THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2003, *56A* (6), 1053–1077

Inspection times and the selection task: What do eye-movements reveal about relevance effects?

Linden J. Ball
Lancaster University, Lancaster, UK
Erica J. Lucas
University of Derby, Derby, UK
Jeremy N. V. Miles
University of York, York, UK
Alastair G. Gale
University of Derby, Derby, UK

Three experiments are reported that used eye-movement tracking to investigate the inspection-time effect predicted by Evans' (1996) heuristic–analytic account of the Wason selection task. Evans' account proposes that card selections are based on the operation of relevance-determining heuristics, whilst analytic processing only rationalizes selections. As such, longer inspection times should be associated with selected cards (which are subjected to rationalization) than with rejected cards. Evidence for this effect has been provided by Evans (1996) using computer-presented selection tasks and instructions for participants to indicate (with a mouse pointer) cards under consideration. Roberts (1998b) has argued that mouse pointing gives rise to artefactual support for Evans' predictions because of biases associated with the task format and the use of mouse pointing. We eradicated all sources of artefact by combining careful task constructions with eye-movement tracking to measure directly on-line attentional processing. All three experiments produced good evidence for the robustness of the inspection-time effect, supporting the predictions of the heuristic–analytic account.

Requests for reprints should be sent to Linden J. Ball, Department of Psychology, Lancaster University, Lancaster, LA1 4YF, UK. Email: l.ball@lancaster.ac.uk

This research was conducted whilst Linden Ball was at the University of Derby, and the reported studies made extensive use of the eye-tracking facilities of the University's Applied Vision Research Unit (AVRU). We are grateful to AVRU staff—particularly Kevin Purdy and Mark Mugglestone—for their expert technical assistance during the acquisition and computer-based analysis of our eye-movement data. We also thank Kevin Paterson and Edward Sutherland for useful discussions on the potential role of eye tracking in research on high-level cognition, and Maxwell Roberts, Simon Handley, and Ruth Byrne for valuable comments on previous versions of this paper. We acknowledge the University of Derby for its financial support.

http://www.tandf.co.uk/journals/pp/02724987.html DOI:10.1080/02724980244000729

Wason's selection task (Wason, 1966) continues to be one of the most investigated reasoning problems, perhaps because of the ongoing divergence of opinion concerning how to explain response patterns that variants of the task elicit. In a standard abstract version of the task, participants are presented a rule of the form *If p then q* and four cards to which the rule refers. For example, the rule might be *If there is an A on one side of the card then there is a 3 on the other side*, with the facing sides of cards showing A, J, 3, and 7 (corresponding to the logical values, *p*, *not-p*, *q*, and *not-q*). Participants are told that each card has a capital letter on one side and a number on the other side, and are instructed to decide which card or cards need to be turned over to determine whether the rule is true or false. The logically correct choices for a conditional reading of the rule are A and 7 (*p* and *not-q*), and for a biconditional reading are all four cards. Few subjects produce either of these response patterns. Most choose A alone (*p*), or A and 3 (*p* and *q*).

Wason's explanation of this selection pattern (Wason & Johnson-Laird, 1972) was that it reflected a verification bias—that is, people were trying to prove the rule true (by finding a card with an A and 3 combination) rather than trying to prove it false, as logic necessitates. However, Evans and Lynch (1973) demonstrated that when negative components are permuted through the rule, people's responses suggest that they are simply selecting cards that match the items named in the rule, irrespective of the presence of negative components. If, for example, the rule is *If there is an A on one side of the card then there is not a 3 on the other side*, then there is still a strong tendency for participants to select the A and the 3 cards (which for this rule are the logically correct choices). This phenomenon is known as *matching bias* (Evans, 1989), and it appears to be a robust finding in selection task studies using abstract materials with the negations paradigm (see Evans, 1998b, for a meta-analysis involving weighted mean selection frequencies across a number of similar experiments).

Three main theories have been forwarded to explain people's choices on the abstract selection task, including the dominant matching pattern. Evans' heuristic–analytic theory (e.g., Evans, 1989, 1996) proposes that reasoning is in general a two-stage process. First, implicit, preconscious *heuristics* determine which aspects of a task appear to be relevant, so that attention is selectively focused on such aspects. Second, explicit, conscious *analytic* processes are applied to task features that have been selectively represented as relevant in order to formulate an inference or judgement. Analytic processes, therefore, include people's capability to demonstrate a degree of deductive competence. Evans' (1998b) specific account of matching bias on the selection task is that it derives from the operation of a linguistically based matching heuristic, which reflects the use of negative terms in natural language—where they are used to deny suppositions rather than to assert new information. Essentially, a negative statement is a comment that does not alter the topic of an assertion (e.g., the topic of a statement such as "There is not a 3" is still about the number 3). Evans' theory of abstract selection task performance is, therefore, attentional in nature. Linguistic cues to relevance determine the locus of attention, and participants choose cards that they are induced to attend to.

Johnson-Laird and Byrne's (1991) mental models theory of the selection task is based around two key assumptions. First, participants are viewed as considering only cards that are explicitly represented in their models of the rule. Second, it is proposed that they only select cards for which the hidden value could bear on the truth or falsity of the rule. To account for matching bias, Johnson-Laird and Byrne argue that when negated components are involved in the conditional statement, then initial models are expanded to include affirmative propositions. For example, *If A then not 3* would be represented as:

[A] ¬3
   3
. . .

In this notation, each line depicts a different model that could be true if one assumes the truth of the conditional. The square brackets around "A" in the first model denote that this proposition is exhausted with respect to ¬3 (i.e., that whenever an A occurs a ¬3 also occurs). The second model is incomplete in that whether or not an A would occur given that there is a 3 has not been represented by the reasoner. In the third line, the ellipses denote an implicit model, meaning that there may be other models consistent with the rule to which reasoners have not assigned explicit content. Central to this theory is the notion that people have limited working-memory capacities and hence will represent as little as possible in their initial model-set to capture the meaning of the connective *if*. People may, however, try to "flesh-out" incomplete or implicit models to make them fully explicit as a result of task demands, including contextual or instructional cues. Within the model-set for the *If A then not 3* rule, it is only the A and the 3 cards whose hidden values could bear on its truth or falsity if it is treated as a conditional. It is therefore these cards that will be selected, so producing a matching response pattern.

Evans (1998b) suggests that the mental models account of the abstract selection task may be viewed as an implementation of the matching heuristic within the model theory framework; what is "relevant" is isomorphic to what is explicitly modelled. This is not strictly true, however, since for the models of the *If A then not 3* rule shown above, it is clear that the *not-3* case is explicitly represented, yet Evans would be unhappy with the idea that this card was perceived as relevant by participants, since they should presumably select it (which, in fact, they generally do not). Indeed, this example illuminates a key difference between the mental models and heuristic–analytic accounts of the selection task. The models account involves an analytic stage that does determine card selections (i.e., whilst cards that are explicitly represented are considered, only those bearing on the truth or falsity of the rule are actually chosen), whilst in the heuristic–analytic account, card selections are viewed as being determined solely by heuristically based relevance judgements. Any subsequent analytic processing applied to cards perceived to be relevant is aimed purely at justifying or rationalizing decisions that have already been made on the basis of relevance (Evans, 1995). In this way, the heuristic–analytic account of "analytic" processing on the selection task may have advantages over the mental models account in that it can readily make sense of the finding (Evans & Wason, 1976) that the retrospective verbal reports that people provide for their selections reflect attempts to justify the logic of their choices in terms of either verification or falsification (depending on the nature of the rule), with no evident insight into the actual basis of selections.

The expected information gain theory of the selection task (e.g., Oaksford & Chater, 1994) explains performance in terms of optimal data selection. It is proposed that card choices are based upon their information value in relation to their potential support for the rule (estimated in the form of expected information gain). The upshot of a mathematical analysis of the information value of cards reveals that deciding to turn the *q* card for an abstract affirmative rule can, under certain circumstances, be more informative than deciding to turn the *not-q* card. This proposal is consistent with the data and explains the matching effect seen on consequent cards with the affirmative rule. As Evans (1998b) notes, one problem for this theory remains its inability to account for the matching effects typically observed on antecedent card choices in the negations paradigm.

The production of converging empirical evidence beyond actual patterns of card selections in order to arbitrate between theories has been limited, but has been attempted with some success by those favouring the heuristic–analytic account. For example, Evans, Ball, and Brooks (1987) examined decision order in a computer-presented version of the selection task, where participants were required to make yes–no decisions about cards by pressing buttons. As predicted by the heuristic–analytic account, people made decisions about matching cards before mismatching ones, and there was a correlation between the frequency of choice of a card and the tendency to decide about it earlier than later. Such decision-order effects are suggestive of attentional biases impacting upon selection behaviour. Evans, Newstead, and Byrne (1993), however, note that these effects may simply indicate a response bias to register "yes" before "no" decisions.

Stronger evidence for the heuristic–analytic account was provided by Evans (1996) in two studies adopting a mouse-pointing technique to record participants' card inspection times. This technique involved participants tackling computer-presented selection tasks whilst simultaneously using a mouse pointer to indicate the card they were considering. A card could be selected by clicking the mouse whilst the pointer was touching it, and no action was required for nonselected cards. The cumulative time that the pointer touched a card was referred to as the card's *inspection time*, and was assumed to indicate the duration spent thinking about that card. In order to understand Evans' inspection-time predictions, it is useful to reconsider the fundamental tenets of the heuristic–analytic theory: (1) Card selections are determined purely by heuristic processes; and (2) any analytic processing that occurs will not affect choices, but will instead be aimed at rationalizing heuristically cued decisions. The consequence of these assumptions is that participants should think more about selected cards than rejected ones. More specifically, Evans specified the following predictions: (P1) Cards associated with higher selection rates will be associated with longer inspection times; and (P2) for any given card, participants who choose it will have longer inspection times than those who do not. Evans' (1996) studies provided good support for both predictions. There were large differences in mean inspection times between selected and nonselected cards, with the former generally being greater than 4 s and the latter typically being less than 2 s.

Despite converging evidence for a heuristic–analytic account of selection-task performance, Roberts (1998b) has expressed a need for caution in interpreting inspection-time findings. He argues that there are methodological aspects of the card-inspection paradigm that could have induced artefactual support for the view that people are spending time rationalizing choices that have been cued by relevance-determining heuristics. It is proposed that these artefacts are derived from three potential sources of bias arising from the format of Evans' (1996) tasks. First, people may have a tendency briefly to pause the mouse pointer over a card before making an active decision about it (clicking on it). Since in Evans' (1996) studies active responses were only required for selected cards, such momentary hesitations could add additional time to the total inspection-time values for selected versus rejected items. Second, people may forget to move the mouse pointer to a new card after considering a previous one, even though attention has shifted to the new card. Given the evidence of Evans et al. (1987) that people have a preference for registering yes before no decisions, forgetting to move the mouse would add spurious additional time to selected cards (decided about earlier) compared to rejected cards (decided about later). Third, sensory leakage—whereby cards may be looked at and even rejected before the mouse pointer has had a chance to reach them—would also tend to add additional time to selected rather than rejected cards.

To investigate the impact of these potential sources of bias within the inspection-time paradigm, Roberts (1998b) systematically manipulated their presence or absence across a series of studies. To eliminate sensory leakage and forgetting to move the mouse, participants were permitted to observe only one card at a time (blackened-out cards became visible when touched by the mouse pointer). To neutralize hesitation biases arising from resting the mouse pointer over a card before making an active decision about it, participants were required to make active "yes–no" choices for all cards, as in the methodology pioneered by Evans et al. (1987). This effectively inflates the inspection times for all cards equally. In his experiments, Roberts (1998b) demonstrated that the size of the inspection-time effect was closely related to the number of sources of bias that were present. In his Experiment 4, where all sources of bias were removed, the inspection-time effect was eradicated. In a final study, Roberts was able to reverse the inspection-time effect by employing a novel de-selection task (where participants had to make active decisions about cards that they felt should not be selected). Evidence for the reversed effect was the finding that inspection times for cards that were not selected (active responses) were now longer than those for cards that were selected (passive responses).

Subsequent to Roberts' (1998b) studies, a debate has ensued about the implications of inspection-time findings for the heuristic–analytic account (Evans, 1998a; Roberts, 1998a). The thrust of Evans' (1998a) argument is that Roberts' results actually cast doubt on the mouse-pointing methodology as a technique for revealing the effects of relevance-determining heuristics in certain task contexts, rather than having any major implications for the heuristic–analytic theory per se. Evans suggests that Roberts' manipulations (blanked-out cards and "yes-no" decision requirements) force participants to attend to irrelevant cards that would normally be ignored. This enforced attention, of necessity, increases the inspection times for irrelevant cards (although it does not alter the fact that these cards are subsequently rejected) and thereby dilutes the inspection-time effect typically observed with selected cards. Roberts counters that Evans has to invoke auxiliary assumptions to enable the heuristic–analytic account to deal with paradoxical inspection-time effects, which potentially weakens the account by making it unfalsifiable.

Although we have some sympathy with Roberts' (1998a) concerns, we generally concur with Evans (1998a) that the key issue that has been put in the foreground in the debate surrounding inspection-time findings is the inadequacy of the mouse-pointing technique for studying the influence of heuristic and analytic processes. In the context of the selection task, the method's fundamental weakness ultimately seems to relate to its lack of sensitivity for monitoring the second-by-second attentional processing of cards. This insensitivity itself appears to derive from the highly indirect measure of attentional processing that mouse pointing provides—that is, participants have actively and effortfully to move the mouse pointer to whatever they are thinking about. Such requirements for conscious self-monitoring of one's own attentional focus suggest that mouse pointing is a poor way to monitor the rapid, moment-by-moment attentional shifts that presumably underlie cognitive performance with multifaceted, display-based problems like the selection task.

Our aim in this paper was to employ a more sensitive and direct method—eye-movement tracking—to examine the on-line locus of attentional processing in the selection task. The relationship between eye movements and what is consciously attended to and thought about has been investigated extensively (e.g., Fisher, 1999; Klein, Kingstone, & Pontefract, 1992; Reuter-Lorenz & Frendrich, 1992). Although it is possible to move attention covertly (i.e., without moving the eyes), it is generally acknowledged that with visually based stimuli it is more efficient to move the eyes rather than merely to move attention (e.g., He & Kowler, 1992; Sclingensiepen, Campbell, Legge, & Walker, 1986). Moreover, there is substantial evidence indicating that attention actually precedes a saccade to a new location (Hoffman & Subramaniam, 1995; Kowler, Anderson, Dosher, & Blaser, 1995), that attentional movements and saccades are obligatorily linked (Deubel & Schneider, 1996), and that the length of a fixation or a gaze (which may involve two or more continuous fixations on the same location) provides a good index of ease of processing (Liversedge, Paterson, & Pickering, 1998). In sum, in complex information-processing tasks such as reading or display-based reasoning, the coupling between eye location and the locus of ongoing attentional processing is likely to be very tight indeed (see Rayner, 1999, for further pertinent arguments).

EXPERIMENT 1

The purpose of our first experiment was to use an eye-tracking technique to explore inspection-time predictions deriving from Evans' (1996) heuristic–analytic account of selection task responses. The direct measure of attentional processing that is afforded by tracking eye movements should eliminate all three sources of task format bias that arise from the mouse-pointing technique, which, by inflating inspection times for selected cards, potentially undermine any apparent support for the heuristic–analytic theory (Roberts, 1998b). Since the present experiment did not involve participants moving a mouse pointer to a card (to indicate that it was being considered) prior to clicking the mouse on the card (to indicate its selection), there was no opportunity for any momentary hesitation bias deriving from the mouse-pointing technique to inflate card inspection times for active choices. In addition, the strong link between attentional processing and eye movements necessarily obviated sensory leakage (thinking about one card whilst looking at another) and the possibility of participants forgetting to move their eyes to a card they were thinking about. In line with the heuristic–analytic theory, it was expected that cards that were more likely to be selected would have longer inspection times, and that for any given card, people selecting it would have longer inspection times than those rejecting it.

Method

Participants

Participants were 34 undergraduate volunteers from the University of Derby who took part in the experiment to gain course credit. Participants had not received any tuition on the psychology of reasoning.

TABLE 1
The four types of conditional rule employed in Experiment 1 and the standard terminology that is used to describe associated cards

	Logical case			
	TA	*FA*	*TC*	*FC*
If A then 3	**A** **(p)**	J (not-p)	**3** **(q)**	7 (not-q)
If E then not 5	**E** **(p)**	L (not-p)	2 (not-q)	**5** **(q)**
If not S then 9	D (not-p)	**S** **(p)**	**9** **(q)**	4 (not-q)
If not N then not 8	T (not-p)	**N** **(p)**	1 (not-q)	**8** **(q)**

Note: TA = true antecedent; FA = false antecedent; TC = true consequent; FC = false consequent. Bold type indicates the elements on cards that match those in the presented rule. The logical combination of cards to select for each rule is TA and FC.

Materials

The experiment involved a standard negations paradigm involving abstract conditional rules. Participants received four versions of the selection task, with negatives permuted through the conditional statement as shown in Table 1 (conventional terminology is used throughout the paper). Our use of four abstract rules diverges from the materials employed by Evans (1996) and Roberts (1998a), which included two thematic rules in addition to an abstract affirmative and an abstract negative rule. Notwithstanding this difference in materials across studies, the inspection-time predictions deriving from the heuristic–analytic theory remain equivalent.

Each problem was presented on an A4 sheet positioned on an easel 60 cm in front of the participant in their direct line of vision. For each problem, the rule appeared at the top of the page, a reminder of the task requirement appeared in the middle of the page, and the pictures of the four cards were presented in the lower half of the page in a two-by-two arrangement, with cards separated from each other vertically and horizontally by 5 cm and 8 cm respectively. The location of cards within the array was always random.

Apparatus

The eye-tracker was an Applied Science Laboratories 4000 system that uses a near-infrared light source and two video cameras: a scene camera, which locates the participant in their environment, and an eye camera, which produces a close-up image of one eye. The light source is guided through an arrangement of mirrors and lenses into the participant's eye and produces a retinal reflex (which makes the pupil appear bright) and a corneal image reflection (a bright, smaller reflection off the front surface of the eye). The relationship between these reflections changes as the eye moves and is used to calculate point-of-gaze coordinates to an accuracy of one degree of visual angle.

Before point-of-gaze recording could occur, the relationship between pupil and corneal reflections had to be determined with the participant looking at known points on the scene plane. This was achieved using a three-by-three grid of nine evenly spaced points placed at the same location as that where problem sheets would be presented. The relationship between corneal and pupil reflections was stored as the participant fixated the nine points. Using these data, it was possible to interpolate the point-of-gaze across the whole scene plane. During the study, a participant's point-of-gaze was superimposed onto the scene video image as a small, dark square. The resulting video recording was later used for the card inspection-time analyses in conjunction with data files holding a participant's continuous point-of-gaze co-ordinates (sampled at 50 times per second).

Procedure

Participants were tested individually. No head restraints were used as the system maintains good accuracy even with a degree of lateral head movement. Once a participant's point-of-gaze coordinates for the nine known points had been determined they were given written instructions to read as follows:

> This study is concerned with people's logical reasoning ability and will entail you having to tackle a total of four problems. Each of these problems will appear on a separate sheet in front of you. Each problem consists of four cards and a rule that applies to those cards. This rule may be true or false. The cards have been constructed so that each one always has a letter on one side and a single-figure number on the other side. Naturally only one side of each card will be visible to you.
>
> For each problem your task is to decide which card or cards need to be turned over in order to discover whether or not the rule is true. When you make a decision in relation to each problem, please point briefly at the card or cards you feel need to be turned over using the hand-held pointer.

Once the participant had read the instructions, the experimenter read them aloud once again to provide an opportunity for the participant to clarify the study requirements. The four problems were then presented in a random order. Participants were instructed to use a 20-cm metal pointer to indicate card choices, to avoid interference to eye-movement recording arising from the participant raising their hand near the scene plane.

Results

Our first analyses determined whether the rules we had used demonstrated the typical selection pattern observed in the literature (more matching than mismatching choices across antecedent and consequent cases). To assess the presence of matching bias we used the procedures adopted by Evans et al. (1987), summarized in Table 2a. An alpha level of .05 was set for all tests reported in this and subsequent experiments. Wilcoxon signed-rank tests (one-tailed) revealed that more antecedent matching cards were selected than mismatching ones, $p = .003$, and that more consequent mismatching cards were selected than mismatching ones, $p = .05$. Our selection data are, therefore, typical of what are observed with the negations paradigm.

The quantification of total inspection times for our subsequent analyses made use of commercially available software (Mugglestone, 1999). Very brief fixations of 100 ms or less were excluded from the calculation of a card's total inspection time. Descriptive analysis of inspection-time data suggested that they were highly positively skewed. A log transformation (subsequent to the addition of a constant of 0.2) was found to stabilize variances successfully. Identical problems with our inspection-time data were encountered in all the experiments we

TABLE 2

(a) Formulae for comparing matching versus mismatching selections pooled across (i) antecedent cases and (ii) consequent cases

(b) Percentage frequencies of card selections and overall mean inspection times[a] for Experiment 1 by items[b]

(a)

One-tailed Wilcoxon signed-rank comparisons for:

(i) matching antecedent (TA:1 + TA:2 + FA:3 + FA:4) versus mismatching antecedent (TA:3 + TA:4 + FA:1 + FA:2), and

(ii) matching consequent (TC:1 + TC:3 + FC:2 + FC:4) versus mismatching consequent (TC:2 + TC:4 + FC:1 + FC:3)

(b)		*Card*							
		TA		*FA*		*TC*		*FC*	
		Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
1. If p then q	Frequency of selections	88.2		8.8		35.3		14.7	
	Natural data: time	1.42	1.22	0.42	0.59	1.29	1.27	0.75	0.96
	Transformed data: log time	0.10	0.32	–0.34	0.32	0.03	0.36	–0.19	0.38
	Transformed data: time	1.06		0.26		0.87		0.45	
2. If p then not q	Frequency of selections	70.6		20.6		20.6		29.4	
	Natural data: time	2.19	2.18	1.14	2.32	0.98	1.73	0.59	0.20
	Transformed data: log time	0.21	0.41	–0.12	0.42	–0.16	0.41	0.10	0.37
	Transformed data: time	1.42		0.56		0.49		1.06	
3. If not p then q	Frequency of selections	52.9		23.5		64.7		29.4	
	Natural data: time	1.39	1.52	1.50	1.06	2.26	2.39	0.98	1.05
	Transformed data: log time	0.02	0.41	0.16	0.25	0.22	0.40	–0.04	0.30
	Transformed data: time	0.85		1.25		1.46		0.71	
4. If not p then not q	Frequency of selections	55.9		35.3		50.0		32.4	
	Natural data: time	1.69	2.12	1.91	2.00	1.58	2.20	1.92	1.69
	Transformed data: log time	0.07	0.43	0.10	0.48	0.02	0.44	0.17	0.39
	Transformed data: time	0.98		1.06		0.85		1.28	
Mean		66.9		20.1		42.7		26.5	

Note: TA = true antecedent; FA = false antecedent; TC = true consequent; FC = false consequent.

[a] Inspection times are given in seconds for both the natural data and the transformed data in original units.

[b] $N = 34$.

TABLE 3
Mean inspection times[a] for selected and nonselected cards for Experiment 1, (a) by items and (b) by participants

(a) by items

		Selected			*Not selected*		
			Mean			*Mean*	
Rule	*Card*	*N*	*TD*	*ND*	*N*	*TD*	*ND*
1. If p then q	TA	30	1.28	1.55	4	0.24	0.43
	FA	3	1.00	1.37	31	0.22	0.33
	TC	12	1.18	1.64	22	0.76	1.11
	FC	5	1.84	2.10	29	0.33	0.52
2. If p then not q	TA	24	2.09	2.80	10	0.49	0.73
	FA	7	1.58	1.81	27	0.42	0.97
	TC	7	1.54	2.41	27	0.35	0.61
	FC	10	1.58	0.51	24	0.90	0.63
3. If not p then q	TA	18	1.58	2.13	16	0.38	0.56
	FA	8	1.12	1.30	26	1.28	1.56
	TC	22	1.94	2.75	12	0.87	1.36
	FC	10	1.22	1.66	24	0.56	0.70
4. If not p then not q	TA	19	1.54	2.30	15	0.52	0.93
	FA	12	1.75	2.67	22	0.80	1.50
	TC	17	1.46	2.27	17	0.46	0.88
	FC	11	2.14	2.70	23	0.98	1.54

(b) by participants

	Selected		*Not selected*	
	Mean	*SD*	*Mean*	*SD*
Natural data: time	2.16	1.40	0.90	0.65
Transformed data: log time	0.24	0.24	–0.11	0.27
Transformed data: time	1.54		0.58	

Note: TD = transformed data (in original units); ND = natural data; TA = true antecedent; FA = false antecedent; TC = true consequent; FC = false consequent.

[a] In s.

report, and analyses were therefore performed on log-transformed data throughout. For clarity of interpretation, where applicable, means after transformation but converted back into their original units are also reported.

Our first inspection-time analysis explored the correlation between overall mean inspection times for each of the 16 cards and their associated selection frequencies. Mean inspection times and the percentage frequency of card selections are shown in Table 2b. In line with P1 of Evans (1996), long inspection times were associated with high selection frequencies, $r = .55$, $N = 16$, $p = .028$ (transformed data).

The second inspection-time analysis aimed to examine Evans' (1996) P2: that for any given card, mean inspection times for individuals selecting it would be greater than those for individuals who did not. We followed Roberts' (1998b) method for assessing P2, which simply involves determining, for each card, whether the mean inspection time is greater for selectors than for nonselectors. Mean inspection times for each card are given in Table 3a. After transformation, mean inspection times for 15 out of 16 cards were greater for selectors than for nonselectors, $p = .001$, two-tailed with the binomial test.

As Roberts (1998b) has argued, a problem with the P1 and P2 analyses is their lack of statistical power as they are item analyses. To overcome this difficulty, he proposes a more powerful participant-level analysis to test the prediction, P3, that for each individual, the mean inspection time should be longer for selected than for nonselected cards. Two mean inspection-time scores were, therefore, calculated for each person from the transformed data (Table 3b). A within-participants analysis of variance (ANOVA) provided good support for P3, $F(1, 33) = 65.13$, $MSE = 2.08$, $p < .001$.

Discussion

The results of Experiment 1 appear to lend very good support for predictions deriving from Evans' (1996) heuristic–analytic account of the selection task. Overall, the analyses demonstrate that selected cards are inspected for longer than nonselected ones. The inspection-time effect thus seems robust—even with the elimination of all three potential sources of task-format bias that could have induced artefactual support for the heuristic–analytic account in Evans' (1996) original mouse-pointing studies. However, one potential criticism of interpreting our observed pattern of results as supporting Evans' heuristic–analytic theory is that although we have presented evidence that people select more matching than mismatching cards, together with evidence that people spend longer looking at selected cards than rejected cards, we have failed to demonstrate unequivocally that these two findings hold conjointly (i.e., that the people who select matching cards are the same people as those who spend longer looking at selected cards). In addition, it might be argued that recent studies of individual differences in selection-task performance (Stanovich & West, 1998) may cast some doubt on the inspection-time predictions of the heuristic–analytic account for reasoners of higher intelligence, who seem to resist the relevance-determined matching responses and instead select in a more logical manner.

In response to these kinds of concerns, however, it is noteworthy that Evans (1996, 1998a) has clarified that "relevance" effects can extend well beyond matching cards in determining selections. For example, a key finding of Evans (1996) was that even when selection patterns vary considerably across different thematic materials they can still be interpreted as arising from relevance judgements. Likewise, on the abstract task, an *if-heuristic* has been proposed by Evans (e.g., 1989) to drive the selection of the true antecedent (TA) card across all rules irrespective of whether or not these cards have matching status. In line with this proposed if-heuristic, our data on card selection frequencies for our Experiments 1 to 3 revealed a highly reliable tendency for participants to select the nonmatching TA cards over all other cases (all ps $< .001$, Friedman tests).

The upshot of these arguments is that the heuristic–analytic theory would predict reliable differences between selected and nonselected inspection times for both matching cards and mismatching cards, even in separate analyses that focused on these distinct card types. It is thus possible to derive two new by-participants predictions from the heuristic–analytic theory: (P4a) For each individual, the mean inspection time will be longer for the matching cards they select than for those they reject; and (P4b) for each individual, the mean inspection time will be longer for the mismatching cards they select than for those they reject. Planned contrasts revealed good support for both predictions. For the matching-selected versus matching-rejected comparison, the respective mean inspection times were 1.71 s and 0.82 s (transformed data in natural units), which was highly reliable, $F(1, 24) = 47.98$, $MSE = 1.84$, $p < .001$. For the mismatching-selected versus mismatching-rejected comparison, the respective mean inspection times were 1.50 s and 0.43 s (transformed data in natural units), which was, again, highly reliable, $F(1, 24) = 59.14$, $MSE = 4.59$, $p < .001$.

We believe, then, that our P4 analyses continue to demonstrate good support for predictions that are derivable from Evans' heuristic–analytic theory. It is noteworthy, too, that Evans' (1996) argument that relevance can extend beyond matching cards in determining selections also lends itself to a relevance-oriented account of individual differences in responding. The point is, all that is happening when people make different card selections from one another (even occasionally unusual selections) is that different linguistic, pragmatic, or attentional factors have cued relevance via heuristic processes.

Before claiming unequivocal support for the predictions of the heuristic–analytic account, however, it is important to consider whether a task-based factor potentially present in Experiment 1 could have inadvertently influenced the results. Our concern is that each participant's total inspection time for chosen cards may also have included an amount of time during which they attended to the card whilst initiating and performing a pointing response to indicate its selection. Since this additional time associated with pointing behaviour would only have been associated with selected cards (rejected cards did not require an active, pointing response), it is possible that the inspection time effect in Experiment 1 is, once again, artefactual in nature. Indeed, given our claims for the sensitivity of the eye-tracking measure of inspection times, it is possible that even a small amount of inflation in the inspection times for selected cards over rejected ones arising from the pointing requirement could provide spurious support for our predictions.

One line of evidence against an artefact-based explanation of our findings would be provided if we could demonstrate that any time inflation that arose from an active decision requirement was, by itself, insufficient to account for the full inspection-time difference between selected and nonselected cards. An examination of the video records of card selection responses revealed that the time taken between a participant initiating the movement of the pointing device and registering a card selection was always very short (averaging approximately 0.5 s and rarely taking longer than 1 s). In contrast, the mean inspection-time difference between selected and nonselected cards was 1.26 s (natural data; 0.96 s transformed data in original units), indicating that something in the region of 0.5 s of the inspection-time difference between selected and nonselected cards cannot be attributed to a pointing artefact. Instead, this small, residual inspection-time effect may well reflect the predicted influence of an analytic rationalization process.

EXPERIMENT 2

The results of Experiment 1 provide support for an inspection-time effect on card selections as predicted by Evans' (1998b) heuristic–analytic account, albeit an effect that may have been inflated by a remaining source of bias arising from the use of a standard selection-task paradigm requiring active, pointing-based decisions only to be registered for selected cards. Experiment 2 aimed to remove this source of bias by fully separating the reasoning and pointing components of the selection task. This was achieved by requiring participants to press a button (which activated a light situated above the problem) in order to alert the experimenter to the fact that they were ready to register their card selections. The light activation was simultaneously captured on the video recording of the participant's eye movements, such that during the inspection-time analyses, any card inspections arising subsequent to light activation could be omitted. If the inspection-time effect observed in Experiment 1 was primarily an artefact of the pointing requirement for selected cards, then a reduced inspection-time effect (if not one that was entirely eliminated) would be expected. On the other hand, if the inspection-time effect remained reliable despite these response format changes, then this would appear to lend further support to the predictions of the heuristic–analytic account.

Method

Participants

Participants were 30 undergraduate and postgraduate volunteers from the University of Derby. The undergraduates took part in the experiment to gain course credit. Participants had not received any tuition on the psychology of reasoning.

Materials and apparatus

The selection task materials and eye-tracking apparatus were identical to those used in Experiment 1. In addition, a small button-box was given to participants at the beginning of the study and positioned in their nonfavoured hand. Pressing the button temporarily activated a light above the problem so as to alert the experimenter to the participant's readiness to register their card selections.

Procedure

The procedure was identical to that of Experiment 1, except that the instructions were modified to include reference to the use of the button-box prior to registering decisions by means of a metal pointer. Thus, the fourth paragraph read:

> Once you have reached a point where you think you know which card or cards need to be turned over then please press the button on your hand-held button-box. This will momentarily activate a light so that the experimenter knows that you are ready to indicate your selections. The experimenter will then ask you to point at the card or cards you feel need to be turned over using the hand-held pointer.

Results

Wilcoxon signed-ranks tests (one-tailed) revealed that more antecedent matching cards were selected than antecedent mismatching ones, $p = .02$, and more consequent matching cards were selected than consequent mismatching ones, $p = .004$.

Data analyses of inspection-time data proceeded in the same way as in Experiment 1. Mean inspection times (before and after transformation) and the percentage frequency of card

TABLE 4

Percentage frequencies of card selections and overall mean inspection times[a] for Experiment 2 by items[b]

		Card							
		TA		*FA*		*TC*		*FC*	
		Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
1. If p then q	Frequency of selections	86.7		6.7		60.0		3.3	
	Natural data: time	3.73	3.80	1.67	1.67	3.57	3.19	2.21	2.64
	Transformed data: log time	0.43	0.40	0.07	0.47	0.44	0.37	0.13	0.49
	Transformed data: time	2.49		0.98		2.55		1.15	
2. If p then not q	Frequency of selections	83.3		23.3		13.3		43.3	
	Natural data: time	5.04	4.75	3.04	5.26	2.39	4.02	4.74	6.00
	Transformed data: log time	0.55	0.41	0.24	0.44	0.18	0.43	0.42	0.50
	Transformed data: time	3.35		1.54		1.31		2.43	
3. If not p then q	Frequency of selections	53.3		16.7		56.7		33.3	
	Natural data: time	5.98	11.15	5.44	7.59	5.70	7.67	4.30	5.87
	Transformed data: log time	0.43	0.53	0.54	0.42	0.51	0.50	0.43	0.47
	Transformed data: time	2.49		3.27		3.04		2.49	
4. If not p then not q	Frequency of selections	60.0		30.0		50.0		36.7	
	Natural data: time	4.46	3.57	5.12	4.91	5.15	5.82	7.15	9.73
	Transformed data: log time	0.53	0.38	0.55	0.42	0.53	0.41	0.60	0.49
	Transformed data: time	3.19		3.35		3.19		3.78	
Mean		57.3		19.2		45.0		29.2	

Note: TA = true antecedent; FA = false antecedent; TC = true consequent; FC = false consequent.

[a] In seconds for both the natural data and the transformed data in original units.

[b] *N* = 30.

selections are presented in Table 4. The correlation between selection frequency and mean inspection time was significant: $r = .62$, $N = 16$, $p = .007$ (transformed data).

Table 5a shows mean inspection times (before and after transformation), for selections versus nonselections, for each of the 16 cards. After transformation, the difference between mean inspection times for selected and nonselected cards was in the expected direction for 13 of 16 cases, which was significant with the binomial test, $p = .004$, two-tailed. The inspection-time effect emerging from this item-level analysis was confirmed in a participant-level analysis, using ANOVA, that compared mean inspection times for each participant's selected versus nonselected cards (Table 5b): $F(1, 29) = 43.25$, $MSE = 0.98$, $p < .001$. Two separate

TABLE 5

Mean inspection times[a] for selected and nonselected cards for Experiment 2, (a) by items and (b) by participants

(a) by items

		Selected			*Not selected*		
			Mean			*Mean*	
Rule	*Card*	*N*	*TD*	*ND*	*N*	*TD*	*ND*
1. If p then q	TA	26	2.43	3.29	4	2.82	6.56
	FA	2	4.93	4.98	28	0.85	1.43
	TC	18	2.62	3.23	12	2.49	4.09
	FC	1	10.98	10.98	29	1.06	1.91
2. If p then not q	TA	25	3.69	5.41	5	2.14	3.18
	FA	7	3.43	5.35	23	1.18	2.34
	TC	4	1.42	1.53	26	1.28	2.52
	FC	13	3.35	6.40	17	1.89	3.47
3. If not p then q	TA	16	3.97	9.72	14	1.50	1.69
	FA	5	3.04	4.58	25	3.35	5.61
	TC	17	3.69	4.73	13	2.37	6.97
	FC	10	4.37	7.39	20	1.84	2.76
4. If not p then not q	TA	18	4.59	5.67	12	1.84	2.66
	FA	9	2.96	3.82	21	3.60	5.67
	TC	15	4.07	5.94	15	2.55	4.36
	FC	11	4.17	8.37	19	3.52	6.44

(b) by participants

	Selected		*Not selected*	
	Mean	*SD*	*Mean*	*SD*
Natural data: time	5.33	4.61	3.34	3.39
Transformed data: log time	0.56	0.29	0.31	0.33
Transformed data: time	3.43		1.84	

Note: TD = transformed data (in original units); ND = natural data; TA = true antecedent; FA = false antecedent; TC = true consequent; FC = false consequent.

[a] In s.

planned contrasts (again by-participants) were undertaken to determine the reliability of the selected versus nonselected difference for matching cards (3.25 s vs. 2.43 s; transformed data in natural units), and for mismatching cards (3.60 s vs. 1.31 s). Both contrasts were reliable: $F(1, 23) = 5.06$, $MSE = 0.41$, $p = .034$, and $F(1, 23) = 41.38$, $MSE = 3.92$, $p < .001$, respectively.

Discussion

The inspection-time effect predicted by Evans' (1996) heuristic–analytic theory remained evident in Experiment 2, despite the eradication of all sources of task format bias, including the bias associated with the requirement to point at cards to register selections. Interestingly, the mean, participant-level, inspection-time difference between selected and nonselected cards in Experiment 2 (1.99 s for natural data; 1.59 s for transformed data in original units) was greater than that observed in Experiment 1, where the pointing bias had potentially been present (1.26 s for natural data; 0.96 s for transformed data in original units). One possible explanation of this finding is that the response collection method in Experiment 2 (where participants alerted the experimenter before registering their decisions) may have amplified the size of the inspection-time effect, since participants may have repeated their processing of cards so as to be sure of their selections before committing themselves. Of course, an effect can only be amplified if it is present to begin with, so this explanation in no way undermines the evidence for a predicted inspection-time imbalance between selected and nonselected cards.

The presence of reliable inspection-time effects in Experiment 2 also runs contrary to the evidence of Roberts (1998b, Exp. 4), which nullified inspection-time effects through the use of task modifications designed to eradicate all sources of task format bias (i.e., sensory leakage, forgetting to move the mouse, and hesitation bias). We would suggest that the disparity in results may be attributable to our use of an eye-tracking method as a direct measure of attentional processing, which affords greater sensitivity in measuring momentary fluctuations in the allocation of attention to cards than that obtained with the more indirect mouse-pointing methodology employed by Evans (1996) and by Roberts (1998b).

EXPERIMENT 3

The way Roberts (1998b) removed sources of task-format bias entailed selection task modifications, such as the presentation of blanked-out cards that only became visible when touched by the mouse pointer, together with the enforcement of a "yes–no" decision requirement for all cards. This latter technique was used by Roberts (Exp. 3 and 4) to standardize any influence of hesitation bias, since having to make a decision about each card will avoid any differential inflation of inspection times associated with standard instructions to register only active decisions for selected cards.

The use of a yes–no decision requirement for all cards is also a good way to neutralize the influence of pointing biases associated with inspection times measured via eye-movement tracking. Within a yes–no decision paradigm, any inflation in the allocation of attention to a card resulting from the need to point at it in order to register a decision would be equal for all cards. If evidence for an inspection-time effect could still be found even in a yes–no paradigm that forced participants to register decisions about cards that they might ordinarily completely ignore, then this would, once again, seem to provide very good support for the predictions of the heuristic–analytic account. Given our general arguments for the sensitivity of the eye-tracking method and its ability to detect small but reliable differences in inspection times across cards, we would be comfortable making such risky predictions. It is these predictions that we set out to test in Experiment 3.

Method

Participants

Participants were 31 undergraduate volunteers from the University of Derby who took part in the experiment in order to gain course credit. Participants had not received tuition on the psychology of reasoning.

Materials and apparatus

The selection task materials and eye-tracking apparatus were identical to those used in Experiment 1, with the exception that each card was now associated with small "yes" and "no" decision boxes (separated horizontally by 0.8 cm) that were placed directly below the card at a distance of about 1 cm from its lower edge.

Procedure

The procedure was identical to that of Experiment 1, except that the instructions were modified to include reference to the presence of yes and no response boxes below each card. Thus, the fourth paragraph of the instructions read as follows:

> If you decide that a card needs to be turned over then please point to the "yes" box below the card. If you decide that a card doesn't need to be turned over then point to the "no" box below the card. You will need to make a decision about each of the cards presented to you.

Results

Wilcoxon signed-ranks tests (one-tailed) revealed that more antecedent matching cards were selected than antecedent mismatching ones, $p = .004$, and more consequent matching cards were selected than consequent mismatching ones, $p < .001$.

Transformed inspection-time data were analysed as in Experiments 1 and 2. Mean inspection times for each card and the percentage frequency of card selections are presented in Table 6. The correlation between selection frequency and mean inspection time was significant: $r = .61$, $N = 16$, $p = .011$ (transformed data).

Mean inspection times (both before and after transformation) for selections and nonselections, for each of the 16 cards, are given in Table 7a. For transformed data, the difference between mean inspection times for selected and nonselected cards was in the expected direction for 9 out of 16 cases, which was not significant with a binomial test, $p = .80$, two-tailed. However, a more powerful participant-level analysis, using ANOVA, revealed a significant difference in the mean inspection times for each participant's selected versus nonselected cards (see Table 7b): $F(1, 30) = 14.87$, $MSE = 0.11$, $p = .001$. Two separate planned contrasts (by-participants) were undertaken to assess the reliability of the selected

TABLE 6

Percentage frequencies of card selections and overall mean inspection times[a] for Experiment 3 by items[b]

		Card							
		TA		*FA*		*TC*		*FC*	
Rule		*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
1. If p then q	Frequency of selections	90.3		16.1		74.2		16.1	
	Natural data: time	2.19	1.53	1.67	1.68	1.68	1.44	1.30	0.92
	Transformed data: log time	0.31	0.25	0.15	0.33	0.18	0.27	0.09	0.30
	Transformed data: time	1.84		1.21		1.31		1.03	
2. If p then not q	Frequency of selections	90.3		29.0		16.1		64.5	
	Natural data: time	3.46	3.78	2.31	2.51	2.84	3.42	3.07	2.46
	Transformed data: log time	0.39	0.38	0.25	0.38	0.28	0.43	0.41	0.31
	Transformed data: time	2.26		1.59		1.71		2.31	
3. If not p then q	Frequency of selections	61.3		25.8		83.9		22.6	
	Natural data: time	2.65	2.50	2.45	2.11	2.90	2.56	2.33	2.16
	Transformed data: log time	0.32	0.36	0.28	0.39	0.36	0.34	0.25	0.38
	Transformed data: time	1.89		1.71		2.09		1.58	
4. If not p then not q	Frequency of selections	48.4		51.6		45.2		51.6	
	Natural data: time	1.96	1.57	2.61	2.34	2.08	1.55	2.56	2.02
	Transformed data: log time	0.22	0.34	0.30	0.38	0.24	0.35	0.34	0.31
	Transformed data: time	1.50		1.80		1.54		1.99	
Mean		72.6		30.6		54.9		38.7	

Note: TA = true antecedent; FA = false antecedent; TC = true consequent; FC = false consequent.

[a] In seconds for both the natural data and the transformed data in original units.

[b] $N = 31$.

TABLE 7
Mean inspection times[a] for selected and nonselected cards for Experiment 3, (a) by items and (b) by participants

(a) by items

		Selected			*Not selected*		
			Mean			*Mean*	
Rule	*Card*	*N*	*TD*	*ND*	*N*	*TD*	*ND*
1. If p then q	TA	28	1.80	2.20	3	2.09	2.09
	FA	5	1.99	2.48	26	1.12	1.51
	TC	23	1.28	1.57	8	1.54	2.01
	FC	5	0.78	0.90	26	1.09	1.38
2. If p then not q	TA	28	2.31	3.56	3	1.84	2.51
	FA	9	3.11	4.26	22	1.18	1.52
	TC	5	2.89	3.16	26	1.54	2.78
	FC	20	2.20	2.55	11	2.68	4.02
3. If not p then q	TA	19	2.04	3.10	12	1.66	1.93
	FA	8	1.42	1.91	23	1.80	2.63
	TC	26	2.31	3.00	5	1.31	2.38
	FC	7	1.31	1.90	24	1.71	2.45
4. If not p then not q	TA	15	1.54	1.75	16	1.39	2.18
	FA	16	2.20	3.26	15	1.46	1.91
	TC	14	1.35	1.84	17	1.71	2.27
	FC	16	2.20	2.84	15	1.75	2.26

(b) by participants

	Selected		*Not selected*	
	Mean	*SD*	*Mean*	*SD*
Natural data: time	2.74	1.63	2.10	1.27
Transformed data: log time	0.33	0.22	0.25	0.21
Transformed data: time	1.94		1.58	

Note: TD = transformed data (in original units); ND = natural data; TA = true antecedent; FA = false antecedent; TC = true consequent; FC = false consequent.

[a] In s.

versus nonselected difference for matching cards (2.20 s vs. 1.71 s; transformed data in natural units), and for mismatching cards (2.09 s vs. 1.46 s). Both contrasts were reliable: $F(1, 19) = 5.82$, $MSE = 0.28$, $p = .026$, and $F(1, 19) = 11.53$, $MSE = 0.53$, $p = .04$, respectively.

Discussion

Experiment 3 supported P1 of Evans (1996), but failed to detect a significant inspection-time effect for Evans' P2, which involved an item-based analysis. However, more powerful, participant-level analyses resulted in strong confirmation for the predicted selected versus nonselected inspection-time imbalance. Taken together, the results of Experiment 3 provide good support for the heuristic–analytic account of selection task performance, in spite of the introduction of task modifications aimed at ensuring that any influence of a pointing bias on inspection-time measures was equalized across all cards.

Interestingly, the adoption of the yes–no decision requirement for all cards appears to have had some impact on the magnitude of the inspection-time effect in Experiment 3 when compared to Experiments 1 and 2. In particular, the participant-level data for Experiment 3 reveal the smallest inspection-time imbalance between selected and rejected cards across the three experiments (0.64 s for natural data; 0.36 s for transformed data in original units). However, this diminution in the size of the inspection-time effect is hardly surprising when considered in the context of the heuristic–analytic account, since the yes–no instructional requirement will, of necessity, compel at least some degree of attention to cards that might normally be completely ignored in the standard selection task paradigm (see Evans, 1998a, and Roberts, 1998a, for related arguments). The key point is that even such *response-compelled attention*—to use Roberts' (1998a) term—does little to undermine the presence of a predicted inspection-time effect.

GENERAL DISCUSSION

Experiments 1 to 3 were motivated by Evans' (1996) heuristic–analytic account of choices on the abstract selection task, which claims that preconscious, heuristic processes direct attention towards cards which appear to be relevant (which end up being selected) and away from ones that seem irrelevant (which are rejected), whilst conscious analytic processes only serve to rationalize decisions that have already been made on the basis of relevance. The predictions that we tested concern the so-called inspection-time effect that Evans (1996) claimed should be observable when people tackle selection tasks—that is, people should spend more time inspecting cards that they end up selecting than those that they end up rejecting because only selected items are associated with analytic thought processes aimed at justifying heuristically determined choices. Our experiments aimed to provide a direct measure of the duration of on-line, attentional processing of cards, whilst simultaneously removing task-induced biases that Roberts (1998b) has suggested may be solely responsible for the inspection-time effect observed by Evans (1996) with a mouse-pointing paradigm. We employed an eye-tracking method to provide a sensitive, moment-by-moment index of the locus of participants' attentional focus during task performance.

Experiment 1 used a standard selection task paradigm where active, pointing responses needed to be registered only for selected cards. Although robust inspection-time effects were observed, the fact that pointing responses were required only for selections meant that this experiment may have inadvertently fallen foul of yet another methodological artefact, since the time taken to point at selected cards may have inflated their inspection times by a small but consistent amount relative to nonselected cards. Experiments 2 and 3 aimed to eradicate any influence of a possible pointing bias. Experiment 2 prevented the bias by separating the reasoning component of the selection task from the decision-registering component (inspection times for cards were monitored only up to the point directly prior to participants registering decisions). Experiment 3 equalized the influence of the pointing bias for all four cards by enforcing a yes–no pointing decision for each card. Experiments 2 and 3 again revealed reliable evidence for an inspection-time effect, although the effect in Experiment 3 was slightly diminished.

Overall, the inspection-time effect predicted by Evans' (1996) heuristic–analytic account of selection task performance shows impressive consistency across the different problem formats and decision requirements employed in our experiments. Moreover, failure to support Evans' predictions could have raised serious questions about key theoretical assumptions underpinning the heuristic–analytic account (Roberts, 1998a, 1998b). It is also interesting to consider our results in relation to the recent findings of Roberts and Newton (2001), who report three new selection task experiments encompassing methodological innovations to improve on the mouse-pointing technique implemented in previous studies. One modification involved the use of "change" tasks, where cards were presented as either selected or nonselected, and participants changed these where necessary. Results demonstrated an association between selection and inspection time independent of one between the act of responding and inspection time. Roberts and Newton accept that this supports the view that heuristic-induced biases influence choices.

There remain, however, a few aspects of our data that at first sight seem a little curious when considered in relation to the heuristic–analytic framework. In particular, our participants generally inspected cards for a very short duration, whether or not they ended up selecting them. The mean participant-level inspection times across experiments ranged from 1.54 s to 3.43 s for selected cards, and from 0.58 s to 1.84 s for nonselected cards (transformed data in original units in all cases). Although the low inspection times for nonselected cards are in line with assumptions of the heuristic–analytic account (since these cards do not get thought about because of their lack of relevance), one might have expected that analytic rationalization processes applied to cards destined to be selected would take more than just a few seconds. The view that analytic rationalization processes should take a lengthy amount of time to apply may, however, be more apparent than real. For example, research by Wason and Evans (1975) exploring justifications that people provide for card choices revealed a phenomenon dubbed *secondary matching bias*, which is a tendency to justify card selections in terms of matching values that might be present on their reverse sides. For example, with the conditional rule *If there is a B on one side of the card then there is a 3 on the other side*, participants typically pursue a rationalization of why the *B* card should be selected by stating how a *3* (a matching value) on the other side would verify the rule. We would argue that this rationalization process, whilst conscious, evaluative, and analytic in intent, may, nevertheless, be guided by the rapid, heuristic cueing of information. As such, there is no reason why analytic rationalizations associated with the inspection-time effect should take longer than a couple of seconds for most people—as was observed in Experiments 1 to 3.

A second possible account of the relatively small inspection-time difference between selected and nonselected items in the selection task is that analytic rationalization processes do not actually occur at all, and that heuristic processing dominates the entirety of task-based activity. As such, the inspection-time effect could simply reflect the action of heuristic processes holding attention on relevant cards for a couple of seconds longer that irrelevant ones. Under such an account, however, it is unclear what kind of processing is occurring during the few seconds of heuristically compelled attention if analytic processing is not taking place. This account is also difficult to mesh with the evidence from verbal-protocol studies

(Evans, 1995) indicating that people spontaneously justify card selections by considering values that may be present on the reverse sides.

A resolution to the issue of what specific aspects of processing give rise to the inspection-time effect on the selection task will no doubt need to await the results of further process-tracing studies using verbal-protocol techniques and eye-movement tracking methods. Given that the inspection-time effect appears to be reliable, however, it is important to question whether other contemporary theories of the selection task can accommodate its existence. Oaksford and Chater's (1994) optimal data selection account would seem to have difficulty in making ready sense of the effect since the derivation of expected-information gains for cards should presumably take an equivalent amount of time whether they end up being selected or rejected. As it is currently formulated, however, the optimal data selection account is a computational-level theory of what needs to be computed by the cognitive system, rather than an algorithmic-level theory that specifies the detailed nature and time-course of the processing steps that underpin card selection. It does, therefore, seem infelicitous to criticize this theory for failing to lend itself to an explanation of findings at a level of specificity below that at which it is formulated. Nonetheless, it would be appealing to see this theory developed at an algorithmic level such that it might accommodate evidence for inspection-time effects.

The mental models account of the selection task (e.g., Johnson-Laird, 1995) also seems to run into difficulties in accounting for the inspection-time effect. Take, for example, a selection task involving a rule with a negated consequent such as *If there is a B on one side of the card then there is not a 3 on the other side*. The initial models set would include an affirmative proposition for the negated consequent, as follows:

$$\begin{array}{ll} [B] & \neg 3 \\ & 3 \\ \ldots & \end{array}$$

On this rule most people make the logically correct selections of B and 3. According to the mental models account, all three cards that are explicitly represented in the initial model set (the B, 3, and ¬3 cards) should be considered to assess their impact on the truth or falsity of the rule—and thus all three cards should be associated with longer inspection times than cards not explicitly represented (the ¬B card). The ¬3 card should be rejected (as indeed it tends to be) as a result of the reasoning process applied to represented cards, since it has no bearing on the truth status of the rule. Crucially, however, this rejection of a considered card should break the link between selected cards and longer inspection times and should thus undermine the emergence of any inspection-time effect. Thinking about a card that is not selected is directly contrary to Evans' (1996) inspection-time predictions, and it is also opposed to what the inspection-time data appear to show. Indeed, our own data (see Tables 2b, 4, and 6) indicate that the true consequent (TC) card for a rule such as the above is in fact not considered for very long at all, and in two out of the three experiments this card has the lowest inspection-time value out of all four cards associated with the rule.

Although at first sight the mental models account seems not to lend itself to an explanation of our inspection-time findings, it is possible to do a more formal analysis to examine the theory's capacity to accommodate inspection-time data. In particular, a by-participants analysis can be conducted that compares people's mean inspection times for cards explicitly represented in models versus their mean inspection times for cards not explicitly represented in models. The key prediction (P5) would be that cards represented in models should be inspected for longer than those not represented in models. More crucially, however, the inspection-time imbalance deriving from these mental models analyses could be directly compared with the inspection-time imbalance deriving from the selected versus nonselected analyses associated with Evans' (1996) heuristic–analytic predictions (P3 in Experiments 1 to 3). In this way, it would be possible to examine which theory leads to the larger effect magnitude.

To explore these issues formally we initially pursued separate mental models analyses for Experiments 1 to 3 in line with P5. Repeated measures ANOVAs revealed good support for P5 in all experiments: $F(1, 33) = 52.14$, $MSE = 1.16$, $p < .001$ (Experiment 1); $F(1, 29) = 74.6$, $MSE = 1.01$, $p < .001$ (Experiment 2); and $F(1, 30) = 13.97$, $MSE = 0.21$, $p = .001$ (Experiment 3). On reflection, finding statistical support for P5 may not be viewed as so surprising owing to the lack of orthogonality between the P5 and P3 inspection-time measures. The point here is that the majority of cards represented in initial mental models should, according to models theory, also be cards that end up being selected, since most of the cards represented in initial models will actually have a bearing on the truth or falsity of the rule (Johnson-Laird & Byrne, 1991). This lack of orthogonality between the two inspection-time measures associated with P5 and P3 does not, however, detract from the ability of the mental models theory to provide a coherent account for our inspection-time data.

In terms of arbitrating between different theoretical accounts of the selection task, however, what is perhaps more interesting is the pattern of effect magnitudes for the P5 and P3 inspection-time imbalances arising across Experiments 1 to 3. Our examination of the magnitude of the inspection-time imbalance for P5 (mean inspection times for cards represented in models minus mean inspection times for cards not represented in models) revealed that it was somewhat smaller in comparison to the magnitude of the P3 imbalance (mean inspection times for selected cards minus mean inspection times for nonselected cards) for two out of three experiments. For Experiment 1 the P5 imbalance was 0.55 s and the P3 imbalance 0.96 s, for Experiment 2 the P5 imbalance was 1.36 s and the P3 imbalance 1.59 s, and for Experiment 3 the P5 imbalance was 0.45 s and the P3 imbalance 0.36 s.

Overall, then, since the effect magnitude associated with the heuristic–analytic account tends to be larger than that for the mental models account in a majority of experiments, it would appear that the former theory may provide a marginally better account of inspection time data. It is noteworthy, however, that the most unbiased of the experiments (Experiment 3) is the one that could be argued to have produced better support for the models theory over the heuristic–analytic theory. This finding suggests that the models account remains a very viable theory of selection task performance and associated inspection time effects. On balance, however, we continue to favour Evans' heuristic–analytic account because of its ability to explain a breadth of selection task findings that derive from a wide range of paradigms (including protocol-based studies based on both concurrent and retrospective reports). In fairness, too, it was Evans' heuristic–analytic framework that promoted the derivation of inspection-time predictions in the first place, and it is to the credit of this account that such risky predictions were supported in the present series of eye-tracking experiments.

REFERENCES

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*, 1827–1837.

Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Evans, J. St. B. T. (1995). Relevance and reasoning. In S. E. Newstead & J. St. B. T. Evans (Eds.), *Perspectives on thinking and reasoning: Essays in honour of Peter Wason* (pp. 147–172). Hove, UK: Lawrence Erlbaum Associates Ltd.

Evans, J. St. B. T. (1996). Deciding before you think: Relevance and reasoning in the selection task. *British Journal of Psychology*, *87*, 223–240.

Evans, J. St. B. T. (1998a). Inspection times, relevance and reasoning: A reply to Roberts. *Quarterly Journal of Experimental Psychology*, *51A*, 811–814.

Evans, J. St. B. T. (1998b). Matching bias in conditional reasoning: Do we understand it after 25 years? *Thinking and Reasoning*, *4*, 45–82.

Evans, J. St. B. T., Ball, L. J., & Brooks, P. G. (1987). Attentional bias and decision order in a reasoning task. *British Journal of Psychology*, *78*, 385–394.

Evans, J. St. B. T., & Lynch, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology*, *64*, 391–397.

Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Evans, J. St. B. T., & Wason, P. W. (1976). Rationalization in a reasoning task. *British Journal of Psychology*, *67*, 479–486.

Fisher, M. H. (1999). An investigation of attention allocation during sequential eye movement tasks. *Quarterly Journal of Experimental Psychology*, *52A*, 649–677.

He, P., & Kowler, E. (1992). The role of saccades in the perception of texture patterns. *Vision Research*, *32*, 2151–2163.

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception and Psychophysics*, *57*, 787–795.

Johnson-Laird, P. N. (1995). Inference and mental models. In S. E. Newstead & J. St. B. T. Evans (Eds.), *Perspectives on thinking and reasoning: Essays in honour of Peter Wason* (pp. 115–146). Hove, UK: Lawrence Erlbaum Associates Ltd.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Klein, R. M., Kingstone, A., & Pontefract, A. (1992). Orienting of visual attention. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 46–65). New York: Springer-Verlag.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, *35*, 1897–1916.

Liversedge, S. P., Paterson, K. B., & Pickering, M. (1998). Eye movements and measures of reading time. In G. Underwood (Ed.), *Eye guidance in reading and scene perception*. Oxford: Elsevier Science.

Mugglestone, M. (1999). EMAT: *Eye-movement analysis tool*. Bedford, MA: Applied Science Laboratories.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*, 608–631.

Rayner, K. (1999). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.

Reuter-Lorenz, P. A., & Frendrich, R. (1992). Oculomotor readiness and covert orienting: Differences between central and peripheral cues. *Perception and Psychophysics*, *52*, 336–344.

Roberts, M. J. (1998a). How should relevance be defined? What does inspection time measure? A reply to Evans. *Quarterly Journal of Experimental Psychology*, *51A*, 815–817.

Roberts, M. J. (1998b). Inspection times and the selection task: Are they relevant? *Quarterly Journal of Experimental Psychology*, *51A*, 781–810.

Roberts, M. J., & Newton, E. J. (2001). Inspection times, the change task, and the rapid response selection task. *Quarterly Journal of Experimental Psychology*, *54A*, 1031–1048.

Sclingensiepen, K. H., Campbell, F. W., Legge, G. E., & Walker, T. D. (1986). The importance of eye movements in the analysis of simple patterns. *Vision Research*, *26*, 1111–1117.

Stanovich, K. E., & West, R. F. (1998). Cognitive ability and variation in selection task performance. *Thinking and Reasoning*, *4*, 193–230.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology* (Vol. I). Harmondsworth, UK: Penguin.

Wason, P. C., & Evans, J. St. B. T. (1975). Dual processes in reasoning? *Cognition*, *3*, 141–154.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. London: Batsford.

Original manuscript received 16 April 2002
Accepted revision received 10 September 2002