead	First look at the centre of the screen. If the left side of the box is shaded, focus on the left half of the screen and if the right side is shaded, focus on the right. You will see a pair of playing cards. Look out for the picture/face card. Below the picture card, there will be a pair of boxes. Touch the box that is filled in (with black) unless you see a broken circle just below the filled in box, in which case touch between the square brackets just above the picture card

titions required before the rules were correctly repeated was between 1 and 3 (*M* = 1.16, SE = 0.04, Mode = 1).

After the instruction period, participants performed a practice block with 12 trials, which included 5 regular trials and one critical trial for each sub-task intermixed randomly. Following this, participants performed 10 blocks of 16 trials each, each block including 6 regular trials and 2 critical trials for each sub-task intermixed randomly.

A final manipulation was added to distinguish effects of task vs. trial complexity. For 5 of the blocks (randomly chosen) within each task (*standard blocks*), all trials involved the simple or complex form of the relevant sub-task, as specified in the instructions for the current condition. Main data analyses concerned just these standard blocks. For the remaining 5 blocks (*mixed form blocks*), a manipulation was introduced to separate task from trial complexity. In these blocks, half of the trials from any complex sub-task in the current condition were actually in the simple form, i.e. they contained only a single display panel. Of course, this manipulation had no effect for sub-tasks that, in the current condition, were already in simple form. Participants were fully informed of these manipulations.

At the end of all performance blocks, participants were again cued to recall the rules and their responses were recorded.

2.1.5. Culture fair test

After the participants had completed all the tasks, the Cattell Culture Fair (Scale 2 Form A) instrument (Cattell, 1971; Cattell & Cattell, 1973) was administered under standard conditions. Participants who had already taken this test within the last 24 months were not administered the instrument again, and the previous score was used.

2.1.6. Scoring

The intention in this experiment was to compare accuracy of response decisions across task conditions. Accordingly, in the complex case, accuracies for response and panel decisions were scored independently. To calculate accuracy for the first decision (left or right side of display), responses were scored as correct if they involved touching any element on the appropriate side of the screen. To calculate accuracy for the second decision, responses were scored as correct if they involved touching any element associated with an appropriate object type (regardless of which side it was on). Finally, to calculate accuracy for the response decision, responses were scored as correct depending on the response information within the selected panel, regardless of whether the panel was the correct one. Separate scores were obtained for critical and regular trials. Trials were classified as regular or critical based on the context cues associated with the selected panel, regardless of whether this panel was correct or not. For

Table 2
Experiment 1. Example session structure.

Task conditions and object types	No. of blocks	Block type	Trials presented
General instructions and practice task	–	–	–
Task 1 – SC condition[b] (S1: Vehicles, C2: Books)	–	Instructions and cued repetition	
	1	Practice	Vehicle (simple)
			Books (complex)
	5[a]	Standard	Vehicles (simple)
			Books (complex)
	5[a]	Mixed form	Vehicles (simple)
			Books (complex, simple)
	–	Cued recall	
Task 2 – CC condition[b] (C1: Faces, C2: Animals)	–	Instructions and cued repetition	
	1	Practice	Faces (complex)
			Animals (complex)
	5[a]	Standard	Faces (complex)
			Animals (complex)
	5[a]	Mixed form	Faces (complex, simple)
			Animals (complex, simple)
	–	Cued recall	
Task 3 – SS condition[b] (S1: Coins, S2: Clocks)	–	Instructions and cued repetition	
	1	Practice	Coins (simple)
			Clocks (simple)
	5[a]	Standard	Coins (simple)
			Clocks (simple)
	5[a]	Mixed form	Coins (simple)
			Clocks (simple)
	–	Cued recall	
Task 4 – CS condition[b] (C1: Words, S2: Cards)	–	Instructions and cued repetition	
	1	Practice	Words (complex)
			Cards (simple)
	5[a]	Standard	Words (complex)
			Cards (simple)
	5[a]	Mixed form	Words (complex, simple)
			Cards (simple)
	–	Cued recall	
Cattell culture fair test	–	–	–

[a] Within each sub-task, standard and mixed form blocks were interleaved, in random order.
[b] SS – simple–simple; SC – simple–complex; CS – complex–simple; CC – complex–complex.

simple sub-tasks, only response decisions were required and scored.

2.2. Results

For major analyses, data were taken only from standard blocks, i.e. those for which all trials were in the form (simple or complex) appropriate to the task condition. For the simple–simple condition, all blocks were included as standard and mixed form blocks were identical. In complex sub-tasks, panel decisions were extremely accurate (Decision 1: $M = 98.2$; Decision 2: $M = 98.1$), and not of primary interest. Major analyses concern accuracy of response decisions, especially for critical trials.

Results are presented in 3 sections. First we consider how errors in the present tasks relate to previous cases of goal neglect, in particular through strong correlation with Culture Fair score, and many instances of extreme failure to obey task rules even when they were understood and remembered. Second we consider the critical effects of complexity. Third we consider the mixed form blocks, confirming that, in this experiment, errors were controlled by task rather than trial complexity.

2.2.1. Goal neglect

Previously, neglect has been shown to be largely restricted to people in the lower part of the fluid intelligence range. In our tasks, for critical trials, the accuracy of response decisions (mean across all sub-tasks) showed a strong correlation with Culture Fair IQ ($r = 0.57$, $p < 001$). The correlation was somewhat weaker for the regular trials ($r = 0.33$, $p = 0.06$). Fig. 4 shows a scatter plot relating response decision accuracies to Culture Fair IQs. Especially for critical trials, these data show a dramatic fall-off in accuracy with IQ scores below 100, with very high error rates in some individual participants. As in previous goal neglect experiments, even though rules were correctly repeated before performance began, they were often poorly followed by low-Culture Fair participants.

Neglect is usually manifest in the form of major failures of performance that are persistent throughout the task. In order to discover such cases in our data, for each participant we calculated response decision accuracy for each sub-task separately, so that each participant contributed eight separate scores. Distributions of accuracy in single sub-tasks are shown in Fig. 5, separately for participants with a Culture Fair IQ below and above 100. The results demonstrate that such major failures of performance did
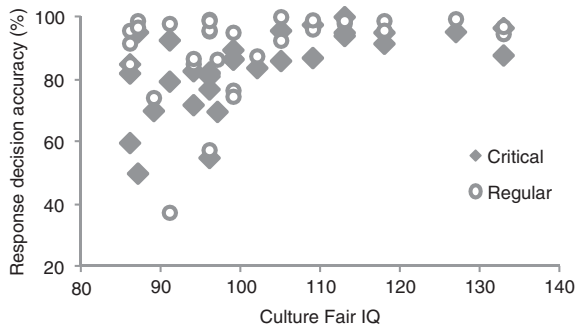
Fig. 4. Experiment 1. Relationship of response decision accuracy to Culture Fair IQ for both critical (diamonds) and regular (open circles) trials. Each point represents data from a single participant averaged across all tasks.

indeed occur in our tasks. In the majority of sub-tasks, performance was in the highest accuracy bin (80–100%). For participants with Culture Fair IQ above 100, performance rarely went below 50%. On the other hand, for participants with a Culture Fair IQ below 100, where neglect is expected, performance fell in the lowest accuracy bin (0–20%) in a significant proportion of sub-tasks. Such neglect cases were seen both on critical and regular trials.

We next looked at the pattern of errors participants made in cases of major performance failure. We examined all cases where response decision accuracy was at 50% or below and classified them on the basis of the most common type of error made. On critical trials, we found 30 cases where accuracy was at or below 50%. The most common error pattern (11 of 30 cases) was the frank neglect of the context cue, with participants responding to all trials as if they were regular trials. Another common error pattern (8 of 30 cases) reflected confusion about the response rule to be used on critical trials, with participants using response rules that were relevant in the other sub-task or

in previously completed tasks. On regular trials, we found 19 cases where accuracy was at 50% or below. By far the most common error pattern, found in 10 of the 19 cases, involved participants treating regular trials as critical trials.

A striking feature of previous goal neglect experiments was the finding that, when probed, participants could often correctly recall a task rule that they had neglected during performance. Therefore, we examined whether response decision errors in our tasks could be explained by the forgetting of task rules. We looked at all cases where performance on regular or critical trials was at 50% or below and asked whether the specific errors made by the participants could be explained by their description of the task rules after the task was completed. We found that only 9 of the 30 cases on critical trials and 8 of the 19 cases on regular trials could be explained by the participant's failure to remember the rules correctly.

To exclude the possibility that complexity effects are driven by rule forgetting, all cases in which post-task recall was incorrect were removed from subsequent accuracy analysis.

2.2.2. Complexity effects

The central interest in this experiment was the effect of task complexity. To address this question, response decisions of each sub-task were scored as a function of both this sub-task's complexity (*same sub-task complexity*) and the complexity of the other sub-task within the task (*other sub-task complexity*). Results are shown in Fig. 6. Values in Fig. 6 are means across first and second sub-tasks; for example, the value for same sub-task simple, other sub-task complex is the mean of data from the first sub-task in the simple–complex condition, and the second sub-task in the complex–simple condition. We used two performance scores, response decision accuracy (percentage
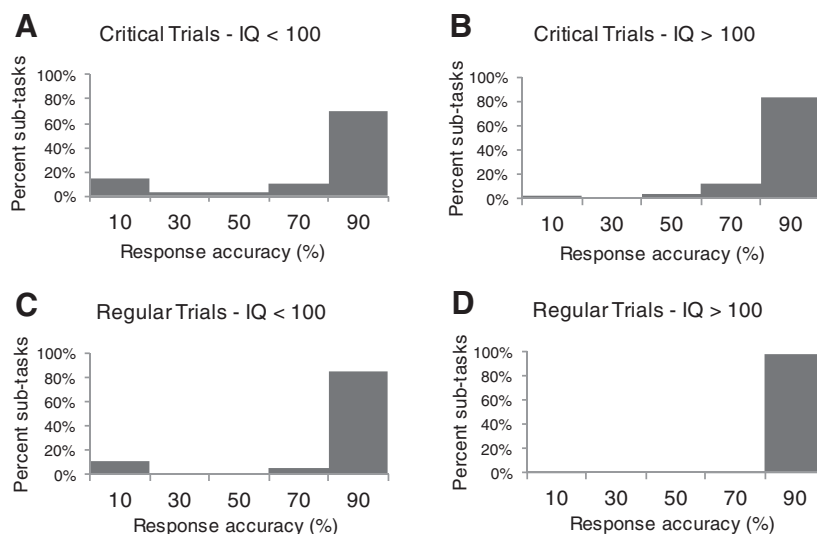

Fig. 5. Experiment 1. Histograms of sub-task performance for critical (A and B) and regular (C and D) trials for low IQ (<100) and high IQ (>100) participants. Bars represent percentage of sub-tasks with performance (response accuracy) in bins.
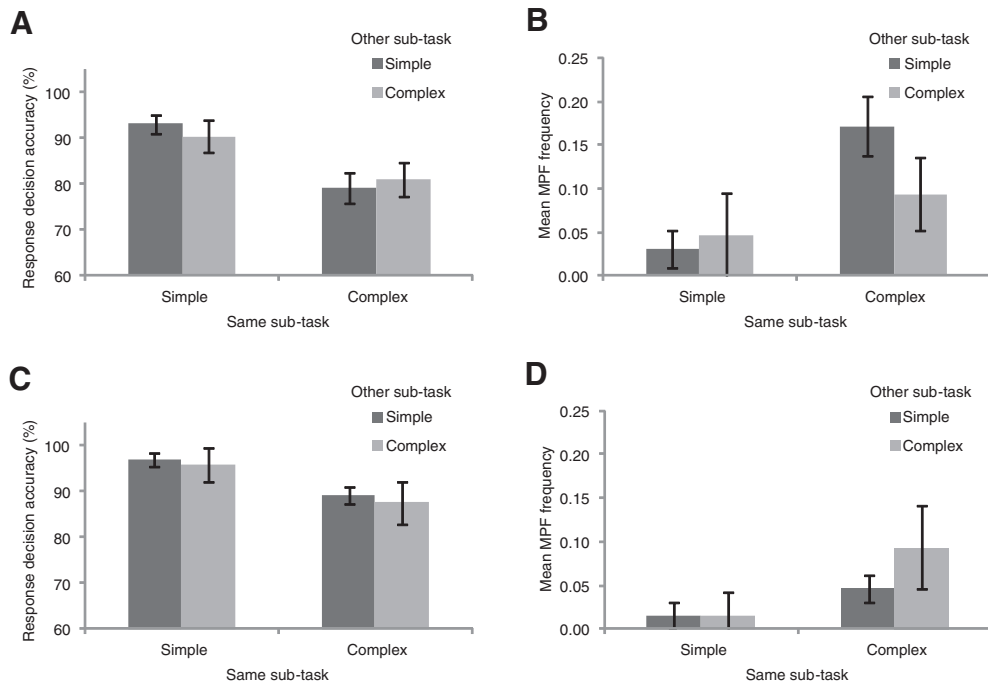
Fig. 6. Experiment 1. Mean response decision accuracies (% correct) and frequency of major performance failures (MPF) for critical (A and B) and regular (C and D) trials as a function of same sub-task and other sub-task complexity.

correct responses), and mean frequency of major performance failure.

For critical trials, the mean response decision accuracy was analysed with a 2 (same sub-task complexity; simple, complex) × 2 (other sub-task complexity; simple, complex) repeated measures analysis of variance (ANOVA). The main effect of same sub-task complexity was significant [$F(1,31) = 12.72$, $p < 0.01$, $\eta_p^2 = 0.29$] confirming poorer accuracy in the complex condition (Fig. 6A). Both the main effect of other sub-task complexity [$F(1,31) = 0.02$, $p > 0.1$, $\eta_p^2 < 0.01$] and the interaction [$F(1,31) = 0.48$, $p > 0.1$, $\eta_p^2 = 0.02$] were non-significant. The mean frequency of major performance failures on critical trials was similarly analysed. The main effect of same sub-task complexity was significant [$F(1,31) = 5.94$, $p < 0.05$, $\eta_p^2 = 0.16$] confirming poorer accuracy in the complex condition (Fig. 6B). Both the main effect of other sub-task complexity [$F(1,31) = 0.80$, $p > 0.1$, $\eta_p^2 = 0.03$] and the interaction [$F(1,31) = 2.35$, $p > 0.1$, $\eta_p^2 = 0.07$] were non-significant.

Similar results were obtained for regular trials. For mean response decision accuracy, the main effect of same sub-task complexity was significant [$F(1,31) = 5.20$, $p < 0.05$, $\eta_p^2 = 0.14$] while the main effect of other sub-task complexity was non-significant [$F(1,31) = 0.48$, $p > 0.1$, $\eta_p^2 = 0.02$]. Their interaction was also non-significant [$F(1,31) = 0.01$, $p > 0.1$, $\eta_p^2 < 0.01$]. For the frequency of major performance failures, the main effect of current sub-task complexity was significant [$F(1,31) = 5.01$, $p < 0.05$, $\eta_p^2 = 0.14$] while the main effect of other sub-task complexity [$F(1,31) = 0.81$, $p > 0.1$, $\eta_p^2 = 0.03$] and the interaction were non-significant [$F(1,31) = 0.59$, $p > 0.1$, $\eta_p^2 = 0.02$].

2.2.3. Mixed form blocks

In this experiment, our aim was to manipulate the complexity of the task model. However, in our tasks, a complex sub-task differs from a simple sub-task not just in terms of requiring a more complex task model, but also presents a greater demand during performance. It is possible, therefore, that the complexity effect we find is driven by the greater performance demand on complex trials. In order to exclude this possibility, we examined performance on mixed form blocks. In these blocks, participants were presented two types of trials. On half the trials, the sub-task appeared in its full (simple or complex) form, as described in task instructions. On the other half of the trials, even sub-tasks that were normally complex appeared in simple form. If our complexity effects are driven by greater performance demand, then response decision accuracies for any complex sub-task should be higher for the simple form trials compared to the full, complex form trials. A paired t-test found no significant difference between mean response accuracies from complex sub-tasks in full, complex form versus simple form, either for critical trials [complex trials: 81.4% ± SE 3.0, simple trials: 84.0% ± SE 3.1, $t(31) = 1.61$, $p > 0.1$, $d = 0.28$], or for regular trials [complex trials: 87.4% ± SE 2.3, simple trials: 89.4% ± SE 3.8, $t(31) = 0.81$, $p > 0.1$, $d = 0.43$]. Results were similar for mean frequency of major performance failures for critical trials [complex trials: 0.14 ± SE 0.04, simple trials: 0.14 ± SE 0.03, $t(31) = 0$, $p > 0.1$, $d = 0.00$], and regular trials [complex trials: 0.09 ± SE 0.02, simple trials: 0.09 ± SE 0.03, $t(31) = 0$, $p > 0.1$, $d = 0.00$]. Matching previous goal neglect findings (Duncan et al., 2008), these data suggest that the complexity effects we found were indeed driven by the complexity of the task model established during task

instructions, not by the performance demands of individual trials.

2.2.4. Response times

Of course, though not of direct interest, response times were also strongly influenced by same sub-task complexity. On critical trials of standard blocks, participants took, on average, 5902 ms to respond in complex sub-tasks and 3908 ms to respond in simple sub-tasks. Corresponding values for regular trials were 3053 ms and 1966 ms respectively.

2.3. Discussion

In Experiment 1 we observed the phenomenon of neglect in a set of novel tasks and reproduced a number of the findings of Duncan et al. (2008). In our tasks, performance on response decisions, especially on critical trials, was closely related to participant's fluid intelligence, with errors much more likely in the lower end of the IQ scale. We also found an effect of task complexity with these errors being significantly more frequent in complex sub-tasks, which had a greater number of task components. This complexity effect was not driven by real-time performance demand during task execution, since performance was similar even on trials involving a simplified version of a sub-task. Errors were also not explained by participants' forgetting task rules during performance, though in the complex tasks used here, final rule recall was far from perfect.

Major performance failures in our tasks took several forms. In some cases there was frank neglect of the context cue, i.e. treatment of critical trials exactly as though they were regular trials. In other cases, errors were made in the decision concerning which trials were regular and which critical, or in choosing the appropriate response on critical trials. In general, however, major performance failures of different kinds all reflected serious failure to use task rules, not usually explained by apparent explicit forgetting. In large part, major performance failures of all kinds were restricted to participants in the lower range of Culture Fair IQ (Fig. 5).

Our observation of neglect in tasks where stimulus presentation was unspeeded and response was self-timed is a novel finding. In previous goal neglect experiments, stimulus presentation was always speeded and responses were also sometimes speeded. Our results suggest that time pressure is not a critical factor in the neglect of task rules, and strengthen the conclusion that neglect is not related to real-time demands during task execution.

Our principal finding is that, on critical trials, both response decision accuracy as well as frequency of major performance failure were strongly sensitive to the complexity of the same sub-task, but not to the complexity of the other sub-task in the overall task. Similar, though weaker, results were seen for regular trials. These findings show that performance is not controlled simply by the total complexity of task knowledge described in initial instructions (Duncan et al., 2008). At the same time, like previous findings, the results also rule out complexity of the current trial as the critical factor, since for complex

sub-tasks in mixed form blocks, performance did not improve when the stimulus on a given trial was simple.

This pattern of results is consistent with the chunking account developed earlier. On this account, when task performance begins, it must be shaped by the knowledge provided in task instructions. As in all complex behaviour, a critical process will be constructing a series of chunks or attentional episodes, each focusing on just one part of the whole space of action alternatives and relevant knowledge (Sacerdoti, 1974). The present tasks were designed such that, at least in principle, response and panel decisions were entirely independent. At the response decision phase, knowledge guiding panel decisions was always irrelevant, either within the same or other sub-task. Accordingly, a successfully-focused attentional episode bearing just on response decisions might have been independent of either same or other sub-task complexity. Instead, added panel decisions within the same sub-task substantially impaired response decisions, while added panel decisions in the other sub-task had little or no effect. Especially in participants with low fluid intelligence, the results suggest a limitation in separating the contents of a single sub-task into effective chunks; so that when response decisions were made, access to the correct rules was impaired by the existence within the sub-task of additional, but now irrelevant rules. As mentioned above, later we shall present evidence that critical failures take place over the first few performance trials, as knowledge is first used to shape task performance.

In our conception of an attentional episode, relevant knowledge from the stored task model is combined with current sensory input to shape correct behaviour. Given the conditions of the current experiment – with complex rules and unlimited decision time – it is appealing to express failure in terms of failed rule retrieval. In these terms, the conclusion might be that response-decision rules are easier to retrieve when they are not strongly linked to other, currently irrelevant rules – specifically, other rules within the same sub-task. This account has some similarity with accounts of strategic retrieval in working memory and recall tasks, which emphasise the importance of episodic context in delimiting search through stored knowledge (Capaldi & Neath, 1995; Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005; Howard & Kahana, 2002; Polyn, Norman, & Kahana, 2009; Unsworth & Engle, 2007a, 2007b). To explain correct recall of neglected rules at the end of a task, it might be thought that the context cues available to participants during performance are very different from the ones available in a final recall test. In many ways, the latter context may better match the encoding context, providing more effective cues and improving recall of task rules.

We would suggest, however, that retrieval of complex task instructions is only one special case of the general problem of dividing complex problems into effective attentional episodes. In prior cases of goal neglect, even simple, easily-retrieved rules may be neglected if they are not strongly cued. In a speeded task, for example, participants may neglect briefly-presented arrows pointing left or right, indicating which part of a screen to monitor for critical events (Duncan et al., 2008). In this case, presumably, the

difficulty is not so much retrieving the rule associated with an arrow, but triggering this rule at the right moment in the task. In other cases of forming useful attentional episodes, it is information in stimulus materials themselves that must be organised into useful parts. Perhaps the most obvious example is the problem-solving of standard fluid intelligence tests, requiring search through a large space of possible cognitive operations for useful parts of a solution. For these reasons, we prefer the more general formulation in terms of attentional episodes to a more specific emphasis on complex rule retrieval.

A number of factors may determine whether a pair of task requirements is chunked into same or different parts of a task representation or model. One factor, especially relevant in the context of goal neglect, may be the format in which task knowledge was presented to participants. Duncan et al. (2008) have linked goal neglect to events that take place as the participant's body of task knowledge is established during instruction. In our experiment, instructions to participants were provided separately for each sub-task, one after the other. It is possible that this may have a bearing on the chunking we observed in our tasks. Experiment 2 was designed to test this hypothesis.

3. Experiment 2

In Experiment 2 we asked whether the chunking of task information into sub-task groups was driven by the format in which the instructions were presented. If the task model is assembled as task instructions are being processed, it is plausible that the format of the instructions has a bearing on chunking. In Experiment 1, instructions for the two sub-tasks were presented in chunks, one after the other, providing participants an opportunity to consolidate the first set of instructions before the second set were presented. It is possible that this format of instructions drives chunking of rules into sub-task groups.

In order to test this possibility, in Experiment 2 we repeated the complexity manipulations of the previous experiment, but between participants, we manipulated the format in which instructions were presented. One group received instructions in a 'chunked' format, grouped by sub-task much as they were in the first experiment. First, all the instructions for one sub-task were read out followed by the instructions for the second sub-task. The other group received instructions in an 'interleaved' format. In this case, rules were grouped orthogonally by levels rather than by sub-task. First, rules for panel decisions from both sub-tasks were presented together, followed by the rules for selecting response boxes in both sub-tasks, and finally the response rules for regular and critical trials.

If the format of the instructions drives chunking of task requirements, the task requirements from the two sub-tasks should be harder to separate in the interleaved instructions group and we should find an effect of other sub-task complexity. In the chunked instructions group, the instructions are separate and should encourage sub-task chunking as before. On the other hand, if chunking is driven by the inherent structure of task requirements, sub-task chunking should be similar for both the groups and the effect of other sub-task complexity should be weak or absent.

3.1. Methods

3.1.1. Participants

Experiment 2 had 48 right-handed adult participants (25 males, 23 females; age-range: 38–69, $M = 55.5$, $SD = 6.9$). 24 were in the chunked instructions group and another 24 were in the interleaved instructions group. Participants had no history of neurological disorder and were recruited from the paid volunteer panel of the MRC Cognition and Brain Sciences Unit. Participants took part after informed consent was obtained.

3.1.2. Protocol

The tasks and stimuli used for Experiment 2 were the same as those from the previous experiment. The only changes from Experiment 1 concerned the way that instructions were presented and the way in which the cued recalls of the rules were conducted. Both of these differed for the chunked and interleaved instruction groups.

Each new task began with an instruction period. For both groups, print-outs of sample trials for both sub-tasks (Fig. 7) were placed before the participant. The experimenter then read out the rules verbatim while pointing to the various stimulus elements on the print-outs. Participants were told that the tasks were self-timed but that they should respond as quickly as they could while being accurate. To illustrate the different instruction formats, consider the example shown in Fig. 7, using Task 1 (see Fig. 2) in the complex–simple condition.

The instructions given to the two groups for this task are shown in Table 3. In both groups, instructions were organised into three 'steps'. The first step involved selecting one of the four panels on the screen. When the sub-task was in simple form, there was only one panel and the first step was not required. The second step involved selecting one of the two labelled response boxes in the panel. The third step involved choosing the appropriate response based on whether the trial was regular or critical. In the chunked instruction group, instructions were organised first by sub-task and, within each sub-task, by step. So, participants first received instructions for steps 1–3 for the first sub-task followed by the corresponding instructions for the second sub-task. This was similar to the format of the instructions used in Experiment 1. In the interleaved instructions group, the instructions were organised by steps first and then by sub-task. So, participants received instructions for step 1 for both the sub-tasks, followed by instructions for step 2 for both sub-tasks and finally for step 3. Thus, for the interleaved instruction group, the format of the instructions encouraged chunking by step rather than by sub-task.

After all the instructions for the task had been read out, the sample trial print-outs were removed and the participants were asked to recall all the rules of the task. To aid recall, participants were cued on each rule by being asked a series of questions. The cued-recall used the same format as the instructions for each group.
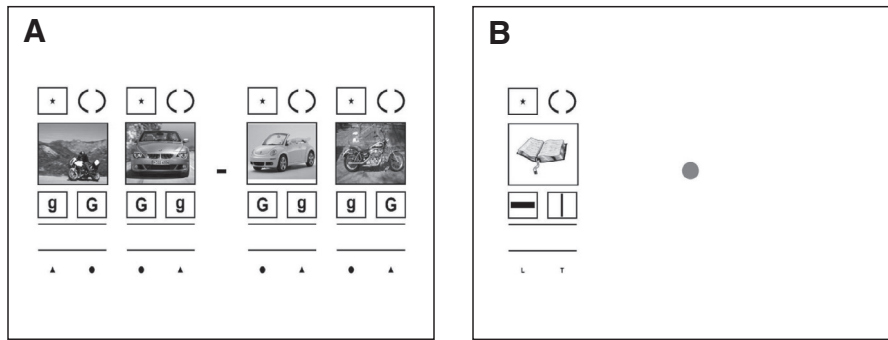
Fig. 7. Sample print out accompanying task instructions (Task 1 from Fig. 2, complex–simple condition). Vehicles (A) are in complex form and books (B) are in simple form. Actual print-outs were in colour.

Table 3
Verbatim instructions accompanying Fig. 7 (Task 1, complex–simple condition).

Chunked group	Interleaved group
In this task you will see screens that look like this (SHOW)	In this task you will see screens that look like this (SHOW)
In some screens there are vehicles, like this one (SHOW). The first step is to select one of the four pictures on the screen. To do that, you first look at the centre of the screen. The symbol is a − or a + sign. If it is a −, you look at the vehicles on the left half of the screen, while for a +, you look to the right. Now, on the selected side, you see a motorbike and a car. You should select the bike. (POINT)	In some screens there are vehicles, like this one (SHOW). The first step is to select one of the four pictures on the screen. To do that, you first look at the centre of the screen. The symbol is a − or a + sign. If it is a −, you look at the vehicles on the left half of the screen, while for a +, you look to the right. Now, on the selected side, you see a motorbike and a car. You should select the bike. (POINT)
The second step is to select one of the boxes below the selected picture. Below the vehicle, you will see a pair of letters (POINT). You should select the lower case letter (POINT)	In other screens, there is a single book so it is already selected. The book can appear in various locations
Finally, the third step is to make a response. It works like this: Normally, you touch the lower case letter, except that occasionally, just beneath the lower the lower case letter you see a dot. When this happens, you should touch the dot instead of the letter (POINTING THROUGHOUT)	The second step is to select one of the boxes below the selected picture. Below the vehicle, you will see a pair of letters (POINT). You should select the lower case letter (POINT). Below the book, you will see a pair of lines (POINT). You should select the horizontal line (POINT)
In other screens, there is a single book so it is already selected. The book can appear in various locations	Finally, the third step is to make a response
The second step is to select one of the boxes below the selected picture. Below the book, you will see a pair of lines (POINT). You should select the horizontal line (POINT)	For vehicles, it works like this: Normally, you touch the lower case letter, except that occasionally, just beneath the lower the lower case letter you see a dot. When this happens, you should touch the dot instead of the letter (POINTING THROUGHOUT)
Finally, the third step is to make a response. It works like this: Normally, you touch the horizontal line, except that occasionally, just beneath the horizontal line you see the letter T. When this happens, you should touch both the horizontal and vertical line (in that order) instead of just the horizontal line (POINTING THROUGHOUT)	For books, it works like this: Normally, you touch the horizontal line, except that occasionally, just beneath the horizontal line you see the letter T. When this happens, you should touch both the horizontal and vertical line (in that order) instead of just the horizontal line (POINTING THROUGHOUT)

To illustrate again using the Vehicles–Books example (Fig. 7), the questions for the chunked instructions group would be: "In step 1, how do you decide whether to focus on the left or the right? Which of the two vehicles do you pick? In step 2, which of the two letters do you pick? Finally, how do you usually respond? What do you do when you see a dot below the lower case letter?" – followed by equivalent questions for books.

For the interleaved instructions group, the questions would be: "For the vehicles, in step 1, how do you decide whether to focus on the left or the right? Which of the two vehicles do you pick? For the books, step 1 has already been done for you. In step 2, for the vehicles, which of the two letters do you pick? And, for the books, which of the two lines do you pick? Finally, in step 3, how do you usually respond in the case of the vehicles? And, what do you

do when you see a dot below the lower case letter? And, for the books, how do you usually respond? What do you do when you see the letter T below the horizontal line?" If any error was made during cued recall, the experimenter would correct the participant, and the entire cued-recall sequence would be repeated until the participant recalled all components of both sub-tasks correctly. Participants were then told that they could not be reminded of the rules again. The total number of cued recalls required before the rules were correctly repeated was similar in both groups. The chunked instructions group required between 1 and 3 repetitions ($M = 1.17$, SE = 0.03, Mode = 1), while the interleaved instructions group required between 1 and 4 repetitions ($M = 1.33$, SE = 0.05, Mode = 1).

As in Experiment 1, after the instruction period, participants performed a practice block with 12 trials, which

included 5 regular trials and one critical trial for each sub-task intermixed randomly. Following this, participants performed 10 blocks of 16 trials each, each block including 6 regular trials and 2 critical trials for each sub-task intermixed randomly. 5 randomly chosen blocks were standard blocks while the rest were mixed form blocks. After all the blocks were completed, participants were again asked to do a single cued recall of all rules.

3.2. Results

Scoring and analysis methods were identical to those used in Experiment 1. Again main analyses concerned accuracy of response decisions in standard blocks. Again, accuracies for panel decisions were very high for the chunked instructions group (Decision 1: $M = 98.0$; Decision 2: $M = 98.0$) as well as the interleaved instructions group (Decision 1: $M = 97.0$; Decision 2: $M = 98.2$).

3.2.1. Goal neglect

For the chunked instructions group, the accuracy of response decisions on critical trials (mean across all sub-tasks) showed a strong correlation with Culture Fair IQ ($r = 0.68$, $p < 0.001$). The correlation was weaker and non-significant for regular trials ($r = 0.25$, $p > 0.1$). Corresponding correlations for the interleaved instructions group were 0.50 ($p < 0.05$) for critical trials and 0.32 ($p > 0.1$) for regular trials. Scatterplots for both groups are shown in Figure 8.

To assess neglect, we again focussed on cases of major performance failures where response decision accuracy was at 50% or lower. Data from the two groups were pooled as they were closely similar. In total there were 48 cases of major performance failure on critical trials and 28 cases on regular trials. On critical trials, there were two common error patterns. In 20 cases, participants confused the two context cues. Another 17 cases reflected confusion about the response rule to be used on critical trials, with participants using response rules that were relevant in the other sub-task or in previously completed tasks. Frank neglect of the context cue occurred in only 6 cases. On regular trials, again, the most common error pattern, seen in 20 cases, involved confusing the two context cues. In summary, error patterns were largely similar to those found in Experiment 1 with the exception

that frank neglect of the context cue was relatively uncommon.

Finally, we examined whether response decision errors in our tasks could be explained by the forgetting of task rules. For all cases where performance on critical or regular trials was at 50% or below, we asked whether the specific errors made by the participants could be explained by their failure to remember the rules correctly. Across the two instruction groups, we found that 23 of the 48 cases on critical trials and 13 of the 28 cases on regular trials could be explained by the rule reported by participants at the end of the task. These cases were excluded from subsequent accuracy analyses.

3.2.2. Complexity effects

In this experiment, our main question was whether the format of instructions influenced complexity. Mean response decision accuracies as a function of same and other sub-task complexity are shown in Figs. 9 (critical trials) and 10 (regular trials).

On critical trials, mean response accuracies were analysed with a 2 (same sub-task complexity; simple, complex) × 2 (other sub-task complexity; simple, complex) × 2 (group; chunked instructions, interleaved instructions) mixed model ANCOVA with group as a between subject factor and Culture Fair IQ as a covariate. The main effect of same sub-task complexity was again significant [$F(1,45) = 19.95$, $p < 0.001$, $\eta_p^2 = 0.31$] showing poorer accuracy in the complex condition (Fig. 9A and C). The main effect of other sub-task complexity was non-significant [$F(1,45) = 0.39$, $p > 0.1$, $\eta_p^2 = 0.01$], as was the interaction [$F(1,45) = 0.06$, $p > 0.1$, $\eta_p^2 < 0.01$]. The main effect of group and all interactions with this factor were non-significant (all $ps > 0.1$).

Mean frequencies of major performance failures showed a similar pattern. The main effect of same sub-task complexity was again significant [$F(1,45) = 10.25$, $p < 0.01$, $\eta_p^2 = 0.19$] showing poorer accuracy in the complex condition (Fig. 9B and D). The main effect of other sub-task complexity was non-significant [$F(1,45) = 2.34$, $p > 0.1$, $\eta_p^2 = 0.05$], as was the interaction [$F(1,45) = 1.19$, $p > 0.1$, $\eta_p^2 < 0.03$]. The main effect of group and all interactions with this factor were non-significant (all $ps > 0.1$).

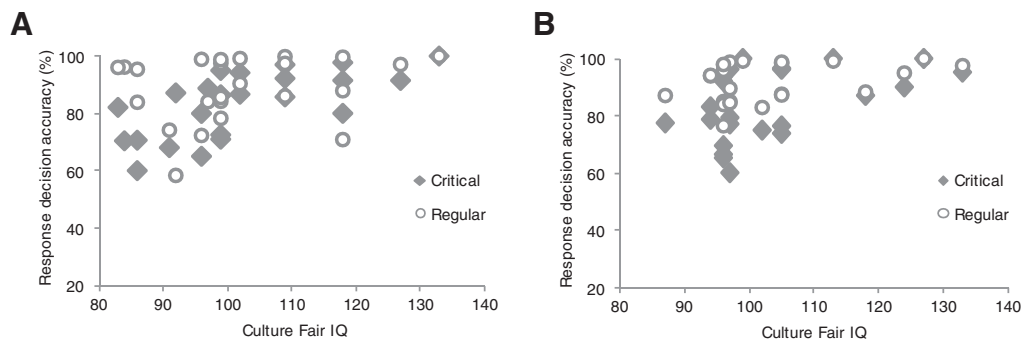On regular trials, mean response accuracies were similarly analysed. This time the main effect of same sub-task

Fig. 8. Experiment 2. Relationship of response decision accuracy to Culture Fair IQ for the chunked (A) and interleaved (B) instructions group in both critical (diamonds) and regular (open circles) trials. Each point represents data from a single participant averaged across all tasks.
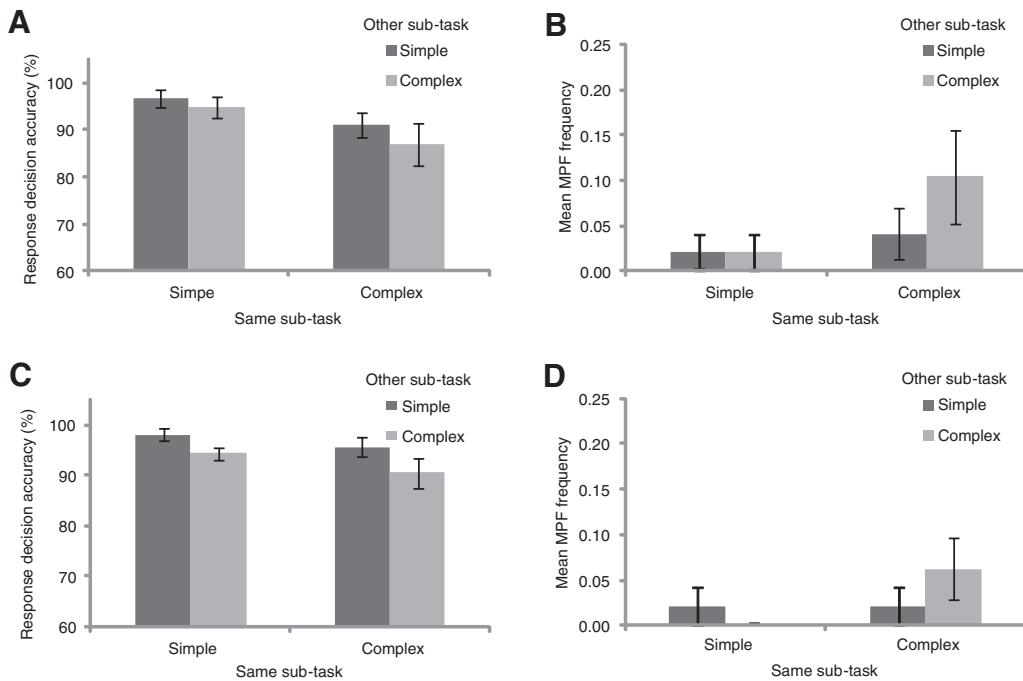
Fig. 9. Experiment 2: Critical trials. Mean response decision accuracies (% correct) and frequency of major performance failures (MPF) for the chunked instruction group (A and B) and the interleaved instructions group (C and D) as a function of same sub-task and other sub-task complexity.

complexity was non-significant [$F(1,45) = 2.99$, $p > 0.05$, $\eta_p^2 = 0.06$] as was the main effect of other sub-task complexity [$F(1,45) = 0.38$, $p > 0.1$, $\eta_p^2 = 0.01$] and the interaction [$F(1,45) = 0.46$, $p > 0.1$, $\eta_p^2 = 0.01$]. The main effect of group and all interactions with this factor were also non-significant (all $ps > 0.1$).

Mean frequency of major performance failures on regular trials were similarly analysed. Again, the main effects of same sub-task complexity [$F(1,45) = 2.15$, $p > 0.1$, $\eta_p^2 = 0.05$] and other sub-task complexity [$F(1,45) = 0.23$, $p > 0.1$, $\eta_p^2 < 0.01$], as well as the interaction [$F(1,45) = 0.69$, $p > 0.1$, $\eta_p^2 = 0.02$] were all non-significant. The main effect of group and all interactions with this factor were also non-significant (all $ps > 0.1$).

3.2.3. Mixed form blocks

Finally we examined performance in mixed form blocks. Data from the two groups were pooled. Paired *t*-tests found no significant difference between mean response decision accuracies from complex sub-tasks in full, complex form versus simple form, either for critical trials [complex trials: 79.7% ± SE 3.1, simple trials: 81.5% ± SE 3.0, $t(47) = 1.41$, $p > 0.1$, $d = 0.20$], or for regular trials [complex trials: 87.6% ± SE 2.7, simple trials: 89.6% ± SE 2.8, $t(47) = 1.62$, $p > 0.1$, $d = 0.23$]. These data suggest that the complexity effects we found were indeed driven by the complexity of the task model established during task instructions, not by the real-time demands of individual trials.

3.2.4. Response times

On critical trials, participants in the chunked instructions group took, on average, 3535 ms to respond in complex sub-tasks and 2291 ms to respond on simple sub-tasks. Corresponding values for the interleaved instructions group were 3602 ms for complex sub-tasks and 2159 ms for simple sub-tasks.

On regular trials, response times for the chunked instructions group were 3389 ms for complex sub-tasks and 2060 ms for simple sub-tasks, with corresponding values of 3361 ms and 2010 ms for the interleaved instructions group.

3.3. Discussion

In Experiment 2, we replicated our findings from Experiment 1. Response decisions were prone to errors and accuracies strongly correlated with Culture Fair IQ, especially on critical trials. Major performance failures were common, with a variety of error types indicating serious failure to use task rules. Again, such failures were not generally explained by failure of explicit recall. We also confirmed the pattern of complexity effects that we observed in Experiment 1. For critical trials, both the accuracy of response decisions as well as the frequency of major performance failures were strongly sensitive to same sub-task complexity, but largely insensitive to the complexity of the other sub-task. For regular trials, unlike in Experiment 1, neither effect was significant. Again, the data suggest that the key factor in these experiments was not the complexity of all task requirements described during instruction but the complexity of a subset of the requirements which are relevant to a single sub-task.

The results from Experiment 2, however, do not support the hypothesis that chunking is driven by the format in which the instructions are presented. Instructions in the
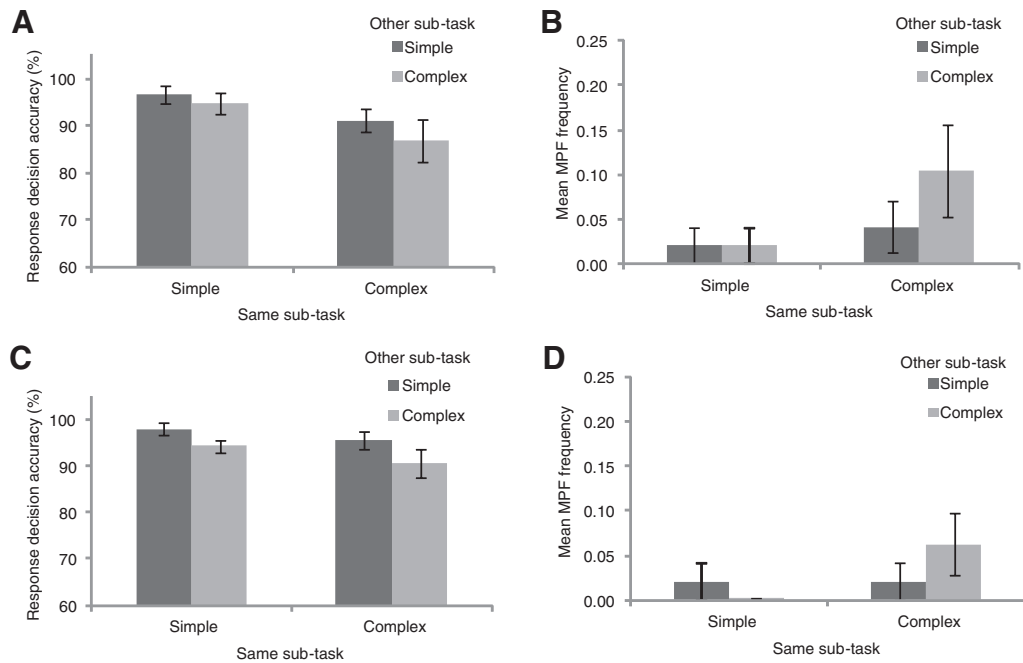
Fig. 10. Experiment 2: Regular trials. Mean response decision accuracies (% correct) and frequency of major performance failures (MPF) for the chunked instruction group (A and B) and the interleaved instructions group (C and D) as a function of same sub-task and other sub-task complexity.

two groups were organised quite differently, to encourage orthogonal forms of chunking. The performance of the two groups, however, was closely similar.

4. Pooled data

The principal concern in our experiments was the differential effect of same sub-task versus other sub-task complexity. The main finding from Experiment 1, replicated in Experiment 2, was a strongly significant main effect of same sub-task complexity, especially on critical trials, but no effect of other sub-task complexity. An inspection of Figs. 6, 9 and 10, however, suggests a small trend toward lower accuracies for complex other sub-tasks compared to simple other sub-tasks. We therefore re-analysed the complexity effects after pooling the data from the standard blocks of Experiments 1 and 2. Given that we found no effect of group in Experiment 2, we also included the data from the interleaved instructions group.

On critical trials, mean response accuracies were analysed using a 2 (same sub-task complexity; simple, complex) $\times$ 2 (other sub-task complexity; simple, complex) repeated measures ANOVA. The main effect of same sub-task complexity was significant [$F(1,79) = 46.82$ $p < 0.001$, $\eta_p^2 = 0.37$]. Both the main effect of other sub-task complexity [$F(1,79) = 2.14$, $p > 0.1$, $\eta_p^2 = 0.03$] and the interaction [$F(1,79) < 0.01$, $p > 0.1$, $\eta_p^2 < 0.01$] were non-significant. The mean frequency of major performance failures was similarly analysed. The main effect of same sub-task complexity was significant [$F(1,79) = 21.58$ $p < 0.001$, $\eta_p^2 = 0.22$]. Both the main effect of other sub-task complexity [$F(1,79) = 0.02$, $p > 0.1$, $\eta_p^2 < 0.01$] and the interaction

[$F(1,79) = 0.21$, $p > 0.1$, $\eta_p^2 < 0.01$] were non-significant. Therefore, even after pooling the data, the pattern of complexity effects for critical trials remained unchanged.

On regular trials, data were analysed using a 2 (same sub-task complexity; simple, complex) $\times$ 2 (other sub-task complexity; simple, complex) repeated measures ANOVA. The main effect of same sub-task complexity was significant [$F(1,79) = 12.67$, $p < 0.01$, $\eta_p^2 = 0.14$]. This time, the main effect of other sub-task complexity was also significant [$F(1,79) = 4.47$, $p < 0.05$, $\eta_p^2 = 0.04$]. Their interaction was non-significant [$F(1,79) = 0.20$, $p > 0.1$, $\eta_p^2 < 0.01$]. The mean frequency of major performance failures was similarly analysed. The main effect of same sub-task complexity was significant [$F(1,79) = 10.01$ $p < 0.01$, $\eta_p^2 = 0.11$]. Both the main effect of other sub-task complexity [$F(1,79) = 1.98$, $p > 0.1$, $\eta_p^2 < 0.02$] and the interaction [$F(1,79) = 2.51$, $p > 0.1$, $\eta_p^2 < 0.03$] were non-significant. After pooling the data, a small effect of other sub-task complexity was now detectable on regular trials.

Importantly, we do not claim that there is no effect of other sub-task complexity, but only that the effect of same sub-task complexity is dominant. A final analysis directly compared the effects of same sub-task complexity and other sub-task complexity. To this end we compared response decision accuracies for two cases: same sub-task simple and other sub-task complex, vs same sub-task complex and other sub-task simple. In these two conditions, complexity of overall task instructions is matched (one simple and one complex sub-task in each case). A comparison of these conditions can be seen as a test of the relative effects of same and other sub-task complexity. For critical trials, mean accuracy in the case of same sub-task simple/other sub-task complex was 92.4%, while

mean accuracy for same sub-task complex/other sub-task simple was 82.8%, a significant difference [t (79) = 3.67, $p < 0.001$, $d = 0.41$]. For regular trials, mean accuracy in the case of same sub-task simple/other sub-task complex was 95.1%, while mean accuracy for same sub-task complex/other sub-task simple was 91.7%. The difference was not significant [$t(47) = 1.6$, $p > 0.1$, $d = 0.20$].

Similar results were obtained for major performance failures. For critical trials, mean frequency of major performance failures in the case of same sub-task simple/other sub-task complex was 0.03, while that for same sub-task complex/other sub-task simple was 0.12, a significant difference [$t(79) = 3.00$, $p < 0.01$, $d = 0.34$]. For regular trials, mean frequency of major performance failures in the case of same sub-task simple/other sub-task complex was 0.01, while mean accuracy for same sub-task complex/other sub-task simple was 0.04. The difference was not significant [$t(47) = 1.7$, $p > 0.1$, $d = 0.18$].

Our results confirm that, for critical trials, it was the complexity of the same sub-task that most strongly influenced performance, with a similar but non-significant trend for regular trials.

5. Task dynamics

Our failure to find an effect of instructional format despite a strong manipulation led us to consider the possibility that the instruction period may be relatively less important in the production of goal neglect. Instead actual task structure, rather than instruction structure, may be critical. On the account we are proposing, neglect reflects a failure to focus on just those aspects of task knowledge relevant to a current decision. This thought led us to consider events on early trials, when task knowledge is first used to control behaviour (Duncan et al., 1996).

Additionally, in both our experiments, the effect of same sub-task complexity on neglect persisted even on the simplified trials of mixed form blocks, where panel decisions were not present. This observation also suggests that a pattern of behaviour is constructed early during the task and then tends to remain stable.

In order to explore the role of early performance in task assembly and neglect, we examined the trial-wise dynamics of task performance in cases of major performance failure, defined as before as response decision accuracy at 50% or below. Across both experiments, for critical trials there were 78 cases, contributed by 46 of the 88 participants. For regular trials there were 37 cases, contributed by 31 of the 88 participants.

In all these cases of major performance failure, participants typically followed a consistent strategy in the performance blocks, with one major error type dominating performance. We asked how participant's strategy evolved over the early trials. For every such case, we defined two major response strategies – the correct strategy and the dominant incorrect strategy. We then computed the probability on every trial, across all cases of major performance failure, of choosing a response consistent with these strategies.

Fig. 11 plots these probabilities separately for critical (Fig. 11A) and regular (Fig. 11B) trials. The horizontal axis shows trial number. Note that the first critical trial and the first 5 regular trials are part of the practice block. Across trials, the figure shows the probability of correct response decision, the probability of making a response that is consistent with the dominant incorrect strategy, and the response time.

In these data, the probability of choosing the correct response starts off near 0.4 and then decreases over the first 10 trials to settle close to zero. Conversely, the probability of choosing a response consistent with the dominant incorrect strategy also starts off near 0.4 and increases over the same period to asymptote close to one. Response time rapidly falls during these trials to settle at a stable level after the first 10 trials.

This pattern of results shows an important role for early performance in shaping behaviour. The data reveal an initial phase during which performance is inconsistent, with participants being just as likely to choose the correct response as an incorrect response. Performance becomes increasingly stable over the first 10 or so trials as one response rule is privileged over others. Such stabilization
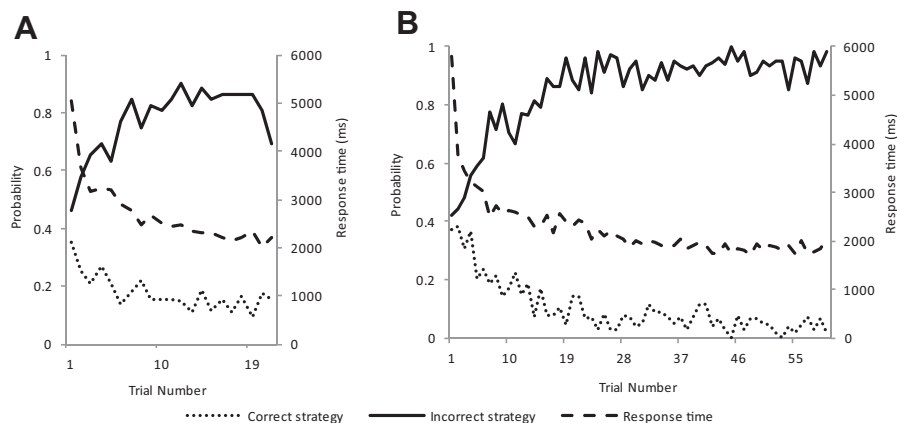

Fig. 11. Pooled data. Probability of following correct strategy (dotted line), probability of following incorrect strategy (solid line) and response time (dashed line), plotted as a function of trial number for critical trials (A) and regular trials (B). Data are averaged across all cases of major performance failure in Experiments 1 and 2.

may reflect a process of automatisation in which initially encoded rules are compiled into procedures thus reducing cognitive demand (Neves & Anderson, 1981). We suggest that this period represents a learning phase during which task performance is shaped. During this period, as participants make response decisions, they have to search through the space of task knowledge for the relevant rules. When a sub-task contains additional panel rules, the chance of implementing the correct response rule is reduced, resulting in an incorrect choice. With every trial, the choice of behaviour becomes more stable. In cases of major performance failure, an incorrect rule is privileged in the final, stable performance stage.

6. General discussion

The results presented here significantly extend the account of goal neglect suggested by Duncan et al. (2008). In that account, the critical factor is the total complexity of task instructions. Duncan et al. (2008) viewed task model assembly as a process of encoding and storing task information into a limited capacity working memory. As more information is entered into the task model and this storage capacity is filled, task rules compete for representation and individual rules may be weakly represented and lost, leading to their neglect during performance.

This simple account does not easily explain the pattern of complexity effects that we observe. For our tasks, this account would predict that the neglect of a requirement should be equally sensitive to same or other sub-task complexity, since in either case, added panel decisions have the same effect on total task complexity. Instead, the effect of same sub-task complexity was substantially stronger.

Our new account builds on a range of previous work suggesting that task complexity is a matter not simply of the total amount of information in a task description. Also critical are structural relations between rules (e.g. Frye et al., 1998), and in particular, chunking into separate task parts (e.g. Halford et al., 2007; Newell, 1990). Reflecting the critical role of information chunking, our account stresses the creation of focused attentional episodes, dividing complex task requirements into simpler, more manageable parts. Such division allows even complex problems to be solved as a series of simpler sub-problems (Sacerdoti, 1974). In the present experiments, the critical task part was the response decision and its sensitivity to additional panel decisions in the overall task model. Though response and panel decisions were independent, response-decision neglect was increased when panel decisions were added. This effect of added complexity, however, was restricted to panel decisions added within the same sub-task. The results suggest a failure in creation of a focused response-decision episode when response and panel rules were conceptually linked. As a result, we suggest, implementation of the correct response rule was impaired, leading in extreme cases to systematic neglect.

Our account shares some ideas with relational complexity theory (Halford et al., 1998, 2007; Waltz et al., 1999) which posits a co-ordinate system or ordering schema for

representing relations necessary for carrying out a cognitive process, and proposes a mechanism of 'segmentation' for parsing a complex task into separate parts. In our experiments, independent parts of a task that were conceptually linked were harder to separate into separate episodes. This suggests the existence of structural representations that relate the various steps of the decision process. Such a representation could serve as a schema and shape not just behaviour but also the way task knowledge is organised. Indeed, in the interleaved instructions group in Experiment 2, participants often spontaneously and rapidly re-organised instructions by sub-task during verbal report suggesting the employment of such a structural schema. A benefit of such a schema is that new tasks with a similar structure can be learnt more quickly, by binding new task-relevant information to the existing co-ordinate system. On the other hand, such a structural representation may also have a deleterious effect when task structure is variable. In the present experiments, and from the perspective of our chunking account, a representation of the structural relations of the rules may have led to a higher rate of neglect in low IQ participants as they failed to separate panel rules from response rules.

Our account is extended by an analysis of the evolution of behaviour during early trials. In this period, action selection initially appears to be unstructured and chaotic, but is rapidly organised into a more automatic stable strategy that is then used for the remainder of the task. We suggest that this initial unstructured period reflects a learning phase and it is in this early phase of task performance that complexity exerts its effects. As the first response decisions are made, significant interference arises from the presence of additional panel rules in the sub-task model. Goal neglect arises when an incorrect plan of behaviour is constructed and stabilized.

In our tasks, what drives the chunking? In Experiment 2, we examined the hypothesis that instructional format drives chunking but found no evidence in its support. Despite a strong manipulation of instructional format to discourage chunking by sub-task, we found that the pattern of complexity effects remained unchanged.

Another factor driving chunking may be temporal context (Badre, Kayser, & D'Esposito, 2010; Botvinick, 2012; Rigotti, Ben Dayan Rubin, Morrison, Salzman, & Fusi, 2010). During initial performance, the attempt to recall task rules may be driven by the problem stimuli encountered. As participants effortfully access stored task rules while solving initial trials, chunking may be driven by the learning of transition probabilities specific to each sub-task. For instance, after the ± rule for vehicles is accessed, the car/motorbike rule will always be required followed by the lower/upper case letter rule and so on. The open/closed book rule will never be needed after the ± rule. Learning these temporal regularities may allow for construction of separate higher level context representations for each sub-task. At the same time, such temporal linkages would be a barrier for the further division of a sub-task into component episodes (Botvinick, Niv, & Barto, 2009).

Chunking may also have been driven by the fact that the two sub-tasks in our tasks had independent sets of rules,

with largely independent stimulus sets. The stimuli themselves could directly activate the appropriate task knowledge in a bottom-up manner. At the same time, the task rules within a sub-task are all functionally linked to each other, describing the various steps in a decision process and thus demanding the top-down influence of a task model for driving the search within the selected set of action alternatives.

As in previous cases of goal neglect, we found that the accuracy of response decisions was strongly correlated with Culture Fair IQ. The results suggest a general approach to the problem of task complexity and its link to fluid intelligence. In any domain, increasing task complexity increases the search space of potential action alternatives, and the requirement for simplifying search through division into separate, simpler sub-tasks. In large part, we suggest, fluid intelligence may concern this search problem, with high fluid intelligence reflecting either the capacity to deal with larger information chunks, or capacity for division of large chunks into smaller, more focused parts. As argued earlier, a close link to goal neglect suggests a central role of frontal lobe control processes in fluid intelligence (Duncan et al., 1996). On many accounts, the frontal lobe is critical in attentional control, integrating processing in multiple brain systems to construct coherent, focused behavioural episodes (Dehaene, Kerszberg, & Changeux, 1998; Desimone & Duncan, 1995; Norman & Shallice, 1980).

In summary, our results demonstrate an important role for chunking in shaping competition between task requirements in the task model. We propose a new account of neglect that emphasises the importance of chunking in navigating the space of task knowledge in goal-directed behaviour, in particular during a labile phase over the first few trials of performance. More broadly, our results suggest a general approach to the well-known link between task complexity and fluid intelligence (Marshalek et al., 1983). At least in large part, fluid intelligence may concern the problem of selecting effective behaviour based on complex knowledge; a problem that must generally be solved by dividing a complex problem space into smaller, more manageable parts.

Acknowledgements

References

Altamirano, L. J., Miyake, A., & Whitmer, A. J. (2010). When mental inflexibility facilitates executive control. *Psychological Science, 21*(10), 1377–1382.

Anderson, J. R. (1983). *The architecture of cognition*, Cambridge, MA.

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

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience, 10*(9), 659–669.

Badre, D., Kayser, A. S., & D'Esposito, M. (2010). Frontal cortex and the discovery of abstract action rules. *Neuron, 66*(2), 315–326.

Botvinick, M. M. (2012). Hierarchical reinforcement learning and decision making. *Current Opinion in Neurobiology*.

Botvinick, M. M., Niv, Y., & Barto, A. C. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition, 113*(3), 262–280.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.

Capaldi, E., & Neath, I. (1995). Remembering and forgetting as context discrimination. *Learning & Memory, 2*(3–4), 107–132.

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

Cattell, R. B., & Cattell, A. (1973). *Measuring intelligence with the culture fair tests*. Institute for Personality and Ability Testing.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87–114.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450–466.

Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review, 112*(1), 3.

De Jong, R., Berendsen, E., & Cools, R. (1999). Goal neglect and inhibitory limitations: Dissociable causes of interference effects in conflict situations. *Acta Psychologica, 101*(2), 379–394.

Dehaene, S., Kerszberg, M., & Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences, 95*(24), 14529–14534.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*(1), 193–222.

Diamond, A., Carlson, S. M., & Beck, D. M. (2005). Preschool children's performance in task switching on the dimensional change card sort task: Separating the dimensions aids the ability to switch. *Developmental Neuropsychology, 28*(2), 689–729.

Duncan, J. (2010). *How intelligence happens*. Yale University Press.

Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia, 33*(3), 261–268.

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*(3), 257–303.

Duncan, J., Parr, A., Woolgar, A., Thompson, R., Bright, P., Cox, S., et al. (2008). Goal neglect and Spearman's g: Competing parts of a complex task. *Journal of Experimental Psychology: General, 137*(1), 131.

Duncan, J., Schramm, M., Thompson, R., & Dumontheil, I. (2012). Task rules, working memory, and fluid intelligence. *Psychonomic Bulletin & Review*, 1–7.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211.

Frye, D., Zelazo, P. D., & Burack, J. A. (1998). Cognitive complexity and control: I. Theory of mind in typical and atypical development. *Current directions in Psychological Science, 7*(4), 116–121.

Gershberg, F. B., & Shimamura, A. P. (1995). Impaired use of organizational strategies in free recall following frontal lobe damage. *Neuropsychologia, 33*(10), 1305–1333.

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences, 11*(6), 236–242.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21*(06), 803–831.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46*(3), 269–299.

Incisa della Rocchetta, A. (1986). Classification and recall of pictures after unilateral frontal or temporal lobectomy. *Cortex, 22*(2), 189–211.

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 & Review, 9*(4), 637–671.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General; Journal of Experimental Psychology: General, 132*(1), 47.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception, 36*(14), 1.1–16.

Kopelman, M. D., & Stanhope, N. (1998). Recall and recognition memory in patients with focal frontal, temporal lobe and diencephalic lesions. *Neuropsychologia, 36*(8), 785–796.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14*(4), 389–433.

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*(2), 107–127.

Milner, B. (1963). Effects of different brain lesions on card sorting: The role of the frontal lobes. *Archives of Neurology, 9*(1), 90.

Neves, D. M., & Anderson, J. R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. *Cognitive Skills and Their Acquisition*, 57–84.

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

Newell, A., Shaw, J. C., & Simon, H. A. (1958). Chess-playing programs and the problem of complexity. *IBM Journal of Research and Development, 2*(4), 320–335.

Norman, D. A., & Shallice, T. (1980). *Attention to action: Willed and automatic control of behavior.* DTIC Document.

Oberauer, K., Süß, H. M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence, 31*(2), 167–193.

Oberauer, K., Süβ, H.-M., Wilhelm, O., & Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence, 36*(6), 641–652.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116*(2), 220.

Piek, J. P., Dyck, M. J., Nieman, A., Anderson, M., Hay, D., Smith, L. M., et al. (2004). The relationship between motor coordination, executive functioning and attention in school aged children. *Archives of Clinical Neuropsychology, 19*(8), 1063–1076.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). Task context and organization in free recall. *Neuropsychologia, 47*(11), 2158–2163.

Raven, J. C., Court, J. H., & Raven, J. (1988). *Manual for Raven's progressive matrices and vocabulary scales*, London.

Rigotti, M., Ben Dayan Rubin, D., Morrison, S. E., Salzman, C. D., & Fusi, S. (2010). Attractor concretion as a mechanism for the formation of context representations. *NeuroImage, 52*(3), 833–847.

Sacerdoti, E. D. (1974). Planning in a hierarchy of abstraction spaces. *Artificial Intelligence, 5*(2), 115–135.

Salthouse, T. A. (1991). Mediation of adult age differences in cognition by reductions in working memory and speed of processing. *Psychological Science, 2*(3), 179–183.

Spearman, C. (1927). The Abilities of Man. New York: Macmillan.

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

Stankov, L. (2000). Complexity, metacognition, and fluid intelligence. *Intelligence, 28*(2), 121–143.

Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence, 30*(3), 261–288.

Towse, J. N., Lewis, C., & Knowles, M. (2007). When knowledge is not enough: The phenomenon of goal neglect in preschool children. *Journal of Experimental Child Psychology, 96*(4), 320–332.

Unsworth, N., & Engle, R. W. (Eds.). (2007b). *Individual differences in working memory capacity and retrieval: A Cue-dependent search approach.*

Unsworth, N., & Engle, R. W. (2007a). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*(1), 104.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., et al. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science, 10*(2), 119–125.

West, R. (2001). The transient nature of executive control processes in younger and older adults. *European Journal of Cognitive Psychology, 13*(1–2), 91–105.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development, 11*(1), 37–63.